Role of surgery in advanced epithelial ovarian cancer: the use of evidence synthesis methodology to support practice guidelines

Andrew Bryant

Thesis submitted for the degree of Doctor of Philosophy

Population Health Sciences Institute, Newcastle University

2023

Abstract

As Statistical Editor of the Cochrane Gynaecological, Neuro-oncology and Orphan Cancers review Group for many years, I have developed a keen interest in gynaecological cancer research, in particular in advanced stage epithelial ovarian cancer. Of women with ovarian cancer, more than 70% have epithelial ovarian cancer. The aim of this PhD was to examine the role of surgery for advanced epithelial ovarian cancer because of a lack of current firm guidelines to support clinical practice. The work described in this thesis uses evidence synthesis and meta-analysis methodology, applied in both traditional and novel ways to attempt to address this main aim. The body of work extends beyond standard approaches to develop, explore, and apply methods that aimed to raise the certainty of the evidence.

I conducted two systematic review publications on the type and radicality of primary surgery. These used Cochrane Intervention Review methodology and met or exceeded the Methodological Expectations of Cochrane Intervention Reviews (MECIR)**.** They followed what is widely considered 'gold standard methods' for this type of review, including the use of a standard pairwise meta-analysis approach. One publication found, with high to moderate-certainty evidence, that there may be little difference between primary debulking surgery and interval debulking surgery in survival outcomes for treatment of epithelial ovarian cancer. The other, found only very low-certainty evidence for all outcomes comparing maximal effort debulking surgery and standard surgery.

To offer a different evidence perspective given the limitations of these reviews, a further prognostic factor review assessed the impact of residual disease on overall survival. This Cochrane prognostic review demonstrated the prognostic effect of debulking to no macroscopic residual disease (0 cm) in a primary debulking surgery setting (moderate-certainty evidence). Evidence for interval debulking surgery was sparse, so further work presented in the thesis focused on primary debulking surgery where there was more available evidence.

I note that the body of work in the thesis did identify some evidence in an interval debulking surgery setting that if a tumour is not debulked to 0 cm, then all other residual disease thresholds may be

sub-optimal and restricting the tumour to <1 cm may not matter. This finding has not been explored or reported in any other guidelines but needs more exploration when more studies adequately report these comparisons.

Given the strong association between residual disease as a prognostic factor after primary debulking surgery and prolonged survival, the thesis then focused on methodologies that aimed to improve, if possible, the confidence in effect estimates presented in the primary analysis. This included a frequentist network meta-analysis and expert elicitation with a Bayesian network meta-analysis application that fully exploited all available evidence. These methods further consolidated the results of the primary meta-analyses reported in the Cochrane prognostic factor systematic review and provided moderate-certainty evidence, based on incorporating a more thorough and informed assessment of the publication bias domain in the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) approach and the overall certainty of the evidence judgement. These methods were applied and developed in the body of work in the thesis to demonstrate added confidence to the certainty of the evidence judgements, but more specifically that the results and conclusions of the primary prognostic factor review can be strengthened to the extent of potentially influencing policy.

The evidence in the thesis suggests there is a clear benefit of achieving cytoreduction to no macroscopic residual disease. It may encourage the surgical community to attempt to increase rates of maximal effort debulking in their centres in order to achieve higher rates of cytoreduction to no macroscopic residual disease. The thesis also outlines several limitations and methodologies that require further development but could be implemented in EOC research in the future. At present, the National Institute for Health and Care Excellence may wish to consider the results of the thesis and the possible adoption of some of the proposed methods in their pending guidelines.

Acknowledgements

I would like to thank all of my coauthors on the included publications in the thesis who have all been a pleasure to collaborate with. I also hold special praise for the CGNOCG team at Bath for support throughout my entire working life at Newcastle University. Gail Quinn, Clare Jess, Jo Morrison, Tracey Bishop, and Jo Platt have been a credit to the review group and Cochrane will rue letting these people go after their restructure in March 2023.

I am very grateful to my supervisors, Professors Luke Vale and Dawn Craig for their excellent guidance and extreme patience with me throughout the whole process. Their expertise has been crucial to the success of the thesis. I also extend this gratitude to Consultant Gynaecological Oncologist Raj Naik who has acted as a clinical supervisor without any formal recognition. He has also been a huge part of my development since I started as a researcher at the University. Senior statisticians Michael Grayling and Thomas Chadwick have also offered invaluable advice and support through the whole process, even while the former was no longer employed at the University.

I would like to express gratitude to Professors Eileen Kaner and Elaine McColl for acting as internal assessors and for their critical appraisal. This gave me the direction and motivation I needed to complete the thesis.

I would also like to thank Consultant Gynaecological Oncologists Ketankumar Gajjar and Ahmed Elattar who have offered expert advice and support, always making themselves available, even if it involved Zoom meetings at midnight!

I am grateful to Shaun Hiu, Eugenie Johnson and Patience Kunonga who have selflessly offered to help and be a part of my research. I also thank Peg Ford for being available for advice on public and patient involvement and ensuring my research is accessible to as many people as possible.

I was also fortunate to conduct some exciting research during the pandemic with some brilliant people including Tess Lawrie, who I have known for years and now has her own excellent research team.

Lastly, I am grateful to Alun Armstrong who I have exchanged knowledge with since the pandemic and who reviewed my thesis as a lay person and has shown me a few New Tricks … I will get a copy of my thesis signed by Brian Lane!

Dedication

I would like to dedicate this thesis to close friends, family and of course football, golf, cricket and an occasional small drink to help get me over the line!

During the completion of my thesis the England cricket team won and lost the Ashes, become world cup ODI and T20 champions and Newcastle United at the time of submission are the richest club in the world and are back playing amongst Europe's elite. There have also been a series of generational events such as the EU referendum and global pandemic. The latter presented an opportunity for some exciting new research and the chance to form new bonds and friendships with some great and interesting people as well as helping to clear out some deadwood!

Abbreviations

AE	Adverse Event
AUC	Area Under Curve
BGCS	British Gynaecological Cancer Society
BIRD	British Ivermectin Recommendation Development
BT	Biggerstaff-Tweedie
CDSR	Cochrane Database of Systematic Reviews
CGNOCG	Cochrane Gynaecological, Neuro-oncology and Orphan Cancers Group
CHARMS-PF	CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies -Prognostic Factor setting
CHSRI	Cochrane Handbook for Systematic Reviews of Interventions
CI	Confidence Interval
CINeMA	Confidence In Network Meta-Analysis
COVID-19	Coronavirus disease 2019
CrI	Credible Interval
CRUK	Cancer Research UK
CTCAE	Common Terminology Criteria for Adverse Events
DECIDE	D: define the problem; E: establish the criteria; C: consider all the alternatives; I: identify the best alternative; D: develop and implement a plan of action; E: evaluate and monitor the solution and feedback when necessary
DL	DerSimonian Laird

EOC	Epithelial Ovarian Cancer
ES	Evidence Synthesis
EtD	Evidence to Decision
FE	Fixed Effects
FIGO	Fédération Internationale de Gynécologie et d'Obstétrique (International Federation of Gynaecology and Obstetrics)
GIV	Generic inverse variance
GOG	Gynaecologic Oncology Group
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HR	Hazard Ratio
IDS	Interval Debulking Surgery
lnHR	$\log_e$(HR) (natural logarithm of the hazard ratio)
IPD	Individual Patient Data
IQR	Inter Quartile Range
IS	Information Size
ITT	Intention To Treat
LVRD	Large Volume Residual Disease
MA	Meta-Analysis
MCMC	Markov Chain Monte Carlo
MD	Mean Difference

MECIR	Methodological Expectations of Cochrane Intervention Reviews
MRC	Medical Research Council
NACT	NeoAdjuvant ChemoTherapy
NHS	National Health Service
NICE	National Institute for health and Care Excellence
NMA	Network Meta-Analysis
NMRD	No Macroscopic Residual Disease
OC	Ovarian Cancer
OS	Overall Survival
P	P-value (significance probability)
PB	Publication Bias
PDS	Primary Debulking Surgery
PFS	Progression-Free Survival
PGI	Patient Generated Index
PhD	Doctor of Philosophy
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
QoL	Quality of Life
QUIPS	QUality In Prognosis Studies
RCT	Randomised Controlled Trial
RD	Residual Disease

ROBINS-I	Risk Of Bias In Non-randomised Studies -of Interventions
RE	Random Effects
RevMan	Review Manager
ROB	Risk Of Bias
ROC	Receiver Operating Characteristic
RR	Risk Ratio
RRR	Relative Risk Reduction
SAE	Serious Adverse Event
SD	Standard Deviation
SE	Standard Error
SHELF	SHeffield ELicitation Framework
SJ	Sidik-Jonkman
SMD	Standardised Mean Difference
SR	Systematic Review
SUCRA	Surface Under the Cumulative RAnking curve
SVRD	Large Volume Residual Disease
TSA	Trial Sequential Analysis
UK	United Kingdom
VOI	Value Of Information
WHO	World Health Organization

Peer-reviewed journal articles submitted in support of PhD

1. Coleridge SL, **Bryant A**, Kehoe S, Morrison J. Neoadjuvant chemotherapy before surgery versus surgery followed by chemotherapy for initial treatment in advanced ovarian epithelial cancer. *Cochrane Database of Systematic Reviews*. 2021;7. doi: 10.1002/14651858.CD005343.pub6.
2. Hiu S[*], **Bryant A**[*], Gajjar K, Kunonga PT, Naik R. Ultra-radical (extensive) surgery versus standard surgery for the primary cytoreduction of advanced epithelial ovarian cancer. *Cochrane Database of Systematic Reviews*. 2022;8. doi: 10.1002/14651858.CD007697.pub3. [*]Joint first author
3. **Bryant A**, Hiu S, Kunonga PT, Gajjar K, Craig D, Vale L, Winter-Roach BA, Elattar A, Naik R. Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery. Cochrane Database of Systematic Reviews 2022;9. doi: 10.1002/14651858.CD015048.pub2.
4. **Bryant A**, Johnson E, Grayling M, Hiu S, Elattar A, Gajjar K, Johnson E, Craig D, Vale L, Naik R. Residual disease threshold after primary surgical treatment for advanced epithelial ovarian cancer, Part 1: A network meta-analysis. *American Journal of Therapeutics. 2023* Jan-Feb 01;30(1):e36-e55. doi: 10.1097/MJT.0000000000001584.
5. **Bryant A**, Grayling M, Hiu S, Gajjar K, Johnson E, Elattar A, Vale L, Craig D, Naik R. Residual disease after primary surgery for advanced epithelial ovarian cancer: expert elicitation exercise to explore opinions about potential impact of publication bias in a planned systematic review and meta-analysis. *BMJ Open*. 2022;12**:**e060183. doi: 10.1136/bmjopen-2021-06018.
6. **Bryant A**, Grayling M, Elattar A, Gajjar K, Craig D, Vale L, Naik R. Residual disease after primary surgical treatment for advanced epithelial ovarian cancer, Part 2: Network meta-analysis incorporating expert elicitation to adjust for publication bias. American Journal of Therapeutics: 2023 Jan-Feb 01;30(1):e56-e71. doi: 10.1097/MJT.0000000000001548.

Tables and Figures

Tables

Figures

Figure 6: Network diagram showing residual disease comparisons after primary debulking surgery for advanced epithelial ovarian cancer

Contents

Role of surgery in advanced epithelial ovarian cancer: the use of evidence synthesis methodology to support practice guidelines

Preface

This thesis aimed to examine the role of surgery in advanced epithelial ovarian cancer. The thesis gives a general background to advanced epithelial ovarian cancer in Chapter 1 along with details of the different types and approaches to primary surgery that are undertaken, as well as a description of residual disease as a prognostic factor. The second chapter then introduces details about the origins of this research and why it is important from my point of view, outlining key aspects that will form a large part of the thesis. The second chapter also describes aspects of evidence synthesis processes and methodology and introduces important biases in meta-analyses and systematic reviews in general, with an emphasis on reporting biases which leads to explanations as to why publication bias plays such an important role in the thesis. Consequently, Chapter 2 includes details about methods for identifying and minimising publication bias. The chapter concludes by formally specifying the research aims and detailed objectives of the thesis.

I feel that it was important to set the clinical scene in the first chapter, so that when the aims and objectives are introduced the reader has a clearer idea about the surgical interventions and management available, as well as a general background to advanced EOC. In Section 2.1 I then give background information as to how this thesis and research came about and introduce the key clinical aims and my motivation for them. The next few subsections discuss biases associated with systematic reviews and why I focus on publication bias in the thesis. The aims and detailed objectives are introduced in Section 2.5, which I feel is the most logical location as all of the main aspects and features of the thesis have then been introduced with appropriate cross-referencing to the objectives where applicable.

The thesis is by publication and a summary of the methods and rationale for the conduct of each one is provided in the third chapter. This methods chapter gives a brief introduction followed by sub-

sections which provide some context of the research in the publication included in the body of work and how the methodology was developed. The remaining sub-sections in the chapter document evidence synthesis methods and inclusion criteria in each of the first three publications covering timing of primary surgery for initial treatment in advanced epithelial ovarian cancer, examination of maximal effort debulking surgery and the impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery. My role and contribution to each included publication in the thesis is provided in Appendix 1.

The body of work in the thesis reports a series of publications that extend beyond the prognostic factor research described in the third publication. This series included an extension to the standard pairwise meta-analyses by reporting the rationale and methods for conducting a network meta-analysis. The other two publications in the series reported methods that aimed to capture the degree of publication bias that was present in the effect estimates and the rationale for the conduct of the chosen methodology. The sixth and final publication in the thesis outlines how the publication bias quantified in the fifth publication was incorporated into the analyses.

The fourth chapter provides a general description of the findings of each of the systematic reviews in the first three publications as well as the results of the network meta-analysis in the fourth publication. The fifth chapter then quantifies publication bias associated with the prognostic factor reviews (from third and fourth publications) in the fifth publication and this is incorporated into effect estimates in the sixth publication. Chapter 5 documents how the results of these publications can be used to raise the certainty of the evidence and give more confidence in the judgements being made.

The thesis concludes with a structured discussion in Chapter 6 which gives summary sub-sections of the clinical findings and details of the evidence synthesis methodology used across the included publications. The discussion also highlighted areas of originality, outlined strengths and limitations, and placed context of this body of research in the thesis with existing guidelines and policy. The discussion chapter also comprehensively outlines areas for future research to identify gaps in the literature arising out of the thesis. The chapter ends with a series of conclusions.

1 Advanced ovarian cancer

1.1 Background

Comprehensive details about ovarian cancer (OC) incidence, aetiology, prognosis, management, and particulars about surgery are provided across the portfolio of published papers included in my thesis,(1-3) however for ease are summarised in the subsections below. The research aims and detailed objectives of the thesis are given in the next chapter (Section 2.5). This chapter introduces the clinical background and context for my thesis before any reference to evidence synthesis (ES) and statistical methodology.

1.1.1 Incidence

Globally, OC is the seventh most common cancer among women up to 64 years of age and is the fifth most common cancer in women of all ages. OC is uncommon in women aged under 40, but its incidence increases with age.(4, 5) In the UK, the estimated lifetime risk of being diagnosed with OC is approximately 2% for females born after 1960.(6, 7) On average there are around 7500 new OC cases in the UK every year, accounting for around 4% of all cancers diagnosed in women in the UK.(8) Globally, there are typically over 200,000 new cases per year, which accounts for 3.9% of all cancers diagnosed worldwide.(9)

It is widely accepted that OC is difficult to diagnose in its earlier stages, as there are generally few symptoms that would be overly concerning in the short term. Therefore, most women present with advanced stage disease.(5, 10-15) Symptoms for all stages include abdominal distension, bloating, indigestion, urinary frequency, urinary urgency, early satiety, weight loss, reduced appetite, abdominal and pelvic pain and, less commonly, vaginal bleeding.(16)

1.1.2 Aetiology and prognosis

Cancers of the ovary are classified according to their cells of origin. Most women have epithelial ovarian cancer (EOC), which is thought to arise from malignant cells in the tissue covering the ovary or the lining of ovarian cysts, or pre-cancerous lesions within the fallopian tube. However, some OCs also arise from the substance of the ovary, called stromal tumours, or from embryological differentiation (sex cord and germ cell tumours).(4, 17, 18) EOC accounts for approximately 90% of all OCs in women(4, 5) so the work presented in this thesis focuses on EOC as other histologies may require different management.(19)

The spread of OC is described using the International Federation of Gynaecology and Obstetrics (FIGO) staging,(20, 21) where: stage I disease is confined to the ovaries; stage II disease is confined to the pelvic cavity (true pelvis, which predominantly contains the urinary bladder, the colon and the internal reproductive organs); stage III disease is abdominal, with spread to the lining (peritoneum) of the abdominal cavity outside the pelvis and/or regional lymph glands; and stage IV disease involves spread to distant organs, such as those in the chest or the liver.(20) (see Appendix 2) Management for early (I-II) and advanced (III-IV) stages differs considerably, and the prognosis is very poor for those with advanced disease. In the UK, 43% of all women with OC are alive five years after diagnosis, but five-year net survival decreases from 68% in stage II to 27% in stage III, and to just over 13% in stage IV disease.(22)

1.1.3 Management

My thesis focuses on surgical treatment of advanced EOC. Other, rarer sub-types of OC, such as germ cell tumours, sex cord stromal or granulosa cell tumours were not considered. Whilst surgery is important in these other histological subtypes, their inherent aetiological and biological differences require them to be considered separately to the more common epithelial subtype, which invariably

presents at an advanced stage.(20) Management of advanced EOCs consists of surgery, either primary (before chemotherapy) or as an interval (mid-way through chemotherapy). The purpose of the surgery is to achieve complete cytoreduction. The chemotherapy is usually given as a combination of carboplatin (alkylating agent) and Taxol (vinca alkaloids) with or without the addition of biological agents including VEG-F receptor antagonists and PARP inhibitors. Recent studies relating to these biological agents have shown the prognostic value of complete cytoreduction.(23, 24) I will not expand on these agents any further as the thesis specifically relates to the surgical aspects of management.

1.1.4 Residual disease thresholds

Traditionally, primary surgery is performed to achieve optimal cytoreduction. However, over time the definition of optimal cytoreduction has varied with respect to the maximal diameter of residual tumour left behind after surgery. In 1994, the Gynaecologic Oncology Group (GOG) defined optimal cytoreduction as having residual disease (RD) < 1 cm.(25) This was updated in 2010 to define optimal as there being no visible residual tumour disease.(26-29) In this thesis, no macroscopic residual disease (NMRD) after primary cytoreduction equates to surgical debulking to no visible RD with the naked eye (RD = 0 cm and is often referred to in the literature as R0 or complete cytoreduction); this is also referred to as 'optimal'. Nevertheless, further definitions of optimal have been advanced in the literature. For the purposes of an ES and any primary research, clear and consistent definitions of optimal and suboptimal RD are needed so that we can reliably evaluate their relative effects on outcomes and agree on subsequent guidelines.

1.2 Types of surgery

Surgery is often the first step in the initial diagnosis and staging of EOC. This thesis focuses on advanced stage disease only, as this involves surgical management; a combination of surgery and chemotherapy with platinum-based agents is exclusively the treatment of choice for advanced disease.(1-3, 28, 29) The timing of initial surgery is one of the questions explored with my thesis (outlined in first objective in Section 2.5.2). Surgery may be given as the first treatment prior to chemotherapy (known as primary debulking surgery (PDS)), or as primary surgery following initial NACT, an approach known as interval debulking surgery (IDS).(2, 30)

To further complicate treatment, women with advanced EOC often develop bowel obstruction; this most frequently occurs in women with recurrent disease. In such circumstances, surgery is palliative in nature(31, 32) and is used solely to relieve the obstruction. A broad range of medical interventions may also be used, including palliative strategies to target pain, nausea, and vomiting. One of the limitations of surgery in this situation is that obstruction normally occurs at multiple levels within the bowel, rather than as a single obstruction lesion; single obstruction lesions often occur in primary bowel cancer.(32) Given that the aim of surgery in this population is symptom relief as opposed to attempts to improve survival, I did not include women with bowel obstruction within their initial primary OC diagnosis in my thesis as these patients would be a large source of heterogeneity and have significantly lower rates of survival.(33)

1.2.1 Upfront primary debulking surgery

Upfront PDS is defined as surgery followed by chemotherapy (usually platinum-based).(34, 35) It is the most common surgery performed as treatment for EOC; it is the focus of the majority of my thesis objectives and publications. Focusing on PDS in this manner allowed a more robust exploration of the methods being applied, since more data were available. PDS aims to debulk the

tumour and reduce the number of cancer cells, minimising the likelihood of developing resistance to chemotherapy after surgery. The main goal is to remove all of the tumour and, if this is not possible, to remove as much of the tumour as possible.(3, 36) PDS delays the start of chemotherapy, as there is the potential for chemotherapy to slow wound healing.(37) It also requires admission to hospital, which may be associated with delays in the start of treatment compared to starting chemotherapy immediately.

1.2.2 Interval debulking surgery

IDS is an alternative to PDS that involves giving platinum-based chemotherapy before PDS for advanced EOC; this is also referred to as NACT.(2, 35, 38) IDS is considered most viable when complete debulking of a tumour is unlikely to be achieved with PDS.(2, 3, 39) IDS may improve survival of women in whom primary surgery was not performed with cytoreductive intent by a gynaecological oncologist.(40)

Proponents of IDS believe that the chemotherapy will shrink the tumour prior to primary surgery enabling better resection and in theory, IDS may help maintain general health and functionality.(41, 42) Also if surgery is not curative, residual tumour cells may proliferate whilst a patient recovers from surgery. On the other hand, clinicians opposing the use of IDS have argued that it delays the removal of the tumour, compromising the overall survival (OS) of a patient. Furthermore, it is suggested in my published body of work (Appendix 3) that chemotherapy might induce fibrosis, which may make complete debulking more difficult.(2) While IDS aims to reduce cancer deposits, it could leave microscopic disease that is then difficult to surgically remove, whereas primary surgery without the use of chemotherapy might make it more visible and, as a consequence, easier to remove. A final issue is that if too many cycles of NACT are given pre-surgery, there is a concern that post-surgery chemo-resistance will increase.(43, 44)

1.3. Approach to surgery

There has been considerable controversy surrounding the use of IDS versus PDS in advanced EOC,(2) and also a lack of consensus on how radical primary surgery should be. How aggressive a surgical approach should be is often difficult for clinicians and patients to decide *a priori* and will be dependent on several factors. These include severity and spread of disease, age, histology, tumour grade and the patient's (functional) performance status, to list but a few. Most surgeons attempt to surgically remove as much tumour as possible. However, some surgeons have a very aggressive ethos, whereas others are more conservative due to uncertainties over the long-term consequences and adverse effects of aggressive surgery. Therefore, part of my published body of work sets out to explore the impact of RD after PDS and ascertain suitable categorisations of RD thresholds for when complete cytoreduction is not possible. The question of how much effort should be made to attempt small volume RD <1cm if a surgeon cannot debulk all the tumour is explored for both IDS and PDS, before more sophisticated statistical methods are applied to PDS. For women, there is uncertainty over disease recurrence, adverse effects of surgery and on how these may affect them both mentally and physically, and on their overall quality of life.(45, 46) These considerations should all be part of the treatment decision making process.

1.3.1 Maximal effort debulking and standard (radical) surgery

Patients with widespread disease, that is, those with upper abdominal disease affecting the diaphragm, liver, spleen and omentum (a large fatty structure which hangs off the middle of the colon and drapes over the intestines inside the abdomen),(47) or widespread disease affecting the bowel, will need much more radical surgery in order to achieve cytoreduction to NMRD. As the extent of disease increases, the complexity of the procedures required to achieve this also undoubtedly increases.

‘Maximal effort debulking surgery’ is now the preferred term for more extensive surgery and this has superseded terms like ‘ultraradical surgery’ that was used in Publication 2.(48) To be consistent with the current preferred term, I therefore use maximal effort debulking surgery throughout the body of the thesis. This surgery, which may include bowel resection, splenectomy, liver resection and diaphragmatic stripping, has been described in the literature as a treatment for advanced OC with low complication rates.(1, 49) Standard (with radical components which are not as extensive as maximal effort debulking) surgery is typically defined as total abdominal hysterectomy, bilateral salpingo-oophorectomy, omentectomy, pelvic and/or para-aortic lymphadenectomy, bowel surgery outside the definition of ‘maximal effort debulking’ (localised colonic resection, non-multiple bowel resection). Maximal effort debulking surgery is an extension of standard (radical) surgery(1, 50) and could include diaphragmatic stripping, extensive peritoneal stripping, multiple resections of the bowel (excluding localised colonic resection), liver resection, partial gastrectomy, cholecystectomy and splenectomy.(51) As would be expected, maximal effort debulking surgery is associated with a prolonged operating time and exposure to anaesthesia. These factors may increase the risk of hypothermia, respiratory complications such as atelectasis (lung collapse), infection, adult respiratory distress syndrome, blood loss and intraoperative ureteric, bowel and bladder injury. In the postoperative period, these women may also require a longer hospital stay and extended recovery time, with an increased risk of infection (chest, wound, urine), venous thromboembolic disease, poorer mobility and poorer nutritional status.(1, 52)

The role of maximal effort debulking and radical surgery is not universally accepted. This in part is due to the risks noted above. However, it has been suggested that the initial extent of advanced disease may reflect the aggressiveness of the tumour, and it is these factors that ultimately dictate treatment success, not the extent of surgery per se.(53) If this argument holds, then seeking cytoreduction to NMRD may not improve survival because there may still be some remaining RD after surgery.(1) Another theory is that these same women may have less biologically aggressive

tumours and that these differences in tumour biology may account for the survival benefits that are reported to be from surgery.(1, 54) However, it has also been argued that women who undergo maximal effort debulking achieving cytoreduction to NMRD have better survival. Cytoreduction to NMRD is recommended as being the aim of any type of primary surgery in UK guidelines,(50) yet many centres continue to have fairly low rates.(55, 56) Maximal effort debulking surgery is explored in the second objective outlined in Section 2.5.2.

1.4 Extent of debulking surgery

It has been argued by some that the goal of surgery, whether IDS or PDS, should be complete resection of all disease.(27-29, 57) As noted above, there is no consensus on this, with many standing by their preference of routinely performing conservative or less complicated and aggressive surgery.(52)

Consequently, there remains divided opinion about the effects of any remaining RD on survival after PDS or IDS and what attempts should be made for maximal efforts at debulking. What has been evidenced is that if surgery is performed by gynaecologists with training in gynaecological oncology who either undertake a high volume of surgeries or are based in high-volume centres, the surgery undertaken is associated with increased likelihood of cytoreduction to NMRD being attempted.(58-62) The value of removing as much tumour as possible after primary surgery is explored in a series of publications in the thesis across objectives 3-5 (outlined in Section 2.5.2).

2 Introduction

2.1 Origins of research interest

Since I started employment at Newcastle University in November 2007, I have been involved with the Cochrane Gynaecological, Neuro-oncology and Orphan Cancers review Group (CGNOCG).(63) Over time I have evolved into the Statistical Editor of the group and have formed strong bonds with CGNOCG affiliates such as Clare Jess, Gail Quinn, and many others listed in the Acknowledgements. From this work I have developed a keen interest in gynaecological cancer research, in particular in OC, which is often described as a silent killer due to the difficulty of recognising symptoms early on in the disease course.(64-66) According to Cancer Research UK (CRUK) there are 7500 OCs diagnosed in the United Kingdom (UK) each year and, of these, it would be expected that more than 70% would be diagnosed at a late stage.(67)

My main motivation for focusing this PhD on surgery for advanced EOC stemmed from a lack of available guidance as to the best course of management for late stage EOC. More specifically, a lack of clarity on what the role of surgery should be within that management. ES methods offer an opportunity to exploit the available evidence and ensure that any further primary (expensive) research is focused in areas where we have a research gap, rather than areas where we have research uncertainty that may be explained or characterised. In my work I have focused on ES approaches to identify and combine evidence that may support decisions that improve the care of those with EOC. I have also used these methods to identify and clarify areas for further research. Throughout, an important goal was also to promote research in advanced EOC by showing how innovative methodological and statistical initiatives could be used to gain confidence in the estimates of effect that formed a part of the evidence being presented. The work described in this

thesis has made extensive use of systematic review (SR) and meta-analysis (MA) methodology. My role and contribution to each included publication in the thesis is provided in Appendix 1.

Publication bias (PB) may cause additional uncertainty as to the true value of debulking to NMRD, and potential concerns regarding the strength of the current evidence base. Although there is now less controversy about the prognostic importance of cytoreduction to NMRD, there remains divided opinion about the effects of any remaining RD after primary surgery and about what attempts should be made for maximal efforts at debulking.(68, 69) Different philosophies are evident within the surgical community, but there are also other important considerations, such as surgical skills, training, the woman's fitness for more radical treatment, morbidity, mortality, and quality of life. These are all considerations when assessing PB and the reliability of the effect estimates in published studies. There is also the possible issue about unreported studies that show "negative" results, which in this context may be a study showing no benefit of cytoreduction to NMRD. Indeed, PB is a well-known threat to the validity of meta-analyses.(70, 71) This, along with other biases associated with SRs are discussed in more detail in Sections 2.2 to 2.4. Some of these other biases can to some degree be minimized by following good ES practice and are probably of a much lesser threat to the validity of the overall findings. For this reason, I was keen to explore PB in the thesis as negative or statistically insignificant findings typically have less chance of being published; therefore, available studies tend to be a biased sample. This leads to an inflation of effect size estimates of unknown degree.(72) Consequently, it has been argued that attempting to correct for bias is typically better than incorporating no correction at all because PB is inevitable in most meta-analyses. This includes when no publication biases are detected, as available tests to ascertain the presence of PB typically have low power.(73) Ultimately, using adequate methods of bias correction can add confidence to the certainty of effect estimates in a MA.

Grading of Recommendations, Assessment, Development and Evaluations (GRADE) assessments are vital to aid the conclusions in any SR.(74) GRADE is one of the most widely used tools for assessing and communicating scientific uncertainty and uses a system for rating the certainty of evidence and grading the strength of recommendations in healthcare. Over 100 organisations around the world are using GRADE or endorse it, including Cochrane and the World Health Organisation (WHO).(75) In order to apply GRADE, typically at least two independent reviewers to make judgements on the certainty of the evidence for each key outcome in a SR following the guidance provided in Chapter 14 of the Cochrane Handbook for Systematic Reviews of Interventions.(76) GRADE uses five considerations (study limitations (risk of bias), unexplained heterogeneity and inconsistency of effect, imprecision, indirectness and PB to assess the certainty of the body of evidence for each outcome. The GRADE system uses criteria ranging from very low to high certainty for assigning grade of evidence.

There are accepted challenges in applying the GRADE approach(77) in SRs and advances in methodology where existing guidance falls short is still sought.(78) I believe that the current priority should be on PB as it is likely to be an issue in most areas of health research;(79) a more thorough assessment is needed to adequately assess this domain in GRADE. When individual studies are not at all published, the results of the SR are potentially biased.(80) PB is a GRADE domain that requires making inferences about missing evidence. Several statistical and visual methods are helpful in detecting PB, but they have serious limitations.(81) These tests can be prone to error and their results should be interpreted with caution as it is extremely difficult to be confident that PB is absent.(81) It is also difficult to judge when to downgrade the certainty of the evidence due to suspicion of PB. Therefore, presenting an adjustment for PB without the need for potentially downgrading the certainty of the evidence would be very appealing. Exploring how to adjust effect estimates in a MA is something I have given considerable thought, especially since the risk of PB may

be higher for SRs of observational studies than for reviews of RCTs. Many studies in my portfolio of gynaecological SRs include these study design. The consolidation of the PB domain would add strength to the confidence in the overall judgements made. Developed from the findings of the thesis, I proposed a correction for PB as an additional domain or as an extension to the current PB domain in GRADE. This was added as one of the objectives in my thesis and had not been pre-defined from the outset (outlined in objective 5 in Section 2.5.2).

Exploration of PB is an important part of a robust SR and should always be considered. At present, there is no consensus on a standard approach for identifying and adjusting for PB, although some methods, particularly around identification, do exist. I wanted to explore adjustments to effect estimates for PB because based on collaborations with clinicians (in all clinical areas) in my work as a trial biostatistician, there is often a distinct lack of equipoise. This is often exacerbated in secondary research as many sceptics refute conclusions in an area despite an ES following rigorous and sound methodology.(82-85) Therefore, it is important to disseminate findings to the wider surgical community by reporting effect estimates that are more likely to be closer to the true effects by removing a degree of bias. It is also important to utilise the most pertinent methodologies, be it through existing ones or making advancements in novel approaches.

There have been recent considerations about how to utilise GRADE when there is limited or no evidence.(86) Areas of sparse evidence often reflect key questions that are critical to address in clinical practice guidelines due to the uncertainty among health care providers. An expert elicitation approach has been suggested as a possible method for panels to transparently deal with the lack of published evidence to directly inform recommendations.(86) I liked the concept of expert elicitation to enhance the evidence base so utilised this approach in the body of work in the thesis, by incorporating it into the PB domain in GRADE. The concept was introduced to me through my

interest in the research conducted by the Cochrane Effective Practice and Organisation of Care Group and further reading of the literature. The expert elicitation in this group is mainly applied in health economics but the concepts can be applied in other settings. Since there was a lot of uncertainty around PB in the prognostic assessment of RD after primary surgery work in the thesis, it made sense to explore expert elicitation in this setting, especially given that there is a general acceptance that PB exists in this area.(87) I knew I could then utilise my knowledge of Bayesian statistics and previous experience of using packages like WINBUGS(88) to enable me to inform prior distributions using the opinions of experts. This was a particularly novel aspect of my research, as this prior information was used to inform adjustment of MAs for PB in a way that, to the best of my knowledge, has been previously unexplored.

Without reliable guidelines based on adequate empirical evidence, polarised views will continue to exist and that was my main motivation for conducting the research in my thesis. Having the most up-to-date and reliable evidence is crucial to the development of clinical guidelines, and thus, it is of paramount importance that optimal analytical methods are used to appraise the available evidence.(89) However, current guidelines related to best management for women undergoing primary EOC surgery are not based on the highest level of evidence.

This thesis is particularly pertinent due to the high level of uncertainty facing women undergoing treatment for advanced EOC, especially given differences in practice between surgeons in the UK and internationally.

2.2 Biases in meta-analyses

The research conduct in the included publications (body of evidence) attempted to use the best available evidence. While little randomised study data were available, by using best practice ES methods along with novel methods, an attempt was made to reduce uncertainties that existed.

The issue of bias in the SR process is addressed in this chapter. Each publication explores different methodologies, both novel and existing, to try and address potential biases that might exist within the evidence and the methods. Methods explored included standard approaches, such as independent identification of studies eligible for inclusion and double data extraction.

Judgements on external and internal validity(90) were made in each included publication. External validity refers to the applicability and generalisability of the findings,(90) such as judging whether the studies in each included publication were representative and could be generalised to other populations and settings. Internal validity was also judged as it was important to determine whether the included body of evidence adequately addressed the research questions and minimised bias. In the included body of evidence, the methodological quality within a study was assessed using different risk of bias tools, which are discussed in Chapter 3. Given sound SR and methodological conduct, the main threats to any MA are reporting biases, especially selective reporting of outcomes and PB.(79, 91-93) It is worth noting that PB could be one of a number of possible explanations for small study effects being observed in a MA. However, PB is likely to be the main plausible threat in this area of research. Therefore, a large part of my thesis explores methods to minimise or overcome the impact of these two biases which are discussed in detail in Section 2.3, with methodology explored more fully in Section 3.6 and Section 3.7 (and in two of my publications).

2.3 Reporting biases

One of the objectives of my thesis was to explore the extent of PB associated with the prognostic factor research and how this might be quantified and incorporated into the analyses (see Section 3.6). The credibility of evidence syntheses can be compromised by reporting biases, which arise when dissemination of research findings is influenced by the nature of the results.(94) The different types of reporting biases comprise how, when and where research findings are reported. The thesis attempted to minimise these through ES methodology in the body of evidence. Specific examples of these methods are outlined in the description of the inclusion criteria in each publication with consideration of how these were likely to reduce the risk.

One form of reporting bias reflects that smaller studies are thought to be more likely to report larger effect sizes than larger studies.(73, 95) The belief being that smaller studies reporting null or small effects being less likely to be published; i.e. there is an element of PB. This has traditionally been seen as the main reason for the observed small-study effects in most MAs.(92, 96) Attempts to minimise other forms of reporting biases associated with small study effects other than PB were made throughout the body of work, and these are listed in Table 1. This was attempted in the design of each ES publication, through carefully set out inclusion criteria a priori and ES methodology. Small-study effects can be induced by a variety of factors besides PB(97) (Table 1), so attempting to minimise other reporting biases in the design and inclusion stage in the ES is important. Table 1 shows all known reporting biases and how I attempted to deal with them in the body of evidence in my thesis either in the review process or using ES methodology. Some biases are of limited relevance to this body of work while others such as PB were likely to be most impactful and therefore were the focus of the application of novel methodology discussed later in this thesis.

Table 1: Potential issues and solutions to reporting biases in included publications in the thesis.

Reporting bias	Potential issue in included publications in the thesis	Attempts to address the bias in the thesis
Publication bias (79)	The possible non-publication of research findings, probably due to non-significant results, or the direction of results does not favour an experimental intervention.	An expert elicitation exercise was conducted in the fifth publication in Appendix 7 and the results were applied using Bayesian methodology to incorporate an adjustment to effect estimates in the sixth publication in Appendix 9.
Selective reporting of outcomes bias (93)	The selective reporting of some outcomes or analyses, but not others, probably due to non-significant results, or the direction of the results does not favour an experimental intervention.	Attempts to minimise this bias was incorporated in the design of the elicitation exercise so that experts could factor this into their judgements. Therefore, this bias was grouped as a component of publication bias. The selective reporting of outcomes also has a domain in risk of bias assessment, so this was considered as a possible bias in all studies that met the inclusion criteria in all included ES publications.
Time-lag bias (98)	Potential delays in publication of research findings, probably due to non-significant results, or the direction of the results does not favour an experimental intervention.	Comprehensive searches and searches of the grey literature were performed in all ES publications in the thesis. A substantial proportion of studies may remain unpublished for a considerable time after study completion but eventually results should become available and the fact that attempts to obtain study data from authors were made meant that some data not available in the published literature was included in some cases. This helped minimise time-lag bias in the included ES publications in the thesis.
Language bias (99)	The publication of research findings in a particular language, possibly due to non-significant results, or the direction of the results does not favour an experimental intervention.	Attempts to translate any pertinent included study that was not published in English were made in all ES publications in the thesis.
Citation bias (100)	The citation or non-citation of research findings, depending on the nature and direction of the results.	Comprehensive searches and searches of the grey literature were performed in all included ES publications in the thesis. The way references are indexed means that most studies would probably be identified, regardless of how poorly it has been cited or disseminated.
Multiple/duplicated publication bias (101)	A study may be published multiple times with different sets of authors, sometimes making it possible that systematic review authors could include a 'unique' study	In the included ES publications in the thesis, included references were thoroughly checked for studies that may have been reported more than once and the references

	several times. This may depend on the nature of the results, or study authors may want multiple publications to boost their academic/research CVs.	were nested within a primary reference for transparency.
Location bias (102)	The publication of research findings in journals has obstructed access or worse levels of indexing in standard databases, probably due to non-significant results, or the direction of the results does not favour an experimental intervention.	Comprehensive searches and searches of the grey literature were performed in each included ES publication, so this bias was not of particular concern in the body of work in the thesis. This was supported by the Information Specialist in the CGNOCG.
Early stopping rules (103)	A study is stopped prematurely and thus lacks the desired power to detect differences in outcomes that may or may not be present, or it was stopped early based on a significant result which was subject to an early treatment effect giving overly 'optimistic' effect estimates in favour of the intervention.	This is probably not a serious form of bias in an advanced EOC setting. While smaller studies may have been more prone to sampling error,(104) even if a study was stopped early, it would only have been included in the advanced EOC publications if it had included the minimum number of participants as part of the inclusion criteria.
Sample size calculation (105, 106)	An inadequately powered study (with small sample size) could lead to spurious findings, while large studies are often statistically significant but not clinically relevant if the sample size calculation is not given full consideration.	Power calculations were not considered as part of any inclusion criteria in any of the included ES publications in the thesis. Sensitivity analyses were conducted when appropriate. A small sample size is often calculated in a pragmatic way based on resources, likelihood of participant availability and funding.
Sponsorship and funding bias (107-109)	The distortion of design and reporting of a study to favour the funder's aims. The presence of funder bias in a commercial setting is where the conduct may be disingenuous.	There are probably no serious concerns about this potential bias in any of the included ES publications in the thesis, as surgery is the mainstay of treatment in advanced EOC and there is little to gain in commercial sponsorship or vested funder interests. Although, in areas of oncology, industry-sponsored trials may be the least published.(110)
Fraud (111)	Papers retracted for fraud (data fabrication or falsification) is a serious bias and is more likely to falsify in favour of positive results towards an experimental intervention. This differs from papers retracted for error, which would likely be corrected and republished.	Data irregularities or inconsistent reporting were scrutinised. Possible research fraud was encountered in the COVID-19publication discussed in Chapter 6.(112) This was not a concern in the body of advanced EOC research in thesis.

2.4 Methods to identify and minimise publication bias

This section focuses on methods to prevent/reduce PB that were used in the body of evidence in the thesis. Good research practice was achieved by following Cochrane methods guidance,(113) as well as adhering to reporting guidelines recommended in PRISMA statements.(114) Furthermore, it was planned in all included ES publications in the thesis to interpret the results of any MA that displayed asymmetric funnel plots as evidence of small-study effects, which could seriously threaten the validity of the results of the MA.(115)

Cochrane methods guidance discusses the various approaches to assessing how the results of MAs vary under different assumptions. However, while techniques such as the Trim and Fill method(116) and selection model techniques(117, 118) have the ability to estimate intervention effects corrected for funnel asymmetry, they are not widely recommended due to their limitations.(117, 119) These methods have been shown to be inconsistent and often report PB where it does not exist particularly when studies in a MA are heterogenous.(120) In each included ES publication in the thesis, if funnel plots had suggested that treatment effects may not have been sampled from a symmetric distribution, as assumed by a random effects (RE) model, sensitivity analysis using fixed effects (FE) models could have been performed. However, it was felt that the assumptions underpinning a FE MA would not be sensible. This is because assuming that there is one true effect size that underlies all studies in the analyses in each ES publication in the thesis, and that all differences in observed effects were due to sampling error, was not felt to be realistic. Furthermore, due to the poor statistical power of all existing tests,(121-123) methods that aimed to raise the certainty of the evidence were preferred and are outlined in subsequent sections of the thesis. These methods included attempts to adjust for PB to obtain more reliable and precise effect estimates from which to draw conclusions.

It has been suggested that attempting to correct for PB is typically better than incorporating no correction at all, as PB is inevitable in most MAs. This includes the scenario when no PB is detected. This is because currently the available tests to ascertain the presence of PB typically have low power so cannot be considered reliable.(73) Ultimately, utilising adequate methods of bias correction can add confidence to the certainty of effect estimates in a MA.

2.5 Research aims and objectives

2.5.1 Aims

To evaluate the clinical effectiveness and safety of different forms of surgical management for women with surgically staged advanced EOC (FIGO stages III and IV). This was accomplished using evidence synthesis methodology. This was used to provide evidence suitable for supporting the development of clinical practice guidelines for the surgical management of advanced EOC.

More specifically I aimed to investigate:

- Primary debulking surgery (PDS) versus interval debulking surgery (IDS), that is surgery that is given as part of the initial treatment phase, either before or mid-way through chemotherapy.
- Maximal effort debulking surgery (more radical surgery beyond what is standard) versus standard surgery.
- How radical should radical surgery be by looking at the relationship between the amount of residual disease left after primary surgery and overall and progression-free survival.

2.5.2 Detailed objectives

These aims were met by undertaking a series of research projects, each of which is reported as a peer reviewed research paper. These papers make up the substantive component of this thesis by publication. These papers report how I met the aims set out above by meeting the following objectives:

1. To assess the effectiveness and safety of treating women with advanced EOC with chemotherapy before IDS (neoadjuvant chemotherapy, NACT) compared with conventional treatment where chemotherapy follows PDS (adjuvant chemotherapy).
2. To evaluate the effectiveness and safety associated with more extensive surgery (maximal effort debulking surgery) in the management of advanced stage EOC.
3. Building upon and addressing the limitations of (2) estimate the impact of the extent of residual disease (i.e., the amount of disease left after primary surgery) on overall and progression-free survival. This analysis was completed for two scenarios explored when addressing objective (1), namely PDS and IDS settings.
4. To address limitations in (3) above, alternative approaches to estimating the prognostic impact of RD on overall survival were used. More specifically:
    a. To explore whether a network meta-analysis (NMA), as an extension to the standard pairwise MA explored in (3), could be used to compare the impact of different RD thresholds in a single analysis.
    b. To explore the extent of publication bias (PB) associated with (3 and 4a) and how this might be quantified and incorporated into the analyses.
5. Use the finding of the method used to quantify PB to modify the current PB domain in the standard Grading of Recommendations, Assessment, Development and Evaluations (GRADE) tool. This aimed to raise the confidence in the GRADE judgement and consequently the overall recommendations being made.
6. Identify gaps in the literature arising out of the thesis and future research needs, in both clinical and methodological capacities, with an emphasis on raising the certainty of the evidence.

3 Methods

3.1 Introduction

This chapter outlines the ES methods used within the body of work presented in this thesis. The methods explored in each included publication are described in detail in Appendices 3-9. This chapter provides a summary of the inclusion criteria and methods used in each of the included publications. All methods were applied with respect to minimising biases in the review process and conduct of analyses. Steps taken to minimise different forms of bias in a SR are outlined throughout the subsections.

3.1.1 Context of research publications in thesis

The publications summarised in the body of work in the thesis explore different methodologies and, as set out in Chapter 2, aim to synthesise the evidence on effects of treatment and make recommendations for policy and management of advanced EOC. This was needed due to the high levels of uncertainty in the current evidence base and the failure of standard methods to go far enough in reducing this uncertainty. The novel methods in this thesis go further and may impact current recommendations. Publications 1-6 describe a progressive evaluation, with each subsequent publication seeking to provide more accurate and reliable effect estimates (by attempting to minimise bias) and so strengthening conclusions. The methods used for each piece of work are given in detail in the subsections below. These methods include the use of a NMA approach in Section 3.5, adapted to consider a prognostic feature (extent of RD) rather than the more typical comparison of interventions, an expert elicitation exercise in Section 3.6, and an adjustment of effect estimates for PB in Section 3.7.

A summary of how the projects are linked to the included publications and the main methods used is given in Figures 1 and 2 below. In summary, Figure 1 shows that Publications 1 and 2 address whether primary surgery should be given before or after chemotherapy and how aggressive the surgery should be. The ordering of these research questions addressing Objectives 1 and 2 is unimportant, but both were required to adequately address the overall research aims. Objective 3 was addressed by conducting the prognostic factor review (Publication 3) which is related to the objectives in the first two publications. However, there is a specific focus on using the results of the subsequent Publications 4-6 as a proxy to address Objective 2 in Publication 2 given an absence of evidence on maximal effort debulking surgery. Publications 4-6 use progressive methodology to attempt to raise the certainty of effect estimates and the confidence in the conclusions drawn. Overall, the methods shown in Figure 2 and applied in this thesis help to achieve these objectives and aim to help inform policy makers to make better evidence-based decisions.

Figure 1: Summary of how projects are linked to the included publications.

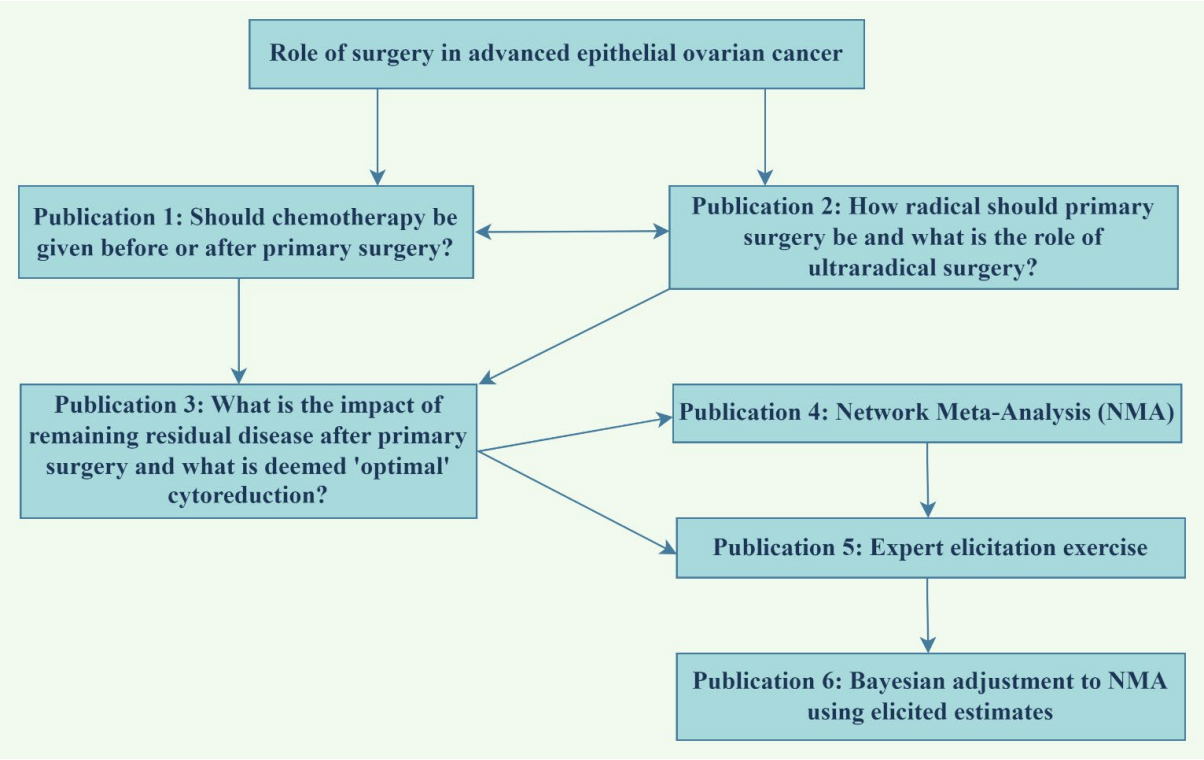
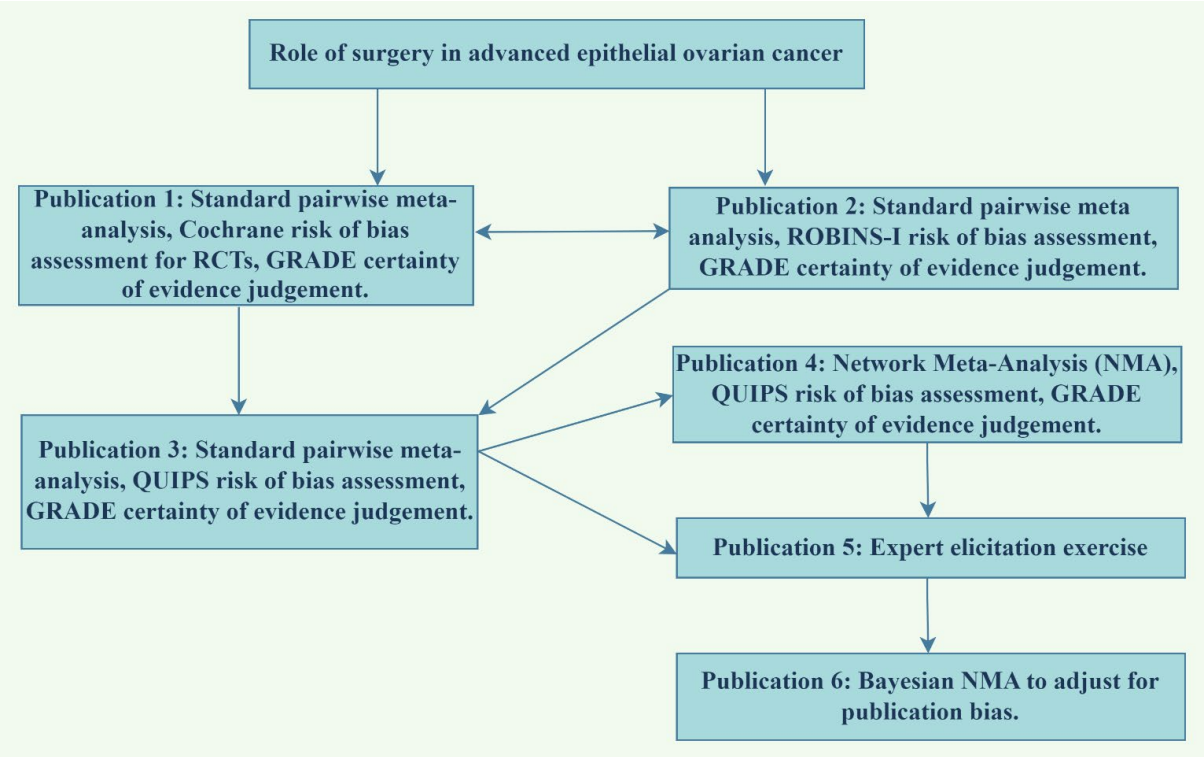


Figure 2: Summary of the methods used in each publication



3.1.2 Development of a priori research methods

A key first stage of the work was to develop research protocols for all the ESs conducted. Each ES was conducted according to internationally accepted principles of good practice by following the recommendations set out in the Cochrane Handbook for Systematic Reviews of Interventions(113) and the general principles of the Centre for Reviews and Dissemination guidance for undertaking reviews in health care.(124) Each relevant ES publication was reported according to PRISMA guidance.(114) While none of the protocols were registered on the international prospective register of systematic reviews (PROSPERO), as is often common practice, they were preregistered through an equivalent process, through the CGNOCG.(125-128)

Thus, all publications fully defined each element of the work that was carried out (the data extraction tool, risk of bias criteria, etc).(125-128) Three publications were ES within a Cochrane framework.(1-3, 113, 129, 130) while three publications (131-133) performed further work based on the original research question that was first introduced in this prognostic factor Cochrane review (third publication in Appendix 5). Each publication was conducted by a project team that had its own multidisciplinary expertise. The teams included senior clinicians, methodologists, and statisticians. All research publications were conducted in consultation between the clinical and methodological experts within the multidisciplinary team used my network of researchers to approach experts to be involved in the research and subsequent publications.

The following sections in this chapter provide a summary of the inclusion criteria and methodology of each included publication reporting on advanced EOC. The full text of these publications is available in the Appendix, but in some cases, additional explanation and justification for the approaches adopted is provided, which was not included in the publications due to strict journal word limits.

3.2 Publication 1: Chemotherapy versus surgery for initial treatment in advanced ovarian epithelial cancer

The research question in this paper addressed the first objective in the thesis, as set out in Section 2.5.2. This assessed the effectiveness and safety of treating women with advanced EOC with chemotherapy before IDS compared with conventional treatment where chemotherapy follows PDS. The following subsections show the inclusion criteria and summarise the methods related to the first publication, which aimed to minimise bias in the research process. See Appendix 3 for the full text of the first publication.

3.2.1 Participants, interventions, comparisons, and outcomes (PICOs)

3.2.1.1 Types of studies

All randomised controlled trials (RCTs) that reported a comparison of chemotherapy versus surgery for the initial treatment of advanced EOC and met the prespecified inclusion criteria were included. Studies were restricted to RCTs as it was known during initial scoping of the extant literature that randomised studies were feasible and they represent the highest level of evidence in the hierarchy of evidence from individual studies.(134) Existing SRs or guidelines were treated as a source to identify primary studies for inclusion. Additional primary studies were identified by conducting searches, as described below.

There were no restrictions for studies with early stopping rules. Given that surgery is the mainstay of treatment,(135, 136) it was expected that any study stopping early would be more likely due to lack of resources and problems with recruitment. The relatively low five-year survival rate for women with advanced EOC also potentially makes early stopping of a study less likely.(14, 20)

There was no consideration of any requirement for a prestated power calculation as part of any inclusion criteria. A standard weighted average approach using a RE model was used in all core MAs using the DerSimonian and Laird (DL) method.(137) RE models were used because the underlying assumption is more plausible in reality (and for this research area). This assumption is that the different studies in each of the included publications are estimating different (but related) intervention effects.(71)

3.2.1.2 Types of participants

Studies that reported adult women (aged 18 years of age or older) with surgically staged primary advanced EOC (FIGO stage III and IV), who had confirmed histological diagnoses were included. Those that included women with other concurrent malignancies (unless these made up only a small proportion (< 10-20%) of the total sample, or if results were reported separately) were excluded. Those studies conducted in a palliative setting, such as those that included women with bowel obstruction, were excluded.

Subgroup analyses were planned and conducted by age (dichotomised to 60 years or less and over 60 years) and by the extent of debulking achieved (complete debulking; residual tumour 1 cm or less; residual tumour greater than 1 cm) to explain any differences between participants. Subgrouping data by FIGO stage (Stage IIIc versus IV) was also planned.

3.2.1.3 Types of interventions

Studies that reported on upfront PDS, with the aim of macroscopic resection or cytoreduction to NMRD, followed by platinum-based chemotherapy, compared to platinum-based NACT followed by IDS, with the aim of resecting disease to the same degree as the PDS group.

Another issue to consider is non-adherence to an intervention, which can introduce heterogeneity into any MA. This could be an issue in any cancer-based surgical setting if participants do not fully comply with chemotherapy schedules. Lower levels of adherence may dilute any intervention effect.(138)

3.2.1.4 Types of outcome measures

Where possible, the following outcomes were assessed. Full details and definitions are given in the publication (see Appendix 3).

Primary outcomes:

- OS; and
- Progression-free survival (PFS) or recurrence-free survival.

Secondary outcomes:

- Morbidity/adverse effects (AEs), classified according to Common terminology criteria for adverse events (CTCAE);(139)
- Quality of life (QoL); and
- Extent of surgical debulking achieved (e.g., NMRD, 0.1 to ≤ 1 cm, > 1 cm).

Different outcomes will vary in importance across individual women with advanced EOC. Since it is possible for either type of surgery in this review to result in a favourable prognosis then OS was thought to be of paramount importance and was the primary outcome, although secondary outcomes such as QoL would be the focus if an individual had a poorer prognosis, such as suboptimal RD after cytoreductive surgery.

With respect to some of the outcomes (e.g., OS and PFS), estimates may be affected by the duration of the study follow-up. Differences in the duration of follow-up across studies in a MA could, in theory, introduce bias if follow-up was associated with the effect estimates.(115) For the analyses of OS and PFS in this review, hazard ratios (HR) were estimated. This statistic takes into account all time points and allows for censoring.(140)

3.2.2 Search methods

Full details on the search methods related to this sub-section are given in the full publication in Appendix 3.

3.2.3 Selection of studies

The sifting of titles and abstracts then full text papers was conducted using the methods recommended by Cochrane.(141) Full details are given in the full publication in Appendix 3.

3.2.4 Data extraction

Where possible, all data extracted were those relevant to an intention-to-treat (ITT) analysis, in which participants were analysed in groups to which they were assigned. Chapter 5 of the Cochrane Handbook for Systematic Reviews of Interventions was consulted regarding any issues with extraction of data.(142)

Data were abstracted onto a data extraction form specially designed for the review. Full details of data extraction are given in the full publication in Appendix 3.

3.2.5 Risk of bias assessment

The risk of bias (and appraised quality) for all included studies was independently assessed by at least two reviewers. For this review (first publication in Appendix 3), which was restricted to the inclusion of RCTs, Cochrane's Risk of Bias (ROB) tool was used.(143, 144) The ROB 1 version of the tool was used rather than ROB 2 because the latter was still under development at the time of production of the first publication.(145) The first pilot Cochrane review was published in November 2021 where ROB 2 was recommended as the preferred tool for assessing RCTs in new reviews, but at present ROB 1 is still almost exclusively used in most reviews. This is due to the lack of implementation within RevMan 5 and switching to ROB 2 after protocol publication, or in updated reviews, being discouraged.(145) The ROB tool assessed risk of bias in the following domains:

- selection bias: random sequence generation and allocation concealment;
- performance and detection bias: blinding of participants, personnel, and outcome assessment;
- attrition bias: incomplete outcome data;
- reporting bias: selective reporting of outcomes; and
- other possible sources of bias.

Results were presented in a risk of bias summary table, which was used to support interpretation of the results of MAs.

3.2.6 Certainty of the evidence

A full assessment was made on the certainty of the evidence for each outcome in each comparison using the GRADE approach.(74) In order to apply GRADE, at least two independent reviewers made judgements on the certainty of the evidence for each key outcome in the review following the

guidance provided in Chapter 14 of the Cochrane Handbook for Systematic Reviews of Interventions.(76)

GRADE uses five considerations (study limitations (risk of bias), unexplained heterogeneity and inconsistency of effect, imprecision, indirectness and PB) to assess the certainty of the body of evidence for each outcome. The GRADE system uses the following criteria for assigning grade of evidence:

- High: we are very confident that the true effect lies close to that of the estimate of the effect;
- Moderate: we are moderately confident that the true effect is likely to be close to the estimate of effect;
- Low: the true effect may be substantially different from the estimate of the effect; and
- Very low: the true effect is likely to be substantially different from the estimate of effect.

Summary of findings tables were created to present the main findings for the following outcomes in the PDS versus IDS review:

- OS;
- PFS;
- Various surgically related side effects (see Appendix 3 for full details);
- Postoperative mortality; and
- QoL.

3.2.7 Data synthesis

If sufficient clinically similar studies were available, their results were pooled in pairwise RE MAs(137) (using direct head to head comparisons) using Review Manager (RevMan) version 5.4.(146) For time-to-event data, hazard ratios (HRs) were pooled using the generic inverse variance facility of RevMan 5.(146) For any dichotomous outcomes, risk ratios (RRs) were calculated for each study and these were then pooled. For continuous outcomes, mean differences (MDs) between the treatment arms at the end of follow-up were pooled as all trials measured the outcome on the same scale, otherwise standardised mean differences (SMDs) were pooled. Ninety-five percent confidence intervals (CI) were calculated for all measures of effect and were reported alongside the point estimates.

3.2.8 Assessment of heterogeneity

Heterogeneity between studies was assessed by:

a) visual inspection of forest plots;(147)
b) estimation of the percentage heterogeneity between trials which cannot be ascribed to sampling variation;(148)
c) a formal statistical test of the significance of the heterogeneity;(149) and, where possible,
d) subgroup analyses (see below).

3.2.9 Assessment of reporting biases

Funnel plots corresponding to MA of the primary outcome were used to assess the potential for small study effects, such as PB.(150) If these plots suggested that treatment effects may not be sampled from a symmetric distribution, as assumed by the RE model, further MAs using FE models would have been performed. Funnel plots were interpreted with caution because visual inspection

of funnel plots is subjective and considered crude in the judgement of asymmetry of small-study effects, which could be interpreted as presence of PB.(150-152) Nevertheless, this approach is commonly adopted due to the simplicity of their interpretation. No formal tests for funnel plot asymmetry were conducted because there is an absence of any consensus of their usefulness.(153)

3.2.10 Handling missing data

Missing outcome data was not imputed for any of the outcomes. If data were missing or only imputed data were reported, study authors were contacted to request data on the outcomes only among participants who were assessed. Applying plausible assumptions to the outcomes of participants with definite missing data, the average change in pooled relative effect estimates could be substantial.(154) To assess the importance of missing data, the incomplete outcome data domain of the risk of bias assessments was used and sensitivity analyses were performed as appropriate (see Appendix 3 for full details). The risk of bias domain judgements were used to inform the overall GRADE profiles.

3.2.11 Subgroup analysis

In this review, the following subgroup analyses were specified:

- age: 60 years or less and over 60 years; and
- extent of debulking achieved: complete debulking; residual tumour 1 cm or less; residual tumour greater than 1 cm.

3.2.12 Sensitivity analysis

The sensitivity analyses mainly focused on excluding studies at high risk of bias.

3.3 Publication 2: Maximal effort debulking (ultraradical) surgery versus standard surgery for the primary cytoreduction of advanced epithelial ovarian cancer

The research question in this publication addressed the second objective in the thesis, set out in Section 2.5.2. This evaluated the effectiveness and safety associated with more extensive surgery (maximal effort debulking surgery) in the management of advanced stage EOC. In the text below, only material differences in methods between this review and the previously described review conducted to address the question of effectiveness of chemotherapy versus surgery for initial treatment in advanced EOC are described. If not stated, the underlying assumption is that the methods were the same. See Appendix 4 for the full text of the second publication.

3.3.1 Participants, interventions, comparisons, and outcomes (PICOs)

3.3.1.1 Types of studies

All relevant studies that assessed primary maximal effort debulking versus standard surgery for the treatment of advanced EOC; this included both RCT and non-randomised studies, as initial scoping of the literature did not identify any RCTs. Consequently, RCTs and non-randomised studies of at least 100 patients with concurrent comparison groups were included. The constraint around the size of the study sample for non-randomised studies was applied(155) in an effort to minimise potential biases, negate small study effects, and because of issues around statistical adjustment. Small studies might be underpowered and designs such as case-control studies might introduce high levels of associated biases.(84, 156, 157) Small sample sizes may have been insufficient for any reliable statistical adjustment to minimise selection biases.(158) Furthermore, null findings might be due to deficiencies in the study design and conduct. Hence, including these studies might not lead to appropriate MA estimates.

For non-randomised trials, prospective and retrospective cohort studies and case series, further criteria were applied, in that, only studies that used statistical adjustment for important baseline characteristics (for example age, stage (III or IV), grade, performance status, etc.) using multivariable analyses were included. This criterion was introduced as it was clear that the inclusion of unadjusted non-randomised studies would be prone to confounding bias, as healthier patients with a probable better prognosis would be at an increased likelihood to be eligible for more radical surgery. Similarly, patients with potentially poorer health may be more likely to receive less extensive or conservative surgery, medical management, or best supportive care only. Therefore, if analyses were not adequately adjusted for important baseline factors, like severity, then the reported results would be prone to selection bias and estimated effect sizes may not be realistic.

3.3.1.2 Types of participants

The inclusion of participants follows that documented in Section 3.2.1.2. Subgroup analyses, such as by stage III and IV disease and by presence or absence of carcinomatosis, were conducted when appropriate to explain any differences between participants.

3.3.1.3 Types of interventions

Studies that reported on primary maximal effort debulking surgery versus standard surgery for the management of advanced EOC.

3.3.1.4 Types of outcome measures

Where possible, the following outcomes were assessed:

Primary outcomes:

- OS.

Secondary outcomes:

- PFS;
- Cytoreduction to NMRD or near-optimal, defined as residual tumour < 1 cm;
- Death within 30 days of intervention;
- AEs, classified according to CTCAE 2017;(139) and
- QoL, measured using a validated scale.

As outlined in Section 3.2.1.4, there were no concerns about the differing duration of follow-up across studies for survival outcomes inducing small study effects.

3.3.2 Search methods

The search methods closely followed that described in the previous section and described in the full publication in Appendix 4. A range of additional efforts to identify relevant studies were attempted due to the lack of randomised studies in this area. This included labour-intensive grey literature searches and hand searches.

3.3.3 Selection of studies

The selection of studies closely followed the methods described in the previous section and described in the full publication in Appendix 4.

3.3.4 Data extraction

Data extraction closely followed the methods described in the previous section and described in the full publication in Appendix 4.

3.3.5 Risk of bias assessments

The risk of bias (and appraised quality) was independently assessed by at least two reviewers. Cochrane's risk of bias tool for RCTs(143, 144) was not applicable to the risk of bias assessment of studies comparing maximal effort debulking versus standard surgery because no RCTs were identified or included. A priori because non-randomised studies were eligible for inclusion, the risk of bias in non-randomised studies tool (ROBINS-I) was used,(159), as recommended and endorsed by the Cochrane Bias Methods Group and Non-Randomised Studies Methods Group. ROBINS-I rates bias across the following seven domains:

1. confounding;
2. selection of participants into the study;
3. classification of interventions;
4. deviation from intended interventions;
5. missing data;
6. measurement of outcomes; and
7. selection of reported result.

Responses to signalling questions in the ROBINS-I lead to the formulation of domain-specific risk of bias ratings: no information; low; moderate; serious; and critical risk of bias. These then guide the judgement for an overall risk of bias rating. Additional signalling questions to the ones in the ROBINS-I domains were added in accordance with additional criteria to assess applicability. This included an assessment of whether there was comparability between treatment groups. This assessed whether there were any differences between the two groups or if differences had been controlled for with respect to important baseline characteristics. Controlling for too few characteristics would potentially result in an inadequate adjustment but controlling for too many is likely unnecessary and may introduce issues of multicollinearity.(160)

To aid signalling questions regarding selection of women into the study, it was assessed whether relevant details of criteria for assignment of patients to treatments were provided, whether the group of women who received each intervention were representative, and if they were not selected by a subset of the population. Results were presented in a risk of bias summary table, which was used to interpret the results of MAs.

3.3.6 Certainty of the evidence

The certainty of the evidence judgements closely followed the methodology described in Section 3.2.6, where a full assessment was made on the quality and certainty of the evidence for each outcome in each comparison using the GRADE approach.(74, 161) The evidence was downgraded by one level due to an absence of any randomised evidence and then assessed using the same principles as previously outlined.

Summary of findings were created to present the main findings from the following outcomes:

- OS;
- PFS;
- QoL; and
- AEs.

3.3.7 Data synthesis

If sufficient clinically similar studies were available, their results were pooled in pairwise MAs (direct head to head comparisons) in RevMan 5.4(146) using the same methodology as outlined in Section 3.2.7. Studies were required to report adjusted effect estimates, as specified in the review inclusion criteria.

3.3.8 Assessment of heterogeneity

Heterogeneity between studies was assessed using the same methods outlined in Section 3.2.8. Whenever there was evidence of substantial heterogeneity, the possible reasons were investigated and reported.

3.3.9 Assessment of reporting biases

Funnel plots corresponding to meta-analysis of the primary outcome were used to assess the potential for small study effects such as PB,(150) as described in Section 3.2.9.

3.3.10 Handling missing data

The handling of missing data in the maximal effort debulking versus standard surgery for advanced EOC review followed the methods specified in Section 3.2.10.

3.3.11 Subgroup analysis

Subgroup analyses were performed, grouping the studies by:

- reporting of survival (overall and disease specific; progression and disease-free); and
- radicality of procedures in the ultra-radical groups.

Factors such as age, FIGO stage, type of surgery (PDS or IDS), type of surgeon and length of follow-up were considered in the interpretation of any heterogeneity.

3.3.12 Sensitivity analysis

Sensitivity analysis was planned, excluding studies at high risk of bias. A sensitivity analysis that included women with more extensive disease (with carcinomatosis) was conducted.

3.4 Publication 3: Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery

My third publication focuses on the prognostic value of RD in this population and addressed the third objective (see Section 2.5.2). This set out to estimate the impact of the extent of RD on overall and progression-free survival. This built upon and addressed the limitations in the review outlined in Section 3.3. This analysis was completed in PDS and IDS settings, as addressed in the review addressing the first objective in Section 3.2. The full publication can be found in Appendix 5. Many of the SR methods are core methods that remain constant regardless of topic or focus, however given the prognostic focus of this review the methods do vary in places and these differences are outlined in the following subsections.

3.4.1 Participants, interventions, comparisons, and outcomes (PICOs)

3.4.1.1 Types of studies

All relevant studies that assessed RD as a prognostic factor for survival in women with advanced EOC after primary surgery and met the inclusion criteria were included. The wider inclusion criteria of studies followed those outlined in Sections 3.2.1.1 and 3.3.1.1 to accommodate both randomised and non-randomised studies.

In each included study, the adequacy of the adjustment factors used in multivariate Cox models, of the 'adjustment for other prognostic factors' and of the 'statistical analysis and reporting', was assessed using domains of the QUality In Prognosis Studies (QUIPS) tool(130) which is discussed in detail in Section 3.4.5. Therefore, the inclusion criteria enabled studies that were less biases to be included.

3.4.1.2 Types of participants

The inclusion of participants follows those documented in Section 3.2.1.2. Women were separated into two distinct groups: those who received upfront PDS and those who received IDS. These distinct groups were analysed separately.

3.4.1.3 Details of prognostic factor

The surgical intervention for which the resulting prognostic factor was assessed was primary surgery. RD thresholds were applied to both PDS and IDS settings:

- No macroscopic residual disease (NMRD) after primary cytoreduction (RD = 0 cm);
- Near-optimal RD after primary cytoreduction (0 < RD ≤ 1 cm, labelled as 0.1-1 cm); and
- Suboptimal RD after cytoreduction (RD > 1 cm).

The three main RD thresholds described NMRD (RD = 0 cm) as above, but RD 0.1-1 cm was categorised as small volume residual disease (SVRD), and RD > 1 cm was categorised as large volume residual disease (LVRD). SVRD and LVRD were used to aid the non-clinical reader, as a combination of numbers and letters can be more challenging. Furthermore, although clinicians may understand the meaning of optimal and sub-optimal, this may be more judgemental to a lay reader.

3.4.1.4 Types of outcome measures

The following outcomes were assessed:

- OS; and
- PFS.

3.4.2 Search methods

The search methods closely followed those described in previous sections and described in the full publication in Appendix 5.

3.4.3 Selection of studies

The selection of studies closely followed the methods described in the previous section and described in the full publication in Appendix 5.

3.4.4 Data extraction

Data extraction closely followed the methods described in Section 3.2.4 but was extended to incorporate the checklist for critical appraisal and data extraction for SRs for prognostic factor studies (CHARMS-PF).(130) This checklist helped with framing the review question and was examined prior to data extraction and critical appraisal. A thorough data extraction for each study is needed to obtain relevant information for the review and the checklist was used in scoping and to inform piloting of data extraction. CHARMS-PF was used in combination with the QUality In Prognostic Studies (QUIPS) tool discussed in Section 3.4.5.

3.4.5 Risk of bias assessments

Risk of bias in studies assessing RD as a prognostic factor was assessed using the QUIPS tool, as per the protocol.(130) For prognostic factor studies, risk of bias for each outcome (OS and PFS) for each study was assessed separately as risk of bias could differ by outcome. QUIPS assesses bias across the six domains listed in Table 2, using intermediate signalling questions to aid the decision-making process.(130) The applicability of each study for four of the domains is set out in Table 2 below. It was important to outline any concerns regarding applicability to accompany the results of the risk of bias assessment.

Table 2: QUIPS tool for assessment of prognostic factor studies.

Domain	Additional considerations
Study participation	This domain was renamed 'Participant selection' as it was more important to explicitly make judgements about how the participants were enrolled in the study and then selected for their type of surgery. Applicability: Are there concerns that the included women do not match the review question?
Study attrition	No further items were added, or revisions made to this domain.
Prognostic factor measurement	Applicability: Are there concerns that residual disease, the way that it is measured, or the way that it is interpreted differ from the review question?
Outcome measurement	Applicability: Are there concerns that the outcome does not match the review question or that follow-up was not of sufficient duration?
Study confounding	This domain was renamed 'Adjustment for other prognostic factors' to distinguish between confounding and adjusting for other important prognostic factors. Adjustment for other sensible and pertinent prognostic factors is important when assessing the independent prognostic ability of residual disease as a prognostic factor. Applicability: Did the prognostic factors adjusted for match the review question?
Statistical analysis and reporting	No further items were added, or revisions made to this domain.

As an amendment to the standard QUIPS existing risk of bias judgements, namely, low, moderate, and high, a fourth 'unclear' option was added. This was added because inconsistent reporting of included studies could lead to misleading and unreliable judgements. For example, in studies that were only reported in abstract form and information was clearly missing, an 'unclear' option was required. Results were presented in a risk of bias summary table, which was used to interpret the results of MAs.

3.4.6 Certainty of the evidence

Guidance on the use of GRADE for prognostic factor studies has not yet been published.(74, 162) However, the quality and certainty of the evidence was still appraised in the review. The certainty of the evidence judgements closely followed the methodology described for the previous two

intervention reviews, where a full assessment was made on the quality and certainty of the evidence for OS and PFS in each key RD comparison using the GRADE approach.(74, 161) Summary of findings tables were created to present the main findings for the following outcomes:

- OS; and
- PFS.

3.4.7 Data synthesis

If sufficient clinically similar studies were available, their results were pooled in pairwise MAs (using direct head to head comparisons) in RevMan 5.4(146). As the main aim of the review was to identify the prognostic impact of RD thresholds to predict survival outcomes (OS and PFS), only time-to-event data were considered applicable to this review. As such, HRs were pooled using the generic inverse variance (GIV) facility of RevMan 5.(146)

3.4.8 Assessment of heterogeneity

Heterogeneity between studies was assessed using the same methods outlined in Section 3.2.8. Whenever there was evidence of substantial heterogeneity, the possible reasons were investigated and reported.

3.4.9 Assessment of reporting biases

Funnel plots corresponding to MA of the primary outcome were used to assess the potential for small study effects such as PB,(150) as described in Section 3.2.9.

3.4.10 Handling missing data

The handling of missing data followed the methods specified in Section 3.2.10.

3.4.11 Subgroup analysis

Subgroup analyses were performed, grouping the studies by women with FIGO stage III versus stage IV disease. Women undergoing PDS and IDS were analysed in separate analyses. Factors such as age, grade, length of follow-up, type and experience of surgeon and type of surgery were considered in the interpretation of any heterogeneity.

3.4.12 Sensitivity analysis

Sensitivity analyses were planned that restricted the analyses to studies that were judged to be at an overall low risk of bias. Several post hoc sensitivity analyses were also conducted which included omitting one study that included a proportion of women with early and unknown stage disease.

3.5 Publication 4: Residual disease threshold after primary surgical treatment for advanced epithelial ovarian cancer: A systematic review and network meta-analysis

The publication described in this section addresses objective 4a in the thesis and builds on the previous prognostic factor Cochrane review by using a more sophisticated analytical approach. This explored whether a network meta-analysis (NMA), as an extension to a standard pairwise MA explored in the review in Section 3.4, could be used to compare the impact of the different RD thresholds in a single analysis. A NMA can incorporate and synthesise multiple treatments, or in this case RD thresholds, allowing for direct and indirect comparisons between groups that have previously not been compared in published studies.(113, 152, 163) See Appendix 6 for the full text of the publication summarised in this section.

3.5.1 Rationale

The first two reviews/publications on surgical approaches and timing of surgery described in earlier paved the way to explore the impact of RD as a prognostic factor after primary surgery for survival in advanced EOC. The emphasis in the thesis from this point is to demonstrate how the use of more sophisticated methodology may offer an opportunity to better use all the available evidence and raise the certainty of that evidence by increasing the confidence in the effect estimates derived from the synthesis. This in turn will impact on the strength of the conclusions being drawn and increase the potential for impact on policy and practice. Having the most up-to-date and reliable evidence is crucial to the development of clinical guidelines. As such, the fourth publication (Appendix 6), outlined in the following sections is a progression of those pieces of research outlined in earlier sections.

The methods used for the core elements of the review are consistent with those reported for the other reviews included in my body of work. However, due to sparsity of evidence in IDS, this review and the more sophisticated methods have been applied to PDS evidence only.

3.5.2 Network meta-analysis

3.5.2.1 Data analysis

The NMA was conducted using a frequentist framework (using Stata statistical software version 15).(164-166) In the NMA, multi-arm studies were included with an adjustment for potential correlation between arms within these studies. Within the frequentist NMA, this was accomplished using the augmented approach.(166) This approach allowed for the inclusion of RD threshold comparisons of any study size. When studies did not include a reference RD threshold, the estimate was augmented. This involved adding an additional artificial reference RD threshold, which has approximately zero weight in the analysis.(166) The augmented command in Stata in the frequentist NMA was used to compute effect estimates.(166) In the augmented format, all RD thresholds are compared with a reference RD threshold (NMRD (RD = 0 cm) was used as the reference). Within each network, RD thresholds are depicted as nodes with lines representing the comparisons and presented as a network diagram.(167)

3.5.2.2 Consistency and transitivity assumption

The methods adopted were designed to limit heterogeneity in the inclusion criteria and so limit inconsistencies. However, in addition to the visual inspection of clinical and methodological characteristics across studies, a network meta-regression was conducted to explore age, stage of disease and histology to determine the similarity of studies for inclusion in the NMA. A meta-regression has low power and is at risk of confounding,(168, 169) so the emphasis here was on checking summary and descriptive characteristics of studies to see if there were any clear systematic differences between them.

An explanation and details about consistency and transitivity is provided in the associated publication (see Appendix 6). Consistency was assessed by node splitting analysis(166, 170) and a formal global test for inconsistency.(165, 166, 171)

A further way of assessing the validity of the NMA was by ascertaining whether or not the transitivity assumption was met.(151, 165) This was performed by examining characteristics across studies. The underlying assumption required for the results of the NMA to be informative was that there was no imbalance in the distribution of effect modifiers across the different types of direct RD comparisons, regardless of the structure of the evidence network.(172) Despite the low power of meta-regression as alluded to above, it did assist in making judgements on the transitivity assumption, along with summaries of characteristics across studies.

3.5.2.3 Effect estimates and methods for identifying best RD thresholds

The results of the NMA were presented as HRs and 95% confidence intervals (CIs).(165) These data were presented alongside analogous results derived from the pairwise MAs previously described in Section 3.4.7. As already noted, all thresholds were reported relative to the NMRD reference threshold.

The results of a NMA may be complex and difficult to interpret for non-statisticians, so to complement the numerical data graphical tools and ranking statistics can be introduced. Some form of ranking is reported in two thirds of all published NMAs, and experts recommend ranking as a form of presentation.(173) Therefore, the probability of each RD threshold being the 'best' in terms of OS was also reported. This was the probability that a RD threshold is at a specific rank (first, second, up to eighth) when compared with the other RD thresholds in the network. These ranking probabilities are estimated with some variability. Inference based solely on the probability of being ranked as the

best, without accounting for the variability, is not recommended.(151) Therefore, emphasis was on plots showing the relative rank of all RD thresholds in terms of OS, which were derived using rankograms. The rankograms rank RD thresholds in order of the highest probability (ranked 1) to the lowest probability (ranked 9) of maximising OS. These rankograms, as well as the cumulative ranking probabilities described below, summarise the estimated probabilities for all possible ranks and, unlike the probability of being best described above, account for uncertainty in relative ranking.

Cumulative ranking probabilities using the surface under the cumulative ranking curve (SUCRA) were also calculated.(173) SUCRA presents a single value associated with each RD threshold. A value of 100% indicates the RD threshold is certain to be the most effective in the network (top ranked), while 0% indicates it is certain to be the least effective (in bottom rank).

3.5.2.4 Certainty of the evidence

The certainty of the evidence was assessed as previously outlined in the Prognostic Factor Cochrane review but extended to incorporate items recommended in the PRISMA NMA reporting guidelines.(174) I did not use the Confidence in Network Meta-Analysis (CINeMA) web application to assess the certainty of the evidence as this does not yet allow for incorporation of time-to-event data.(175) It also essentially covers items recommended in the PRISMA NMA reporting guidelines(174) and is broadly based on the GRADE approach.(74)

3.6 Publication 5: Residual disease after primary surgery for advanced epithelial ovarian cancer: Expert elicitation exercise to explore opinions about potential impact of publication bias in a systematic review and meta-analysis

The fifth paper (Appendix 7) in my body of work addresses Objective 4b in the thesis. This used expert elicitation as a means of exploring the opinions of clinicians with gynaecological expertise about the potential impact of PB, related to the publications discussed in the previous two sections (which address Objectives 3 and 4a) and how this might be quantified and incorporated into the analyses.

3.6.1 Rationale

PB is discussed in detail in Chapter 2, with Section 2.1 outlining why this is an important component to the body of research in the thesis as well as introducing the concept of the GRADE tool.(77) Given the nature of the evidence base related to Publications 3-4, PB could be hypothesised to lead to a bias in favour of more complete removal of RD as described in the elicitation survey in Appendix 8. Exploration of PB is an important part of a robust SR and should always be considered. At present, there is no consensus on a standard approach for identifying and adjusting for PB, although some methods, particularly around identification, do exist. I attempted to reduce PB throughout my body of work by adherence to good review practice, such as thorough searches of the grey literature,(176-178) adequate handling of missing data(71) and post hoc statistical approaches such as funnel plots.(179) Methods such as trim and fill,(116, 180) and file drawer number(181) could also be used but these are crude and limited in nature. Furthermore, when there is evidence for PB or this bias is highly suspected, selection models(182, 183) might be used to investigate how the results of a meta-analysis may be affected by PB. However, these usually require a large number of included studies in the analysis(184, 185) and any adjustment generally requires an assumption of the underlying selection model.(182, 184) A potentially more practical approach is to incorporate external

information into the MA. This external information could be gathered from various sources and incorporated using a Bayesian framework.(186-188) However, this approach would only be useful if the external information is obtained from a reliable source. Therefore, the publication outlined in this section proposed an approach that has hitherto received little attention in MA, namely the consideration of expert opinion and the incorporation of their views and opinions into the MA to inform the adjustment. The elicitation exercise relates to the conduct of the prognostic factor SR and was designed a priori.

This section outlines the derivation of informative priors from the elicitation exercise to use within a Bayesian framework to attempt to adjust for PB within the MAs reported in the previous two publications. Section 3.7 then describes the methods used to adjust for PB using this framework. The method employed in this component of my research was novel. It comprised the design of an elicitation exercise and the subsequent incorporation of this expert opinion in a manner that allowed the overall interpretation of the findings reported from the NMA analysis outlined in the fourth publication (Appendix 6). This overall aim of this methodology was to raise the confidence in effect estimates in MAs and in the conclusions that could be drawn. The underlying rationale for this work is to try to raise the certainty of the evidence. A Bayesian adjustment for PB is not a new concept but the methods proposed here are original.

Expert elicitation attempts to quantify scientific consensus. In simple terms, expert elicitation uses an "educated guess" from experts and this was used in my study to obtain parameter estimates. The fifth publication in Appendix 7 considered the use of expert opinion and how it could be incorporated into the MA to inform the adjustment. The elicitation exercise was conducted amongst eligible British Gynaecological Cancer Society (BGCS) members (based on a pertinent job title and expertise in gynaecology) and sought to identify expert opinions on the potential nature and extent

of PB in the prognostic factor SRs outlined in the previous two sections. More specifically, the expert elicitation aimed to identify the sort of studies that have been conducted but not published, the plausible magnitude and direction of any PB and possible explanations for why and how the PB occurs.

3.6.2 Expert elicitation exercise

Details of the case study are described in Appendix 7, along with a description of the instructions given to expert participants who took part in the elicitation exercise.

3.6.2.1 Design of elicitation exercise

The specific elicitation exercise completed by the experts is given in Appendix 8, but in summary the elicitation exercise asked respondents their opinions on l different scenarios related to the likelihood of different studies not being published. These hypothetical (unpublished) studies varied in both the size of the study population and the RD thresholds being compared. Usually, expert opinions are elicited either directly using interview methods or via an elicitation exercise. It has been suggested that elicited opinions potentially only need to be provided by as few as four experts,(189, 190) given the controversy surrounding radical surgery I wanted to utilise more experts to capture a broader range of views and provide generalisability.(191, 192) The BGCS provided the vehicle for identifying professional expert clinical members who had the correct knowledge and experience to participate in the elicitation exercise. Utilising this cohort ensured that disagreement amongst the experts (i.e., the respondents) in the management of advanced EOC, would be captured and reflected in the uncertainty of elicited estimates. The elicitation exercise was conducted in three parts.

3.6.2.2 Expert elicitation part A

In Part A, for each comparison of different macroscopic RD thresholds the sample size of each hypothetical study was varied between a minimum sample size (n = 100) and a maximum sample size (based on observed studies across each comparison). Respondents were asked their perspective regarding the chance a study reporting each comparison would be published on a scale of 0 (no chance of publication) to 100 (certainly published).

3.6.2.3 Expert elicitation part B

Part B aimed to obtain the opinion of respondents on the estimated number of conducted-but-unpublished studies that might exist. For each question, participants were asked to consider a particular macroscopic RD threshold (RD < 1 cm; RD > 1 cm; RD > 2 cm) and compare it with RD 0 cm. Next, respondents were asked how likely it was that a study that found no evidence of a statistically significant difference (P > 0.05) in survival between macroscopic disease and NMRD (RD 0 cm) would be published. Estimated likelihood was recorded using a Likert scale from 1 (not likely at all) to 5 (extremely likely).

Respondents were then asked to give an estimate of how many studies of a certain size and magnitude of effect might be unpublished, along with a rationale for their answer. The sample size of unpublished studies was varied in increments of 100 from 100 to > 500. The effect size, reported as the adjusted HR, was likewise varied in decrements of 0.1, between 1 and ≤ 0.5. In total, respondents were asked to think about the number of unpublished studies for 36 different hypothetical combinations of sample sizes and effect sizes. The questions were repeated for scenarios involving suboptimal RD thresholds (> 1 cm and > 2 cm) compared with NMRD.
The responses to the elicitation exercise were summarised using descriptive statistics.

3.6.2.4 Expert elicitation part C

Part C was used to gauge the attitudes of the respondents to reporting biases more generally. These data were used as additional information when formulating the priors used in the Bayesian analyses outlined in Section 3.7 (publication in Appendix 9).

3.7 Publication 6: Residual disease threshold after primary surgical treatment for advanced epithelial ovarian cancer: A network meta-analysis incorporating expert elicitation to adjust for publication bias

This paper uses the results of the expert elicitation exercise and methodology proposed in Section 3.6 to make an adjustment to effect estimates in MAs conducted in the previously described prognostic factor reviews using Bayesian methods. The work reported in this section on PB was used to address the fifth objective; modifying the PB domain in GRADE.(77) This aimed to raise the confidence in the GRADE judgement and consequently the overall recommendations being made. A summary of concepts and statistical methods is presented in this section, with full details provided in Appendix 9.

3.7.1 Rationale

This section describes how the results of the expert elicitation exercise described above were used in the NMA assessing RD as a prognostic factor for OS in advanced EOC. The work evolved statistical methodology presented earlier on the conduct of the NMA by applying an adjustment for PB within the NMA. To assess the potential effects of PB, a modified version of the selection model described in Part A of the elicitation exercise in Section 3.6 was used. This approach extends the Copas selection model for a conventional two-group meta-analysis to the general NMA setting.(193-195) An additional and alternative approach was also used to adjust for PB in the NMA. This approach leverages informative priors in an otherwise conventional Bayesian NMA. The former approach was closely followed but an extension to an existing approach and the second proposed method was completely novel.

3.7.2 Adjustment for publication bias

3.7.2.1 Part A: Copas model approach

Part A of the elicitation exercise asked clinicians about their perceived probability of publication of individual studies. The probability of publication was related to the standard error (SE) of the study effect sizes. Part A was conducted to facilitate an extension to the conduct of a previously proposed method of adjusting for PB in a NMA.(118) This publication included an outline of the measurement model, selection model and prior distributions for model selection parameters; however, it did not describe how to approach the complexities of four-arm trials. Therefore, the exact methodology and code for the analyses in this publication is novel.

3.7.2.2 Part B: Alternative novel approach

Part B of the elicitation involved an alternative approach that asked clinicians to estimate the number of studies for key comparisons they thought would be conducted but unpublished and, thus, unidentified in the NMA I conducted. It went on to ask clinicians to specify sample and effect sizes for each of the studies they believed were missing. This is a particularly novel aspect of my research, as it can be used as prior information to inform adjustment of MAs for PB in a way that, to the best of my knowledge, has been previously unexplored.

Part B of the elicitation exercise required several assumptions starting with an estimate of 5-year survival for advanced EOC. This survival rate can be assumed to be 36%(14, 196, 197) and given the minimum sample size constraint of $n = 100$ to meet the criteria for inclusion in the NMA, a minimum 64 events (deaths, d) would be expected in a study of this size, with 36 participants being alive and censored at the end of the study:

$$d = 100(1 - 0.36) = 64.$$

Generalising this result, it was assumed that d can be related to n in general through the following formula.

$$n = \frac{d}{1 - (5 \text{ year suvival rate})} = \frac{d}{0.64}.$$

The standard error of the log hazard ratio $(\mathrm{SE}(\log HR))$ can then be related to n by rearranging the following formula:

$$d = \frac{4}{\mathrm{SE}(\log HR)^2},$$

$$\Rightarrow \mathrm{SE}(\log HR) = \sqrt{\frac{4}{d}} = \sqrt{\frac{4}{0.64n}} = \sqrt{\frac{6.25}{n}}.$$

Next, m_{cij} can be denoted as the number of missing studies according to expert responder $c = 1, \dots, C$, with a HR of HR_j and a sample size of n_i, where:

$$n_1 = 100, n_2 = 200, n_3 = 300, n_4 = 400, n_5 = 500, n_6 = 625,$$
$$HR_1 = 1, HR_2 = 0.9, HR_3 = 0.8, HR_4 = 0.7, HR_5 = 0.6, HR_6 = 0.5.$$

The average number of missing studies of type ij across the responders is denoted as:

$$m_{ij} = \frac{1}{C}\sum_{c=1}^{C} m_{cij}.$$

This can be used to form an average sample size of missing studies with a HR of HR_j through:

$$m_j = \frac{\sum_i n_i m_{ij}}{\sum_i m_{ij}}.$$

With this, it can be assumed that information from missing studies with a HR of HR_j can be categorised through the following distribution:

$$P_j \sim N\left(\log HR_j, \frac{6.25}{m_j}\right).$$

The P_j can then be combined in a weighted manner, giving more weight to those values of j with a larger value of m_j, via conflation. This gives an elicited prior of:

$$P \sim N\left(\frac{\sum_j \frac{m_j \log HR_j}{6.25}}{\sum_j \frac{m_j}{6.25}}, \frac{1}{\sum_j \frac{m_j}{6.25}}\right) = N\left(\frac{\sum_j m_j \log HR_j}{\sum_j m_j}, \frac{6.25}{\sum_j m_j}\right).$$

This elicited estimate can then be used as prior information and be applied in a Bayesian analysis(186-188) that reflects the results of the expert opinion in the elicitation exercise.

3.7.2.3 Data analysis

Here the methodology for the NMA utilising the expert elicitation exercise to adjust the effect estimates for PB. The NMA was conducted within a Bayesian framework in R statistical software (v4.1.2; R Core Team 2021) when incorporating parameters from part A(198) and WinBUGS 1.4.3 using parameters from part B (MRC Biostatistics Unit, Cambridge, UK),(88, 199) of the elicitation exercise. Two chains each with 100,000 Markov Chain Monte Carlo (MCMC) simulations and with a burn-in period of 30,000 simulations was undertaken for both analyses. The base case Bayesian analysis (analogous to the frequentist analysis) performed in WinBUGS 1.4.3 used vague non-

informative priors and adjusted for multi-arm trials using conditional distributions. A network diagram of the thresholds (nodes) and comparisons (lines) is presented in Section 3.5.(167) In addition, a summary of designs in the network is presented in Appendix 9. Convergence of the model in the two chains was assessed using Brooks-Gelman-Rubin, trace and autocorrelation plots.(199)

Transitivity and design consistency measured in terms of agreement of direct and indirect evidence was previously described in Section 3.5. In the Bayesian setting, consistency was assessed by comparing the individual data point's posterior mean deviance contributions for the consistency and inconsistency model.(170, 171, 200)

The results of the Bayesian NMA of different RD thresholds were presented using effect sizes, reported as posterior median HRs and 95% credible intervals (CrIs). All thresholds are relative to the NMRD reference threshold (RD 0 cm). Rankograms, probability of being best, and SUCRA were also presented, and have previously been explained in Section 3.5.(173)

A number of sensitivity analyses (SA) were reported that attempt to adjust the base case estimates for PB. The *a priori* focus was on NMRD, RD < 1 cm and suboptimal RD > 1 cm. Other thresholds would add strength to the network but were not of direct interest, such as the comparison of suboptimal RD > 2 cm versus NMRD. The base case Bayesian analysis above was repeated and the elicitation exercise was used, as described above, to employ the Copas selection model (Part A) and incorporate informative priors (Part B) in place of the vague (non-informative) ones. To assess consistency the Bayesian NMA results were compared to those obtained from the Frequentist NMA discussed earlier.

All statistical code for the main Bayesian NMAs described in this section is given in Appendix 10.

3.8 Summary

In this chapter, methods have been presented for the work that I conducted and reported across six publications. The body of work includes standard methods that gradually evolve with each subsequent publication, with the aim to present more reliable estimate of effects. The body of work includes the most pertinent forms of risk of bias assessment across the different SRs. The statistical methodology described in this chapter evolved from a standard pairwise meta-analytic approach to the conduct of NMAs, which were extended in a novel manner through the use of expert opinion to make an adjustment for PB using a Bayesian framework; again, aiming to minimise bias and raise the certainty of the evidence.

In the following chapter, the results of these analyses are presented in a way that demonstrates the reduction in uncertainty of the estimates obtained using the existing evidence base and the novel, originality of the body of methods work underpinning my thesis.

4 Results

4.1 Introduction

This chapter provides an overview of the results of the four MAs that provide the findings to address Objectives 1-4a of the thesis:

- MAs for PDS versus IDS (First publication in Appendix 3);
- Maximal effort debulking surgery versus standard surgery for advanced EOC (Second publication in Appendix 4);
- The impact of RD after PDS and IDS and the application of pairwise MAs (Third publication in Appendix 5);
- The impact of RD after PDS and the application of NMAs (Fourth publication in Appendix 6).

4.2 Publication 1: Chemotherapy versus surgery for initial treatment in advanced ovarian epithelial cancer

A summary of the results is presented here; for full details see the full publication in Appendix 3 which was conducted to meet the first objective.

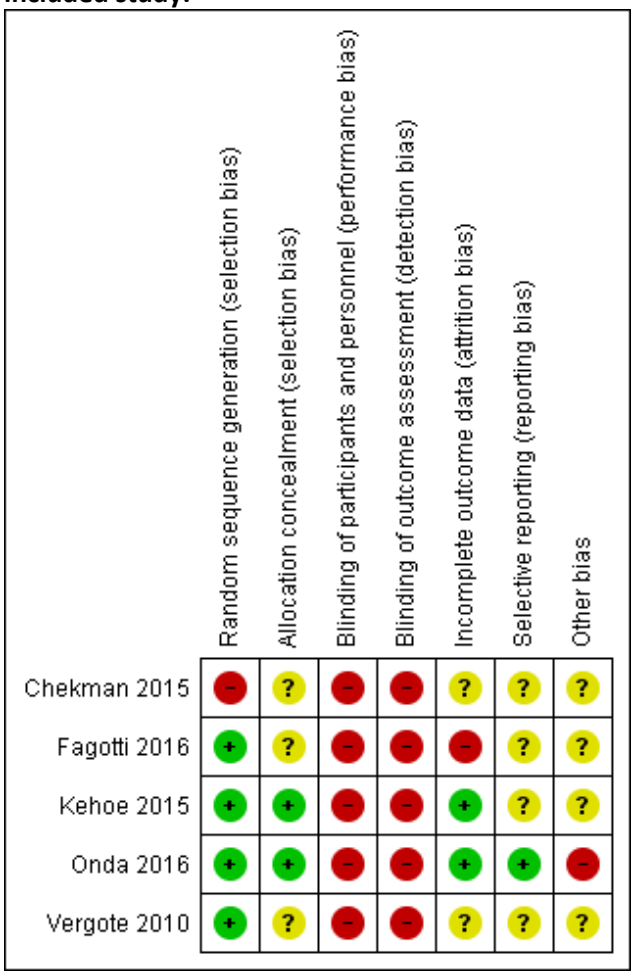
4.2.1 Description of included studies

A total of 23 references reporting on five RCTs met the inclusion criteria.(2) The characteristics of the five included RCTs are summarised in the full publication in Appendix 3.

4.2.2 Risk of bias in included studies

Risk of bias was assessed using the Cochrane Risk of Bias (ROB 1) tool(143, 144) and judgements are depicted in the risk of bias tables and in Figure 3 below. The five included studies were open-label studies and outcome assessment was not blinded. This is not an issue for primary outcomes such as survival, which is the major outcome reported in the thesis, due to the potential for reasonable 5-year survival rates in women with the disease. However, it may lead to detection bias with regard to other outcomes or subgroups (e.g., extent of debulking achieved). The importance of blinding of outcome assessment in OC studies had been raised in a Gynecologic Cancer Inter Group consensus statement.(26) Data for such outcomes were interpreted with caution.

Figure 3: Risk of bias summary graph showing judgements about each risk of bias item for each included study.



4.2.3 Effects of interventions

Analyses were interpreted in terms of the certainty of the evidence using the GRADE approach.(77).

This section presents a summary of the main results, with full results available in the publication in Appendix 3.

Overall survival (OS)

The primary MA of four studies, assessing 1692 participants, demonstrated little or no difference in OS between NACT and PDS for initial treatment in advanced OC (HR = 0.96, 95% CI 0.86 to 1.08; I^2 = 0%; high-certainty evidence). The results were robust when trials were subgrouped by age (< 50, 50-70 and 70+ years), extent of RD (up to 0.5 mm, 0.5-1 cm, > 1 cm) and FIGO stage (III and IV).

Progression-free survival (PFS)

MA of four studies, assessing 1692 participants, found there is probably little or no difference in risk of disease progression between NACT and PDS for initial treatment in advanced OC (HR = 0.98, 95% CI 0.88 to 1.08; I^2 = 0%; moderate-certainty evidence).

Severe adverse effects (SAEs) and surgical morbidity

AEs and surgical morbidity outcomes were variable and incompletely reported across studies. Findings show that there are likely clinically meaningful differences in favour of NACT and IDS compared to upfront PDS with regard to overall postoperative serious adverse effects (SAE grade 3+): 6% in NACT and IDS group, versus 29% in PDS group (RR = 0.22, 95% CI 0.13 to 0.38; participants = 435; studies = 2; I^2 = 0%; moderate-certainty evidence).

NACT and IDS likely result in a large reduction in the need for stoma formation: 5.9% in NACT and IDS group, versus 20.4% in PDS group (RR = 0.29, 95% CI 0.12 to 0.74; participants = 632; studies = 2; I^2 = 70%; moderate-certainty evidence).

The risk of needing bowel resection at the time of surgery is likely reduced: 13.0% in NACT and IDS group versus 26.6% in PDS group (RR = 0.49, 95% CI 0.30 to 0.79; participants = 1565; studies = 4; I^2 = 79%; moderate-certainty evidence).

NACT and IDS reduces postoperative mortality: 0.6% in NACT and IDS group, versus 3.6% in PDS group, (RR = 0.16, 95% CI 0.06 to 0.46; participants = 1623; studies = 5; I^2 = 0%; high-certainty evidence).

Quality of life (QoL)

QoL on the European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) scale produced inconsistent and imprecise results at six months in three studies (MD = -0.29, 95% CI -2.77 to 2.20; participants = 524; studies = 3; I^2 = 81%; very low-certainty evidence) but the evidence is very uncertain and should be interpreted with caution.

Other outcomes

Other outcomes reported include extent of RD, duration of operation and length of stay following surgery.

4.3 Publication 2: Maximal effort debulking (ultraradical) surgery versus standard surgery for the primary cytoreduction of advanced epithelial ovarian cancer

A summary of the results is presented here; for full details see the publication in Appendix 4 which was conducted to meet the second objective.

4.3.1 Description of included studies

A total of four references reporting three studies met the inclusion criteria. Searches of the grey literature did not identify any additional relevant studies. The characteristics of the three non-randomised studies are summarised in the full publication.

4.3.2 Risk of bias in included studies

No randomised studies were identified, so the ROBINS-I tool was used to assess bias in the included studies.(159) Judgements are depicted in the risk of bias tables in in Table 3 below.

Table 3: Risk of bias summary showing judgements about each ROBINS-I risk of bias domain for each included study.

Author	Confounding	Selection bias	Classification of interventions	Deviation	Missing data	Measuring outcomes	Reporting bias
Aletti et al(201)	Critical	Low	Low	Unclear	Moderate /high	Critical	Critical
Chang et al(202)	Critical	Low	Low	Unclear	Moderate /high	Critical	Unclear
Luyckx et al(203)	Critical	Low	Low	Unclear	Moderate /high	Critical	Unclear

4.3.3 Effects of interventions

Analyses were interpreted in terms of the certainty of the evidence using the GRADE approach.(74) Very low certainty evidence was identified for all outcomes reported, mainly due to relatively few women being included due to stringent inclusion criteria in the SR. A breakdown of AEs was not

adequately reported in two studies and QoL was not reported in any of the three included studies. Only two of the three included studies were included in MAs.

4.3.3.1 Survival (Overall and disease specific)

Upfront primary debulking surgery (PDS)
MA of two studies, assessing 397 participants, found that women who underwent radical procedures (maximal effort debulking) as part of PDS had 40% less chance of mortality compared to women who underwent standard surgery (adjusted HR = 0.60, 95% CI 0.43 to 0.82, I^2 = 0%), but the certainty of the evidence was very low. One of these studies reported the 5-year disease specific survival rather than categorising deaths by any cause.

The results were robust to sensitivity analysis, which assessed 283 participants with more extensive disease (carcinomatosis). This sensitivity analysis found that women who underwent radical procedures as part of PDS had 39% less chance of mortality compared to women who underwent standard surgery (adjusted HR = 0.61, 95% CI 0.44 to 0.85, I^2 = 0%), but the certainty of the evidence was very low.

4.3.3.2 Progression-free survival

Upfront primary debulking surgery
One study, which assessed 203 participants, found that women who underwent radical procedures as part of PDS had nearly 40% less chance of disease progression or death compared to women who underwent standard surgery (adjusted HR = 0.62, 95% CI 0.42 to 0.92), but the certainty of the evidence was very low.

The results were robust to a sensitivity analysis assessing a subset of 139 women with carcinomatosis, which found that women who underwent radical procedures as part of PDS had nearly 50% less chance of disease progression or death compared to women who underwent standard surgery (adjusted HR = 0.52, 95% CI 0.33 to 0.82), but the certainty of the evidence was very low.

4.3.3.3 Disease-free survival (DFS)

Upfront primary debulking and interval debulking surgery
One study, which included only women with stage IIIC disease, reported disease-free survival for a comparison of radical versus standard surgical procedures associated with both primary upfront and interval debulking surgical procedures. A combined analysis in this study (allowing for a pooled estimate), assessing 527 women, found that those who underwent radical procedures were associated with significantly increased chance of disease progression or death than those who received standard surgery (adjusted HR = 1.60, 95% CI 1.11 to 2.31, I^2 = 0%), but the certainty of the evidence was very low.

4.3.3.4 Perioperative mortality (death within 30 days of surgery)

Upfront primary debulking surgery
None of the studies reporting this outcome used any statistical adjustment.

In total, there were only four deaths within 30 days of surgery in two studies that reported this outcome and none in the maximal effort debulking surgery group, so RR was not reported as to not provide potentially misleading results with so few deaths (very low certainty evidence).

4.3.3.5 Adverse events

Upfront primary debulking surgery

One included study reported adverse events but did not use any statistical adjustment.

In this study, significant postoperative morbidity occurred in 32/84 (38.1%) versus 14/119 (11.8%) women (RR 3.24; 95% CI 1.84 to 5.68) in maximal effort debulking and standard surgery respectively, but the certainty of the evidence was very low.

Women who underwent maximal effort debulking surgery had significantly larger median estimated blood loss (800 vs. 500ml, $p = 0.03$), were more likely to receive an intra- or post-operative blood transfusion (Intraoperative: 25% vs. 17.6%; Postoperative: 39.3% vs. 26.1%; $p = 0.01$), had longer median days in the intensive care unit (1.5 vs. 0.8; $p < 0.01$), and were more likely to experience postoperative morbidity (38% vs. 11.8%; $p < 0.01$) than those who underwent standard surgery. The certainty of the evidence was low.

4.3.3.6 Operative time

Upfront primary debulking surgery

One included study did not use any statistical adjustment for operative times between groups but was the only study to report this outcome.

In this study, women who underwent maximal effort debulking surgery had significantly longer median operative time than those who had standard surgery (307 vs. 235 minutes; $p < 0.01$).

4.4 Publication 3: Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery

A summary of the results is presented here; for full details see the publication in Appendix 5 which was conducted to meet the third objective.

4.4.1 Description of included studies

A total of sixty-seven references reporting on 45 unique studies (with one study reporting two separate analyses) met the inclusion criteria. Searches of the grey literature did not identify any additional relevant studies.

4.4.2 Risk of bias in included studies

Risk of bias was assessed at the outcome level for OS and PFS for each study using the QUIPS tool.(130) Most studies reported OS in both PDS and IDS settings. The detailed assessments are depicted in Figure 4 below for OS. The assessment for PFS is provided in the third publication in Appendix 5. Most studies included in the review were judged as being at an overall 'moderate' risk of bias as they satisfied some but not all of the domains using the QUIPS tool.

Figure 4: Risk of bias summary graph showing judgements about each QUIPS risk of bias domain for overall survival in each included study reporting PDS.

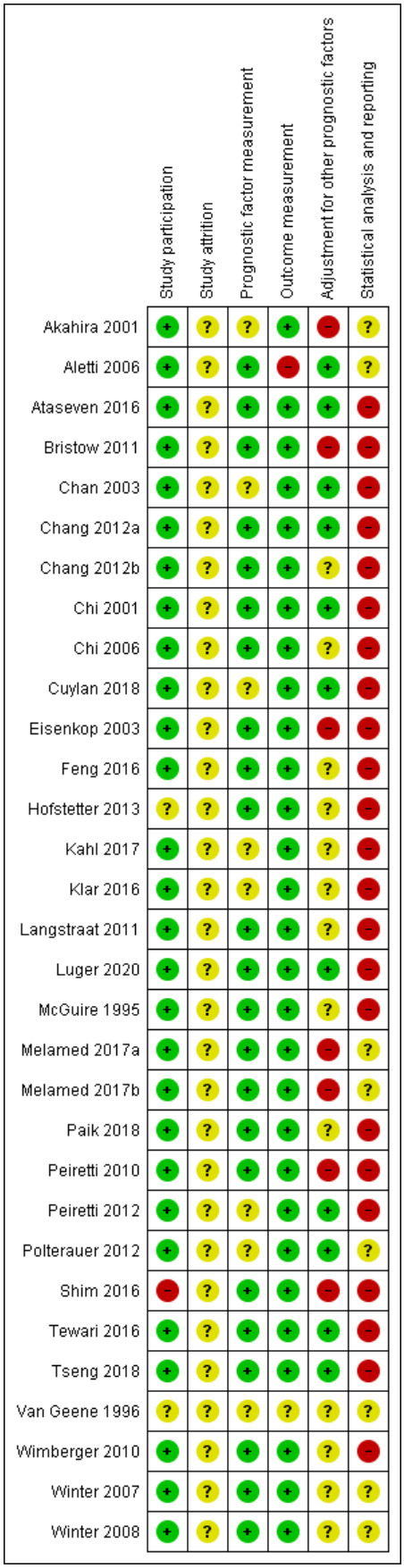
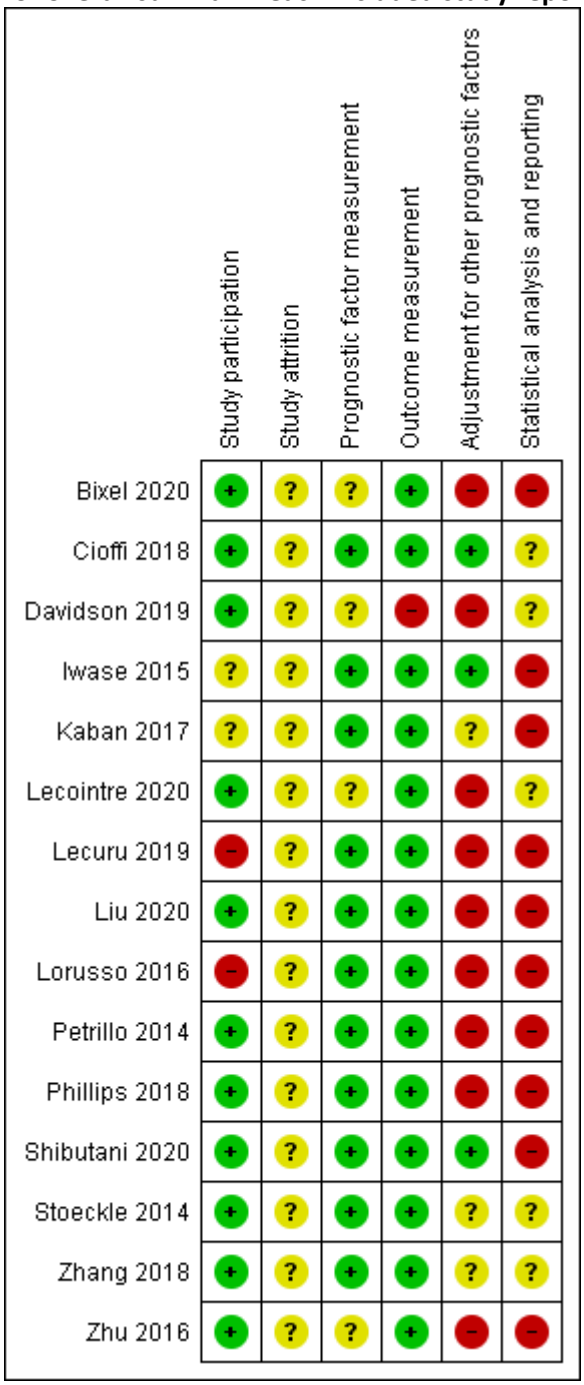


Figure 5: Risk of bias summary graph showing judgements about each QUIPS risk of bias domain for overall survival in each included study reporting IDS.



4.4.3 Findings

Forty-five studies reporting on 46 unique multivariate prognostic analyses (referred to as 46 studies for ease of reporting) including RD as a prognostic factor met the inclusion criteria. The review

included 22,376 women who underwent PDS and 3697 who underwent IDS, all with varying levels of RD.

While a range of different RD thresholds were identified, only the most pertinent comparisons are summarised in this section as these are the main focus of clinical uncertainty. These comparisons involved NMRD (RD=0cm), RD < 1 cm (categorised as 'near-optimal') and RD > 1 cm (categorised as 'suboptimal'). The comparison involving any remaining macroscopic disease (RD > 0 cm) and NMRD was also an important comparison.

4.4.3.1 Upfront primary debulking surgery setting

Most PDS studies showed an increased risk of death in all RD groups when those with macroscopic RD were compared with NMRD.

RD < 1 cm versus NMRD

Overall survival

MA of 17 studies assessing 9404 women found that those who were near-optimally debulked (0 < RD ≤ 1 cm, labelled as 0.1-1 cm) after PDS had more than twice the risk of death compared to women with NMRD (HR = 2.03, 95% CI 1.80 to 2.29; I^2 = 50%). The certainty of the evidence was moderate.

Progression-free survival

MA of 10 studies assessing 6596 women found that those who were near-optimally debulked (RD 0.1-1 cm) after PDS had nearly twice the risk of disease progression or death compared to women with NMRD (HR = 1.88, 95% CI 1.63 to 2.16; I^2 = 63%). The certainty of the evidence was moderate.

RD > 0 cm versus NMRD

The certainty of the evidence was not assessed for this outcome as it was not specified a priori to report in a summary of findings table.

Overall survival

MA of four studies assessing 1220 participants found that women who had RD greater than 0 cm after PDS were associated with a two-fold increase in the risk of death compared to women with NMRD (HR = 1.96, 95% CI 1.44 to 2.67, I^2 = 49%).

Progression-free survival

MA of three studies assessing 1029 participants found that women who had RD greater than 0cm after PDS had more than one and a half times the risk of disease progression or death compared to women with NMRD (HR = 1.60, 95% CI 1.36 to 1.89; I^2 = 0%).

RD > 1 cm versus NMRD

Overall survival

MA of 14 studies assessing 7988 participants found that women with suboptimal RD > 1 cm after PDS was associated with two and a half times the risk of death compared to women with NMRD (HR = 2.50, 95% CI 2.13 to 2.94; I^2 = 63%). The certainty of the evidence was moderate.

Progression-free survival

MA of six studies assessing 2629 participants found that women with suboptimal RD > 1 cm after PDS had more than twice the risk of disease progression or death compared to women with NMRD (HR = 2.10, 95% CI 1.84 to 2.40; I^2 = 24%). The certainty of the evidence was moderate.

RD > 1 cm versus RD < 1 cm

Overall survival
When suboptimal (> 1 cm) versus near-optimal (RD < 1 cm defined as 0.1-1 cm) cytoreduction was compared in a MA of five studies including 6000 women, the estimates were attenuated compared to those with NMRD. All analyses showed a survival benefit in women who had been near-optimally debulked (HR = 1.22, 95% CI 1.13 to 1.32, I^2 = 0%). The certainty of the evidence was moderate.

Progression-free survival
MA of two studies assessing 3402 participants found that women with suboptimal RD > 1 cm after PDS had a greater risk of disease progression or death compared to women with near-optimal RD < 1 cm (HR = 1.30, 95% CI 1.08 to 1.56; I^2 = 53%). The certainty of the evidence was low.

4.4.3.2 Interval debulking surgery setting

RD < 1 cm versus NMRD

Overall survival
One study that included 310 women and appropriately defined the categories as NMRD, RD=0.1-1 cm and RD > 1 cm showed that women who were near-optimally (RD=0.1-1 cm) debulked after IDS had more than twice the risk of death compared to women who had NMRD (HR = 2.09, 95% CI 1.20 to 3.66, I^2 = 56%). The certainty of the evidence was very low.

Progression-free survival
MA of two studies assessing 248 women found no difference in disease progression or death in women with near-optimal RD < 1 cm after IDS and those with NMRD (HR = 3.03, 95% CI 0.81 to 11.38, I^2 = 94%). The certainty of the evidence was very low.

RD > 0 cm versus NMRD

The comparison involving any remaining macroscopic disease (RD > 0 cm) and NMRD in an IDS setting was also an important comparison, so this was additionally given a certainty of the evidence judgement.

Overall survival

A MA of four studies that included 906 women found that those who had any amount of visible RD after IDS had more than twice the risk of death compared to women with NMRD (HR 2.11, 95% CI 1.35 to 3.29, I^2 = 81%). The certainty of the evidence was very low.

Progression-free survival

One study assessing 471 women found that RD > 0 cm after IDS was associated with an increased risk of disease progression or death compared those in whom NMRD was achieved (HR = 1.36, 95% CI 1.05 to 1.76). The certainty of the evidence was very low.

The authors of a different study found that the risk of disease progression for women with RD > 0 cm after IDS was higher than those with NMRD (n = 163, P < 0.01) but the magnitude of effect was not reported.

RD > 1 cm versus NMRD

Overall survival

One study that included a allowed a combined analysis of 343 women showed that those who were suboptimally (RD > 1 cm) debulked after IDS had more than twice the risk of death compared to women who had NMRD (HR = 2.23, 95% CI 1.49 to 3.34, I^2 = 35%). The certainty of the evidence was very low.

Progression-free survival
PFS was not reported for this comparison.

RD > 1 cm versus RD < 1 cm

Overall survival
A MA including six studies that assessed 1572 women found that those who were suboptimally debulked (RD > 1 cm) had a significantly greater risk of death compared to women with RD < 1cm (but inclusive of NMRD) (HR = 1.60, 95% CI 1.21 to 2.11, I^2 = 58%). The certainty of the evidence was very low. However, this result is biased as it was not possible to distinguish NMRD within the < 1 cm thresholds in all but one study. Only one study separated NMRD from RD=0.1-1 cm. All other studies included NMRD in the RD=0.1-1 cm group, resulting in serious bias. Inclusion of NMRD in the RD=0.1-1 cm category is a clear reporting bias when comparing suboptimal RD, making comparisons challenging.

Progression-free survival
MA of four studies assessing 1145 women found that achieving suboptimal RD > 1 cm after IDS was associated with a greater risk of disease progression compared to women in whom RD < 1 cm was achieved after surgery (HR = 1.76, 95% CI 1.23 to 2.52, I^2 = 60%). The certainty of the evidence was low. These four studies included NMRD in the RD < 1 cm category so was prone to the same serious reporting bias as alluded to above.

4.5 Publication 4: Residual disease threshold after primary surgical treatment for advanced epithelial ovarian cancer: A systematic review and network meta-analysis

A summary of the results is presented here; for full details see the fourth publication in Appendix 6 which was conducted to meet part a) of the fourth objective. This involved conducting a NMA as an extension beyond standard pairwise analyses for reasons described in Section 3.5.1.

4.5.1 Description of included studies

The results of the search were the same as outlined for the prognostic factor reviews presented previously but restricted to studies assessing PDS. Fifty-two references reporting on 30 unique studies included RD as a prognostic factor after PDS. Five of these studies were excluded from the NMA because the studies did not report a magnitude of effect, leaving 46 references reporting on 25 primary studies that met the inclusion criteria for the NMA. The 25 included studies assessed a total of 20,927 women, with the majority having stage III disease. The risk of bias in included studies was the same as outlined in Section 4.4.2 but restricted to studies assessing PDS.

4.5.2 Network meta-analysis

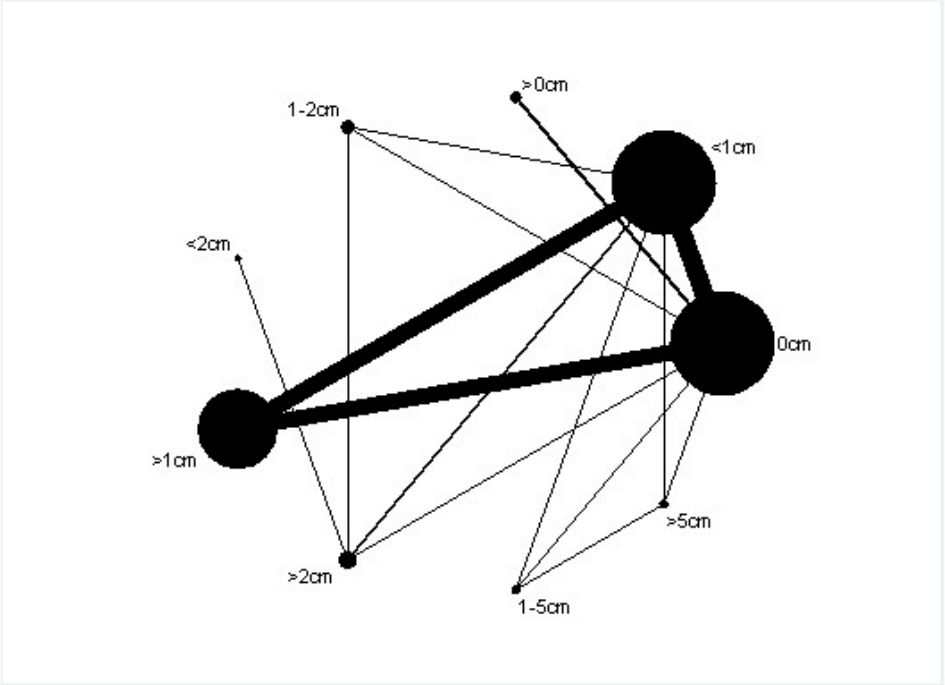
The results of the NMA are an extension to the pairwise MAs presented in the prognostic review. The following sections provide a summary of the main results.

4.5.2.1 Network

The network includes the same RD thresholds as previously outlined for the prognostic review, with one additional comparison which was created as a result of conducting the NMA (RD < 2 cm versus RD 0cm). While this comparison is not of great importance, it is the additional power of the indirect

comparisons across the entire network that offers the real benefit of the NMA. Further details about the network and the RD thresholds presented in the NMA are given in the network diagram in Figure 6.

Figure 6: Network diagram showing residual disease comparisons after primary debulking surgery for advanced epithelial ovarian cancer.



4.5.2.2 Results

The fourth publication in Appendix 6 presents the results of the NMA, which shows a comparison of direct and indirect effect sizes of different RD thresholds. The results appear consistent across all split RD comparisons (sides) and there was no evidence of inconsistency in the network (P = 0.48).

Table 4 shows the results of the NMA demonstrate prolonged survival if primary cytoreductive surgery reduced the tumour to NMRD (RD=0 cm) compared to any other RD threshold. Complete primary cytoreduction to NMRD was overwhelmingly the best ranked threshold, as it was consistently ranked first with a very high probability of being the best RD threshold (SUCRA and P-best of 99.9% and 99%, respectively). as previously mentioned, P-best is estimated with some

variability, so inference was based on the entirety of the evidence, which included SUCRA values, which are more appropriate.

Table 4: Results of NMA and pairwise analyses of different RD threshold after primary debulking surgery for advanced EOC

RD threshold vs. 0 cm (reference)	NMA	Pairwise		Mean rank	P (best) %	SUCRA %
	HR (95% CI)	HR (95% CI)	n studies (participants)			
0 cm	Reference			1	99	99.9
<1 cm	1.98 (1.76 to 2.24)	2.03 (1.80 to 2.29)	17 (9404)	3.4	0	70.2
>0 cm	1.95 (1.48 to 2.58)	1.96 (1.44 to 2.67)	4 (1220)	3.4	0	70.6
1-2 cm	3.34 (2.04 to 5.47)	3.95 (1.33 to 11.78)	1 (68)	7.3	0	21.8
<2 cm	2.82 (1.58 to 5.04)	No direct estimate		6.0	0	36.9
>1 cm	2.57 (2.26 to 2.93)	2.50 (2.13 to 2.94)	14 (7988)	5.8	0	40.0
>2 cm	4.36 (2.69 to 7.04)	8.24 (2.68 to 25.33)	1 (87)	8.7	0	3.4
1-5 cm	1.85 (1.11 to 3.08)	1.83 (1.14 to 2.94)	1 (193)	3.2	1	72.0
>5 cm	2.75 (1.62 to 4.67)	2.72 (1.65 to 4.47)	1 (118)	6.2	0	35.3

NMA: network meta-analysis; RD: residual disease; EOC: epithelial ovarian cancer; HR: hazard ratio; CI: confidence interval; P (best): probability that RD threshold is the best; SUCRA: surface under the cumulative ranking curve

4.5.2.3 Sensitivity analysis

A sensitivity analysis incorporating the results of eight studies which adequately adjusted for extent of disease at primary surgery saw an increase in the magnitude of effect estimates. It showed significantly prolonged survival in those with cytoreduction to NMRD. The results of this sensitivity analysis also appear to be consistent across all sides in the network and there was no evidence of overall inconsistency (P = 0.31). Other key probability and ranking statistics continued to provide

strong evidence that RD 0 cm is the best threshold (P-best = 99.4%) and the SUCRA value remained very high (99.9%). Adjustment for extent of disease included: type (aggressive versus standard) and extent of surgery; surgical complexity score; and progressively extensive tumour involvement in anatomic regions.

4.6 Summary

This chapter presented the results from my body of work published across publications 1 to 4. Thus far the methods described and used to achieve these findings have enabled me to address objectives 1 to 4a of my thesis (see Section 2.5.2).

In summary, high to moderate-certainty evidence in my first review suggested there is little or no difference in primary survival outcomes between PDS and IDS, with IDS potentially being better for some risks of adverse events. Following this, my second review found only very low certainty evidence for the comparison of maximal effort debulking (involving more radical procedures) and standard (radical) surgery in women with advanced EOC. The evidence was limited to retrospective, non-randomised studies and so is at high risk of bias. A third review explored the categorisation of the prognosis of varying RD thresholds. In a PDS setting, it was determined with moderate certainty evidence that there were three distinct categories for RD for survival outcomes including NMRD (RD=0 cm), RD=0.1-1 cm (near-optimal) and RD > 1 cm (sub-optimal).

The conclusions of the first two publications, which examined the timing and radicality of primary surgery for advanced EOC, prompted the investigation of RD after primary surgery as a prognostic factor for survival. Whether surgery is given before or after chemotherapy remains an area of controversy in the gynaecological oncology community and there is an absence of good quality evidence on maximal effort debulking surgery. Consequently, it seemed natural to explore the impact of removing as much tumour as possible after primary surgery in a prognostic setting. The results could then be used as a proxy to address the impact of maximal effort debulking surgery and potentially reduce the level of clinical equipoise and polarised views in this area.

Given review one found little difference in survival between PDS and IDS and review two assessing radicality of primary surgery, found very low certainty evidence for all outcomes, the results of the prognostic review of RD thresholds and the subsequent NMA are likely to resonate with gynae-oncologists, as management is likely to be individualised to each patient. It may mean that, if debulking to NMRD is not possible, they may still attempt a more radical approach to surgery if near-optimal cytoreduction is plausible. Otherwise, with suboptimal cytoreduction being defined as RD > 1 cm, they may place emphasis on minimising morbidity and focus on QoL.

After IDS, there may be only two categories, although this is based on very low certainty evidence, as all but one study included NMRD in the RD=0.1-1 cm category. Due to such sparse data and very low certainty evidence in the IDS setting, the focus of subsequent publications was on PDS. The NMA presented in the fourth publication, consolidates the results shown in the first prognostic factor review and pair-wise MA (Publication 3). The more sophisticated method of NMA, used in Publication 4, has increased statistical power providing results that gave more confidence in the certainty of the overall evidence. While there was very little difference between the NMA and pairwise MA results, the NMA increases the precision without introducing any further bias due to the direct and indirect evidence being very similar in terms of their effect estimates. Therefore, the NMA adds value in strengthening the confidence in the original conclusions found in the MA reported in Publication 3, as well as adding in an additional comparison that was not estimable in the pairwise analyses due to the lack of direct evidence.

The next chapter presents the results of the elicitation exercise and the subsequent use of the results to enable adjustment for PB.

5 Raising the certainty of the evidence

5.1 Introduction

This chapter outlines a further method that can be incorporated into the current PB GRADE domain(77) or as a separate stand-alone item to ensure the current GRADE methodology is more robust. More specifically, this extension to GRADE offered an opportunity to improve the confidence in the effect estimates being presented and strengthen the certainty of the evidence judgements by making a comprehensive assessment and adjustments for PB.

The certainty of the evidence was assessed using a standard GRADE approach(74) (and additionally by incorporating additional items from PRISMA reporting guidelines for NMAs where applicable) in all included SR publications. The expert elicitation exercise described in the publication in Appendix 7 was designed to estimate parameters to adjust for PB in a Bayesian NMA framework (publication in Appendix 9). Since these estimates are part of a series of sensitivity analyses, it would not be appropriate to present them as outright analyses. Instead, these can be used as an additional component to GRADE. If the results are robust to the primary analysis, then it will strengthen the original certainty of evidence judgement. Conversely, it could also create further uncertainty and the original GRADE judgement could potentially be downgraded by a level, but would still present a more reliable judgement.

5.2 Publication 5: Residual disease after primary surgery for advanced epithelial ovarian cancer: Expert elicitation exercise to explore opinions about potential impact of publication bias in a systematic review and meta-analysis

This section describes a publication that reports the results of an expert elicitation exercise that explored opinions on potential PB related to the two prognostic reviews presented as part of my body of research. This elicitation exercise was designed to meet Objective 4b. Experts then completed the exercise so that levels of PB could be ascertained and then quantified. A summary of the key components of the publication and a summary of the results are given in the sections below. Full details are available in the associated publication (See Appendix 7).

5.2.1 Expert elicitation exercise

5.2.1.1 Characteristics of respondents

The elicitation exercise was sent to all 455 BGCS members, and it was estimated that around 80% would be eligible to complete based on the breakdown of membership. A total of 98 BGCS members opened the link for the exercise and 28 proceeded past the participant information sheet. Of these, 18 respondents fully completed the elicitation exercise, with the remaining 10 respondents not adequately contributing to the exercise to be included in the analysis.

Most responders were consultant gynaecological oncologists (11/18; 61%) or sub-specialist consultants (4/18; 22%). The median time to complete the exercise was 18 minutes (inter-quartile range (IQR) 16 to 27 minutes), with a range of 8 to 61 minutes. The mean completion time was 23 minutes (standard deviation (SD) 14 minutes).

While all of the BGCS members were experts in gynaecology, I did not explicitly ask about their level of research experience. However, the majority of respondents (11/18 (61%)) chose to waive their anonymity, and for those whose identity was known their biographies, publication outputs and other research credentials were reviewed. From this it was clear that respondents were highly experienced in research. Nevertheless, if the elicitation exercise was to be repeated in the future (or the same methods used in another area), it would be useful to request this information more formally. This would allow an exploration of the impact of research experience on responses. Although the elicitation exercise was restricted to practicing clinical experts in the UK, it would seem reasonable to assume that similar views would have been observed in a cohort of purposively selected international experts with knowledge experience of research in this field. This is because the topic under investigation, advanced EOC, has international relevance and the evidence base, as identified by the SR, is the same for all.

5.2.1.2 Expert elicitation part A

The specific expert elicitation exercise showing Parts A, B and C is reported in Appendix 8. Responses in Part A of the elicitation exercise suggested that PB may be quite likely in studies where the sample size was just 100. However, there was generally quite widespread variation in the responses, indicating that some responders thought the probability of publication was much higher than others (e.g., a range of 0-100%). Responders appeared to indicate that the probability of publication was lowest for comparisons involving greater macroscopic disease volume.

Responders did not view PB to be a major threat for comparisons of RD < 1 cm versus NMRD and RD > 1 cm versus NMRD, which were the main comparisons of interest. Mean and median probabilities of publication were higher and close to 100%, indicating that respondents were highly certain that a study would be published. Comparisons involving higher volume suboptimal RD (greater

macroscopic disease volume) versus NMRD were considered to have a low probability of being published for larger studies, which was consistent with the results for smaller studies.

5.2.1.3 Expert elicitation part B

Most responders acknowledged that the likelihood of PB is 'somewhat' or 'quite' likely (72.5%) in the comparison of RD < 1 cm with NMRD, with only one responder (5.5%) thinking it was not likely at all. These results are in line with the set of studies, which reported effect estimates (or significance) for overall survival in univariate analyses in favour of NMRD but did not report in multivariable analyses.(204-211) This may or may not have been due to a lack of significance. Additionally, further studies that reported no statistically significant difference between NMRD and RD <1cm might not be published. The view of respondents was completely reversed for comparisons involving suboptimal RD > 1 cm with NMRD, where most responders thought PB was 'not likely at all'. This finding is consistent with what would have been expected and justifies the focus on comparisons of lower volume disease.

The mean and median number of missing studies estimated by responders for comparison of RD < 1 cm versus NMRD was 17 (SD 16.5) and 10 (IQR 5-20), respectively. The average number of estimated missing studies was lower for the comparisons involving suboptimal RD thresholds that are > 1 cm.

The respondents seemed to suggest that the number of studies that might be missing may be influenced by the effect size those studies detected. For example, for the comparison of RD < 1 cm versus NMRD, on average 9.4 of the 17 studies would be associated with a HR of 1. As the HR increased, fewer studies were felt to be missing such that, when the detected HR was 0.5, the average number of studies felt to be missing was 0.83 (95% CI 0.77 to 0.90) for the comparison of RD < 1 cm compared with NMRD.

5.2.1.4 Expert elicitation part C

The results of Part C of the elicitation exercise were not provided in the fifth publication in Appendix 7 but are presented in Table 5 below. This part of the work was not reported in the paper due to journal word limit constraints. Some questions in Part C were used as a pseudo verification for the responses given in Parts A and B of the elicitation exercise. Responders clearly thought that many study authors only report univariate analyses to maximise the magnitude of effect estimates, as indicated by average (mean and median) scores being around 70 (range 18-100, where 100 represents a total belief that univariate analyses are reported to maximise statistical significance).

On average, respondents thought that four or five journal submissions should be made to publish the results of any study they were involved in (mean 4.7, SD 2.2). Respondents did not think that the number of attempts to get a paper published would differ if the results were not statistically significant ($p > 0.05$; mean 4.4, SD 2.5). Most respondents (83%) would consider submission of their study to a journal with a low impact factor (as low as 1 or less), regardless of the significance of their results. Three respondents (17%) would not accept an impact factor less than 6 (Table 5). This suggests that some of the clinical experts are complicit in PB, and study results potentially not being disseminated.

Table 5: Survey responders views on submission and publication of studies

Question	n	Mean (SD)	Median (IQR)	Observed range
To what extent do you think that the reason study authors only report univariate analyses is to maximise the magnitude in effect estimates to favour either an experimental or comparator group? (0-100)[a]	18	68.5 (23.2)	70.5 (51-82)	18-100
How many attempted submissions should you make to journals to publish the results of your study? (1-10)	18	4.7 (2.2)	4 (3-5)	2-10
How many attempted submissions should you make to journals to publish the results of your study if it is not statistically sig. (p > 0.05)? (1-10)	18	4.4 (2.5)	4 (2-5)	1-10
What is lowest impact factor in a journal that you would consider submission of your work, regardless of the significance of your results? n (%)	< 1: 3 (17) 1-5: 12 (67) 6-10: 2 (11) 11-14: 1 (5)			< 1-14
To what extent do you think it is important to publish results of a study even if the impact factor of the accepting journal is perceived to be very low? (0-100)	18	69.6 (17.5)	70 (60-81)	30-90

n: Number of expert respondents; SD: Standard deviation; IQR: Inter-quartile range

[a]100 represents a total belief or to greatest extent and 0 represents the reverse (i.e., no belief at all)

5.3 Publication 6: Residual disease threshold after primary surgical treatment for advanced epithelial ovarian cancer: A network meta-analysis incorporating expert elicitation to adjust for publication bias

Utilising adequate methods of bias correction can add confidence to certainty of effect estimates in a MA as previously outlined in Section 2.1. The research in this section was used to meet Objective 5 and essentially had two main aims. Firstly, to compare the results of a Bayesian NMA using a non-informative prior(212) with the frequentist one reported in the fourth publication in Appendix 6. Secondary, a series of sensitivity analyses were conducted using NMAs that adjusted for PB by incorporating the results of the expert elicitation exercise outlined in Section 5.2 These results were used to make a more reliable judgement of the uncertainty in the current PB GRADE domain and improved confidence in the effect estimates presented. This ultimately strengthened the overall certainty of the evidence judgements and consequently the overall recommendations being made. A summary of the main results are given in Section 5.3.1 and detailed results are given in the full publication in Appendix 9.

5.3.1 Bayesian analyses

5.3.1.1 Base case analysis

As expected, the results of the base case Bayesian NMA were consistent with the frequentist analysis reported in the fourth publication in Appendix 6 and summarised in Section 4.5.3.2. The base case analysis refers to the Bayesian analysis which does not use any informative priors to adjust for PB and is (approximately) equivalent to the frequentist NMA presented in the publication in Appendix 6. Similarly, there was also no evidence of inconsistency in the network or in any of the convergence diagnostics reported in the sixth publication in Appendix 9.

5.3.1.2 Part A: Copas model approach

The results of the selection model analyses show that the introduction of increasing levels of PB adjustment typically shows greater reductions in the estimated OS benefit for the NMRD (RD 0 cm) reference category. However, in almost all instances the results changed little compared to the base case frequentist (Section 4.6) and base case Bayesian analyses (Section 5.3.1.1). The NMRD category retains at least an 87.5% estimated chance of providing the best OS. See Table 4 in the sixth publication in Appendix 9 for the full results.

5.3.1.3 Part B: Alternative novel approach

Seven sensitivity analyses were conducted that incorporated prior information using the estimates derived from the elicitation exercise, namely N(-0.24, 0.06) for RD < 1 cm versus RD > 0 cm, N(-0.26, 0.05) for RD > 1 cm versus NMRD, and N(-0.24, 0.06) for RD > 2 cm versus NMRD. The sensitivity analyses varied from applying adjustments to some of the RD comparisons to applying an adjustment for all comparisons (Sensitivity Analysis 5), which was the most extreme sensitivity analysis considered. See Table 5 in Appendix 9 for the full set of sensitivity analyses.

All sensitivity analyses were in line with the base case analysis and demonstrated prolonged OS if PDS achieved NMRD compared to any other RD threshold. However, the effect estimates were attenuated in comparisons involving NMRD, though not to any suggestion of changing the existing conclusions. This was even the case for the most extreme sensitivity analysis, which utilised all RD thresholds, including ones that would not have been expected to have been widely reported in reality.

5.3.2 Bayesian findings in context

Given that experts perceived there to be a strong likelihood that PB would be present, a single adjustment would not adequately have reflected the wide range of possible adjustments that could be made. Therefore, a series of sensitivity analyses were important to conduct and present. These sensitivity analyses ranged from ones that appeared to best reflect the experts' views to more extreme scenarios that fully tested the robustness of the base case analysis described in Section 5.3.1.1.

Publications 3 and 4 had shown a survival benefit of surgical debulking to NMRD after primary surgery in women with advanced EOC. In the Bayesian framework, extreme value sensitivity analyses were used to examine the plausibility of overturning conclusions obtained from the base case analyses (analyses in Publications 3 and 4 and the Bayesian analysis which used a non-informative prior in Publication 6). There seemed to be little likelihood that the existing conclusions could be overturned. The selection model described in Section 3.7.2.1 indicates that the findings are robust to large levels of PB (see Section 5.3.1.2). Similarly, the elicited estimate used in Part B of the elicitation exercise, which were used as an adjustment for PB, and did not change the conclusion from those drawn from the base case results. However, the elicited estimate and subsequent results may be more representative of the strength of feeling in the experts' opinions. For example, the mean number of missing studies estimated by experts for the comparison of RD <1 cm versus RD=0 cm was 17.8, corresponding to the derivation of an informative prior (N(20.24, 0.06)). This reduced the magnitude of effect estimates and is reflective of the likely omission of unpublished studies in the base case MA. Further research is now extremely unlikely to change the confidence in the existing estimates of effect.

5.3.3 Categorisation of RD following primary debulking surgery

This section has shown that it can be said with more confidence that there are three clear and distinct categories of RD thresholds after PDS, a result which agreed with the findings reported in the frequentist NMA presented in Section 4.6. Surgical debulking to NMRD was the most effective surgical option, but the results suggest that every effort should be made to achieve near-optimal (< 1 cm) debulking if this is not possible.

The results of the analysis reported in Section 5.3 also reinforced the earlier finding shown in sections 4.4. and 4.5 that suboptimal cytoreduction can be defined as RD > 1 cm. The overall certainty of the evidence remains moderate despite a reduction in the magnitude of the effect estimates in comparisons involving RD=0 cm. This is because effect estimates are now likely to be more reliable as missing studies have likely been accounted for. Arguably, both opponents and proponents of maximal effort debulking surgery are now more likely to accept these findings as previously.

5.4 Conclusion

This chapter set out to demonstrate the use of several novel methodologies, which aimed to raise the certainty of the evidence. The methods presented were either novel in their own right or demonstrated originality in their unique application. The elicitation exercise explored the opinion of experts by probing their perceptions. Specifically, participants were asked to account for the sort of studies that have been conducted but not published, the plausible magnitude and direction of any PB, and the possible explanations for why and how the PB occurs. The thesis showed that these data could be used to adjust the results of a NMA for PB. The conclusions drawn after this adjustment had the same certainty of the evidence judgement found in my earlier review, but the justification for this judgement was strengthened.

To convince those who may still be opposed to the findings of the previous publications and perceive maximal effort debulking studies to report a biased set of analyses, extensions to the earlier publications was needed to report "fair" effect estimates. Making an adjustment for PB using elicited views of gynaecological experts was arguably the best approach to achieving this, especially since the experts in the elicitation exercise appeared to be representative with a range of different views on maximal effort debulking surgery. The benefit of obtaining prior information on experts' views on PB in this case is obvious because MAs are almost all exclusively subject to some degree of reporting bias. Therefore, the improved reliability of effect estimates by adjusting for PB strengthened the certainty of the evidence judgments to a wider audience, when previously there may still have been some doubt in the conclusions reported in Publications 3 and 4 by some of the surgical community.

The sophisticated selection models used in the analyses using results from part A of the elicitation exercise may also not have made the kind of adjustment for PB that reflected the opinions of the

experts who participated. This was because the adjustment in part A was minimal. The novel methodology used in part B of the exercise where a prior was formulated from the average number of missing studies with their effect sizes may offer a simple and highly desirable approach. The adjustments in parts A and B do not give different results leading to different conclusions, but Part B seemed to adjust effect estimates in a way that seemed to more reflect the opinions of the experts. Consequently, there could be more scope for the results in Parts A and B to differ if this exercise was repeated in the future. In either approach, it is important to specify methods a priori as to not abuse the results by making post hoc adjustments.

6 Discussion

6.1 Summary

The main aim of the thesis was to explore the role of primary surgery for advanced EOC. To do this I evaluated the clinical effectiveness and safety of different forms of surgical management for women with surgically staged advanced EOC (FIGO stages III and IV) using ES methodology. For the management of advanced EOC, the focus was on the role of surgery as part of the initial treatment (PDS or IDS) of advanced EOC and then on the extent of this primary surgery. That is how aggressive that primary surgery should be in terms of the extent of residual disease that would be left. This latter aspect was explored by reinterpreting the existing empirical evidence to look at how estimates of overall survival varied by the extent or residual disease that remained. The impact of RD was treated as a prognostic factor for OS and interpreted as a proxy for the need for maximal effort debulking surgery. As the research evolved the limitations with the approaches adopted were identified and these were progressively addressed using the novel methods. These novel methods aimed at raising the certainty of the evidence and confidence in the overall conclusions.

Six publications were included in the thesis. These were identified and reviewed using well-established methods. The data from these studies were used to progressively explore the EOC evidence base. To do this I applied alternative, incrementally more complex methods to the evidence base, each additional analysis sought to address a limitation of an earlier analysis. Concurrently, I have exploited and maximised the usefulness of the existing evidence base to decision makers, offering methods to highlight areas of uncertainty. I argue that Publication 6 presents results that appear to be more reliable and trustworthy to both opponents and proponents of maximal debulking surgery for reasons outlined and critiqued in Section 5.3. The basis for my assertion is mainly founded on the fact that despite a reduction in the magnitude of effect estimates, the results are likely to be more reliable given missing studies have been accounted for in

the analysis. The surgical community may now be more likely to accept these findings than previously as the certainty of the evidence judgement has been more adequately assessed. The individual PB domain in GRADE is greatly strengthened given the adjustment for PB, even though it may not necessarily result in upgrading the overall GRADE certainty of evidence judgement from moderate. If a GRADE judgement is moderate, it means that we are moderately certain in the effect estimate. This translates to the true effect being likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.(74) Clearly the series of sensitivity analyses reported in Publication 6 strengthens the conviction of the overall judgement being made.

Most women usually have widespread disease so surgery alone is unlikely to cure the disease, and most will also need chemotherapy.(2) Traditionally, chemotherapy was given after surgery (PDS), but can be used before surgery (IDS) with the aim of shrinking the cancer and allowing women to get better prior to undertaking major surgery. Women who receive IDS complete the remaining cycles of chemotherapy following surgery. High to moderate-certainty evidence in my body of work suggests there is little or no difference in primary survival outcomes between PDS and IDS, but that IDS probably reduces the risk of serious adverse events, especially those around the time of surgery.

At the time of submitting this thesis, the role of IDS versus PDS remains an area of controversy in the gynaecological oncology community,(2) despite a high quality and comprehensive review in this area, with Publication 1 reporting on four well-conducted trials. It is an area which often suffers from a distinct lack of equipoise. This is most often directed as criticism of the results of the included studies, largely based on concerns regarding low rates of achieving cytoreduction to NMRD in two of the core trials.(213, 214) Consequently, my next focus centred on investigating whether women with advanced EOC had better outcomes if they received 'maximal effort debulking' surgery, which is much more extensive than standard surgery, to remove tumours. Standard surgery in an advanced

disease setting still has an element of radicality and comprises as a minimum many of the surgical procedures involved in more radical surgery. Maximal effort debulking surgery is an extension of standard surgery and may include at least one additional extensive surgical procedure.

My second publication found only very low-certainty evidence comparing maximal effort debulking surgery and standard surgery in women with advanced EOC. The evidence was limited to retrospective, non-randomised studies and is at critical risk of bias. The results may suggest that maximal effort debulking surgery could result in improved OS, but results are based on very few women who were chosen to undergo each intervention, and consequently the evidence is very uncertain. In the absence of any clear evidence of a difference in most outcomes including overall survival in the body of work involving more extensive primary surgery and a failure to establish the most beneficial time to perform this, a series of separate prognostic reviews assessing RD as a prognostic factor in this area were conducted. The series demonstrated the survival benefit of macroscopic debulking to NMRD after primary surgery. This result could potentially be used as a proxy to strive towards higher rates of maximal effort debulking surgery in order to achieve this. A detailed summary of the main findings is provided in Sections 6.1.1 and 6.1.2, with Table 6 depicted below summarising each research question and the main findings.

Table 6: Summary of research objectives and main findings

Research objective	Main findings
To assess the effectiveness and safety of treating women with advanced EOC with IDS compared with PDS.	High to moderate-certainty evidence suggests there is little or no difference in primary survival outcomes between PDS and IDS. IDS probably reduces the risk of serious adverse events, especially those around the time of surgery, and reduces the risk of postoperative mortality and the need for stoma formation.
To evaluate the effectiveness and safety associated with maximal effort debulking surgery in the management of advanced EOC.	Only very low-certainty evidence comparing maximal effort debulking surgery and standard surgery in women with advanced EOC was identified. The evidence was limited to retrospective, NRSs and so is at critical risk of bias. The results may suggest that maximal effort surgery could result in improved OS, but results are based on very few women who were chosen to undergo each intervention, rather than a randomised study, and so the evidence is very uncertain. Results for PFS were inconsistent and evidence was sparse. QoL and morbidity was incompletely or not reported in the three included studies.
Building upon and addressing the limitations of the evidence in the maximal effort debulking SR, estimate the impact of the extent of residual disease on overall and progression-free survival in PDS and IDS settings.	In a PDS setting, there is moderate-certainty evidence that the amount of RD after primary surgery is a prognostic factor for OS and PFS in women with advanced EOC. The analyses were separated into three distinct categories for the survival outcome including NMRD (RD=0 cm), near-optimal (RD < 1 cm) and sub-optimal (RD > 1 cm) cytoreduction. After IDS, there may be only two categories required, although this is based on very low-certainty evidence, as all but one study included NMRD in the near-optimal category. The one study that separated NMRD from near-optimal showed no improved survival outcome in the near-optimal category, compared to sub-optimal. Further low-certainty evidence also supported restricting to two categories, where women who had any amount of macroscopic RD after IDS had a significantly greater risk of death compared to women with NMRD.
To address limitations in the prognostic factor SR, explore whether a NMA, as an extension to the previous standard pairwise MA approach.	The results confirm the strong association between cytoreduction to NMRD and improved OS that was established in the previous (third) objective. A NMA approach forms part of the methods guidance underpinning policy making in many jurisdictions. The analyses that addressed this objective present an extension to the previous work in this area.
Exploring the extent of PB associated with the prognostic factor SR and NMA and how this might be quantified and incorporated into the analyses.	The results suggest that a degree of scepticism may be needed when reviewing studies comparing RD < 1 cm versus RD=0 cm. There is also a belief among respondents that comparisons involving RD=0 cm and suboptimal thresholds (RD > 1 cm) are likely to be impacted by PB, but this is unlikely to attenuate effect estimates in MAs.
Use the finding of the method used to quantify PB to revise, or use as an	There remains a strong association between the achievement of cytoreduction to NMRD and improved OS even after

add-on to the current PB domain in the GRADE tool. This aimed to raise the confidence in the GRADE judgement and consequently the overall recommendations being made.	adjustment for PB using strong informative priors formed from the expert elicitation exercise explored in the previous objective.
Identify gaps in the literature arising out of the thesis and future research needs, in both clinical and methodological capacities, with an emphasis on raising the certainty of the evidence.	Several clinical and methodological limitations in advanced EOC research have been identified in this thesis; many of these are discussed in detail in Section 6.5. Several extensions to existing methods and to strengthen the standard GRADE approach have been proposed. The rationale was to strengthen the confidence in the overall GRADE judgements being made. It was proposed that all of the suggested methods were incorporated into GRADE as an additional item, rather than as stand-alone items to attempt to minimise independent judgmental decisions and raise confidence in the overall conclusions. Incorporating additional items as part of GRADE such as a trial sequential analysis (TSA), value of information (VOI), expert elicitation, or some other methodology component, especially any that adds an element of objectivity, is worthy of consideration to methodologists who may want to further develop the tool.

6.1.1 Detailed summary of clinical results

The thesis' first objective which aimed to compare PDS and IDS for women with advanced EOC was accomplished by conducting the SR publication in Appendix 3. It found high to moderate-certainty evidence that there was little difference in primary survival outcomes between PDS and IDS. IDS may reduce the risk of some SAEs, especially those occurring around the time of surgery, but the benefit of having PDS and having an additional schedule of chemotherapy to administer at a later date has perceived benefits. The findings of this review imply that there is no need to change current practice regarding timing of initial primary surgery.

The second review assesses the impact of extent of surgery by comparing maximal effort debulking surgery with standard surgery. The review found only very low certainty evidence for the comparison of maximal effort debulking versus standard surgery in women with advanced EOC. The evidence available was limited to retrospective, non-randomised studies and was at high risk of bias, especially selection and confounding biases. Given the very low certainty of the evidence, it is unclear what impact maximal effort debulking surgery has on OS compared with standard surgery. The conclusion is unsurprising, as the results are based on very few women. Therefore, the impact on clinical practice based on this review is expected to be minimal.

The first two reviews within my thesis do not provide clear recommendations for any one surgical technique or course of management. Likewise, the latest version of the guidelines by the National Institute for Health and Care Excellence (NICE), last updated in 2013, did not make any clear recommendation on this matter.(50) Other existing guidelines from Europe(54, 215) largely agree with the NICE guideline. Given the lack of a clear recommendation regarding maximal effort debulking surgery, it is unsurprising that there is widespread variation in surgical practice globally, with varying rates of survival. Despite the limited evidence base, the guidelines from Belgium(215)

suggested that it supports the use of radical surgical techniques (such as diaphragm resection, peritoneal stripping, splenectomy etc.) to obtain complete resection of all macroscopic tumour.

Whilst direct head-to-head evidence would be preferred to support guideline recommendations, one way to infer if there is any value in removal of RD in terms of OS is to assess the prognostic impact of RD after primary surgery on survival outcomes. This was the focus of my third review. The main aim of this work was to attempt to inform surgeons on the desirability, or not, of their approach to primary cytoreductive surgery, particularly if cytoreduction to NMRD is at all plausible. In a PDS setting, there was moderate-certainty evidence that RD is a strong prognostic factor for survival in women with advanced EOC.

The analysis was separated into three distinct categories for the extent of RD, to assess their relative impact on OS. These categories were NMRD (RD=0 cm), RD=0.1-1 cm (near-optimal) and RD > 1 cm (sub-optimal). After IDS, the extant evidence only allowed the consideration of two categories. For these, there was very low certainty evidence, as all but one study included NMRD in the RD=0.1-1 cm category. Since data were sparse for IDS, the focus in subsequent publications was on PDS.

My next piece of work explored methods that facilitated more reliable estimate of effects and consequently raise the certainty of the evidence. This work repeated and updated the prognostic factor review but applied a more sophisticated NMA approach to the analyses of the data. The results consolidated the association between achievement of cytoreduction to NMRD and improved OS, as established in the third publication in Appendix 5. However, this area appears to have genuine clinical equipoise.(55) The more complex analysis and findings of the NMA that I reported (see fourth publication in Appendix 6) may encourage some of the surgical community to strive towards improving rates of cytoreduction to NMRD. However, there remained some doubt that the

estimates of effect presented in this publication were as reliable as they could be. Therefore, to some clinicians, this would likely limit the authenticity of the results and the wider evidence base.

The evidence base available may be subject to PB, which may favour studies reporting beneficial results for cytoreduction to NMRD compared to other RD groups. Therefore, my next piece of work explored the opinions of gynae-oncologists and other clinicians with expertise in gynaecology on their views about PB in this area (Appendix 7). This work explored whether, and to what extent, the evidence included in both of my earlier prognostic factor reviews could be subject to material PB (i.e., a bias that is sufficient to alter the conclusions drawn).

An elicitation exercise involving 18 experts from the BGCS was conducted (Appendix 8). This exercise suggested potential concern about the nature and extent of PB associated with the two prognostic factor reviews (Appendix 7). The concerns were such that the unpublished evidence may have substantially reduced or even removed the suggested OS benefit from cytoreduction to NMRD compared to RD < 1 cm. The results showed that a degree of scepticism may be needed when reviewing studies comparing RD < 1 cm versus NMRD. Experts considered that PB for comparisons of NMRD versus suboptimal RD thresholds would not translate into many elicited missing studies in Part B of the elicitation exercise.

The results for Parts A and B of my elicitation exercise appear to correspond and be largely in agreement. Responses suggested that PB may be quite likely in studies where the sample size was just n = 100. Respondents in Part A of the exercise also indicated that there was potential for PB in some comparisons when studies had larger sample sizes but to a much lesser extent. In Part B, most respondents acknowledged that the likelihood of PB is ‘somewhat’ or ‘quite’ likely (72.5%) in the comparison of RD < 1cm with NMRD. Most respondents considered that it would be small studies

that would be missing (n = 100) in Part B of the elicitation exercise, which is consistent with their opinions in Part A.

Building upon the above, I utilised the findings from the elicitation exercise to adjust for PB in a Bayesian NMA (see the sixth publication in Appendix 9). This method showed that there remained a strong association between the achievement of cytoreduction to NMRD and improved OS, even after adjustment for PB using strong informative priors. Using a Bayesian framework, a series of sensitivity analyses were conducted to gauge the plausibility of overturning conclusions presented in earlier prognostic factor review publications. These included considering the amount of unpublished data and the method used to adjust for PB. Over the sensitivity analyses conducted, the conclusions about the effect of cytoreduction to NMRD on OS were robust. This even holds true for one of the 'extreme' sensitivity analyses, which incorporated a strong prior belief that the NMRD group was assumed to be favourable against all other comparisons in the network. This situation was considered by clinical experts to be unrealistic in practice. The adjustments made on the basis of the responses in Part A had a minimal impact on the effect estimates. Part B appears to have scope for more widespread priors. and further research in the prognostic factor EOC setting is unlikely to change confidence in the existing estimates of effect. The results of the 'extreme value' sensitivity analysis were consistent with the findings in the primary NMA (reported in the fourth publication in Appendix 6). Therefore, the existing conclusions appear to be reliable.

6.1.2 Summary of evidence synthesis methodology in thesis

There was a natural and sequential progression in statistical and ES methodology in the included publications. The initial review used standard pairwise analyses, as only two interventions were being compared, and Cochrane SR methodology for inclusion of RCTs. Pairwise MA simply pools evidence from a direct head-to-head comparison of studies that compare the same two

interventions. The methods set out to attempt to address whether there was any difference in various outcomes between IDS and PDS for women with EOC.

The second review evaluated the effectiveness and safety associated with maximal effort debulking surgery in the management of advanced EOC, using methods that extended to analysis of non-randomised studies in the absence of RCTs. This included a full risk of bias assessment using ROBINS-I.(84) The ROBINS-I tool is the preferred risk of bias tool to be used in Cochrane Reviews for non-randomised studies of interventions. The first two reviews used guidance from the Cochrane Handbook for Systematic Reviews of Interventions (CHSRI) and the certainty of the evidence was assessed using the GRADE approach.(77) I ensured that key aspects of this guidance was adhered to in each publication to meet the Methodological Expectations for Cochrane Intervention Reviews (MECIR), as adherence to MECIR is a requirement of a Cochrane review to meet core standards.(113, 216)

These initial two reviews did not provide any clear evidence of best surgical management, so an assessment of the prognostic impact of RD after primary surgery on survival outcomes was made in the third publication in Appendix 5. While the same pairwise approach to the MA was used, guidance for the conduct of SRs of prognostic factors was followed.(130) This included a risk of bias assessment using QUIPS and a modified approach to assessing the certainty of the evidence. The Cochrane Prognosis Methods Group recommends the use of the QUIPS tool to assess risk of bias in prognostic factor studies. The fourth publication in Appendix 6 extends beyond standard pairwise MAs by using a NMA. The NMA is a better way of utilising the available data as the approach forms part of the methods guidance underpinning policy making in many jurisdictions. A NMA simultaneously allowed estimates of relative effectiveness for any pair of RD thresholds forming the evidence network.

The thesis then attempts to explore different approaches aimed at strengthening the confidence in effect estimates presented in the NMA. This was achieved by enabling a more informed judgement of the PB domain to be made which facilitated the overall certainty of the evidence judgements to be more reliable. In the first instance, an expert elicitation exercise was conducted as a means of quantifying and incorporating views on the nature and extent of PB associated with the prognostic factor SRs. Although expert elicitation is not an objective exercise, it makes the (arguably reasonable) assumption that any consensus judgements will raise certainty.

The results of the expert elicitation exercise were used as a framework for a novel extension to a previous statistical approach that used elicitation methods to fit a selection model.(118) The elicitation exercise also used an alternative novel approach that asked respondents to estimate the number of missing studies and their corresponding effect sizes. The practical application of the results of both parts of the elicitation exercise was demonstrated in the sixth publication in Appendix 9 using a Bayesian NMA. The Bayesian NMA has potential advantages over a frequentist NMA as it can incorporate external evidence (in this case from the expert elicitation exercise) and can more easily perform complex analyses due to the flexibility of the simulation approach.(217) The most extreme sensitivity analysis in the sixth publication (see Section 5.3.1.3), which diluted the effect estimates for PB the most, also demonstrated the extent to which existing data had the potential to convince those who were sceptical about the results of the primary analyses.

6.2 Originality and novel methodology

All publications were unique in their own right, but certain aspects were novel in their application or in the methodology components that were developed. The thesis demonstrates several novel approaches involving the adjustment of effect estimates in a MA, as well as methods to fully utilise the assessment of the certainty of the evidence.

While most of the methodology in my first three publications was applied in a standard way, several aspects were original such as the modification of some risk of bias domains and the application of GRADE to a prognostic factor framework. The fourth publication in Appendix 6 demonstrates the application of a NMA; it represents an extension beyond pairwise analyses that has not been previously conducted in this area.

Publications 4-6 explored procedures to raise the certainty of the evidence by presenting effect estimates that are more reliable. The NMA findings were also used to demonstrate novel methods for addressing uncertainties that may reduce the impact of the findings on practice. Whilst not an aim of my thesis it is worth noting that it is uncommon to present a NMA using both frequentist and Bayesian frameworks. There appears to be a lack of literature that compares these two approaches empirically, especially for time-to-event data. The analyses presented in my body of work offers a route to understanding the differences between the two approaches.

The results of the frequentist NMA reported the in the publication in Appendix 6 and that of the Bayesian one using non-informative priors (referred to as 'base case' in the Bayesian analyses) in the publication in Appendix 9 appeared to be near identical other than the different interpretations across frameworks.

I gave some focus on contributing to the debate surrounding how to explore the potential impact of PB on study findings and conclusions. There is no consensus on a standard approach for identifying and adjusting for PB, with many methods focusing on its identification alone.(218, 219) As such, I have proposed an approach that has hitherto received little attention in MAs: the consideration of expert opinion and the incorporation of their views and opinions into the MA to inform the adjustment. The information gained in the elicitation exercise was utilised to adjust the MAs for any perceived PB in a Bayesian NMA (sixth publication in Appendix 9).

The analyses that adjust for PB represents a major update and extension to the analyses presented in Sections 4.4.3 and 4.5.2. Copas selection models (used in Publication 6 and based on the methodology described in Section 3.7.2.1 relating to Part A of the elicitation exercise) have been used to adjust for PB in other areas (specifically single comparison MAs).(118, 191) The framework was applied in the thesis as an extension to the NMA by incorporating multi-arm studies, something I am not aware of being used elsewhere. Furthermore, the use of Copas selection models had only previously been proposed for networks of trials with less than or equal to three network nodes. Therefore, the methodology and code for the analyses reported is included in the sixth publication (Appendix 9). This code can be readily adapted to other NMAs which include studies with more than three arms.

In Part B of the elicitation exercise, I generated new knowledge by applying a new method for adjusting for PB in the EOC clinical data. This approach could be useful for MAs in the wider advanced EOC research area as well as more generally in other areas. The approach may be particularly pertinent in oncology where survival estimates are readily available and reliable.

6.3 Strengths and limitations

The obvious strength of the publications included in the thesis is the rigour and comprehensive nature of the research conducted. Systematic guidance was followed, as well as general good clinical practice, while all research papers were published in peer reviewed journals. Common strengths and limitations as well as guidance followed in ES methodologies are outlined in the individual publications that form the body of evidence in this thesis. This section outlines the main strengths and limitations of the body of research and focuses on non-standard items that have not been previously mentioned in these publications or in any great detail in the thesis.

ES guidance encourages attempts to be made to minimise bias in the review process, which includes trying to minimise selection and confounding biases. The second publication in Appendix 4 demonstrated the first instance of the application of the minimum n=100 inclusion constraint. The SR included non-randomised studies of at least 100 patients, where it can be argued that this could minimise potential biases, negate small study effects, and avoid issues around statistical adjustment. While it was hypothesised that including small studies might not lead to appropriate MA estimates, the reverse may also be true and the pragmatic approach in the thesis may have introduced small study effects in their omission. However, only including non-randomised studies that used statistical adjustment for important baseline characteristics using multivariate analyses seems wholly sensible.

A further threat to the validity of any SR is the possibility of PB. Studies that did not find a statistically significant difference between treatments may not have been published. Standard approaches are available in ES guidance to identify PB but very little is available on making an adjustment to effect estimates in a MA. I was able to suggest alternative methods in the body of work in the thesis to the limited existing ones. The methods and their application in two of the publications report estimates of effect that should be considered more reliable than any that do not

attempt to adjust for PB. I argue that they should be considered in their application in any area where PB is considered a serious threat to the validity of the results in a ES. The relative simplicity of the application of the novel methodology proposed in Part B of the elicitation exercise could be particularly appealing to researchers conducting a review. Part A of the exercise extends beyond the existing method to allow the inclusion of studies including up to four arms. The expert elicitation exercise described in the publication in Appendix 7 was designed in consultation with several experienced senior gynae-oncologists and it was expected *a priori* that the number of respondents who could provide a basis for meaningful conclusions to be drawn would be as few as four to 16 experts.(189-192) The sample size achieved in the elicitation exercise (n = 18) was comfortably above this so provided a firm basis to form priors in which to use in the Bayesian analyses.

The thesis did not account for the experience of patient representatives which can be viewed as a limitation. It was felt that the elicitation exercise should be restricted to experts in the field and the views of patients in this case would not be useful. However, in the absence of any adequate QoL data in the overall body of research in the thesis it is clearly a limitation and represents a gap in the evidence base. I make several suggestions in the following section as to how this could potentially be addressed.

I was mindful that in the elicitation exercise, answers given by the experts to open ended questions could be prone to an "extreme answer bias". Therefore, the instructions that accompanied the elicitation exercise were deliberately quite extensive. This was discussed in detail when the exercise was designed, and it was felt that more biases would be introduced if a ceiling of the number of estimated studies had been applied. It was anticipated that the experts would give ‘sensible’ answers given their expertise in the subject area and the thorough briefing and instructions given

prior to respondents completing the exercise. From the observed responses there was no reason to question this assumption as most seemed to be within an acceptable range.

Whilst the elicitation approach utilised in the body of research appeared to be a success, there are alternative approaches that could also have been considered. The SHeffield ELicitation Framework (SHELF),(220) which describes the process of elicitation through discussion among groups of experts might be an alternative. There are potential advantages from a group discussion, notably the ability to help a respondent who is unfamiliar with the concept of making an informed guess or estimating a probability.(220) The limitation of the SHELF approach is that individuals must devote much more of their time and has the potential for high costs. The elicitation approach preferred in the body of research in the thesis was by comparison inexpensive and allowed a degree of flexibility, as the exercise could be completed over several sessions.

6.4 Context of body of research in thesis with existing guidelines and policy

This thesis includes publications that should be considered in any review of the management of advanced EOC, whilst these now represent the most up-to-date reviews of relevant evidence to underpin new guidance. In addition, the novel methods applied allow the certainty of the evidence to be raised.

6.4.1 Summary of existing studies and guidelines identified

NICE are currently in consultation with experts in developing an interventional procedure document reporting an overview of extensive surgery for advanced EOC,(48) which makes this thesis very timely. The guidance that is currently in preparation has discouraged the use of the term 'ultraradical' favouring instead, the term 'maximal effort debulking (cytoreductive) surgery' as an alternative which I have incorporated into my thesis. The current guidelines which were last updated in 2013, state that evidence on the efficacy and safety of maximal effort debulking surgery for advanced EOC is inadequate.(50) The guidance recommends that maximal effort debulking surgery should not be done except with special arrangements for clinical governance, consent and audit or research. The NICE Committee for the 2013 guidance noted that cytoreduction of OC to NMRD, together with all intra-abdominal metastases, is the best prognostic factor for improving survival. However, this potential survival advantage needs to be weighed against the morbidity and risks of very extensive surgery when considering the balance of QoL and survival.(50) NICE state that standard (radical) surgery should as a minimum comprise, total hysterectomy, bilateral adnexectomy with excision of the pelvic peritoneum, total omentectomy including the supracolic omentum, removal of bulky pelvic and lumbo-aortic nodes, simple peritonectomies and/or localised colonic resection. Procedures such as appendicectomy may have previously been considered part of standard surgery but evidence now suggests that this could be unnecessary and may cause harm. Maximal effort debulking surgery is an extension of standard (radical) surgery including at least one

of the following: stripping of the diaphragm(s), extensive stripping of the peritoneum, multiple resections of the bowel (excluding localised colonic resection), liver resection, partial gastrectomy, cholecystectomy and splenectomy (with or without resection of the tail of the pancreas).(50)

A more recent review of clinical guidelines(54) showed clear international differences in OC survival and these differences in treatment could be contributing to survival disparities. The objective of the review by Norell et al(54) was to compare clinical practice guidelines and patterns of care across seven high-income countries. Guidelines widely used in routine OC treatment were included. The review also included an expert questionnaire component with questions on surgical practice which were validated and tested by an expert clinical working group. Twenty-seven guidelines were compared, and 119 clinicians completed the survey. Guideline-related measures varied between countries but did not correlate with survival internationally. Given that Norell et al looked at seven countries it is clear that within countries, there is likely to be a lack of consensus and differing guidelines. Indeed, Norell et al reports patterns of surgery to vary internationally, including rates of performing more extensive surgery and achieving high levels of cytoreduction to NMRD. Consensus amongst Norwegian and Australian clinicians in the survey was for the use of maximal effort debulking surgery, whereas clinicians from Canada and the UK agreed with the use of maximal effort debulking surgery to a lesser extent, with some respondents opposing a more extensive approach. There was a reasonable proportion of clinicians in the survey who appeared willing to undertake maximal effort debulking surgery and this belief correlated with 3-year survival in advanced OC (r_s = 0.94, P = 0.017). However, Norell et al(54) found that most guidelines identified did not explicitly recommend maximal effort debulking surgery. Clinicians from countries performing higher rates of cytoreduction to NMRD were more likely than countries performing lower rates to be proponents of maximal effort debulking surgery. Norwegian clinicians were least likely to perceive age of the patient as a barrier to achieving cytoreduction to NMRD and Norway demonstrated the highest survival in elderly patients with advanced stage disease. In the UK, where clinicians had a perceived

lack of supportive care, survival for these older patients was lower. Norell et al reported that patients with advanced OC are more likely to have severe co-morbidities and higher mortality, and historically, elderly patients were shown to be less likely to receive comprehensive surgical treatment. It was also noted in the clinical guidelines that available resources and operating theatre time may influence a surgeons' ability to perform extensive surgery and could impact patient outcomes.

Another guidelines report from Belgium in 2016 by Vergote et al(215) provided recommendations based on scientific evidence for the diagnosis, treatment and follow-up of EOC. The report stated that clinicians were encouraged to interpret their recommendations in the context of the individual patient's situation and her own values and preferences. Furthermore, in the absence of good quality evidence on optimal treatment options, patient participation in clinical trials was to be encouraged as much as possible. The guidelines reported by Vergote et al(215) acknowledged that the evidence was limited but supported the use of radical surgical techniques (such as diaphragm resection, peritoneal stripping, splenectomy etc) to obtain resection of all macroscopic tumour to 0 cm. The guidelines showed the prognostic value of debulking to NMRD at the end of surgery and supporting evidence from the use of radical surgery. The guidelines formulated a strong recommendation (despite low level of evidence) that cytoreduction to NMRD should be the aim of primary surgery (upfront or interval debulking surgery) and that the term optimal should no longer be used as old definitions of surgery resulting in debulking to < 2cm or < 1cm.

A before and after uncontrolled cohort study of maximal effort debulking surgery after it was introduced at a population level, did not demonstrate improved outcomes(69). The shift to a more radical surgical approach led to a reduction in the proportion of women who had surgery as part of their treatment (10% fewer), either due to the general poor health of patients or the lack of surgical

expertise to perform more radical surgery. This study should be repeated to incorporate a control component in order to make meaningful inferences.

While not part of any existing guidelines yet, QoL was reported as a primary outcome in the Surgery in Ovarian Cancer - Quality of life Evaluation Research (SOCQER-2) study,(221) which was a prospective, non-randomised multicentre observational study conducted in the UK, India and Australia. This study found that patients with advanced stage EOC had no important differences in EORTC QLQ-C30 global scores measured across six weeks, six months and 12 months post- surgery when undergoing surgery of varying complexity, despite a higher preoperative disease burden in patients undergoing more radical surgical procedures. Across all groups of women receiving all forms of complex surgery (categorised by surgery complexity scores (SCS) and grouped into low, intermediate, and high), global QoL showed a small but significant improvement by 12 months postoperatively. Patients who underwent the most complex surgery (high-SCS group) had small to moderate detriments in EORTC QLQ-C30 physical function, role function and emotional function at 6 weeks post-surgery compared with patients undergoing less extensive surgery (intermediate- and low-SCS groups), but by 6 to 12 months post-surgery these functions are comparable across all SCS categories. A majority of women undergoing high-SCS surgery without disease progression experienced a positive change in QoL by 12 months post-surgery. There were no clinically meaningful differences in QoL among patients undergoing surgery of different complexities. The authors of the study concluded that patients undergoing high-complexity surgery can be reassured that by 12 months post-surgery most will have better QoL after than immediately before surgery.(221)

The SOCQER-2 study (221) also found that patients who underwent low-complexity surgery had higher rates of RD and lower survival compared with those with a similar disease burden undergoing

surgery of intermediate complexity. However, no statistical adjustment was performed in these analyses. Postoperative RD was associated with poorer OS, particularly in patients undergoing low-complexity surgery, but again no statistical adjustment was made. The SOCQER-2 study(221) acknowledged potential selection bias, but since recruitment to the study was carried out by research nurses, that systematic bias introduced by surgeons recruiting patients whom they believed would recover well after extensive surgery was unlikely.

6.4.2 Agreements and disagreements between body of research in thesis and existing studies and guidelines

Evidence from the body of evidence in the thesis suggested that it may not matter too much about the timing of the surgery, rather how successful the surgery is in terms of cytoreduction to NMRD. My research found only very low certainty evidence comparing maximal effort debulking and standard surgery in women with advanced EOC and also subgroups with carcinomatosis. In isolation, the evidence from this publication in the thesis generally supports the current 2013 NICE guidance appraisal.(50) There is a suggestion in the body of evidence in the thesis that maximal effort debulking surgery may result in better survival, but results are based on retrospective studies, at high risk of bias, in relatively few women.

The NICE guidance found cytoreduction of OC to NMRD, together with all intra-abdominal metastases, as being the best prognostic factor for improving survival.(50) Therefore, given a lack of consensus in guidelines and an absence of good quality evidence from interventional studies, reviews and clinical guidelines, the thesis focused on a prognostic setting to assess the importance of RD as a prognostic factor. This was used as a proxy to determine the value in cytoreduction to NMRD and whether or not maximal effort to achieve this should be made if at all possible. The body of evidence in the thesis culminated in presenting a Bayesian NMA that adjusted for PB. The results of

these analyses were robust to sensitivity analyses which incorporated a high level of perceived PB which diluted the effect estimates, but not to a large enough extent to overturn the primary conclusions. Therefore, this should remove some uncertainty about attempts at achieving cytoreduction to NMRD and may encourage more aggressive approaches in order to do so in cases where it may potentially be more difficult to achieve. The guidelines by Norell et al(54) show an apparent willingness by clinicians to undertake a more radical approach.

The body of evidence in the thesis supported Belgian clinical guidelines from 2016 (215) by demonstrating the strong association between OS and debulking to NMRD after PDS, even after adjustments for PB. In a PDS setting, the research in the thesis provided moderate-certainty evidence that RD after PDS is a strong prognostic factor for OS (and PFS in the publication in Appendix 5) in women with advanced EOC. The certainty of the evidence for these outcomes was very low for studies involving IDS. The evidence in the thesis supported the Belgian guidelines that there should be three distinct categories of RD after PDS including NMRD (labelled as optimal), < 1 cm (labelled as near-optimal and strictly meaning 0.1-1 cm), and > 1 cm (suboptimal). This consolidates previous findings but provides more compelling evidence since the NMA in my body of work adjusts for PB and allows the moderate certainty of the evidence to be a more robust judgement. Although evidence was limited, the findings of the prognostic factor assessment of IDS was still of interest because it was not reported in any other literature. The prognostic factor Cochrane SR (publication in Appendix 5) suggested that after IDS there may be only two categories of RD for OS, although this is based on very low certainty evidence. Only one study in the SR adequately separated NMRD from the RD < 1 cm category with all other studies including NMRD in this category. I did not identify any other research that examined this and the fact that prognostic factor study designs seem to include NMRD in the RD < 1 cm category suggests that there may already be an acceptance that two categories hold.

Any potential survival advantage of more extensive surgery needs to be offset against the risks of morbidity and QoL.(50) Assessment of QoL was not widely available in interventional studies included in publications within the body of evidence in the thesis. Therefore, survival was the focus when in a prognostic setting. SOCQER-2 is an important study(221) as it was the first of its size and conduct to critically address QoL in radical surgery and plugs a big research gap in the thesis. Although the main focus of my thesis was on survival outcomes, QoL and surgical morbidity are of high importance, so the results of the SOCQER-2 study are timely.

The publications that form the body of work in the thesis contribute to the evidence base available for guidelines in this area. Policy makers and guideline developers should give the results full consideration given some of the contemporary and novel methods that were employed. The concept of raising confidence in the PB GRADE domain and the overall certainty of the evidence judgement should be appealing to policy makers and guideline developers. In the UK the latest guidance from NICE on maximal debulking (cytoreductive) surgery for advanced ovarian cancer is still under consultation(48). Therefore, the thesis could be used as an additional source in any considerations and can be made available to the guideline panel. At present, maximal effort debulking surgical procedures are recommended to be performed by a team of surgeons with appropriate expertise and should be restricted to accredited specialised centres. However, the core evidence in the current NICE guidelines comes from unadjusted cohort studies,(51, 222-227) small case series(228), uncontrolled before and after studies(69, 229) and SRs.(1, 3, 230, 231) The guideline document used the evidence reported in the maximal effort debulking SR (Publication 2) and the prognostic factor SR (Publication 3) and these publications reported adjusted analyses. Adjusting for baseline characteristics and other important covariates is important when only non-randomised studies are available. The associated biases of such studies are comprehensively discussed in Sections 2.2 to 2.4. The NICE guidelines acknowledge evidence from various sources, but failed to extend beyond the findings of Publications 3 and 4. They would likely benefit from

documenting the extensions reported in Publications 5-6 in Chapter 5, for reasons previously outlined in Section 6.1. There is scope for the current NICE guideline(48) to reflect the data provided in the thesis. If Publications 5 and 6 were incorporated into their guideline it would increase confidence in the conclusions drawn from the prognostic risk factor SR evidence (Publication 3). It would also potentially allow for the extent of tumour debulking to be considered a proxy for the expected impact of maximal effort debulking surgery on overall survival. This may appeal to guideline developers wanting to move away from individual clinician preferences regarding choice of initial primary surgery.

6.5 Areas for future research

While I have highlighted various limitations throughout the thesis, this section identifies gaps in the literature arising out of the thesis and future research needs to meet the sixth objective. This includes both clinical and methodological capacities, with an emphasis on raising the certainty of the evidence.

6.5.1. Areas for clinical research

For the first publication in Appendix 3, which compared PDS and IDS, there are currently five ongoing studies, but none of these studies will likely report adequate data on QoL outcomes.(232-236) Reuss et al(233) aims to assess the role of maximal effort PDS (to achieve higher rates of cytoreduction to NMRD) versus IDS. The results of this trial will hopefully address questions raised by studies with lower optimal and near-optimal debulking rates. Collection of QoL data is an important patient-centred outcome in advanced ovarian disease, especially if there is minimal difference in survival between treatment options. QoL outcomes were poorly and incompletely reported across included studies in the interventional SRs included in the thesis. One of the main research gaps in the literature arising out of the thesis was an absence of substantial or robust QoL data. In order to explore this further in women with advanced EOC, a patient generated index (PGI) survey could be conducted.(237) It is known that EOC can profoundly impact women's quality of life but my body of work presented in this thesis was unable to adequately address this outcome due to a scarcity of data. Therefore, the aim of a PGI survey would be to provide evidence on outcomes of importance to women who have advanced EOC. The concept and methodology of a PGI is discussed in Section 6.5.3.4.

The second publication in Appendix 4 confirmed that there has never been a randomised study to address the role of maximal effort debulking surgery in advanced EOC. To date, most studies in this

area have assessed RD as an outcome rather than survival. In order to support existing guidelines, the role of maximal effort debulking surgery in the management of advanced stage OC could be addressed through conducting a sufficiently powered RCT comparing maximal effort debulking and standard surgery. If RCTs are not feasible, high quality non-randomised studies should be designed. Such studies should include all patients diagnosed within a fixed population and agree criteria for prognostic factors that will form the key adjustment in analyses. Population-level, multi-centre studies are important in this area, as what works or does not work in one institution may be very different from what works elsewhere.(221) Multivariable analysis should allow for baseline prognostic factors but not for variables (such as extent of RD or operating time) that would be recorded until after decisions were made on which type of surgery to adopt. The experience of the treating surgeon should also be factored into any analysis.(223)

Future primary research should focus on investigations that determine whether increasing attempts at debulking to NMRD have a direct effect in improving survival outcomes using methodologies and trial designs that reduce or eliminate confounding effects, such as the women's performance status, disease spread and tumour biology. The body of evidence in the thesis used secondary research to demonstrate an acceptable level of remaining RD after PDS to significantly prolong survival and what threshold should result in the surgeon being more conservative in their approach if debulking to NMRD is not plausible. The next piece of research in this area could focus on developing a prognostic model for accurately predicting important outcomes to determine the performance for predicting the risk stratification of women with this disease.(238, 239) The results of the prognostic factor analyses in the thesis showed the relative benefit of debulking to NMRD and while the analyses adjusted for various important baseline characteristics, they do not give the prognosis of women after their primary surgery. It would be useful for patients and clinicians alike to predict their survival based on their own individual risk factors to allow planning for future further line treatment options. Risk prediction models have also been previously proposed for predicting events such as unplanned

hospital admissions, which are costly and potentially preventable.(240) Measures of a risk prediction review should include discrimination, the area under the (curve) (AUC) receiver operating characteristic (ROC), calibration and overall model performance.(241, 242) A prediction model for survival would allow women and their families to plan for the future, as well as aid future decisions on their subsequent further line treatment care pathway.

6.5.2 Public and Patient Involvement and Engagement

The guidelines reported by Norell et al(54) also accounted for the experience of patient representatives which was not done in the thesis. The influence of more radical surgery on long term QoL was reported to not be a major drawback, the survival benefit weighing more importantly in the overall balance. In the absence of any good quality QoL data in the body of evidence in the thesis, any external reported outcomes should be considered in the wider evidence in any implementation and policy decisions. Any future guidelines should incorporate Patient and Public Involvement and Engagement (PPIE).(243-245) PPIE is essential to the success of research, with continuous involvement at every stage of the research cycle from applying for funding, through to sharing and disseminating study findings. Through active partnerships with a diverse group of patients, their family and carers and members of the public, who have personal experiences and an interest in the relevant research, PPIE can help make sure research stays relevant as well as being high quality.(243)

The CGNOCG provided valuable support with PPIE which ensured key sections of the Cochrane SRs were peer reviewed by patient representatives so that they were understandable by a lay reader. In the UK, organisations such as Ovacome,(246) Target Ovarian Cancer(247) and CRUK(248) can also be approached to assist with the formulation of patient and public advisory groups for ovarian cancer specific projects to ensure that the public's perspective is considered throughout. I have previously had excellent support in my ovarian cancer research from Peg Ford who is co-Founder and Emeritus President of the Ovarian Cancer Alliance of San Diego (OCASD)(249). As an ovarian cancer survivor

and cancer research advocate for over 15 years, her commitment matches many other informed and involved advocates to be a bridge between the patient community and the scientific/medical world, working hand-in-hand to advance evidence-based research to better serve both communities.(250) This research has varied from applying for grant applications to assisting with my SR publications to ensure key sections are understandable to a lay person. An organisation like OCASD could provide a PPIE advisor on a project to offer invaluable input into any future reviews or guidelines and be a part of any patient advisory groups. Patients could also be identified from sources such as the Managing Editors of the CGNOCG and via the Cochrane Consumer Network,(251) and patient advisors could assist with dissemination and ensuring the guidelines are accessible to patients. An organisation such as the OCASD would be able to assist in dissemination to all partner members throughout the USA. Patient advisors are usually recompensed for their time in accordance with INVOLVE rates.(252-254)

6.5.3 Areas for further methodological development

A number of methods have been applied and developed in this thesis and any limitations and ease of application have been documented. The focus of the thesis was to identify, apply and develop tools to help facilitate ES methods such as NMAs and adjustment for PB. While the thesis has acknowledged existing methods, many are either not easy to use or are not currently implemented in certain situations, especially when analysing time to event data. It was clear that more concise guidance across the board is still required such as whether to conduct a NMA in a frequentist or Bayesian framework in different settings, how to conduct a trial sequential analysis (TSA)(255-257) and further techniques for assessing PB to list but a few. This sub-section focuses on the methodological areas in the greatest need of development to strengthen the evidence base and ES more generally.

6.5.3.1 Risk of bias assessment

The domains in ROB2 are unchanged from those included in ROB1 and the signalling questions in ROB2 are aimed at getting better responses, but these may be unlikely to change overall judgements if experienced reviewers are the ones making the assessments. Therefore, it is unlikely that the risk of bias judgements would differ in any of the six domains used to assess risk of bias in the publication in Appendix 3. I was aware of the developmental signalling questions so used these when making judgements. There are also lots of tools for assessment of risk of bias in non-randomised studies so any further development should be monitored to utilise in any updated research in the future.(157, 258)

6.5.3.2 GRADE assessment in a prognostic factor review setting

Guidance on the use of GRADE for prognostic factor studies has not yet been published.(74, 162) Therefore, it will be important to adhere to full published prognostic factor ES guidelines in the future when guidance is made available. While attempts to appraise the quality and certainty of the evidence were made in the publication in Appendix 5 and in the subsequent NMA publications, expert guidance is likely to improve the confidence in the judgements. The CINeMA web application to assess the certainty of the evidence in a NMA needs to be developed to incorporate time-to-event data.(175) Section 3.5.2.4 documents that the methodological framework of CINeMA essentially covers items recommended in the PRISMA NMA reporting guidelines(174) and is broadly based on the GRADE approach,(74) but specific guidance for survival outcomes that could be used in the publication in Appendix 6 is sought.

6.5.3.3 Replication of future elicitation exercise

One expert elicitation exercise may not be sufficient to change practice, but the collection of more evidence through routine expert elicitation should be encouraged, not only in EOC research, but in

other disease areas. Every ES should aim to improve precision and reliability of effect estimates and consequently raise confidence in the estimates and the certainty of the overall evidence. The Bayesian application of this should also become more routine in areas where PB is likely to be an issue.

6.5.3.4 Patient generated index for quality of life outcome

In the absence of good quality data on QoL and morbidity in the thesis, an evaluation of the impact of advanced EOC on QoL, as perceived by women with EOC, using a PGI could be made. A PGI survey could be developed to include women with EOC to provide information on outcomes of importance to them, as well as aiming to identify additional outcomes that ought to be collected in future primary studies or define relevant outcomes for further SRs of the literature. Existing cancer specific questionnaires may not fully capture the impact of EOC on the QoL of women. Yet understanding the outcomes of importance to this patient group could inform the tailoring of interventions available to this group of women and could help better meet the need of a particular individual. The PGI is an individualised patient-reported health instrument which allows the respondent to select, weight, and rate the importance of a particular health outcome.(259) It was designed with the aim of producing a valid measure of outcome that reflects areas of importance to patients' lives(237) and involves the respondent deciding what factors are important to them. The aim is therefore to capture the diverse range of concerns or priorities of respondents. Using the PGI, respondents can vary the weight they attach to these concerns or priorities. An overall score for each respondent can then be calculated by multiplying the rating for each health area by the proportion of points allocated to that particular area. The participant sample for the PGI would consist of women who have advanced stage EOC. Patients could be recruited via the NHS (e.g., via hospital attendance, clinical contacts) and cancer charities, such as Ovacome as discussed in Section 6.5.2. A postal survey could be sent out to women and a recruitment target over several waves of recruitment could be set

in order to be able to conduct the PGI. Data analysis would include sample demographics, grouping of the areas of importance of an individual's life that they report to be affected by their condition and the frequency that the area was mentioned within the sample as a whole. PGI scores of overall QoL could also be reported alongside existing questionnaires such as EQ-5D-5L. FS-12/36 and EORTC QLQ-C30 and Correlation between these could be performed.

6.5.3.5 Individual participant data meta-analysis

An extension to the analyses included in the thesis could utilise an individual patient data (IPD) MA,(260) rather than using aggregate data. IPD MAs are recognised as providing high-quality clinical evidence that could be considered when the ES is updated in the future. They can more easily incorporate a consistent selection of confounders to adjust for, which would reduce the impact of selective reporting of analyses and outcomes. An IPD MA would also allow for a further comprehensive exploration of confounders, which could include looking at possible interaction effects between confounders.(261) IPD MAs may be particularly important for chronic and other diseases where treatment effects may depend on the length of follow-up. This is especially the case where there are risks and benefits that vary differently over time.(262) This may be applicable to the surgical interventions involving more radical procedures which have short term high risk but potential long-term benefit.

While an IPD may improve the quality of the data and analyses through the inclusion of all trials and all participants and detailed checking, they are not plausible in many areas. They are exceptionally time-consuming and can typically take many months to retrieve data from trial administrators. IPD MAs often include a mix of aggregate data too where it is not possible to obtain IPD datasets within an acceptable timeframe. That is the primary reason such analyses were not considered in this thesis

and might not be worth consideration until there is a sufficient change in practice regarding the administrative processes in dealing with IPD.

6.5.3.6 Living systematic review

The body of work in this thesis has made extensive use of SR and MA methodology. A concern with such work arises from the frequency with which a SR or MA is updated. An update has been defined as "a new edition of a published SR with changes that can include new data, new methods, or new analyses to the previous edition".(263) Cochrane reviews are typically updated within two to three years, but the recommendation is that a review should be habitually updated to include the most recent evidence.(264) Indeed, Cochrane has recently encouraged the use of living SRs, which adopt a continual updating process, such as monthly searching followed by rapid incorporation of new evidence into the published review. Living SRs are claimed to be more appropriate for questions that are of high importance to decision makers and for which new evidence is likely to be frequently published that would have an important impact on the review's findings.(265)

The updating approach that is an integral part of standard Cochrane and living SRs is, essentially, unadjusted cumulative MA.(266) Reanalysing a MA after adding each newly identified study may lead to spurious statistically significant results due to repeated testing of significance as study data accumulate. This is analogous to the biases associated with conducting multiple interim analyses in trials without adjustment.(255, 267) Therefore, further consideration about ways of accounting for this is needed before living reviews should be considered in future EOC research.

6.5.3.7 Trial sequential analysis

To deal with the issues associated with repeated testing of significance discussed in Section 6.5.3.6, a TSA(255-257) could be conducted. One of the limitations of my body of work was the inability to assess the impact of a MA being subjected to repeated statistical evaluation. In a MA, it is important to minimise the risk of making a false-positive or false-negative conclusion. A TSA is similar in nature to a group sequential analysis of a single trial and may be applied to a MA to evaluate the evidence. When a MA is subjected to repeated statistical evaluation, there is an increased risk that point estimates and confidence intervals will yield spurious inferences. More precisely, reanalysing a MA after adding each newly identified study, or adding in batches of studies, is prone to spurious p values and significant results because of repeated testing of significance as study data accumulate. This is analogous to biases associated with conducting interim analyses in a trial without inferential adjustment.(267) The aim of a TSA is to attempt to avoid wrongly concluding treatment differences in the absence of a benefit (i.e., true versus false positive). Therefore, a TSA aims to conceptualise the results of a MA, avoid the need to continually update and potentially accelerate the decision-making process.

In a TSA, the required information size (IS) is calculated to demonstrate or reject a relative risk reduction (RRR) of an event between two groups in a primary MA.(255-257) IS in a MA should at least equal the sample size of an adequately powered trial. In the primary framework,(268) decisions are largely based on statistical significance, where power and stopping rules are crucial to the future success of treatment and avoidance of continuing treatments that are not effective or are harmful. In the context of policy and guidelines, a TSA in a secondary analysis setting would be useful to test the robustness of existing evidence from a MA. At a review level, it may also aid futility (acceptance that an intervention is unlikely to be any better than the comparator) in the same way that adaptive designs would with certain treatments. To avoid making potentially false conclusions based on

statistically significant MAs, I would prefer in any future exercise to utilise a TSA in my EOC research as an additional component to GRADE, rather than as an isolated statistical analysis.

It is important to note that substantial or considerable heterogeneity would have a large impact on the TSA, so a model variance-based estimate should be used to correct for heterogeneity. A heterogeneity-corrected required IS should be used to construct sequential monitoring boundaries based on the O'Brien-Fleming type alpha-spending function for cumulative z-scores (corresponding to the cumulative MA).(257) This is analogous to interim monitoring in an RCT to determine when sufficient evidence of a treatment effect has been accrued.(267) These monitoring boundaries are relatively insensitive to the number of repeated significance tests. The required IS is the sample size required for a reliable and conclusive MA and is at least as large as that needed in a single powered RCT. Futility boundaries should also be considered (to test for no statistically significant difference as well as a significant benefit in a two-sided test) and the possibility that an intervention could also harm.

A TSA would be useful as an additional component of GRADE to potentially increase confidence in the overall certainty of the evidence judgements and potentially lead to more reliable conclusions. Currently only one TSA has been conducted in ovarian cancer research and this was identified in women in remission after initial surgery, which assessed maintenance chemotherapy.(269) Unfortunately a TSA could not be conducted in my body of work as the methodology needs further development so that it can be routinely used for time-to-event data and be applied to the EOC research. At present the TSA framework and guidance is restricted to dichotomous and continuous outcomes.(257)

The TSA approach has been criticised by some, largely based on concerns about the abuse of significance probabilities (p values). TSA or other sequential methods have not been recommended in Cochrane reviews,(270) as there has been a general move away from significance testing.(271) In this sense a TSA can be seen as a step backwards. Previous research found that a TSA may produce results consistent with a standard MA, and where this is not the case, it may be because the studies are small and heterogeneous, so any interpretation should be cautious.(272, 273) I am a strong advocate of moving beyond statistical significance in the main analyses but feel adding a TSA as a component of GRADE adds an element of objectivity and makes the overall certainty of the evidence judgements more robust. In this context, the main argument against a TSA being the misuse of statistical significance is not persuasive, if expert guidance is adhered to.(257) Therefore, to demonstrate the full TSA methodology for consideration in future EOC research, I conducted one in a different setting(112) and outline a full description of the required parameters and estimates that would be required in Table 7. The most important consideration when choosing the parameters used to calculate the required IS is that they are all defined *a priori*.

Table 7: Description of required parameters in a trial sequential analysis and estimates used in case study publication and proposed future epithelial ovarian cancer meta-analysis

Parameter	Case study estimate	Future advanced epithelial ovarian cancer estimate	Comments
Pc(death)	From the MA that was conducted (but estimate also considered the other existing MAs that had been published in this area)	Previous MA of RCTs; previous RCT; previous MA of observational studies; Previous observational study; from the meta-analysis that was conducted	It is important that the Pc(death) (as well as the rest of the parameters to calculate the required IS are defined *a priori* in a protocol, never after.
α	5%	1-5%	5% is usually reported but could consider having a lower value.
β	10% (90% power)	10% (90% power)	A β of up to 20% could be used in theory, but 10% is more sensible so as not to compromise power.
RRR	From the MA that was conducted (but estimate also considered the other existing MAs that had been published in this area)	Based on meta-analysis	Be careful to consider a RRR that is plausible. Surgery, certain vaccines, and antibiotics could potentially have large RRRs in certain settings but, generally, powering on a smaller RRR is more realistic.
D^2	Model variance-based SAs using different values of D^2	Model variance-based SAs using different values of D^2	D^2 in a MA is the relative variance reduction when the MA model is changed from a RE into a FE model. D^2 can readily adjust the required IS in any RE model meta-analysis. $D^2 \geq I^2$
Continuity correction	0.01	0.01	Used in trials that reported zero events in one or both arms.

TSA: Trial sequential analysis; SR: Systematic review; Pc(death): Proportion of deaths in control group; α: Type I error rate; β: Type II error rate; RRR: Relative risk reduction; IS: Information size; I^2: The percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance); D^2: Diversity (measure of heterogeneity) is described as the proportion of the total variance in a random-effects model contributed by the between trial variation despite the chosen between trial variance estimator.; RCT: Randomised controlled trial; MA: Meta-analysis; CI: Confidence interval; RE: Random effects model; FE: Fixed effects model

6.5.3.8 Value of information analysis

Additionally, a value of information (VOI) analysis(274, 275) could be undertaken to identify whether further research is worthwhile, which could also be incorporated as an additional domain in GRADE rather than as a separate analysis in its own right. In an advanced EOC setting, a VOI could be used in parallel with a TSA or as an alternative approach to determine the real value of conducting further research, especially if dealing with time-to-event data which cannot currently be implemented within the TSA framework. The results of a VOI would be presented at both an individual and population level for all scenarios. Furthermore, the use of expected value of partial perfect information (EVPPI) could be explored to ascertain what type of additional evidence would be most valuable. This is a research gap unlikely to have been addressed before so presents a good opportunity for detailed discussion and recommendations.

6.5.3.9 Evidence to decision framework

An evidence to decision (EtD) using DECIDE framework(276) could also be applied to my EOC research in the future to make recommendations that consider the entirety of the evidence base. An EtD is an extension beyond GRADE by widening the overall approach and considering elements outside of the SR evidence. This builds on the practical application of GRADE(74) as well as additional considerations to assess the evidence on a policy and implementation level. DECIDE is based on six domains involved in the decision-making process (D: define the problem; E: establish the criteria; C: consider all the alternatives; I: identify the best alternative; D: develop and implement a plan of action; E: evaluate and monitor the solution and feedback when necessary). This approach utilises the five main considerations used in the GRADE approach(74) (study limitations (risk of bias), unexplained heterogeneity and inconsistency of effect, imprecision, indirectness, and PB), along with the wider certainty of the body of evidence, considering other aspects that will affect policy and implementation.

While the opportunity to conduct an EtD did not materialise during the completion of my body of work, I did lead on the recommendation development process of a technical working group for an EtD framework for the British Ivermectin Recommendation Development (BIRD) steering group that utilised the DECIDE approach.(276) The purpose of this working group was to reach consensus for decision making and I use this experience to outline the potential for repeating a similar exercise in my EOC research if the opportunity presents itself in the future.(277)

The design of an EtD exercise should involve a formal panel of clinical experts, methodologists and stakeholders that will evaluate the evidence to address priority questions. The target audience in an EOC setting should include national and local policymakers, health care professionals, implementers, patients, and the public. An EtD exercise should utilise standard procedures for guideline development and could be developed as described in the World Health Organization (WHO) Handbook for Guideline Development (2014).(278) The following procedures should be given consideration:

(i) Identification of priority questions and outcomes;

(ii) Evidence retrieval and synthesis;

(iii) Assessment of the evidence;

(iv) Formulation of the recommendation; and

(v) Planning for implementation, dissemination, impact evaluation and updating.

The GRADE approach for quantitative evidence should be applied, to ensure the quality of the scientific evidence that forms the basis of the recommendation. If applied to advanced EOC research in the future each of the following criteria should be examined:

- Intervention effects: the benefits and harms associated with the intervention for advanced EOC;
- Values: the importance that those affected by advanced EOC assign to the outcomes associated with the intervention;
- Resources: the resource implications (costs and cost-effectiveness) of the intervention implementation;
- Equity: considers the health equity implications associated with the intervention;
- Acceptability: how acceptable the intervention would be to relevant stakeholders, including health care workers and patients; and
- Feasibility: how feasible it would be to implement.

Any panel would make judgements on these different decision-making criteria, which would lead to possible recommendations which could be endorsing or rejecting the intervention based on the wider evidence or offering a conditional recommendation in certain contexts, such as a research context or specific populations or settings.

The added time, cost, and complexity of an EtD framework could be viewed as too burdensome and labour intensive compared with using the GRADE approach for assessing the certainty of the evidence alone. Healthcare decisions tend to be complex and any system for moving from evidence

to decisions requires a balance between simplicity and fully transparent consideration of all important factors.(276, 279) Although EtD frameworks are complex, they can add clarity and make the judgments underlying a decision more explicit.(279)

6.5.3.10 Summary

This section has outlined various limitations of the work reported in this thesis, as well as identifying gaps in the evidence base that remain, and from these the future research needs. Several extensions to existing methods to consolidate the standard GRADE approach(74) have been proposed in this section. The rationale for consolidating existing domains or introducing additional ones was to strengthen the confidence in the overall GRADE judgements being made. I proposed that all of the suggested methods should be considered as additional items within an existing or additional GRADE domain, rather than reported separately. Within the GRADE framework, this would potentially minimise independent judgemental decisions and raise the confidence in the overall conclusions. Incorporating additional items as part of GRADE such as a TSA,(255-257) VOI,(274, 275) expert elicitation,(132) or some other methodological component, especially any that adds an element of objectivity, is worthy of consideration to methodologists who may want to further develop the GRADE tool.(74)

6.6 Conclusions

The body of work described in this thesis includes ES and MA methodology, applied in both traditional and novel ways. The research that I conducted extends beyond standard approaches to develop, explore, and apply methods that aimed to raise the certainty of the evidence. The aim of my thesis was to evaluate the role of surgery in advanced EOC, where ES methodology was applied and developed to help establish the evidence base to support practice guidelines.

A combination of chemotherapy and debulking surgery with maximal tolerable effort, is standard treatment for women with advanced EOC. The timing of surgery appears to have little or no effect on survival outcomes for the overall population. Publication 1 found evidence to support the role of PDS as treatment for advanced (stage IIIc/ IV) ovarian cancer, where achieving a macroscopic debulk to NMRD can be reasonably expected. IDS may be a reasonable (or preferred) alternative for women with stage IV disease, poor performance status or co-morbidities. Compared to PDS, IDS may increase the rate of macroscopic cytoreduction, but this does not appear to translate into an increase in OS. Evidence on extent of surgery which was examined in Publication 2 was much less certain as it was limited and provided only very low-certainty evidence. In order to support existing guidelines, the role of maximal effort debulking surgery in the management of advanced stage EOC could be addressed through the conduct of a sufficiently powered RCT comparing maximal effort debulking with standard surgery, or a well-designed non-randomised study, if a RCT is not possible.

It is acknowledged that there is considerable variation in practice with respect to debulking to small volume disease (achieving NMRD or near-optimal cytoreduction to < 1 cm) after primary surgery between different surgeons and centres. Predicting surgical debulking to small volume disease prior to surgery will be dependent on this variation. This is likely to present problems when attempting to develop models which aim to predict prognosis. This affects decisions on whether to perform PDS or IDS as the initial surgery. At present this decision is dictated by clinician preference.

NMRD remains a key prognosticator of survival in advanced EOC. This result was largely consolidated through the research that used elicitation methods to adjust the effect estimates in the NMA for PB. The series of sensitivity analyses that used various adjustments showed that the result of the primary unadjusted MA holds. Moving forward, PB should be considered in the interpretation of any MA and an expert elicitation exercise could be used as one of several ways of obtaining external information to make an adjustment. According to the body of research, whether PDS or IDS is the primary treatment, the surgical goal should be to completely remove all visible disease. Although RD of < 1 cm (near-optimal) improves survival, as shown in this thesis, this is not clear following IDS. The survival advantage should be considered with any potential morbidity or AE trade-offs.

The concepts of the expert elicitation exercise in my body of research should be strongly considered for utilisation in other meta-analyses, particularly in areas of oncology. Empirical evidence that incorporates expert elicitation in areas of uncertainty may assist in the development of clinical guidelines, enabling the disadvantages of contemporary statistical methodologies to be combined with previously implicit expert consensus. The findings in the thesis may inform clinical guidelines and assist the shared decision-making process between patients, carers, and clinicians in routine practice on selecting the most appropriate choice of primary surgical approach for women with advanced EOC. This work may represent “the best available evidence” at this time while trials are conducted. Waiting for ‘perfect’ evidence will only further delay decisions regarding the delivery of health care in the management of advanced EOC. Other methods explored in this thesis (but not confined to the included publications) aimed at raising the certainty of the evidence, such as a TSA and EtD framework. A TSA as an outright analysis is designed to determine whether or not further evidence is required, but I suggest in future work that this is utilised as an additional component to the GRADE domains to test how robust the judgements were. These methods should be related to future advanced EOC research when sufficient data are available and methodological advances are made to incorporate time-to-event outcomes.

The evidence presented in the thesis and from identification of wider clinical guidelines in advanced EOC suggests there is a clear benefit of achieving cytoreduction to NMRD after primary surgery. Other guidelines(215) make similar recommendations based upon the evidence on RD as a prognostic factor for advanced EOC after primary surgery, but they do not report the most reliable estimates. They also appear to overstate the volume and strength of evidence on maximal effort debulking surgery in the interventional ES that they identified. The added value of my body of work in this thesis is that it demonstrates that there is almost a complete absence of good quality evidence on maximal effort debulking surgery. My research however found a very strong association between cytoreduction to NMRD and increased survival. Such evidence may encourage the surgical community to attempt to increase the rate that maximal effort debulking is attempted. My research can be focused on exploring whether the extent of tumour debulking was a prognostic factor predicting survival. This in turn has been interpreted as a proxy for the expected impact of maximal effort debulking surgery on overall survival, as such data has not hitherto been (and is unlikely) to come from the future RCTs of one debulking strategy over another. The body of work in the thesis also showed that it may not matter if the form of debulking surgery was PDS or IDS if the outcome was cytoreduction to NMRD. However, my research showed that if it is unlikely that the tumour could be cytoreduced to NMRD, but it was plausible it could be debulked to near-optimal (RD<1cm), then PDS is likely to be the optimal choice of treatment. The thesis found evidence that if a tumour is not completely debulked to NMRD after IDS then all other RD thresholds may be sub-optimal. While this needs further exploration, this finding has not been explored or reported in any guidelines.

Arguably, NICE should consider the body of evidence in the thesis in their pending guideline update(48) as documented in Section 6.4.2. They should also consider the adoption of some of the purposed methods that aimed to raise the confidence in the certainty of the evidence.
In an absence of any firm guidance which has remained largely unchanged in the last decade, there is a need to be more directive about treatment rather than relying solely on the treatment

preference of the clinicians performing the surgery. An EtD framework exercise may also be worth consideration by NICE. This approach could add further strength to the overall confidence in the findings drawn from the evidence base. It could also widen the scope by introducing the six DECIDE domains(276) which could be used to address priority questions about maximal debulking surgery. These questions may ultimately dictate whether the evidence from the thesis gets implemented because policy decisions are likely to be dependent on factors beyond the effectiveness evidence base. The most obvious of these being costs and available resources within the NHS.

References

1. Hiu S, Bryant A, Gajjar K, Kunonga PT, Naik R. Ultra-radical (extensive) surgery versus standard surgery for the primary cytoreduction of advanced epithelial ovarian cancer. *Cochrane Database of Systematic Reviews*. 2022(8). doi: 10.1002/14651858.CD007697.pub3.

2. Coleridge SL, Bryant A, Kehoe S, Morrison J. Neoadjuvant chemotherapy before surgery versus surgery followed by chemotherapy for initial treatment in advanced ovarian epithelial cancer. *Cochrane Database of Systematic Reviews*. 2021(7). doi: 10.1002/14651858.CD005343.pub6.

3. Bryant A, Hiu S, Kunonga P, Gajjar K, Craig D, Vale L, et al. Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery. *Cochrane Database of Systematic Reviews*. 2022(9). doi: 10.1002/14651858.CD015048.pub2.

4. Kurman RJ, Carcangiu ML, Herrington CS. WHO classification of tumours of female reproductive organs. 4 ed. Lyon: WHO Press; 2014.

5. Webb PM, Jordan SJ. Epidemiology of epithelial ovarian cancer, Best Practice & Research. Clinical Obstetrics & Gynaecology. 2017;41:3-14.

6. Smittenaar CR, Petersen KA, Stewart K, Moitt N. Cancer incidence and mortality projections in the UK until 2035. Br J Cancer. 2016;115(9):1147-55.

7. Sasieni PD, Shelton J, Ormiston-Smith N, Thomson CS, Silcocks PB. What is the lifetime risk of developing cancer?: The effect of adjusting for multiple primaries. Br J Cancer. 2011;105(3):460-5.

8. CRUK. Ovarian Cancer Survival Statistics London, UK: Cancer Research UK,; 2018 [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer#heading-Zero].

9. Bray F, Ferlay J, Soerjomataram, I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. 2018. p. 394-424.

10. Bast RCJ, Lu Z, Han CY, Lu KH, Anderson KS, Drescher CW, et al. Biomarkers and Strategies for Early Detection of Ovarian Cancer. Cancer Epidemiol Biomarkers Prev. 2020;29(12):2504-12.

11. Kirby T. Confronting a rare ovarian cancer during lockdown. The Lancet Respiratory Medicine. 2020;8(12):1176-8.

12. Kurman RJ, Visvanathan K, Roden R, Wu TC, Shih IM. Early detection and treatment of ovarian cancer: shifting from early stage to minimal volume of disease based on a new model of carcinogenesis. American Journal of Obstetrics and Gynecology. 2008;198(4):351-6.

13. Lancet Editorial. An experiment in earlier detection of ovarian cancer. Lancet. 2007;369(9579):2051.

14. Siegel RL, Miller KD, Jemal A. Cancer statistics. CA A Cancer J Clin. 2020;70:7-30.

15. Visintin I, Feng Z, Longton G, Ward DC, Alvero AB, Lai Y, et al. Diagnostic markers for early detection of ovarian cancer. Clinical Cancer Research. 2008;14:1065-72.

16. Rani G, Bandupadhyay S, Medhi AC, Shafi F. Ovarian Cancer Screening. Journal of Medical Science and Clinical Research. 2018;6(5):1042-44.

17. American Cancer Society. Cancer Facts & Figures 2020. Atlanta. 2020 [Available from: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2020/cancer-facts-and-figures-2020.pdf].

18. CRUK. Types of ovarian cancer: Cancer Research UK; 2021 [Available from: https://www.cancerresearchuk.org/about-cancer/ovarian-cancer/types].

19. Coppleson M. Gynaecologic Oncology. London, UK: Churchill Livingstone; 1981.

20. Berek JS, Kehoe ST, Kumar L, Friedlander M. Cancer of the ovary, fallopian tube, and peritoneum. International Journal of Gynaecology and Obstetrics. 2018;143:59-78.

21. Prat J, (FIGO Committee on Gynecologic Oncology). FIGO's staging classification for cancer of the ovary, fallopian tube, and peritoneum: abridged republication. Journal of Gynecologic Oncology. 2015;26(2):87-9.

22. ONS. Cancer survival in England: adult, stage at diagnosis and childhood - patients followed up to 2018. London: Office for National Statistics (ONS); 2019.

23. Oza AM, Cook AD, Pfisterer J, Embleton A, Ledermann JA, Pujade-Lauraine E, et al. Standard chemotherapy with or without bevacizumab for women with newly diagnosed ovarian cancer (ICON7): overall survival results of a phase 3 randomised trial. Lancet Oncol. 2015;16(8):928-36.

24. Ray-Coquard I, Pautier P, Pignata S, Pérol D, González-Martín A, Berger R, et al. Olaparib plus Bevacizumab as First-Line Maintenance in Ovarian Cancer. N Engl J Med. 2019;381(25):2416-28.

25. Hoskins WJ, McGuire WP, Brady MF, Homesley HD, Creasman WT, Berman M, et al. The effect of diameter of largest residual disease on survival after primary cytoreductive surgery in patients with suboptimal residual epithelial ovarian carcinoma: a Gynecologic Oncology Group study. American Journal of Obstetrics and Gynecology. 1994;170(4):974-80.

26. Stuart GC, Kitchener H, Bacon M, et al. Gynecologic Cancer InterGroup (GCIG) consensus statement on clinical trials in ovarian cancer: report from the Fourth Ovarian Cancer Consensus Conference. International Journal of Gynecological Cancer. 2011;21(4):750–5.

27. Fotopoulou C, Hall M, Cruickshank D, Gabra H, Ganesan R, Hughes C, et al. British Gynaecological Cancer Society (BGCS) Epithelial Ovarian / Fallopian Tube / Primary Peritoneal Cancer Guidelines: Recommendations for Practice. 2017.

28. Colombo N, Sessa C, du Bois A, et al. ESMO-ESGO consensus conference recommendations on ovarian cancer: pathology and molecular biology, early and advanced stages, borderline tumours and recurrent disease. Ann Oncol. 2019;30(5):672–705.

29. National Comprehensive Cancer Network (NCCN). Clinical Practice Guidelines in Oncology, Ovarian Cancer Including Fallopian Tube Cancer and Primary Peritoneal Cancer 2020 [Available from: http://www.nccn.org/professionals/physician_gls/pdf/ovarian.pdf].

30. Tangjitgamol S, Manusirivithaya S, Laopaiboon M, Lumbiganon P, Bryant A. Interval debulking surgery for advanced epithelial ovarian cancer. *Cochrane Database of Systematic Reviews*. 2016(1). doi: 10.1002/14651858.CD006014.pub7

31. Caceres A. Colorectal stents for palliation of large bowel obstructions in recurrent gynecologic cancer: an updated series. Gynecologic Oncology. 2008;108(3):482-5.

32. Kucukmetin A, Naik R, Galaal K, Bryant A, Dickinson HO. Palliative surgery versus medical management for bowel obstruction in ovarian cancer. *Cochrane Database of Systematic Reviews*. 2010(7).

33. Tigert M, Lau C, Mackay H, L'Heureux S, Gien LT. Factors impacting length of stay and survival in patients with advanced gynecologic malignancies and malignant bowel obstruction. Int J Gynecol Cancer. 2021;31(5):727-32.

34. Lheureux S, Braunstein M, Oza AM. Epithelial ovarian cancer: Evolution of management in the era of precision medicine. CA Cancer J Clin. 2019;69:280-304.

35. Zheng F, Hao W, Ruimin L, Shuai L, Yi F, Chen X, et al. Comparison of Survival Between Primary Debulking Surgery Versus Neoadjuvant Chemotherapy for Ovarian Cancers in a Personalized Treatment Cohort. Frontiers in Oncology. 2021;10.

36. Querleu D, Planchamp F, Chiva L, et al. European Society of Gynaecological Oncology (ESGO) Guidelines for Ovarian Cancer Surgery. International Journal of Gynecologic Cancer. 2017;27:1534-42.

37. Deptuła M, Zieliński J, Wardowska A, Pikuła M. Wound healing complications in oncological patients: perspectives for cellular therapy. Postepy Dermatol Alergol. 2019;36(2):139–46.

38. Gao Y, Li Y, Zhang C, et al. Evaluating the benefits of neoadjuvant chemotherapy for advanced epithelial ovarian cancer: a retrospective study. J Ovarian Res. 2019;12(85).

39. Swart PE. Contemporary considerations for neoadjuvant chemotherapy in primary ovarian cancer. Current Oncology Reports. 2009;11:457-65.

40. Ghirardi V, Moruzzi MC, Bizzarri N, Vargiu V, D'Indinosante M. Garganese G, et al. Minimal residual disease at primary debulking surgery versus complete tumor resection at interval debulking surgery in advanced epithelial ovarian cancer: A survival analysis. Gynecologic Oncology. 2020;157(1):209-13.

41. Sato S, Itamochi H. Neoadjuvant chemotherapy in advanced ovarian cancer: latest results and place in therapy. Ther Adv Med Oncol. 2014;6(6):293-304.

42. Moschetta M, Boussios S, Rassy E, Samartzis EP, Funingana G, Uccello M. Neoadjuvant treatment for newly diagnosed advanced ovarian cancer: where do we stand and where are we going? Ann Transl Med. 2020;8(24):1710.

43. Pokhriyal R, Hariprasad R, Kumar L, Hariprasad G. Chemotherapy Resistance in Advanced Ovarian Cancer Patients. Biomark Cancer. Biomark Cancer. 2019;11:1–19.

44. Akilli H, Rahatli S, Tohma YA, Karakas LA, Altundag O, Ayhan A. Effect of increased number of neoadjuvant chemotherapy cycles on tumor resectability and pathologic response in advanced stage epithelial ovarian cancer. J BUON. 2018;23(7):111-5.

45. Schulman-Green D, Ercolano E, Dowd M, Schwartz P, McCorkle R. Quality of life among women after surgery for ovarian cancer. Palliat Support Care. 2008;6(3):239-47.

46. O'Sullivan CK, Bowles KH, Jeon S, Ercolano E, McCorkle R. Psychological Distress during Ovarian Cancer Treatment: Improving Quality by Examining Patient Problems and Advanced Practice Nursing Interventions. Nurs Res Pract. 2011;351642.

47. Vasilev S, Brar G. The Omentum and Metastatic Ovarian Cancer: Verywell Health; 2022 [Available from: https://www.verywellhealth.com/omentum-definition-and-example-2553408.

48. NICE. Maximal cytoreductive surgery for advanced ovarian cancer. Interventional procedure guidance (IPG757). National Institute for Health and Care Excellence (NICE). 05 April 2023.

49. Chang SJ, Bristow RE, Chi DS, Cliby WA. Role of aggressive surgical cytoreduction in advanced ovarian cancer. J Gynecol Oncol. 2015;26(4):336-42.

50. NICE. Ultra-radical (extensive) surgery for advanced ovarian cancer. NICE Interventional procedure guidance 470. National Institute for Health and Care Excellence (NICE); 2013.

51. Phillips A, Sundar S, Singh K, Pounds R, Nevin J, Kehoe S, et al. The NICE classification for 'Ultra-radical (extensive) surgery for advanced ovarian cancer' guidance does not meaningfully predict postoperative complications: a cohort study. BJOG. 2019;126(1):96-104.

52. Buruiana FE, Ismail L, Ferrari F, Majd HS. The Role of Ultra-Radical Surgery in the Management of Advanced Ovarian Cancer: State of the Art. Ho G, Webber K, editors. London: IntechOpen; 2021.

53. Borley J, Wilhelm-Benartzi C, Brown R, Ghaem-Maghami S. Does tumour biology determine surgical success in the treatment of epithelial ovarian cancer? A systematic literature review. Br J Cancer. 2012;107(7):1069-74.

54. Norell CH, Butler J, Farrell R, Altman A, Bentley J, Cabasag CJ, et al. Exploring international differences in ovarian cancer treatment: a comparison of clinical practice guidelines and patterns of care. Int J Gynecol Cancer. 2020;30(11):1748-56.

55. Naik R, Bayne L, Founta C, Kehoe S, Rustin G, Fotopoulou C. Patient Support Groups Identifying Clinical Equipoise in UK Gynaecological Oncology Surgeons as the Basis for Trials in Ultraradical Surgery for Advanced Ovarian Cancer. Int J Gynecol Cancer. 2016;26(1):91-4.

56. Nordin A, Jones A, Rennison R, Wakefield C, Platt M-C, Sundar S, et al. Ovarian cancer audit feasibility pilot: Disease profile in England: Incidence, mortality, stage and survival for ovary, fallopian tube and primary peritoneal carcinomas. Public Health England; 2020.

57. Onda T, Yoshikawa H, Yasugi T, Matsumoto K, Taketani Y. The optimal debulking after neoadjuvant chemotherapy in ovarian cancer: proposal based on interval look during upfront surgery setting treatment. Japanese Journal of Clinical Oncology. 2010;40(1):36-41.

58. Bristow RE, Zahurak ML, Diaz-Montes TP, et al. Impact of surgeon and hospital ovarian cancer surgical case volume on in hospital mortality and related short-term outcomes. Gynecologic Oncology. 2009;115(3):334e8.

59. Greggi S, Falcone F, Carputo R, et al. Primary surgical cytoreduction in advanced ovarian cancer: An outcome analysis within the MITO (Multicentre Italian Trials in Ovarian Cancer and Gynecologic Malignancies) Group. Gynecologic Oncology. 2016;140(3):425-9.

60. Bristow RE, Chang J, Ziogas A, Campos B, Chavez LR, Anton-Culver H. Impact of National Cancer Institute Comprehensive Cancer Centers on ovarian cancer treatment and survival. J Am Coll Surg. 2015;220(5):940-50.

61. Vernooij F, Heintz AP, Coebergh JW, Massuger LF, Witteveen PO, van der Graaf Y. Specialized and high-volume care leads to better outcomes of ovarian cancer treatment in the Netherlands. Gynecol Oncol. 2009;112(3):455-61.

62. Vernooij F, Heintz P, Witteveen E, van der Graaf Y. The outcomes of ovarian cancer treatment are better when provided by gynecologic oncologists and in specialized hospitals: a systematic review. Gynecol Oncol. 2007;105(3):801-12.

63. CGNOCG. Cochrane Gynaecological, Neuro-oncology and Orphan Cancers Group [Available from: https://gnoc.cochrane.org/].

64. Jasen P. From the "silent killer" to the "whispering disease": ovarian cancer and the uses of metaphor. Med Hist. 2009;52(4):489-512.

65. Bankhead CR, Collins C, Stokes-Lampard H, Rose P, Wilson S, Clements A, et al. Identifying symptoms of ovarian cancer: a qualitative and quantitative study. BJOG. 2008;115(8):1008-14.

66. Schorge JO, McCann C, Del Carmen MG. Surgical debulking of ovarian cancer: what difference does it make? Rev Obstet Gynecol. 2010;3(3):111-7.

67. Torre LA, Trabert B, DeSantis CE, Miller K, Samimi G, Runowicz C, et al. Ovarian cancer statistics. CA: a Cancer Journal for Clinicians. 2018;68(4):284–96.

68. Hall M, Savvatis K, Nixon K, Kyrgiou M, Hariharan K, Padwick M. Maximal-effort cytoreductive surgery for ovarian cancer patients with a high tumor burden: variations in practice and impact on outcome. Annals of Surgical Oncology. 2019;26(9):2943-51.

69. Falconer H, Joneborg U, Krawiec K, Palsdottir K, Bottai M, Salehi S. Ultra-radical upfront surgery does not improve survival in women with advanced epithelial ovarian cancer; a natural experiment in a complete population. Gynecol Oncol. 2020;159(1):58-65.

70. Marks-Anglin A, Chen Y. A historical review of publication bias. Res Syn Meth. 2020;11:725-42.

71. Deeks JJ, Higgins JPT, Altman DG. Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.3*. 2022.

72. Ropovik I, Adamkovic M, Greger D. Neglect of publication bias compromises meta-analyses of educational research. PLoS ONE. 2021;16:e0252415.

73. Page MJ, Higgins JPT, Sterne JAC. Chapter 13: Assessing risk of bias due to missing results in a synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.3*. 2022.

74. GRADE working Group. Grading quality of evidence and strength of recommendations. BMJ. 2004;328:1490-4.

75. Anttila S, Persson J, Vareman N, Sahlin N-E. Challenge of communicating uncertainty in systematic reviews when applying GRADE ratings. BMJ Evidence-Based Medicine. 2018;23:125–6.

76. Schünemann HJ, Higgins JPT, Vist GE, Glasziou P, Akl EA, Skoetz N, et al. Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.3*. 2022.

77. Schünemann H, Guyatt G, Oxman A, (editors). GRADE Handbook for Grading Quality of Evidence and Strength of Recommendations. Grading of Recommendations, Assessment, Development and Evaluation (GRADE) Working Group; 2013.

78. Hilton Boon M, Thomson H, Shaw B, Akl EA, Lhachimi SK, López-Alcalde J. Challenges in applying the GRADE approach in public health guidelines and systematic reviews: a concept article from the GRADE Public Health Group. Journal of Clinical Epidemiology. 2021;135:42-53.

79. Page MJ, Sterne JAC, Higgins JPT, Egger M. Investigating and dealing with publication bias and other reporting biases in meta-analyses of health research: A review. Res Syn Meth. 2021;12:248– 59.

80. Malmivaara A. Methodological considerations of the GRADE method. Ann Med. 2015;47(1):1-5.

81. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J. GRADE guidelines: 5. Rating the quality of evidence--publication bias. Journal of clinical epidemiology. 2011;64(12):1277-82.

82. CMPHU. Guideline on Adjustment for Baseline Covariates in Clinical Trials. In: Committee for Medicinal Products for Human Use (CMPHU). Editor: European Medicines Agency; 2015. p. 1-11.

83. Altman DG. Covariate Imbalance, Adjustment for Encyclopedia of Biostatistics; 2005.

84. Reeves BC, Deeks JJ, Higgins JPT, Shea B, Tugwell P, Wells GA. Chapter 24: Including non-randomized studies on intervention effects. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.3.* 2022.

85. McKenzie JE, Brennan SE, Ryan RE, Thomson HJ, Johnston RV, Thomas J. Chapter 3: defining the criteria for including studies and how they will be grouped for the synthesis. 2021. In: *Cochrane Handbook for Systematic Reviews of Interventions Version 6.2* [Internet].

86. Mustafa RA, Garcia CAC, Bhatt M, Riva JJ, Vesely S, Wiercioch W, et al. GRADE notes: How to use GRADE when there is "no" evidence? A case study of the expert evidence approach. J Clin Epidemiol. 2021;137:231-5.

87. Naik R, Spirtos N, Pomel C, Bristow R, Chi D, Vergote I, et al. The “definitive” trial of surgical cytoreduction in advanced stage ovarian cancer. IJGC. 2013;23(4):588-91.

88. Lunn DJ, Thomas A, Best N. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing. 2000;10:325–37.

89. Leucht S, Chaimani A, Cipriani AS, Davis JM, Furukawa TA, Salanti GP. Network meta-analyses should be the highest level of evidence in treatment guidelines. Eur Arch Psychiatry Clin Neurosci. 2016;266(6):477-80.

90. Avellar SA, Thomas J, Kleinman R, Sama-Miller E, Woodruff SE, Coughlin R, et al. External Validity: The Next Step for Systematic Reviews? Eval Rev. 2017;41(4):283-325.

91. Dwan K, Altman DG, Clarke M, Gamble C, Higgins JP, Sterne JA, et al. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. PLoS Med. 2014;11(6):e1001666.

92. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS ONE. 2008;3:e3081.

93. Dwan KK, Williamson PR, Gamble C. Selective reporting of outcomes in randomised controlled trials in systematic reviews of cystic fibrosis. BMJ open. 2013;3(6):e002709.

94. Page MJ, McKenzie JE, Higgins JPT. Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review. BMJ Open. 2018;8:e019703.

95. Zhang Z, Xu X, Ni H. Small studies may overestimate the effect sizes in critical care meta-analyses: a meta-epidemiological study. Crit Care. 2013; 17(1). doi: 10.1186/cc11919. Crit Care. 2013;17(1).

96. Williamson PR, Gamble C. Identification and impact of outcome selection bias in meta-analysis. Stat Med. 2005;24(10):1547-61.

97. Rücker G, Carpenter JR, Schwarzer G. Detecting and adjusting for small-study effects in meta-analysis. Biom J. 2011;53(2):351-68.

98. Reyes MM, Panza KE, Martin A, Bloch MH. Time-lag bias in trials of pediatric antidepressants: a systematic review and meta-analysis. J Am Acad Child Adolesc Psychiatry. 2011;50(1):63-72.

99. Stern C, Kleijnen J. Language bias in systematic reviews: you only get out what you put in. JBI Evidence Synthesis. 2020;18(9):1818-9.

100. Urlings MJE, Duyx B, Swaen GMH, Bouter LM, Zeegers MP. Citation bias and other determinants of citation in biomedical research: findings from six citation networks. Journal of Clinical Epidemiology. 2021;132:71-8.

101. Fairfield CJ, Harrison EM, Wigmore SJ. Duplicate publication bias weakens the validity of meta-analysis of immunosuppression after transplantation. World J Gastroenterol. 2017;23(39):7198-200.

102. Egger M, Smith GD. Bias in location and selection of studies. BMJ. 1998;316(7124):61-6.

103. Pignon JP, Arriagada R. Early stopping rules and long-term follow-up in phase III trials. Lung Cancer. 1994;10:S151-9.

104. Lin L. Bias caused by sampling error in meta-analysis with small sample sizes. PLoS One. 2018;13(9):e0204056.

105. Jones SR, Carley S, Harrison M. An introduction to power and sample size estimation. Emergency Medicine Journal. 2003;20:453-8.

106. Turner RM, Bird SM, Higgins JP. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. PLoS One. 2013;8(3):e59202.

107. Lexchin J. Sponsorship bias in clinical research. Int J Risk Saf Med. 2012;24(4):233-42.

108. Wareham KJ, Hyde RM, Grindlay D, et al. Sponsorship bias and quality of randomised controlled trials in veterinary medicine. BMC Vet Res. 2017;13(234).

109. Liang F, Zhu J, Mo M, Zhou CM, Jia HX, Xie L, et al. Role of industry funders in oncology RCTs published in high-impact journals and its association with trial conclusions and time to publication. Annals of Oncology. 2018;29(10):2129-34.

110. Ramsey S, Scoggins J. Commentary: Practicing on the tip of an information iceberg? Evidence of underpublication of registered clinical trials in oncology. Oncologist. 2008;13:925-9.

111. Gupta A. Fraud and misconduct in clinical research: A concern. Perspect Clin Res. 2013;4(2):144-7.

112. Bryant A, Lawrie TA, Dowswell T, Fordham EJ, Mitchell S, Hill SR, et al. Ivermectin for Prevention and Treatment of COVID-19 Infection: A Systematic Review, Meta-analysis, and Trial Sequential Analysis to Inform Clinical Guidelines. Am J Ther. 2021;28(4):e434-e60.

113. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. *Cochrane Handbook for Systematic Reviews of Interventions version 6.3*; 2022.

114. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.

115. Egger M, Smith GD, Schneider M, Minder C. Bias in Meta-Analysis Detected by a Simple, Graphical Test. BMJ: British Medical Journal. 1997;315(7109):629–34.

116. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000;56(2):455-63.

117. Hedges LV. Modeling Publication Selection Effects in Meta-Analysis. Statistical Science. 1992;7(2):246–55.

118. Mavridis DA, Sutton A, Cipriani A, Salanti G. A Fully Bayesian Application of the Copas Selection Model for Publication Bias Extended to Network Meta-Analysis. Statistics in Medicine. 2013;32(1):51–66.

119. McShane BB, Bockenholt U, Hansen KT. Adjusting for Publication Bias in Meta-Analysis: An Evaluation of Selection Methods and Some Cautionary Notes. Perspect Psychol Sci. 2016;11(5):730-49.

120. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Performance of the trim and fill method in the presence of publication bias and between study heterogeneity. Statistics in Medicine. 2007;26:4544-62.

121. Lin L, Chu H. Quantifying publication bias in meta-analysis. Biometrics. 2018;74(3):785-94.

122. Lin L, Chu H, Murad MH, et al. Empirical Comparison of Publication Bias Tests in Meta-Analysis. J Gen Intern Med. 2018;33:1260–7.

123. Hayashino Y, Noguchi Y, Fukui T. Systematic evaluation and comparison of statistical tests for publication bias. J Epidemiol. 2005;15(6):235-43.

124. Akers J, Aguiar-Ibáñez R, Baba-Akbari A. Systematic reviews: CRD's guidance for undertaking reviews in health care. York, UK: Centre for Reviews and Dissemination, University of York; 2009.

125. Morrison J, Swanton A, Kehoe S. Chemotherapy versus surgery for initial treatment in advanced ovarian epithelial cancer (protocol). *Cochrane Database of Systematic Reviews*. 2005(3).

126. Ang C, Chan KKL, Naik R, Bryant A, Dickinson HO. Ultraradical surgery for the primary debulking of epithelial ovarian cancer (protocol). *Cochrane Database of Systematic Reviews*. 2009(2).

127. Bryant A, Hiu S, Kunonga P, Gajjar K, Craig D, Vale L. Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery (protocol). *Cochrane Database of Systematic Reviews*. 2021(9).

128. Bryant A, Lawrie T, Dowsell T, et al. Ivermectin for Prevention and Treatment of Covid-19 (Protocol): The Evidence-Based Medical Consultancy Ltd; 2021 [Available from: https://tinyurl.com/cx7pnaxa].

129. PMG. Cochrane Prognosis Methods Group [Available from: https://methods.cochrane.org/prognosis/].

130. Riley RD, Moons KGM, Snell KIE, Ensor J, Hooft L, Altman DG, et al. A guide to systematic review and meta-analysis of prognostic factor studies. BMJ. 2019;364:k4597.

131. Bryant A, Grayling M, Elattar A, Gajjar K, Craig D, Vale L, et al. Residual disease after primary surgical treatment for advanced epithelial ovarian cancer, Part 2: Network meta-analysis incorporating expert elicitation to adjust for publication bias. American Journal of Therapeutics. 2023 Jan-Feb 01;30(1):e56-e71. doi: 10.1097/MJT.0000000000001548.

132. Bryant A, Grayling M, Hiu S, Gajjar K, Johnson E, Elattar A, et al. Residual disease after primary surgery for advanced epithelial ovarian cancer: expert elicitation exercise to explore opinions about potential impact of publication bias in a planned systematic review and meta-analysis. BMJ Open. 2022;12:e060183. doi: 10.1136/bmjopen-2021-06018.

133. Bryant A, Johnson E, Grayling M, Hiu S, Ellatar A, Gajjar K, et al. Residual disease threshold after primary surgical treatment for advanced epithelial ovarian cancer, Part 1: A network meta-analysis. American Journal of Therapeutics. *2023* Jan-Feb 01;30(1):e36-e55. doi: 10.1097/MJT.0000000000001584.

134. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evid Based Med. 2016;21(4):125-7.

135. National Comprehensive Cancer Network (NCCN). Guidelines for patients: Ovarian Cancer 2021 [Available from: https://www.nccn.org/patientresources/patient-resources/guidelines-for-patients/guidelines-for-patients-details?patientGuidelineId=32].

136. NICE. Contributing to clinical guidelines – a guide for patients and carers: National Institute for Health and Care Excellence 2013 [Available from: https://www.nice.org.uk/media/default/About/NICE-Communities/Public-involvement/Developing-NICE-guidance/Factsheet-1-contribute-to-developing-clinical-guidelines.pdf].

137. DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled Clinical Trials. 1986;7:177-88.

138. Hugtenburg JG, Timmers L, Elders PJ, Vervloet M, van Dijk L. Definitions, variants, and causes of nonadherence with medication: a challenge for tailored interventions. Patient Prefer Adherence. 2013;10(7):675-82.

139. CTCAE. Common terminology criteria for adverse events. Version 5.0. In: US Department of Health and Human Services, editor. 2017.

140. Barraclough H, Simms L, Govindan R. Biostatistics Primer: What a Clinician Ought to Know: Hazard Ratios. Journal of Thoracic Oncology. 2011;6(6):978-82.

141. Lefebvre C, Glanville J, Briscoe S, Featherstone R, Littlewood A, Marshall C, et al. Chapter 4: Searching for and selecting studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.3*. 2022.

142. Li T, Higgins JPT, Deeks JJ. Chapter 5: Collecting data. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.3*. 2022.

143. Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. Chapter 8: Assessing risk of bias in a randomized trial. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.3*. 2022.

144. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ (Clinical research ed). 2011;343:d5928.

145. Cochrane. Status and expectation of implementation of Risk of Bias 2 in Cochrane intervention reviews: Cochrane; 2021 [Available from: http://community.cochrane.org/news/status-and-expectations-implementation-risk-bias-2-cochrane-intervention-reviews].

146. Review Manager (RevMan) [Computer program]. Version 5.4, The Cochrane Collaboration. 2020.

147. Alavi M, Hunt GE, Visentin DC, Watson R, Thapa DK, Cleary M. Seeing the forest for the trees: How to interpret a meta-analysis forest plot. J Adv Nurs. 2021;77:1097-101.

148. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327:557-60.

149. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, Altman DG (eds). Second ed. Systematic Reviews in Health Care: Meta-Analysis in Context, editor. London: BMJ Publication Group; 2001.

150. Debray TPA, Moons KGM, Riley RD. Detecting small-study effects and funnel plot asymmetry in meta-analysis of survival data: A comparison of new and existing tests. Res Syn Meth. 2018;9:41–50.

151. Chaimani A, Caldwell DM, Li T, Higgins JPT, Salanti G. Chapter 11: Undertaking network meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.3*. 2022.

152. Rouse B, Chaimani A, Li T. Network Meta-Analysis: An Introduction for Clinicians. Intern Emerg Med. 2017;12(1):103–11.

153. Mueller KF, Meerpohl JJ, Briel M, Antes G, von Elm E, Lang B, et al. Methods for detecting, quantifying, and adjusting for dissemination bias in meta-analysis are described. J Clin Epidemiol. 2016;80:25-33.

154. Kahale LA, Khamis AM, Diab B, Chang Y, Lopes LC, Agarwal A, et al. Potential impact of missing outcome data on treatment effects in systematic reviews: imputation study. BMJ. 2020;370:m2898.

155. Kenworthy J, et al. Systematic Review Approaches for HTA: Horses for Courses? Value in Health. 2013;16(7):A614-A5.

156. Geneletti S, Richardson S, Best N. Adjusting for selection bias in retrospective, case–control studies. Biostatistics. 2009;10(1):17–31.

157. Sterne JAC, Hernán MA, McAleenan A, Reeves BC, Higgins JPT. Chapter 25: Assessing risk of bias in a non-randomized study. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.3*. 2022.

158. Jenkins DG, Quintana-Ascencio PF. A solution to minimum sample size for regressions. PLOS ONE. 2020;15(2):e0229345.

159. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;12(355):i4919.

160. Kim JH. Multicollinearity and misleading statistical results. Korean J Anesthesiol. 2019;72(6):558-69.

161. Mustafa RA, Santesso N, Brozek J, Akl EA, Walter SD, Norman G, et al. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. Journal of clinical epidemiology. 2013;66(7):736-42.

162. Foroutan F, Guyatt G, Zuk V, Vandvik PO, Alba AC, Mustafa R, et al. GRADE Guidelines 28: Use of GRADE for the assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks. Journal of Clinical Epidemiology. 2020;121:62-70.

163. Higgins JPT. Network meta-analysis: A norm for comparative effectiveness? Lancet. 2015;386(9994):628-30.

164. StataCorp. Stata Statistical Software: Release 15. In: Stata Corporation LLC, editor. College Station, Texas.2017.

165. Shim S, Yoon BH, Shin IS, Bae JM. Network meta-analysis: application and practice using Stata. Epidemiology and Health. 2017;39:e2017047.

166. White IR. Network meta-analysis. The Stata Journal. 2015;15(4):951–85.

167. Chaimani A, Higgins JPT, Mavridis D, Spyridonos P, Salanti G. Graphical Tools for Network Meta-Analysis in STATA. PLoS ONE. 2013;8(10):e76654.

168. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Stat Med. 2002;21(11):1559-73.

169. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Notes on Subgroup Analyses and Meta-Regression. In Introduction to Meta-Analysis. In: eds M Borenstein LV Hedges, JPT Higgins and HR Rothstein, editor.: John Wiley & Sons, Ltd; 2009.

170. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials. London, UK: National Institute for Health and Care Excellence (NICE); 2014.

171. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med. 2010;29:932–44.

172. Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. BMC Medicine. 2013;11(159):1-8.

173. Mbuagbaw L, Rochwerg B, Jaeschke R, Heels-Andsell D, Alhazzani W, Thabane L, et al. Approaches to interpreting and choosing the best treatments in network meta-analyses. Syst Rev. 2017;6(1):79.

174. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med. 2015;162(11):777-84.

175. Papakonstantinou T, Nikolakopoulou A, Higgins JPT, Egger M, Salanti G. CINeMA: Software for semiautomated assessment of the confidence in the results of network meta-analysis. Campbell Systematic Reviews. 2020;16:e1080.

176. Egger M, Jüni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technology Assesment. 2003;7(1).

177. Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database of Systematic Reviews*. 2007(2):MR000010.

178. Mallett S, Hopewell S, Clarke M, editors. Grey literature in systematic reviews: The first 1000 Cochrane systematic reviews. 4th Symposium on Systematic Reviews: Pushing the Boundaries; 2002 2002; Oxford, UK.

179. Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. J Clin Epidemiol. 2001;54(10):1046-55.

180. Duval S, Tweedie R. Practical estimates of the effect of publication bias in meta-analysis. Australasian Epidemiologist. 1998;5(4):14-7.

181. Rosenthal R. The file drawer problem and tolerance for null results. Psychological Bulletin. 1979;86(3):638-41.

182. Copas JB, Malley PF. A robust P-value for treatment effect in meta-analysis with publication bias. Stat Med. 2008;27(21):4267-78.

183. Sutton AJ, Song F, Gilbody SM, Abrams KR. Modelling publication bias in meta-analysis: a review. Statistical Methods in Medical Research. 2000;9(5):421-45.

184. Sterne JAC, Egger M, Moher D. Chapter 10: Addressing reporting biases. 2011. In: *Cochrane Handbook for Systematic Reviews of Interventions* [Internet]. The Cochrane Collaboration. Version 5.1.0 [updated March 2011]. Available from: https://handbook-5-1.cochrane.org/.

185. Egger M, Smith GD, Altman DG, (Eds.). Systematic revews in health care: Meta-analysis in context. London, UK: BMJ Publishing Group; 2001.

186. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Chichester, UK: John Wiley & Sons; 2004.

187. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. Stat Methods Med Res. 2001;10(4):277-303.

188. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Methods for Meta-analysis in Medical Research. Chichester, UK: John Wiley & Sons; 2000.

189. Colson AR, Cooke RM. Expert Elicitation: Using the Classical Model to Validate Experts' Judgments. Review of Environmental Economics and Policy. 2018;12(1):113-32.

190. Cooke RM, Goossens LJH. Procedures Guide for Structured Expert Judgment. European Commission; 2000.

191. Mavridis D, Welton NJ, Sutton A, Salanti G. A selection model for accounting for publication bias in a full network meta-analysis. Stat Med. 2014;33(30):5399-412.

192. Wilson ECF, Usher-Smith JA, Emery J, Corrie PG, Walter FM. Expert Elicitation of Multinomial Probabilities for Decision-Analytic Modeling: An Application to Rates of Disease Progression in Undiagnosed and Untreated Melanoma. Value in Health. 2018;21(6):669-76.

193. Copas J, Shi JQ. Meta-analysis, funnel plots and sensitivity analysis. Biostatistics. 2000;1(3):247-62.

194. Copas JB. What works? Selectivity models and meta-analysis. Journal of the Royal Statistical Society (Series A). 1999;162(1):95-109.

195. Copas JB, Shi JQ. A sensitivity analysis for publication bias in systematic reviews. Stat Meth Med Res. 2001;10(4):251-65.

196. OCRA. What is the survival rate for Stage 3 ovarian cancer? Ovarian Cancer Research Alliance (OCRA). 2022 [Available from: https://ocrahope.org/patients/about-ovarian-cancer/staging/#:~:text=Most%20women%20diagnosed%20with%20Stage%20III%20ovarian%20cancer%20have%20a,survival%20rate%20of%20approximately%2039%25].

197. American cancer society. Survival Rates for Ovarian Cancer [Available from: https://www.cancer.org/cancer/ovarian-cancer/detection-diagnosis-staging/survival-rates.html].

198. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.

199. Dias S, Welton NJ, Sutton AJ, et al. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. In: NICE Decision Support Unit, editor. London, UK: National Institute for Health and Care Excellence (NICE); 2014.

200. Higgins JP, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. Res Synth Methods. 2012;3:98–110.

201. Aletti GD, Dowdy SC, Gostout BS, Jones MB, Stanhope CR, Wilson TO. Aggressive surgical effort and improved survival in advanced-stage ovarian cancer. Obstetrics and Gynecology. 2006;107(1):77-85.

202. Chang SJ, Bristow RE, Ryu HS. Impact of complete cytoreduction leaving no gross residual disease associated with radical cytoreductive surgical procedures on survival in advanced ovarian cancer. Annals of surgical oncology. 2012;19(13):4059-67.

203. Luyckx M, Leblanc E, Filleron T, Morice P, Darai E, Classe JM. Maximal cytoreduction in patients with FIGO stage IIIC to stage IV ovarian, fallopian, and peritoneal cancer in day-to-day

practice: a Retrospective French Multicentric Study. International journal of gynecological cancer. 2012;22(8):1337-43.

204. Heitz F, Harter P, Alesina PF, Walz MK, Lorenz D, Groeben H. Pattern of and reason for postoperative residual disease in patients with advanced ovarian cancer following upfront radical debulking surgery. Gynecologic Oncology. 2016;141(2):264-70.

205. Kessous R, Laskov I, Abitbol J, Bitharas J, Yasmeen A, Salvador S. Clinical outcome of neoadjuvant chemotherapy for advanced ovarian cancer. Gynecologic Oncology. 2017;144(3):474-9.

206. Keyver-Paik MD, Abramian A, Domröse C, Döser A, Höller T, Friedrich M. Integrated care in ovarian cancer "IgV Ovar": results of a German pilot for higher quality in treatment of ovarian cancer. Journal of Cancer Research and Clinical Oncology. 2016;142(2):481-7.

207. Lee YY, Lee JW, Lu L, Xu W, Kollara A, Brown T. Impact of interval from primary cytoreductive surgery to initiation of adjuvant chemotherapy in advanced epithelial ovarian cancer. International Journal of Gynaecology and Obstetrics. 2018;143(3):325-32.

208. McGuire WP, Hoskins WJ, Brady MF, Kucera PR, Partridge EE, Look KY. Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer. New England Journal of Medicine. 1996;334(1):1-6.

209. Omura GA, Bundy BN, Berek JS, Curry S, Delgado G, Mortel R. Randomized trial of cyclophosphamide plus cisplatin with or without doxorubicin in ovarian carcinoma: a Gynecologic Oncology Group Study. Journal of Clinical Oncology. 1989;7(4):457-65.

210. Raspagliesi F, Bogani G, Matteucci L, Casarin J, Sabatucci I, Tamberi S. Surgical efforts might mitigate diGerence in response to neoadjuvant chemotherapy in stage IIIC– IV unresectable ovarian cancer: a case-control multiinstitutional study. International Journal of Gynecologic Cancer. 2018;28(9):1706-13.

211. Stewart JM, Tone AA, Jiang H, Bernardini MQ, Ferguson S, Laframboise S. The optimal time for surgery in women with serous ovarian cancer. Canadian Journal of Surgery. 2016;59(4):223-32.

212. Spiegelhalter DJ. Incorporating Bayesian Ideas into Health-Care Evaluation. Statistical Science. 2004;19(1):156–74.

213. Kehoe S, Hook J, Nankivell M, Jayson GC, Kitchener H, Lopes T. Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer (CHORUS): an open-label, randomised, controlled, non-inferiority trial. Lancet. 2015;386(9990):249-57.

214. Vergote I, Trope CG, Amant F, Kristensen GB, Ehlen T, Johnson N. Neoadjuvant chemotherapy or primary surgery in stage IIIC or IV ovarian cancer. New England Journal of Medicine. 2010;363(10):943-53.

215. Vergote I, Vlayen J, Heus P, Hoogendam JP, Damen JA, Van de Wetering FT, et al. Ovarian cancer: diagnosis, treatment and follow-up: KCE Report 268. In: Belgian Health Care knowledge Centre, editor. 2016.

216. Higgins J, Lasserson T, Chandler J, Tovey D, Thomas J, Flemyng E, et al. Methodological Expectations of Cochrane Intervention Reviews (MECIR) Standards for the conduct and reporting of new Cochrane Intervention Reviews, reporting of protocols and the planning, conduct and reporting of updates 2021 [Available from: https://methods.cochrane.org/methodological-expectations-cochrane-intervention-reviews]

217. Shim SR, Kim SJ, Lee J, Rücker G. Network meta-analysis: application and practice using R software. Epidemiol Health. 2019;41:e2019013.

218. Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. BMJ. 2001;323(7304):101-5.

219. van Aert RCM, Wicherts JM, van Assen MALM. Publication bias examined in meta-analyses from psychology and medicine: A meta-meta-analysis. PLoS One. 2019;14(4):e0215052.

220. Gosling JP. SHELF: The Sheffield Elicitation Framework. 2017. Springer International Publishing. International Series in Operations Research & Management Science; [61-94].

221. Sundar S, Cummins C, Kumar S, Long J, Arora V, Balega J, et al. Quality of life from cytoreductive surgery in advanced ovarian cancer: Investigating the association between disease

burden and surgical complexity in the international, prospective, SOCQER-2 cohort study. BJOG. 2022;129(7):1122-32.

222. Horowitz NS, Miller A, Rungruang B. Does aggressive surgery improve outcomes? Interaction between preoperative disease burden and complex surgery in patients with advanced-stage ovarian cancer: an analysis of GOG 182. Journal of Clinical Oncology. 2015;33:937–43.

223. Cummins C, Kumar S, Long J, Balega J, Broadhead T, Edmondson R, et al. Investigating impact of ultra-radical surgery on survival in advanced ovarian cancer using population based data in a multicentre UK study. Cancers. 2022;4362.

224. Tseng JH, Cowan RA, Zhou Q. Continuous improvement in primary debulking surgery for advanced ovarian cancer: do increased complete gross resection rates independently lead to increased progression-free and overall survival? Gynecological Oncology. 2018;151:24–31.

225. Angeles MA, Hernandez A, Perez-Benavente A. The effect of major postoperative complications on recurrence and long-term survival after cytoreductive surgery for ovarian cancer. Gynecologic Oncology. 2022;166:8–17.

226. Palmqvist C, Michaelsson H, Staf C. Complications after advanced ovarian cancer surgery-A population-based cohort study. Acta Obstetricia et Gynecologica Scandinavica. 2022;101:747–57.

227. Ekmann-Gade AW, Schnack TH, Seibaek L. Impact of surgery and chemotherapy timing on outcomes in older versus younger epithelial ovarian cancer patients: A nationwide Danish cohort study. Journal of Geriatric Oncology. 2022;14(1); 101359.

228. Ehmann S, Aviki EM, Sonoda Y. Diaphragm hernia after debulking surgery in patients with ovarian cancer. Gynecologic Oncology. 2021;Reports 36(100759).

229. Dahm-Kähler P, Holmberg E, Holtenman M. Implementation of National Guidelines increased survival in advanced ovarian cancer - A population-based nationwide SweGCG study. Gynecologic Oncology. 2021;161:244–50.

230. Di Donato V, Kontopantelis E, Aletti G. Trends in mortality after primary cytoreductive surgery for ovarian cancer: a systematic review and metaregression of randomized clinical trials and observational studies. Annals of Surgical Oncology. 2017;24:1688–97.

231. Kengsakul M, Nieuwenhuyzen-de Boer G, Udomkarnjananun S. Factors predicting postoperative morbidity after cytoreductive surgery for ovarian cancer: a systematic review and meta-analysis. Journal of Gynecologic Oncology. 2022;33:e53.

232. Kumar L, Hariprasad R, Kumar S, Bhatla N, Thulkar S, Shukla NJ. Neo-adjuvant chemotherapy in advanced epithelial ovarian cancer (EOC): a prospective, randomized study. Journal of Medical and Paediatric Oncology. 2009;30(1):15.

233. Reuss A, Du Bois A, Harter P, Fotopoulou C, Sehouli J, Aletti G, et al. TRUST: Trial of Radical Upfront Surgical Therapy in advanced ovarian cancer (ENGOT ov33/AGO-OVAR OP7). Journal of Gynecological Cancer. 2019;29(8):1327-31.

234. NCT04257786. Primary cyto-reductive surgery vs neoadjuvant chemotherapy (NAC) in epithelial ovarian cancer [Available from: Clinicaltrials.gov].

235. NCT04515602. Stratified evaluation of PDS and NACT-IDS in ovarian cancer [Available from: Clinicaltrials.gov].

236. SUNNY (NCT02859038). Study of upfront surgery versus neoadjuvant chemotherapy in patients with advanced ovarian cancer (SUNNY) [Available from: ClinicalTrials.gov.

237. Ruta DA, Garratt AM, Leng M, Russell IT, MacDonald LM. A new approach to the measurement of quality of life. The Patient-Generated Index. Med Care. 1994;32(11):1109-26.

238. Moons KGM, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart. 2012;98:683-90.

239. Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. 2012;98(9):691-8.

240. Lewis G, Curry N, Bardsley M. Choosing a predictive risk model: a guide for commissioners in England. In: Nuffield Trust, editor. London, UK. 2011.

241. Steyerberg EW. Clinical prediction models. USA, New York: Springer; 2009.

242. Debray TP, Koffijberg H, Nieboer D, et al. Meta-analysis and aggregation of multiple published prediction models. Stat Med. 2014;33(14):2341-62.

243. Newcastle Clinical Trials Unit. Patient and Public Involvement and Engagement (PPIE) Newcastle University, UK. 2022 [Available from: https://www.ncl.ac.uk/nctu/services/patient/].

244. NIHR. 5-year Strategy for Patient and Public Involvement Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology: National Institute for Health Research (NIHR); 2022 [Available from: https://www.moorfields.nhs.uk/sites/default/files/2017-2022%20five%20year%20PPIE%20strategy%20for%20Moorfields%20BRC%20%26%20CRF.pdf].

245. UCL. Patient and Public Involvement and Engagement (PPIE) University College London (UCL), UK: Institute of Epidemiology & Health Care; [Available from: https://www.ucl.ac.uk/epidemiology-health-care/research/primary-care-and-population-health/research/methodological-themes/patient-and-public].

246. Ovacome. Ovacome Ovarian Cancer London, UK. 2022 [Available from: https://www.ovacome.org.uk/].

247. TOC. Target Ovarian Cancer London, UK. 2022 [Available from: https://targetovariancancer.org.uk/].

248. CRUK. Cancer Research UK London, UK. 2022 [Available from: https://www.cancerresearchuk.org/].

249. OCASD. Ovarian Cancer Alliance of San Diego San Diego, USA. 2022 [Available from: https://www.ocaofsd.org/].

250. Bryan AM. Coronado Philanthropist and Young Scientists Working Together. The Coronado Times. 2016.

251. Cochrane. Cochrane consumer network London, UK. 2022 [Available from: https://consumers.cochrane.org/].

252. NIHR. Payment guidance for researchers and professionals. National Institute for Health and Research (NIHR); 2022.

253. INVOLVE. Policy on payment of fees and expenses for members of the public actively involved with INVOLVE. In: National Instutite for Health Research (NIHR), editor. Southampton, UK. 2016.

254. INVOLVE. Starting Out – essential information for members of the public getting started in involvement. Southampton, UK. 2017.

255. Brok J, Thorlund K, Gluud C, et al. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. J Clin Epidemiol. 2008;61:763–9.

256. Wetterslev J, Thorlund K, Brok J, et al. Estimating required information size by quantifying diversity in random-effects model meta-analyses. BMC Med Res Methodol. 2009;9(86).

257. Thorlund K, Engstrøm J, Wetterslev J, et al. User Manual for Trial Sequential Analysis (TSA). 2011.

258. Ma LL, Wang YY, Yang ZH, et al. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? Military Med Res. 2020;7(7).

259. Martin F, Camfield L, Rodham K, Kliempt P, Ruta D. Twelve years' experience with the Patient Generated Index (PGI) of quality of life: a graded structured review. Qual Life Res. 2007;16(4):705-15.

260. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ. 2010;340:c221.

261. Riley RD, Debray TPA, Fisher D, Hattle M, Marlin N, Hoogland J, et al. Individual participant data meta-analysis to examine interactions between treatment effect and participant-level covariates: Statistical recommendations for conduct and planning. Stat Med. 2020;39(15):2115-37.

262. Tierney JF, Stewart LA, Clarke M. Chapter 26: Individual participant data. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors): Cochrane; 2022 [Available from: www.training.cochrane.org/handbook].

263. Garner P, Hopewell S, Chandler J, MacLehose H, Schünemann HJ, Akl EA, et al. Panel for Updating Guidance for Systematic Reviews (PUGs). When and how to update systematic reviews: consensus and checklist. BMJ. 2016;354: i3507.

264. Cumpston M, Chandler J. Chapter IV: Updating a review. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.3* (updated February 2022). Cochrane. 2022.

265. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living Systematic Review Network. Living systematic review: 1. Introduction-the why, what, when, and how. Journal of Clinical Epidemiology. 2017;91:23-30.

266. Clarke M, Brice A, Chalmers I. Accumulating Research: A Systematic Account of How Cumulative Meta-Analyses Would Have Provided Knowledge, Improved Health, Reduced Harm and Saved Resources. PLoS ONE. 2014;9(7):e102670.

267. Jennison C, Turnbull BW. Group sequential methods with applications to clinical trials: Chapman & Hall; 1999.

268. Pallmann P, Bedding A, Choodari-Oskooei B, Dimairo M, Flight L, Hampson L, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. BMC Medicine. 2018;16(29).

269. Messori A, Fadda V, Maratea D, Trippoli S. Maintenance Chemotherapy in Ovarian Cancer: A Trial-Sequential Analysis. Journal of Cancer Therapy. 2013;4(7):1242-3.

270. Schmid C, Senn S, Sterne J, Kulinskaya E, Posch M, Roes K, et al. Should Cochrane apply error-adjustment methods when conducting repeated meta-analyses? : Expert Panel Recommendation and guidance points; 2018.

271. Wasserstein RL, Schirm AL, Lazar NA. Moving to a World Beyond "p < 0.05". The American Statistician. 2019;73:1-19.

272. Simmonds M, Salanti G, McKenzie J, Elliott J. Living Systematic Review Network. Living systematic reviews: 3. Statistical methods for updating meta-analyses. J Clin Epidemiol. 2017;91:38-46.

273. Higgins JP, Whitehead A, Simmonds M. Sequential methods for random-effects meta-analysis. Stat Med. 2011;30(9):903-21.

274. Briggs A, Claxton C, Sculpher M. Decision Modelling for Health Economic Evaluation. Oxford: Oxford University Press; 2006.

275. Claxton KP, Sculpher MJ. Using value of information analysis to prioritise health research: some lessons from recent UK experience. Pharmacoeconomics. 2006;24(11):1055-68.

276. Guo K. DECIDE: A Decision-Making Model for More Effective Decision Making by Health Care Managers. Health Care Manag (Frederick). 2020;39(3):133-41.

277. BIRD Steering Group. The BIRD Recommendation on the Use of Ivermectin for Covid-19: Executive Summary: British Ivermectin Recommendation Development; [Available from: https://bird-group.org/wp-content/uploads/2021/03/bird-executive-summary-1.pdf].

278. WHO. WHO handbook for guideline development, 2nd ed. In: World Health Organization, editor. 2014.

279. Moberg J, Oxman A, Rosenbaum S, et al. The GRADE Evidence to Decision (EtD) framework for health system and public health decisions. Health Res Policy Sys. 2018;16(45).

Appendices

Appendix 1: My role and contribution to each included publication in the thesis

Coleridge SL, **Bryant A**, Kehoe S, Morrison J. Neoadjuvant chemotherapy before surgery versus surgery followed by chemotherapy for initial treatment in advanced ovarian epithelial cancer. *Cochrane Database of Systematic Reviews*. 2021;7. doi: 10.1002/14651858.CD005343.pub6. https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD005343.pub6/epdf/full

Hiu S*, **Bryant A***, Gajjar K, Kunonga PT, Naik R. Ultra-radical (extensive) surgery versus standard surgery for the primary cytoreduction of advanced epithelial ovarian cancer. *Cochrane Database of Systematic Reviews*. 2022;8. doi: 10.1002/14651858.CD007697.pub3. *Joint first author https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD007697.pub3/epdf/full

Bryant A, Hiu S, Kunonga PT, Gajjar K, Craig D, Vale L, Winter-Roach BA, Elattar A, Naik R. Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery. *Cochrane Database of Systematic Reviews* 2022;9. doi: 10.1002/14651858.CD015048.pub2. https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD015048.pub2/epdf/full

***Bryant A**, Johnson E, Grayling M, Hiu S, Elattar A, Gajjar K, Craig D, Vale L, Naik R.* Residual disease threshold after primary surgical treatment for advanced epithelial ovarian cancer, Part 1: A systematic review and network meta-analysis. *American Journal of Therapeutics. 2023*. Jan-Feb 01;30(1):e36-e55. doi: 10.1097/MJT.0000000000001584. https://journals.lww.com/americantherapeutics/Fulltext/2023/02000/Residual_Disease_After_Primary_Surgical_Treatment.6.aspx

***Bryant A**, Grayling M, Hiu S, Elattar A, Johnson E, Gajjar K, Craig D, Vale L, Naik R.* Residual disease after primary surgery for advanced epithelial ovarian cancer: expert elicitation exercise to explore opinions about potential impact of publication bias in a planned systematic review and meta-analysis. *BMJ Open*. 2022;12:e060183. doi: 10.1136/bmjopen-2021-060183 https://bmjopen.bmj.com/content/bmjopen/12/8/e060183.full.pdf

***Bryant A**, Grayling M, Elattar A, Gajjar K, Craig D, Vale L, Naik R.* Residual disease after primary surgical treatment for advanced epithelial ovarian cancer, Part 2: Network

meta-analysis incorporating expert elicitation to adjust for publication bias. *American Journal of Therapeutics*. 2023. Jan-Feb 01;30(1):e56-e71. doi: 10.1097/MJT.0000000000001548. https://journals.lww.com/americantherapeutics/fulltext/2023/02000/residual_disease_after_primary_surgical_treatment.6.aspx

SUBMISSION BY STAFF CANDIDATES FOR THE DEGREE OF PHD BY PUBLISHED WORK

CO-AUTHORSHIP FORM

This form must accompany any submission of a joint authored publication for the degree of Doctor of Philosophy on the basis of published work.

A candidate should submit a separate form for each jointly authored work which is submitted for the degree.

TITLE OF PUBLICATION (article, book, chapter, monograph)

Neoadjuvant chemotherapy before surgery versus surgery followed by chemotherapy for initial treatment in advanced ovarian epithelial cancer

DATE OF PUBLICATION 2021

NAME AND VOLUME OF JOURNAL (where appropriate)

Cochrane Database of Systematic Reviews

PUBLISHER (for book, chapter or monograph)

EDITORS (chapter only)

ISBN (where appropriate)

If the work has not been published but has been accepted for publication please attach a statement from the Editor or Publisher which confirms the intention to publish the work.

NAMES OF JOINT AUTHORS	INSTITUTION
Bryant A	Newcastle University
1. Morrison J	Somerset NHS Foundation Trust
2. Coleridge SL*	Nottingham University Hospitals NHS Trust
3. Kehoe S	University of Birmingham

*First author

CONTRIBUTION OF THE CANDIDATE TO THIS WORK (%)

Design of investigation	10
Conduct of research	30
Analysis of outcome	50
Preparation for publication	30
TOTAL	30 *(To be an average of, or at least consistent with, the above figures)*

This statement should be endorsed by all of the co-authors.

I confirm that the above is a true estimate of the candidate's contribution to this work.

Signature 1 J. M

Signature 2 Dr Sarah Coleridge Sarah Coleridge

Signature 3 SEAN KEHOE Le Locke

SUBMISSION BY STAFF CANDIDATES FOR THE DEGREE OF PHD BY PUBLISHED WORK

CO-AUTHORSHIP FORM

This form must accompany any submission of a joint authored publication for the degree of Doctor of Philosophy on the basis of published work.

A candidate should submit a separate form for each jointly authored work which is submitted for the degree.

TITLE OF PUBLICATION (article, book, chapter, monograph)

Ultra-radical (extensive) surgery versus standard surgery for the primary cytoreduction of advanced epithelial ovarian cancer

DATE OF PUBLICATION 2022

NAME AND VOLUME OF JOURNAL (where appropriate)

Cochrane Database of Systematic Reviews

PUBLISHER (for book, chapter or monograph)

EDITORS (chapter only)

ISBN (where appropriate)

If the work has not been published but has been accepted for publication please attach a statement from the Editor or Publisher which confirms the intention to publish the work.

NAMES OF JOINT AUTHORS	INSTITUTION
Bryant A*	Newcastle University
1. Naik R	Gateshead Health NHS Foundation Trust
2. Hui S*	Newcastle University
3. Kunonga PT	Newcastle University
4. Gajjar K	Nottingham University Hospitals NHS Trust

*Joint first authors

CONTRIBUTION OF THE CANDIDATE TO THIS WORK (%)

Design of investigation	30
Conduct of research	40
Analysis of outcome	50
Preparation for publication	50
TOTAL	45 *(To be an average of, or at least consistent with, the above figures)*

This statement should be endorsed by all of the co-authors.

I confirm that the above is a true estimate of the candidate's contribution to this work.

Signature 1 _____ Khan Raj WALK

Signature 2 _____ Shaun Hill

Signature 3 _____ TAFADZWA PATIENCE KUNONGA

Signature 4 _____ KETAN GAJJAR

SUBMISSION BY STAFF CANDIDATES FOR THE DEGREE OF PHD BY PUBLISHED WORK

CO-AUTHORSHIP FORM

This form must accompany any submission of a joint authored publication for the degree of Doctor of Philosophy on the basis of published work.

A candidate should submit a separate form for each jointly authored work which is submitted for the degree.

TITLE OF PUBLICATION (article, book, chapter, monograph)

Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery

DATE OF PUBLICATION 2022

NAME AND VOLUME OF JOURNAL (where appropriate)

Cochrane Database of Systematic Reviews

PUBLISHER (for book, chapter or monograph)

EDITORS (chapter only)

ISBN (where appropriate)

If the work has not been published but has been accepted for publication please attach a statement from the Editor or Publisher which confirms the intention to publish the work.

NAMES OF JOINT AUTHORS	INSTITUTION
Bryant A*	Newcastle University
1. Naik R	Gateshead Health NHS foundation Trust
2. Vale L	Newcastle University
3. Hiu S	Newcastle University
4. Kunonga PT	Newcastle University
5. Craig D	Newcastle University
6. Winter-Roach BA	Christie Hospital NHS Foundation Trust
7. Elattar A	Sandwell and West Birmingham NHS Trust
8. Gajjar K	Nottingham University Hospital NHS Trust

*First author

CONTRIBUTION OF THE CANDIDATE TO THIS WORK (%)

Design of investigation	40
Conduct of research	50
Analysis of outcome	70
Preparation for publication	70
TOTAL	60 *(To be an average of, or at least consistent with, the above figures)*

This statement should be endorsed by all of the co-authors.

I confirm that the above is a true estimate of the candidate's contribution to this work.

Signature 1 _____ Raj NAIK

Signature 2 _____ Luke Vale

Signature 3 _____ Shaun Hiu

Signature 4 _____ PATIENCE TAFADZWA KUNONGA.

Signature 5 _____ DAWN CRAIG

Signature 6 Brett A Winter-Roach _____ Brett A Winter-Roach

Signature 7 _____ Ahmed Elattar

Signature 8 _____ KETAN GAJJAR

SUBMISSION BY STAFF CANDIDATES FOR THE DEGREE OF PHD BY PUBLISHED WORK

CO-AUTHORSHIP FORM

This form must accompany any submission of a joint authored publication for the degree of Doctor of Philosophy on the basis of published work.

A candidate should submit a separate form for each jointly authored work which is submitted for the degree.

TITLE OF PUBLICATION (article, book, chapter, monograph)

Residual disease threshold after primary surgical treatment for advanced epithelial ovarian cancer; Part 1: A systematic review and network meta-analysis

DATE OF PUBLICATION 2022

NAME AND VOLUME OF JOURNAL (where appropriate)

American Journal of Therapeutics

PUBLISHER (for book, chapter or monograph)

_____

EDITORS (chapter only)

_____

ISBN (where appropriate)

_____

If the work has not been published but has been accepted for publication please attach a statement from the Editor or Publisher which confirms the intention to publish the work.

NAMES OF JOINT AUTHORS	INSTITUTION
*Bryant A	Newcastle University
1. Naik R	Gateshead Health NHS foundation Trust
2. Johnson E	Newcastle University
3. Vale L	Newcastle University
4. Hiu S	Newcastle University
5. Grayling M	Newcastle University
6. Elattar A	Sandwell and West Birmingham NHS Trust
7. Craig D	Newcastle University
8. Gajjar K	Nottingham University Hospitals NHS Trust

*First author

CONTRIBUTION OF THE CANDIDATE TO THIS WORK (%)

Design of investigation	60
Conduct of research	80
Analysis of outcome	90
Preparation for publication	80
TOTAL	80 *(To be an average of, or at least consistent with, the above figures)*

This statement should be endorsed by all of the co-authors.

I confirm that the above is a true estimate of the candidate's contribution to this work.

Signature 1 _____ Raj NAIK

Signature 2 _____ EUGENIE JOHNSON

Signature 3 _____ Luke VALE

Signature 4 _____ Shaun Hiu

Signature 5 _____ MICHAEL GRAYLING

Signature 6 _____ Ahmed Elattar

Signature 7 _____ DAWN CRAIG

Signature 8 _____ KETAN GAJJAR

SUBMISSION BY STAFF CANDIDATES FOR THE DEGREE OF PHD BY PUBLISHED WORK

CO-AUTHORSHIP FORM

This form must accompany any submission of a joint authored publication for the degree of Doctor of Philosophy on the basis of published work.

A candidate should submit a separate form for each jointly authored work which is submitted for the degree.

TITLE OF PUBLICATION (article, book, chapter, monograph)

Residual disease after primary surgery for advanced epithelial ovarian cancer: expert elicitation exercise to explore opinions about potential impact of publication bias in a planned systematic review and meta-analysis

DATE OF PUBLICATION 2022

NAME AND VOLUME OF JOURNAL (where appropriate)

BMJ Open

PUBLISHER (for book, chapter or monograph)

_____

EDITORS (chapter only)

_____

ISBN (where appropriate)

_____

If the work has not been published but has been accepted for publication please attach a statement from the Editor or Publisher which confirms the intention to publish the work.

NAMES OF JOINT AUTHORS	INSTITUTION
*Bryant A	Newcastle University
1. Naik R	Gateshead Health NHS foundation Trust
2. Johnson E	Newcastle University
3. Vale L	Newcastle University
4. Hiu S	Newcastle University
5. Grayling M	Newcastle University
6. Elattar A	Sandwell and West Birmingham NHS Trust
7. Craig D	Newcastle University
8. Gajjar K	Nottingham University Hospitals NHS Trust

*First author

CONTRIBUTION OF THE CANDIDATE TO THIS WORK (%)

Design of investigation	60
Conduct of research	80
Analysis of outcome	90
Preparation for publication	80
TOTAL	80 *(To be an average of, or at least consistent with, the above figures)*

This statement should be endorsed by all of the co-authors.

I confirm that the above is a true estimate of the candidate's contribution to this work.

Signature 1 _____ Raj NAIK

Signature 2 _____ EUGENIE JOHNSON

Signature 3 _____ LUKE VINE

Signature 4 _____ Shaun Hiu

Signature 5 _____ MICHAEL GRAYLING

Signature 6 _____ Ahmed Elattar

Signature 7 _____ DAWN CRAIG

Signature 8 _____ KETAN GAJJAR

SUBMISSION BY STAFF CANDIDATES FOR THE DEGREE OF PHD BY PUBLISHED WORK

CO-AUTHORSHIP FORM

This form must accompany any submission of a joint authored publication for the degree of Doctor of Philosophy on the basis of published work.

A candidate should submit a separate form for each jointly authored work which is submitted for the degree.

TITLE OF PUBLICATION (article, book, chapter, monograph)

Residual disease after primary surgical treatment for advanced epithelial ovarian cancer; Part 2: Network meta-analysis incorporating expert elicitation to adjust for publication bias

DATE OF PUBLICATION 2022

NAME AND VOLUME OF JOURNAL (where appropriate)

American Journal of Therapeutics

PUBLISHER (for book, chapter or monograph)

EDITORS (chapter only)

ISBN (where appropriate)

If the work has not been published but has been accepted for publication please attach a statement from the Editor or Publisher which confirms the intention to publish the work.

NAMES OF JOINT AUTHORS	INSTITUTION
*Bryant A	Newcastle University
1. Naik R	Gateshead Health NHS foundation Trust
2. Vale L	Newcastle University
3. Grayling M	Newcastle University
4. Elattar A	Sandwell and West Birmingham NHS Trust
5. Craig D	Newcastle University
6. Gajjar K	Nottingham University Hospitals NHS Trust

*First author

CONTRIBUTION OF THE CANDIDATE TO THIS WORK (%)

Design of investigation	90
Conduct of research	80
Analysis of outcome	70
Preparation for publication	80
TOTAL	80 *(To be an average of, or at least consistent with, the above figures)*

This statement should be endorsed by all of the co-authors.

I confirm that the above is a true estimate of the candidate's contribution to this work.

Signature 1 _____ Raj NAIK

Signature 2 _____ Luke VALE

Signature 3 _____ MICHAEL GRAYLING

Signature 4 _____ Ahmed Elattar

Signature 5 _____ DAWN CRAIG

Signature 6 _____ KETAN GAJJAR

Appendix 2: FIGO classification for ovarian cancer 2014 FIGO ovarian, fallopian tube, and peritoneal cancer staging system and corresponding TNM

Stage	Extent of tumour	Substage	Details
I (T1-N0-M0)	Tumour confined to ovaries or fallopian tube(s)	IA (T1a-N0-M0)	Tumour limited to one ovary (capsule intact) or fallopian tube; no tumour on ovarian or fallopian tube surface; no malignant cells in the ascites or peritoneal washings
		IB (T1b-N0-M0)	Tumour limited to both ovaries (capsules intact) or fallopian tubes; no tumour on ovarian or fallopian tube surface; no malignant cells in the ascites or peritoneal washings
		IC (T1c1-N0-M0)	Tumour limited to one or both ovaries or fallopian tubes, with any of the following: IC1: surgical spill IC2: capsule ruptured before surgery or tumour on ovarian or fallopian tube surface IC3: malignant cells in the ascites or peritoneal washings
II (T2-N0-M0)	Tumour involves one or both ovaries or fallopian tubes with pelvic extension (below pelvic brim) or primary peritoneal cancer	IIA (T2a-N0-M0)	Extension and/or implants on uterus and/or fallopian tubes and/or ovaries
		IIB (T2b-N0-M0)	Extension to other pelvic intraperitoneal tissues
III (T1/T2-N1-M0)	Tumour involves one or both ovaries or fallopian tubes, or primary peritoneal cancer, with cytologically or histologically confirmed spread to the peritoneum outside the pelvis and/or metastasis to the retroperitoneal lymph nodes	IIIA IIIA1: (T1/T2-N1-M0) IIIA2: (T3a2-N0/N1-M0)	IIIA1: Positive retroperitoneal lymph nodes only (cytologically or histologically proven): IIIA1(i) Metastasis up to 10 mm in greatest dimension IIIA1(ii) Metastasis more than 10 mm in greatest dimension IIIA2: microscopic extra-pelvic (above the pelvic brim) peritoneal involvement with or without positive retroperitoneal lymph nodes
		IIIB (T3b-N0/N1-M0)	Macroscopic peritoneal metastasis beyond the pelvis up to 2 cm in greatest dimension, with or without metastasis to the retroperitoneal lymph nodes
		IIIc (T3c-N0/N1-M0)	Macroscopic peritoneal metastasis beyond the pelvis more than 2 cm in greatest dimension, with or without metastasis to the retroperitoneal lymph nodes (includes extension of tumour to capsule of liver and spleen without parenchymal involvement of either organ)
IV (Any T, any N, M1)	Distant metastasis excluding peritoneal metastases	IVA	Pleural effusion with positive cytology
		IVB	Parenchymal metastases and metastases to extra-abdominal organs (including inguinal

			lymph nodes and lymph nodes outside of the abdominal cavity)

FIGO: The International Federation of Gynecology and Obstetrics; T: Extent of the primary tumour; N: Whether the cancer is present in the lymph nodes; M: whether the cancer has metastasised

Appendix 3: Publication 1: Chemotherapy versus surgery for initial treatment in advanced ovarian epithelial cancer

Cochrane Library

Cochrane Database of Systematic Reviews

Neoadjuvant chemotherapy before surgery versus surgery followed by chemotherapy for initial treatment in advanced ovarian epithelial cancer (Review)

Coleridge SL, Bryant A, Kehoe S, Morrison J

Coleridge SL, Bryant A, Kehoe S, Morrison J.
Neoadjuvant chemotherapy before surgery versus surgery followed by chemotherapy for initial treatment in advanced ovarian epithelial cancer.
Cochrane Database of Systematic Reviews 2021, Issue 7. Art. No.: CD005343.
DOI: 10.1002/14651858.CD005343.pub6.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Review]

Neoadjuvant chemotherapy before surgery versus surgery followed by chemotherapy for initial treatment in advanced ovarian epithelial cancer

Sarah L Coleridge[1], Andrew Bryant[2], Sean Kehoe[3], Jo Morrison[4]

[1]Department of Obstetrics and Gynaecology, Nottingham University Hospitals NHS Trust, Nottingham, UK. [2]Institute of Health & Society, Newcastle University, Newcastle upon Tyne, UK. [3]Institute of Cancer and Genomics, University of Birmingham, Birmingham, UK. [4]Department of Gynaecological Oncology, Musgrove Park Hospital, Taunton, UK

Contact address: Jo Morrison, jo_morrison@doctors.org.uk.

Editorial group: Cochrane Gynaecological, Neuro-oncology and Orphan Cancer Group.
Publication status and date: New search for studies and content updated (no change to conclusions), published in Issue 7, 2021.

Citation: Coleridge SL, Bryant A, Kehoe S, Morrison J. Neoadjuvant chemotherapy before surgery versus surgery followed by chemotherapy for initial treatment in advanced ovarian epithelial cancer. *Cochrane Database of Systematic Reviews* 2021, Issue 7. Art. No.: CD005343. DOI: 10.1002/14651858.CD005343.pub6.

ABSTRACT

Background

Epithelial ovarian cancer presents at an advanced stage in the majority of women. These women require a combination of surgery and chemotherapy for optimal treatment. Conventional treatment has been to perform surgery first and then give chemotherapy. However, there may be advantages to using chemotherapy before surgery.

Objectives

To assess whether there is an advantage to treating women with advanced epithelial ovarian cancer with chemotherapy before debulking surgery (neoadjuvant chemotherapy (NACT)) compared with conventional treatment where chemotherapy follows debulking surgery (primary debulking surgery (PDS)).

Search methods

We searched the following databases up to 9 October 2020: the Cochrane Central Register of Controlled Trials (CENTRAL), Embase via Ovid, MEDLINE (Silver Platter/Ovid), PDQ and MetaRegister. We also checked the reference lists of relevant papers that were identified to search for further studies. The main investigators of relevant trials were contacted for further information.

Selection criteria

Randomised controlled trials (RCTs) of women with advanced epithelial ovarian cancer (Federation of International Gynaecologists and Obstetricians (FIGO) stage III/IV) who were randomly allocated to treatment groups that compared platinum-based chemotherapy before cytoreductive surgery with platinum-based chemotherapy following cytoreductive surgery.

Data collection and analysis

Two review authors independently extracted data and assessed risk of bias in each included trial. We extracted data of overall (OS) and progression-free survival (PFS), adverse events, surgically-related mortality and morbidity and quality of life outcomes. We used GRADE methods to determine the certainty of evidence.

Main results

We identified 2227 titles and abstracts through our searches, of which five RCTs of varying quality and size met the inclusion criteria. These studies assessed a total of 1774 women with stage IIIc/IV ovarian cancer randomised to NACT followed by interval debulking surgery (IDS) or PDS followed by chemotherapy. We pooled results of the four studies where data were available and found little or no difference with regard to overall survival (OS) (Hazard Ratio (HR) 0.96, 95% CI 0.86 to 1.08; participants = 1692; studies = 4; high-certainty evidence) or progression-free survival in four trials where we were able to pool data (Hazard Ratio 0.98, 95% CI 0.88 to 1.08; participants = 1692; studies = 4; moderate-certainty evidence).

Adverse events, surgical morbidity and quality of life (QoL) outcomes were variably and incompletely reported across studies. There are probably clinically meaningful differences in favour of NACT compared to PDS with regard to overall postoperative serious adverse effects (SAE grade 3+): 6% in NACT group, versus 29% in PDS group, (risk ratio (RR) 0.22, 95% CI 0.13 to 0.38; participants = 435; studies = 2; heterogeneity index (I^2) = 0%; moderate-certainty evidence). NACT probably results in a large reduction in the need for stoma formation: 5.9% in NACT group, versus 20.4% in PDS group, (RR 0.29, 95% CI 0.12 to 0.74; participants = 632; studies = 2; I^2 = 70%; moderate-certainty evidence), and probably reduces the risk of needing bowel resection at the time of surgery: 13.0% in NACT group versus 26.6% in PDS group (RR 0.49, 95% CI 0.30 to 0.79; participants = 1565; studies = 4; I^2 = 79%; moderate-certainty evidence). NACT reduces postoperative mortality: 0.6% in NACT group, versus 3.6% in PDS group, (RR 0.16, 95% CI 0.06 to 0.46; participants = 1623; studies = 5; I^2 = 0%; high-certainty evidence). QoL on the European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) scale produced inconsistent and imprecise results in three studies (MD -0.29, 95% CI -2.77 to 2.20; participants = 524; studies = 3; I^2 = 81%; very low-certainty evidence) but the evidence is very uncertain and should be interpreted with caution.

Authors' conclusions

The available high to moderate-certainty evidence suggests there is little or no difference in primary survival outcomes between PDS and NACT. NACT probably reduces the risk of serious adverse events, especially those around the time of surgery, and reduces the risk of postoperative mortality and the need for stoma formation. These data will inform women and clinicians (involving specialist gynaecological multidisciplinary teams) and allow treatment to be tailored to the person, taking into account surgical resectability, age, histology, stage and performance status. Data from an unpublished study and ongoing studies are awaited.

PLAIN LANGUAGE SUMMARY

Does giving chemotherapy before surgery improve survival or quality of life for women with advanced ovarian epithelial cancer?

What is the issue?

Epithelial ovarian cancer, arising from the surface layer of the ovaries or lining of the fallopian tubes, is the ninth most common cancer worldwide in women, and is the most common form of ovarian cancer (approximately 90% of ovarian cancers). Unfortunately, most women with ovarian cancer present at a late stage, when their disease has spread throughout the abdomen. This is because ovarian cancer often arises from the ends of the fallopian tubes, from where single cells can drop out into the abdominal cavity even when the primary tumour is microscopic. These tumour cells circulate around the abdominal cavity in the lubricating peritoneal fluid, implant on other surfaces and grow over time until they cause symptoms. Even then, symptoms, such as bloating and bowel disturbance (most commonly constipation), are nonspecific and easily attributed to more common benign conditions. In Europe and the UK, just over a third of women diagnosed with ovarian cancer are alive five years after diagnosis.

Conventional treatment for ovarian cancer involves two modalities of treatment: surgery and chemotherapy. The intention of surgery is to stage the disease (assess where the cancer has spread to) and remove as much of the visible (macroscopic) cancer as possible (known as debulking or cytoreduction), preferably to the point where the surgical team is not able to see any visible residual disease in the abdominal cavity. However, since most women will have widespread disease, surgery alone is unlikely to cure the disease and most will also need chemotherapy. Chemotherapy for ovarian cancer uses platinum-based drugs to treat cells that cannot be removed by surgery (macroscopic disease) or are too small to be seen (microscopic disease). Traditionally, chemotherapy was given after surgery (primary debulking surgery (PDS) and adjuvant chemotherapy) . However, chemotherapy can be used before surgery (known as neoadjuvant chemotherapy (NACT) and interval debulking surgery (IDS)) with the aim of shrinking the cancer and allowing women to get better prior to undertaking major surgery. Women who receive NACT and IDS complete the remaining cycles of chemotherapy following surgery.

What did we do?

We searched electronic databases up to October 2020 and conducted handsearches for unpublished reports of trials. We included randomised controlled trials (RCTs) of NACT and IDS versus surgery (primary debulking surgery (PDS)) followed by chemotherapy in women diagnosed with advanced stage epithelial ovarian cancer and pooled study outcome data, where appropriate.

What did we find?

We identified 2227 titles and abstracts from the search. From these, we found five RCTs which met our inclusion criteria, including a total of 1774 women with advanced ovarian cancer. We were able to pool data from four studies. These trials compared women who were given chemotherapy prior to surgery (NACT) with women who underwent surgery first (PDS) prior to chemotherapy. We found little or no difference between the two treatments with respect to the time to death and probably little or no difference in the time to progression of the disease. We found that giving NACT reduces the risk of postoperative mortality and need for stoma formation, for which we have

high certainty. NACT probably reduces the risk of some severe complications of surgery, but some of these data were less well reported in the included studies and so we have moderate to low certainty about these results. The studies only enrolled women with stage IIIc/IV ovarian cancer i.e. those who had advanced disease; a large proportion of women in this review had very bulky tumours. We are currently awaiting results of three ongoing studies and one unpublished full publication of a study that is awaiting classification that will hopefully contribute more evidence to guide clinical practice in this area in the future.

What does this mean?

Overall, the evidence was of moderate to high certainty. There is little or no difference in how long women with advanced epithelial ovarian cancer will survive, if they have chemotherapy or surgery first, where both treatments are planned. There is probably little or no difference in how long it will take for the cancer to regrow after initial treatment. NACT probably reduces some of the risks of surgery, probably halves the risk of needing the bowel removed, and probably has a large reduction in the risk of needing the bowel diverted through the abdominal wall via a stoma (a bag attached to the abdominal wall to collect bowel contents). NACT/IDS is an alternative to PDS followed by chemotherapy in women with bulky stage IIIc/IV disease. Individual decisions about which treatment to have first will depend on the individual woman's wishes, how well she is at the time of diagnosis, the risks of surgery and the burden and distribution of disease.

SUMMARY OF FINDINGS

Summary of findings 1. Neoadjuvant chemotherapy prior to interval surgery (NACT) compared to surgery followed by chemotherapy (PDS) for initial treatment in advanced ovarian epithelial cancer

NACT/IDS compared to PDS/adjuvant chemotherapy for initial treatment in advanced ovarian epithelial cancer

Women or population: women with advanced ovarian epithelial cancer

Settings: hospital-based care in countries including Algeria, Argentina, Austria, Belgium, Canada, Ireland, Italy, Japan, Norway, the Netherlands, Portugal, Spain, Sweden, the UK and New Zealand

Intervention: platinum-based chemotherapy followed by debulking surgery (neoadjuvant chemotherapy)

Comparison: primary debulking surgery followed by platinum-based chemotherapy (adjuvant chemotherapy)

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with PDS	Risk with NACT				
Overall survival (follow-up 4.4 to 6 years)	Study population 757 per 1,000	743 per 1000 (704 to 783)	HR 0.96 (0.86 to 1.08)	2000 (4 RCTs)	⊕⊕⊕⊕ HIGH	NACT results in little to no difference in overall survival. Absolute risk of death at 4 years demonstrated for absolute effects using formula of corresponding intervention risk per 1000 = 1000 - (exp[ln(1 - proportion of patients with event) x HR]) x 1000 (Tierney 2007). Baseline risk of death at 4 years taken from PDS outcomes for combined Vergote 2010 and Kehoe 2015 data published in Vergote 2018
Progression-free survival (follow-up 4.4 to 6 years)	Study population 858 per 1,000	852 per 1000 (821 to 879)	HR 0.98 (0.88 to 1.08)	1847 (4 RCTs)	⊕⊕⊕⊖ MODERATE [2]	NACT probably results in little to no difference in progression-free survival. Absolute risk of recurrence at 1 year demonstrated for absolute effects using formula of corresponding intervention risk per 1000 = 1000 - (exp[ln(1 - proportion of patients with event) x HR]) x 1000 (Tierney 2007). Baseline risk of recurrence in PDS taken from combination of Vergote 2010 and Kehoe 2015 data published in Vergote 2018
Surgically-related severe adverse effects (grade 3+) - stoma for-	Study population 204 per 1,000	59 per 1000 (24 to 151)	RR 0.29 (0.12 to 0.74)	632 (2 RCTs)	⊕⊕⊕⊖ MODERATE [2]	NACT probably results in a large reduction in rate of stoma formation.

mation (within 30 days of surgery)						
Surgically-related severe adverse effects (grade 3+) - bowel resection (within 30 days of surgery)	Study population 266 per 1,000	130 per 1000 (80 to 210)	RR 0.49 (0.30 to 0.79)	1565 (4 RCTs)	⊕⊕⊕⊖ MODERATE	NACT probably reduces surgically-related severe adverse effects (grade 3+) - bowel resection.
Surgically-related severe adverse effects (grade 3+) - postoperative G3+ events (within 30 days of surgery)	Study population 294 per 1,000	65 per 1,000 (38 to 112)	RR 0.22 (0.13 to 0.38)	435 (2 RCTs)	⊕⊕⊕⊖ MODERATE [2]	NACT probably reduces surgically-related severe adverse effects (grade 3+) - postoperative G3+ events.
Surgically-related postoperative mortality (28 days to 6 months of surgery[4])	Study population 36 per 1,000	6 per 1000 (2 to 17)	RR 0.16 (0.06 to 0.46)	1623 (5 RCTs)	⊕⊕⊕⊕ HIGH	NACT reduces postoperative mortality.
EORTC QLQ-C30 QoL at 6 months - global health	The mean EORTC QLQ-C30 QoL at 6 months - global health was 66.5	MD 0.29 lower (2.77 lower to 2.2 higher)	-	524 (3 RCTs)	⊕⊖⊖⊖ VERY LOW [2]	NACT may reduce/have little to no effect on EORTC QLQ-C30 QoL at 6 months - global health but the evidence is very uncertain.

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the __relative effect__ of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk ratio; **OR:** Odds ratio;

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1] Outcome unlikely to have been seriously affected by lack of blinding in the study and so not downgraded for risk of bias due to lack of blinding
[2] Downgraded by 1 level for risk of bias due to unblinded study designs, which may have had an effect on some outcomes
[3] Downgraded by 3 levels due to concerns about overall risk of bias, concerns about imprecision, inconsistencies in results and general heterogeneity. QoL outcome was based on a selected number of institutions with better QoL compliance in largest study. While the trial authors offer justification for their approach, several differences were found when comparing the outcomes of the 404 selected women (of which only 212 were assessed in QoL domains) to the overall populations of 670 women. Women from the selected institutions had significantly better OS and PFS when compared to women treated in institutions which were excluded because of poor compliance rates.

4 Most postoperative deaths were within 28-30 days of surgery, but there were four late surgically-related deaths in Fagotti 2016. Definition of postoperative period varied between studies.

BACKGROUND

Description of the condition

Ovarian cancer is now the ninth most common cancer in females, affecting 313,959 women globally in 2020 (GLOBOCAN 2020). In Europe and the UK, just over a third of women with ovarian cancer are alive five years after diagnosis (CRUK 2018; EUROCARE 2015), largely because most women with ovarian cancer are diagnosed when the cancer is already at an advanced stage (Siegel 2018). Symptoms are often vague and of short duration and, as yet, there are no effective screening programmes. In early-stage disease (Federation of International Gynaecologists and Obstetricians (FIGO) stage I/IIa; Table 1), radical surgery will cure most women, although a proportion of women benefit from adjuvant chemotherapy (Lawrie 2015). In advanced cancer, even radical surgery cannot remove all microscopic disease and so survival is dependent upon chemo sensitivity. Unfortunately, around 75% of women present when the disease has spread outside the pelvis (FIGO stage III/IV), when surgery alone cannot be curative and the role of surgery is less clear.

The standard treatment of advanced ovarian cancer (FIGO stage III/IV) is a staging laparotomy with primary debulking surgery (PDS) followed by platinum-based chemotherapy. The extent of tumour cytoreduction is considered the most important prognostic factor. Griffiths 1975 was the first to report a relationship between the size of residual disease and survival. Meta-analyses of nonrandomised studies (NRS) have since concurred that survival correlates positively with the extent of tumour debulking achieved (Allen 1995; Bristow 2002; Hunter 1992). The extent of debulking achievable, however, may be directly related to tumour biology, which would strongly bias results from nonrandomised controlled trials (RCTs). Tumours that have also spread to the para-aortic or scalene lymph nodes may be less likely to be optimally debulked intra-abdominally at surgery (Burghardt 1991; Petru 1991). Thus, the ability to achieve successful debulking may in part reflect tumour biology. One exploratory analysis of three prospectively randomised trials in advanced ovarian cancer suggested that surgical debulking can partially overcome these biological factors (Du Bois 2009). Other independent prognostic factors for overall survival (OS) were shown to be age, performance status, grade, FIGO stage and histology (Du Bois 2009). Interestingly, a recent study demonstrated that routinely removing non-bulky lymph nodes in epithelial ovarian cancer (EOC) does not improve survival (Harter 2019).

The definition of what constitutes 'optimal' or 'maximal' debulking has changed since the 1980s, originally considered to be no residual tumour deposit of greater than 2 cm in diameter, and more recently as residual tumour of ≤ 1 cm; the current aim is to leave no macroscopic disease (no disease left visible to the naked eye - so called 'complete' or 'R0' surgery) (Thigpen 2011). This is somewhat misleading in advanced ovarian cancer, since in other cancers an 'R0 resection' indicates that the tumour has been removed with proven microscopically normal margins. In advanced ovarian cancer, due the pattern of spread via the intra-abdominal cavity, microscopic disease is likely to remain, even after a macroscopic debulk is achieved, hence the terms 'complete' and 'R0' will not be used in this review.

In the past, some investigators had not shown a benefit to maximal debulking in women with high-volume, advanced disease (Hoskins 1992; Vergote 1998). However, this may have been because some were very unwell prior to surgery and not fit enough at that stage to withstand a major operation. Vergote 1998, therefore, introduced a policy of treating women with primary chemotherapy (neoadjuvant chemotherapy (NACT)) or primary debulking surgery (PDS), depending on the extent of the disease and performance status. Following the change in patient management, they reported an overall improvement in survival, despite a reduction in primary debulking rates from 82% to 57%.

The role of so-called ultra-radical surgery in ovarian cancer, with extensive surgical effort often involving the upper abdomen, is a separate question and this review does not seek to question the value or extent of surgery, rather its timing in respect to its combination with chemotherapy. However, a nonrandomised study demonstrated the importance of the combination of surgery and chemotherapy, with a reduced survival in those who had chemotherapy alone and did not go on to have interval debulking surgery (IDS) (Hall 2019). This is supported by findings from a recent cohort study from Sweden, which demonstrated no improvement in survival with system-wide introduction of ultra-radical surgery for ovarian cancer, associated with a reduction in those undergoing surgery by around 10% (Falconer 2020). Studies that do not use whole population cohorts are at critical risk of bias and may overestimate the benefits of upfront surgery (e.g. Mueller 2016).

Description of the intervention

NACT involves giving chemotherapy before attempting cytoreductive surgery for advanced ovarian cancer and is a rationale used in other tumour types. It has evolved from the practice of IDS, a secondary attempt at tumour cytoreduction performed after a suboptimal attempt at primary cytoreduction and adjuvant chemotherapy. In a Cochrane Review (Tangjitgamol 2010), additional IDS performed by gynaecological oncologists secondary to PDS and adjuvant chemotherapy was found to offer no additional survival benefit compared with standard treatment of advanced ovarian cancer. However, IDS may improve survival of women in whom primary surgery was not performed with cytoreductive intent by a gynaecological oncologist and in those who have had suboptimal PDS.

Bristow 2007 reviewed 26 nonrandomised studies (NRS) comparing NACT with PDS and concluded that, while NACT might be a viable option for those unsuitable for an attempt at primary cytoreduction, because of significant comorbidities, current poor performance status or impossibility of surgery, survival outcomes with NACT may be inferior to PDS. However, this was based on highly selected data, at critical risk of bias, as women with worse disease were more likely to have received NACT/IDS rather than PDS. Thus, platinum-based NACT may be an alternative to PDS, particularly where complete cytoreduction at PDS is considered unlikely (Swart 2009). Tumour resectability depends on the patient's age, disease burden, comorbidities, location of metastatic sites, performance status and stage (Vergote 2011a), as well as the skill and philosophy of the surgical team (Chi 2010; Kehoe 1994; Vergote 2011b). Retrospective data suggest that the optimal time for IDS may be after three cycles of chemotherapy, followed by a further three cycles, and that delaying to four cycles might worsen OS (Bogani 2017). However, these data are based on retrospective analysis of NRS data and are therefore at critical risk of bias, as women who are less well are more likely to have delayed surgery. On multivariate analysis, only the Eastern Co-operative

Oncology Group performance status correlated with OS (hazard ratio (HR), 1.76; 95% confidence interval (CI), 1.2 to 2.49; P = 0.001).

The goal of surgery, whether IDS or PDS, should be complete resection of all disease (Onda 2010). A review of 21 NRS (Kang 2009) found that, compared with PDS, NACT improved the rate of optimal cytoreduction. However, this did not seem to influence survival.

How the intervention might work

There are several reasons why NACT may be preferable to PDS:

- NACT may decrease the size and extent of the tumour such that complete resection is more feasible;
- NACT may improve patient performance status;
- PDS necessitates hospital admission, whereas chemotherapy can be administered in an outpatient setting and started immediately;
- PDS delays starting chemotherapy as there is the potential for chemotherapy to interfere with wound healing;
- if surgery is not curative, residual tumour cells may multiply while the individual awaits recovery from surgery.

Concerns about using NACT include the following:

- NACT delays the removal of the tumour and, thereby, may compromise women's survival;
- chemotherapy induces fibrosis, which may make complete cytoreduction more difficult;
- NACT may effectively shrink cancer deposits but leave microscopic disease that is then not surgically removed, whereas the whole deposit might have been removed had it been visible;
- if too many cycles of NACT are given pre-surgery, there is a concern regarding the possibility of chemo-resistance post-surgery. One meta-analysis found a negative association between OS and the number of NACT cycles given (Bristow 2006);
- PDS reduces the tumour bulk and number of cancer cells, thereby reducing the chance of developing chemo-resistance.

Why it is important to do this review

There has been considerable controversy in the literature surrounding the use of NACT in advanced ovarian cancer (Chi 2011; Du Bois 2011; Vergote 2011a). In one overview, Onda 2011 stated "NACT is expected to become standard treatment for unselected women with advanced ovarian cancer when favourable results are confirmed by Phase III studies and several problems are resolved". However, surveys among members of the US Society of Gynecologic Oncology (Dewdney 2010), and the European Society of Gynaecologic Oncology (Vergote 2011b) suggest a large discrepancy in acceptance and use of NACT as a treatment option for advanced ovarian cancer. Many investigators agree that NACT has a place, at the very least, in women with lesions that cannot be optimally resected, or in those too unwell to undergo major surgery at diagnosis (Bristow 2007; Chi 2010; Swart 2009; Vergote 2011a). To our knowledge, nine randomised trials of NACT versus PDS have been started or completed in the past two decades (Fagotti 2016; Kumar 2009; Kehoe 2015; Mahner 2017; NCT04257786; NCT04515602; Onda 2016; SUNNY; Vergote 2010). Since RCTs are the 'gold standard' of evidence-based medical research, we hope that a review of randomised evidence may clarify what the benefits and risks are of using NACT for women with advanced ovarian cancer, compared with the standard treatment of PDS.

This review updates previous analyses in this area, incorporating additional data from previously published studies.

OBJECTIVES

To assess whether there is an advantage to treating women with advanced epithelial ovarian cancer (EOC) with chemotherapy before debulking surgery (neoadjuvant chemotherapy (NACT)) compared with conventional treatment where chemotherapy follows debulking surgery (primary debulking surgery (PDS)).

METHODS

Criteria for considering studies for this review

Types of studies

Randomised controlled trials (RCTs).

Types of participants

Women with advanced epithelial ovarian cancer (EOC) (FIGO stage III/IV).

Types of interventions

Primary debulking surgery (PDS), with the aim of macroscopic resection or optimal debulking (as defined by the investigators), followed by platinum-based chemotherapy, compared to platinum-based neoadjuvant chemotherapy (NACT) followed by interval debulking surgery (IDS), with the aim of resecting disease to the same degree as the PDS group.

Types of outcome measures

We extracted data for direct outcome measures, relevant to patients and clinicians, including benefits, harms and quality of life data, as detailed below.

Primary outcomes

- Overall survival (OS): defined as death from any cause from time of randomisation
- Progression-free survival (PFS): defined as time free of disease progression or death from time of randomisation

Secondary outcomes

- Morbidity/adverse effects classified according to CTCAE 2017:
  - direct surgical morbidity (e.g. bladder injury, intestinal obstruction, haematoma, local infection, duration of operation, need for blood transfusion; need for bowel resection and/or stoma formation);
  - surgically-related systemic morbidity and mortality (e.g. deep vein thrombosis (DVT), pulmonary embolism (PE), chest infection, cardiac events, need for blood transfusion);
  - recovery, including duration of hospital stay;
  - toxicity related to chemotherapy; grouped as haematological, gastrointestinal, genitourinary, skin and neurological toxicity.
- QoL measured using a validated scale (e.g. QLQ-C30 (Osaba 1994), QLQ-OV28 (Greimel 2003)).

- Extent of surgical debulking achieved (e.g. macroscopic, 0.1 to ≤ 1 cm, > 1 cm and combined macroscopic and 0.1 to ≤ 1 cm, i.e. 'optimal').

We will present a summary of findings table reporting the following outcomes listed in order of priority:

1. Overall survival
2. Progression-free survival
3. Surgically-related side effects: need for blood transfusion
4. Surgically-related side effects: stoma formation
5. Surgically-related side effects: bowel resection
6. Surgically-related side effects: postoperative grade 3+ events
7. Postoperative mortality ; postoperative grade 5 event
8. EORTC QLQ-C30 QoL at 6 months

Search methods for identification of studies

Electronic searches

The following electronic databases were searched on 9th October 2020:

- Embase via Ovid (1980 to 2020 week 40) (Appendix 1);
- MEDLINE (Silver Platter/Ovid, 1966 to October week 1 2020) (Appendix 2);
- Cochrane Central Register of Controlled Trials (CENTRAL; 2020, Issue 10) (Appendix 3);
- PDQ and MetaRegister (October 2020).

Searching other resources

The reference lists of the relevant papers found were searched for further studies and we contacted the authors of relevant trials to request information relating to their participation in unpublished trials. Papers in all languages were sought and translations carried out, if necessary.

All relevant articles found were entered into PubMed and, using the 'related articles' feature, a further search was carried out for any other published articles. Meta-register and PDQ were searched for ongoing trials. We contacted the main investigators of relevant trials for further information.

Data collection and analysis

Selection of studies

Two review authors independently selected trials from the results of the searches according to the inclusion criteria specified above (JM and SC, for this update). Disagreements were resolved by discussion and referral to a third author (AB), if required.

Data extraction and management

Two review authors (SC and JM) independently extracted data from the included trials onto a specifically designed data-collection form. Where there were disagreements, these were resolved by discussion. No attempt was made to blind review authors to authors of articles or to journals.

For included studies, we recorded details of trial methodology, the study population and sample size, inclusion and exclusion criteria, intervention and comparison, duration of follow-up and risks of bias. We extracted data relating to participant characteristics (age, histology, grade, extent of disease, previous therapies) and outcomes. For each outcome, we extracted the outcome definition and unit of measurement.

Results were extracted as follows:

- for time-to-event data (survival and disease progression), we extracted the log of the hazard ratio [log(HR)] and its standard error. If these were not reported, we would have estimated the log (HR) and its standard error using the methods of Parmar 1998;
- for dichotomous outcomes (e.g. adverse events or deaths), we extracted the number of women in each treatment arm who experienced the outcome of interest and the number of women assessed at the end point, in order to estimate a risk ratio (RR);
- for continuous outcomes (e.g. quality of life (QoL) measures), we extracted the final value and standard deviation of the outcome of interest and the number of women assessed at the end point in each treatment arm, in order to estimate the mean difference (MD) between treatment arms and its standard error.

Where data were missing or methods were unclear, we contacted the authors for further information. We entered data into Review Manager software (RevMan 2014) and three review authors (SC, AB, JM) checked the data for accuracy.

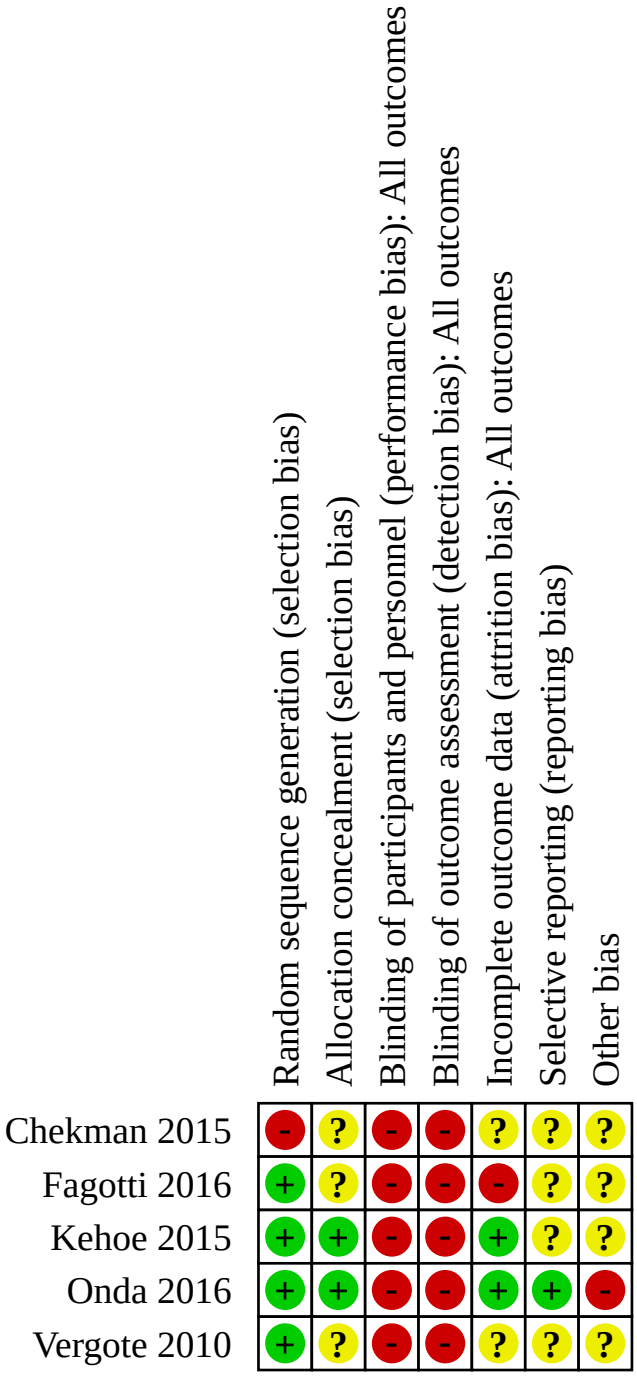
Assessment of risk of bias in included studies

Using Cochrane's risk of bias tool (Higgins 2011), we re-assessed the following for the included studies:

- selection bias: random sequence generation and allocation concealment;
- Blinding of patietns and assessors: performance and detection bias;
- attrition bias: incomplete outcome data;
- reporting bias: selective reporting of outcomes;
- other possible sources of bias.

The risk of bias tool (Appendix 4) was applied independently by up to two review authors (SC and JM) and differences of opinion were resolved by discussion. Results were summarised in a risk of bias graph (Figure 1).

Figure 1. Risk of bias summary: review authors' judgements about each risk of bias item for each included study.



Measures of treatment effect

We used the following measures of the effect of treatment:

- for time-to-event data, we used the hazard ratio (HR);
- for dichotomous outcomes, we used the risk ratio (RR);

- for continuous outcomes, we used the mean difference (MD) between treatment arms.

Unit of analysis issues

No issues were noted.

Dealing with missing data

We noted levels of attrition. We did not impute missing outcome data for any of the outcomes.

Assessment of heterogeneity

We assessed heterogeneity between studies by visual inspection of forest plots, by estimation of the percentage heterogeneity between trials that could not be ascribed to sampling variation (Higgins 2003), by a formal statistical test of the significance of the heterogeneity (Deeks 2001) and, where possible, by subgroup analyses (see below). If there was evidence of substantial heterogeneity, the possible reasons for this were investigated and reported.

Assessment of reporting biases

We did not produce funnel plots to assess the potential for small-study effects as there were only five included trials.

Data synthesis

If sufficient clinically similar studies were available, their adjusted results were pooled in meta-analyses.

- for time-to-event data, hazard ratios (HRs) were pooled using the generic inverse variance facility of RevMan 5;
- for any dichotomous outcomes, RRs were calculated for each study and these were then pooled;
- for continuous outcomes, the MDs between the treatment arms at the end of follow-up were pooled as all trials measured the outcome on the same scale, otherwise standardised MDs would have been pooled.

Random-effects models with inverse variance weighting were used for all meta-analyses (DerSimonian 1986).

Subgroup analysis and investigation of heterogeneity

For this updated review, we included the following subgroup analyses:

- age: 60 years or less and over 60 years;
- extent of debulking achieved: complete debulking; residual tumour 1 cm or less; residual tumour greater than 1 cm.

These subgroups were not prespecified in the original protocol (see Differences between protocol and review), and were evaluated with respect to primary outcomes only. In future versions of this review, we plan to subgroup data by FIGO stage (Stage 3c versus 4).

Sensitivity analysis

In future versions of this review, where possible and with the inclusion of additional studies, sensitivity analyses will be performed where there is a risk of bias associated with the quality of any of the included trials.

Summary of findings and assessment of the certainty of the evidence

We presented the overall certainty of the evidence for each outcome (Types of outcome measures) according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, which takes into account issues not only related to internal validity (risk of bias, inconsistency, imprecision, publication bias) but also to external validity such as directness of results (Langendam 2013). We created a summary of findings table (Summary of findings 1) based on the methods described in the *Cochrane Handbook for Systematic Reviews of Interventions* (Schünemann 2020) and using GRADEpro GDT 2015 (GRADEpro GDT). We used the GRADE checklist and GRADE Working Group certainty of evidence definitions (Meader 2014). We downgraded the evidence from 'high' certainty by one level for serious (or by two for very serious) concerns for each limitation.

- High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
- Moderate certainty: we are moderately confident in the effect estimate. The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
- Low certainty: our confidence in the effect estimate is limited. The true effect may be substantially different from the estimate of the effect.
- Very low certainty: we have very little confidence in the effect estimate. The true effect is likely to be substantially different from the estimate of effect.

RESULTS

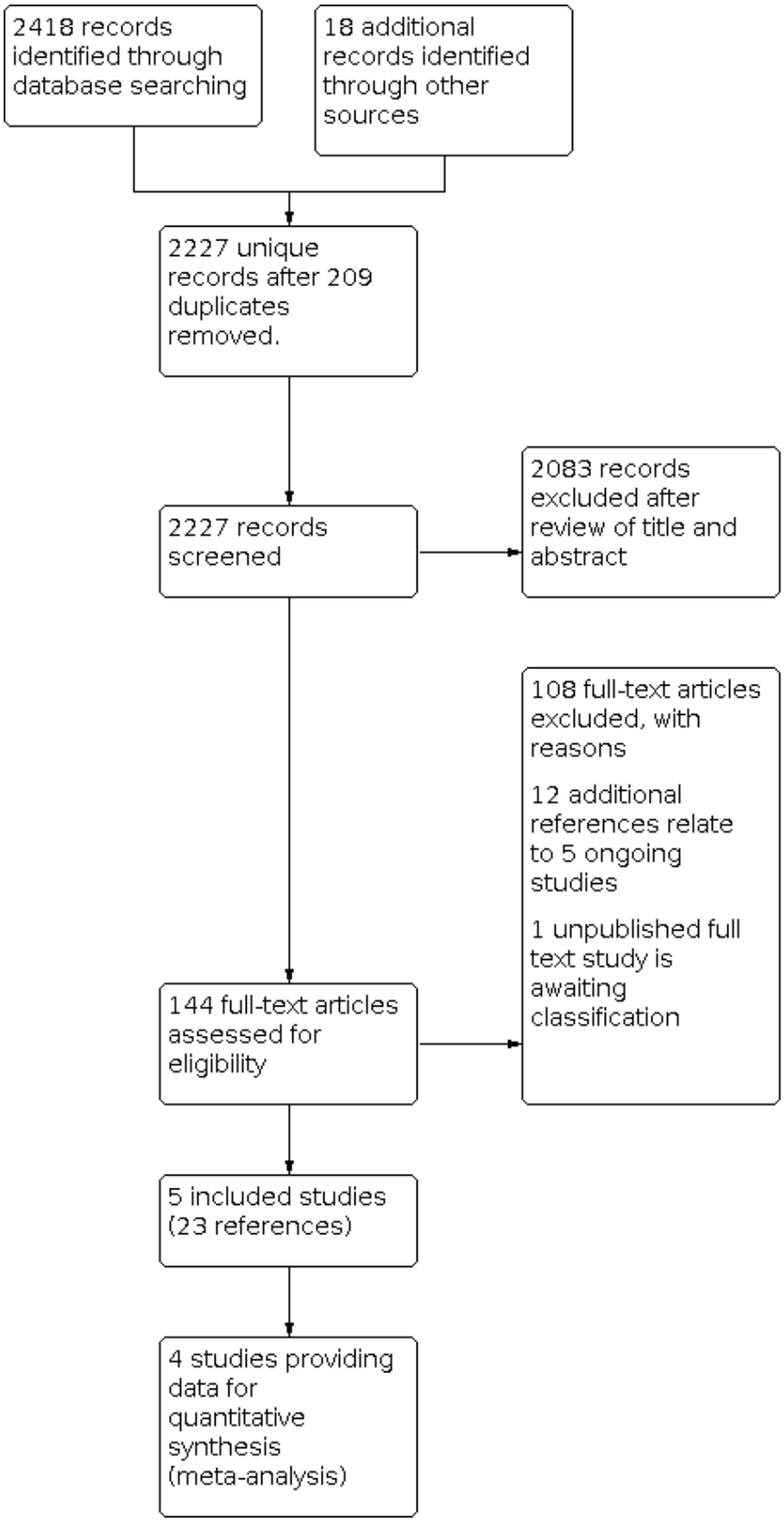
Description of studies

Results of the search

For details of the search strategies, see Appendix 1 and Appendix 2.

Our search identified 2227 unique references, excluding duplicates (Figure 2). At least two review authors (JM, SC) independently screened each abstract in this update of the review; 2083 articles that obviously did not meet the inclusion criteria were excluded at this stage. We retrieved 144 references in full and translated these into English, where appropriate. We found 23 references, reporting on five randomised controlled trials (RCTs), that met our inclusion criteria (Chekman 2015; Fagotti 2016; Kehoe 2015; Onda 2016; Vergote 2010); 12 references reporting on five ongoing trials (Kumar 2009, Mahner 2017 NCT04257786; NCT04515602; SUNNY). We excluded the remaining 108 references (see Excluded studies).

Figure 2. Study flow diagram of the search (up to October 2020).



Kumar 2009 had reported interim analyses in abstract form, but the outcomes were inadequately reported and the risk of bias profile was unclear, so we briefly discussed this trial in the Agreements and disagreements with other studies or reviews in the Discussion and included it with the list of ongoing studies (see Ongoing studies) rather than give it any weight in the main body of the review. Despite contacting the author, unfortunately, no further data have been provided to date for inclusion in the review.

One full-text study (Jiang 2018) is awaiting classification (see Characteristics of studies awaiting classification). Jiang 2018 described the study as a retrospective, cross-sectional study. However, the two groups (NACT versus PDS) were described as 'randomised' and ethical approval and informed consent were sought from study participants. There were significant differences in surgical outcomes between the two groups, but no significant differences in survival outcomes. Despite contacting the author, unfortunately, no further data have been provided to date for inclusion in the review.

Included studies

See Characteristics of included studies.

Chekman 2015 was a randomised controlled trial (RCT), conducted in Algeria between 2008 and 2014. The study enrolled 90 women with FIGO stage IIIc ovarian carcinoma who were randomised to either primary debulking surgery (PDS) followed by chemotherapy or neoadjuvant chemotherapy (NACT) followed by interval debulking surgery (IDS). The same surgeon operated on all women in both intervention arms. It would appear that all women had surgery as well as chemotherapy. Nine women were excluded (reasons not stated) and only data for those who had their disease resected to < 1 cm (including no macroscopic residual disease) were reported, i.e. there did not appear to be an intention-to-treat analysis. The diagnosis of stage IIIC ovarian carcinoma was confirmed by laparoscopic exploration in all but three cases. The number of cycles of chemotherapy in the NACT arm was six cycles (Carboplatin AU5/7.5 mg/mL/minute + Paclitaxel 175 mg/m^2/3 hours every three weeks) on average with 44% having six cycles (range 3 to 7 cycles). Women in the PDS arm had six cycles of chemotherapy on average (78%) (range: 4 to 9) and followed the same chemotherapy protocol as in the NACT arm. The mean duration of follow-up was 254.2 months (range: 69 to 480 months). The trial reported on < 1 cm residual tumour nodules (optimal debulk) or macroscopic resection, overall survival (OS), recurrence-free survival (RFS), morbidity and discussed the role of lumboaortic lymphadenectomy. The study was in abstract form only, but Professor Chekman kindly provided us with more information on request. Unfortunately, survival outcomes could not be analysed, as data for time-to-event outcomes were not provided in an appropriate format for inclusion.

Kehoe 2015 (CHORUS) was a multicentre, non-inferiority phase III RCT, conducted in 87 institutions in the UK and New Zealand. Inclusion criteria were women with clinical or radiological evidence of a pelvic mass with extra-pelvic disease compatible with stage III or IV ovarian, fallopian tube or primary peritoneal cancer who were fit for surgery and chemotherapy. All women had clinical assessment including serum tumour markers and radiological imaging and 552 women were randomised to undergo treatment; two women were subsequently excluded due to being randomised in error. In the PDS arm, 276 women were assigned to undergo PDS followed by six cycles of platinum-based chemotherapy within six weeks of surgery. In the PDS arm, women with residual tumour deposits > 1 cm were eligible to undergo an additional cytoreductive surgery after three cycles of chemotherapy. In the NACT arm, 274 women were assigned to undergo NACT for three cycles with platinum-based chemotherapy and then have IDS and to recommence chemotherapy within six weeks of surgery. Women in the NACT arm had histological or cytological confirmation of diagnosis before commencing chemotherapy. The primary outcome measure was OS; secondary outcomes were progression-free survival and quality of life (QoL). QLQC-30 and QLQ-Ov28 QoL questionnaires were used. The published QoL data provided only the global score at baseline (pretreatment), six months and 12 months post-treatment.

In the NACT arm, 253 (92%) of 274 women started treatment as allocated and 217/274 (79%) had IDS. Nineteen of the 274 (6.9%) women in the NACT arm had no treatment; 36 women had no surgery following chemotherapy; 17 women had no postoperative chemotherapy (one of whom had primary surgery). In the PDS arm, 251 (91%) of 276 women started treatment as allocated; 212 (77%) had adjuvant chemotherapy. Ten of the 276 (3.6%) women had no treatment; 11 women had chemotherapy first with no surgery afterwards; 39 women had no postoperative chemotherapy (one of whom had preoperative chemotherapy); one woman had an unknown postoperative treatment status. See Characteristics of included studies for further details.

Vergote 2010 (EORTC 55971/NCIC OV13) was a large, international, multicentre, non-inferiority phase III RCT. In total, 718 women were enrolled between 1998 and 2006; however, 48 were excluded after randomisation owing to authorisation irregularities at the Argentinian centre. Thus, 670 women with stage IIIc/IV epithelial ovarian cancer (EOC), primary peritoneal cancer or fallopian tube cancer were evaluated. For inclusion, an extra-pelvic tumour needed to be 2 cm or more and treatment needed to begin within three weeks of the initial biopsy. The experimental group (334 women) were allocated to receive three cycles of platinum-based NACT, followed by IDS and then at least three more cycles of chemotherapy (CT). The control group (336 women) received 'standard' treatment (i.e. PDS plus at least six cycles of platinum-based CT ± IDS). The primary outcome was OS. Secondary outcomes were progression-free survival (PFS), surgical morbidity and mortality, QoL and adverse effects. The investigators performed subgroup analyses on OS with respect to age, FIGO stage and extent of residual tumour. Subgroups of age were: age under 50 years, age 50 to 70 years and age over 70 years; subgroups of extent of residual tumour were: no residual tumour, residual tumour of 1 mm to 10 mm, and residual tumour greater than 10 mm. QoL data from the Vergote 2010 trial were subsequently reported by Greimel 2013 (see nested references in Vergote 2010).

Of the 334 women assigned to NACT, 326 (98%) started chemotherapy and 295 (88%) underwent IDS. Of the 336 women assigned to the PDS group, 315 (94.3%) had PDS and 88.4% started chemotherapy. See Characteristics of included studies for further details.

Onda 2016 (JCOG0602) was a multicentre, non-inferiority, phase III RCT conducted in Japan. The authors enrolled 301 women between 2006 and 2011. For inclusion, women had stage III/IV ovarian, tubal and peritoneal cancers diagnosed by clinical findings, radiological imaging and cytology. CA125 had to be > 200 U/mL and CEA < 2

ng/mL to exclude malignancies of other anatomical sites. Women assigned to the control group (149) underwent PDS followed by eight cycles of platinum-based chemotherapy. An additional debulking operation was performed after PDS, if PDS left > 1 cm of residual tumour. An additional debulking operation was mandatory if the uterus, adnexa or omentum had not been removed at PDS, unless disease progression occurred. Women assigned to the experimental group (152) received four cycles of platinum-based NACT, then underwent IDS followed by a further four cycles of chemotherapy. The primary outcome of the study was OS, with survival data published in a peer reviewed journal in 2020, having previously been presented in conference proceedings. Secondary outcomes were invasiveness of surgery in terms of adverse events; these data have been published. No QoL assessment was performed.

Fagotti 2016 (SCORPION) was a single institution, superiority, phase III RCT. In total, 280 women with advanced ovarian cancer were enrolled into the study but, in order to be eligible for randomisation to the study arms, women had to undergo a staging laparoscopy. This was to obtain histology and confirm diagnosis, as well as assess the tumour load. Tumour load was assessed using a predictive index (PI). Only women with a PI score >/= to 8 and </ = 12, corresponding to a high tumour load, were eligible for randomisation. If it was deemed not possible to perform a staging laparoscopy due to large masses occupying the abdominal cavity infiltrating the abdominal wall or the presence of mesenteric retraction, women were withdrawn from the study. After recruitment reached 110 women in order to achieve statistical power for the analysis of the first co-primary end point of major perioperative morbidity, further women were recruited to attain statistical power on PFS (more details are given in Risk of bias in included studies). Two hundred and twenty-five women underwent staging laparoscopy in total, but only 171 went on to be randomised. In the control group, 84 women were assigned to PDS followed by six cycles of platinum-based chemotherapy started within four weeks of surgery. Once women in the control arm had undergone PDS they were not allowed to have an additional cytoreductive procedure. In the experimental group, 87 women were assigned to three or four cycles of platinum-based NACT and to undergo surgery within four weeks of the last cycle, if disease progression was excluded on imaging. The final cycles of chemotherapy in the experimental arm were resumed within four weeks of IDS. The mean and median time to the start of chemotherapy was 42.7 (SD = 18.3) and 41 days (range: 18-169) in the PDS arm, respectively. In the NACT arm, the mean time to chemotherapy was 26.4 days (SD = 11.5) and the median was 26 days (range: 3-79). The mean and median time to start adjuvant treatment after IDS in the NACT arm was 39 (SD = 10.8) and 37 (range: 14-71) days, respectively. Co-primary outcomes were PFS survival and postoperative complications. Secondary outcomes were OS and QoL. Further data were kindly provided by Professor Fagotti. Some outcomes were reported based on the initially published cohort of 110 patients, whereas others were reported for the final 171 participants in the randomised cohort.

Excluded studies

See Characteristics of excluded studies.

One hundred and eight references were excluded for the following reasons:

- Non-RCTs (77);
- Eleven RCTs without a surgical arm comparison (Bertelsen 1990; Chan 2017; Deval 2003; Dutta 2005; Liu 2017; Lotze 1987; Mackay 2011; Mahner 2006; Polcher 2009; Rutten 2012; Trope 1997);
- Three RCTs of IDS following PDS (Redman 1994; Van der Burg 1995; Varma 1990);
- One RCT of non-platinum-based NACT versus surgery (Evdokimova 1982);
- One RCT of chemotherapy plus iliac artery embolisation versus surgery (Liu 2004);
- Fourteen reviews or systematic reviews (Baekelandt 2003; Bristow 2001; Dai-yuan 2013; Fujiwara 2013; Kumar 2015; Lyngstadaas 2005; Mahner 2014; Makar 2016; Qin 2018; Sato 2014; Schorge 2014; Xiao 2018; Yang 2017; Zeng 2016);
- Two pooled analyses of studies included in the review (Vergote 2018; Vergote 2019);
- One RCT comparing early IDS after 3 cycles of NACT with late IDS after 6 cycles of NACT (Kumari 2020).

Liu 2004, an RCT comparing NACT plus iliac artery embolisation versus PDS, was originally an 'included study' in the 2006 version of this review. In a previous update of the review, we revised our assessment of this study and excluded it, as the study findings might have been attributable to NACT versus PDS, iliac artery embolisation, or the combination, because NACT versus PDS was not the only variable in the study and iliac artery embolisation was not delivered in both arms.

Risk of bias in included studies

For this update of the review, a combination of two out of three review authors (from SC, AB, JM) independently re-assessed the risk of bias in each included trial according to pre-defined criteria stated in the methods section (Figure 1).

Allocation

The Chekman 2015 study selection bias was judged to be at high risk, especially when compared to other studies with centralised randomisation, although allocation concealment was unclear due to lack of information. Ninety women with FIGO stage IIIc ovarian carcinoma were enrolled and underwent surgery, but only 82 women were randomised: 41 to PDS/chemotherapy and 41 to NACT/IDS. The randomisation was performed in the operating room by random draw by someone other than the surgeon, once verification of inclusion criteria and resectability under laparoscopy or laparotomy had been confirmed.

The Fagotti 2016 study was deemed to be at low risk of selection bias, albeit from a highly selected population. A centrally-performed, computer-generated list for block randomisation (1:1 ratio) was used. Women were randomly (maximum allowable percentage deviation = 10%) allocated to PDS + systemic adjuvant chemotherapy (arm A, control) or to NACT + IDS (arm B, experimental). Women were only eligible for randomisation into the study once they had undergone a staging laparoscopy to assess disease burden. The staging laparoscopy was used as a triage tool to assess eligibility for the study. If a staging laparoscopy was unfeasible, women were removed from the study. If the staging laparoscopy was successful, a predictive index (PI) value was calculated based upon seven parameters: presence or absence of omental cake, extensive carcinomatosis of the peritoneal or

diaphragmatic surfaces, mesenteric retraction, infiltration of the stomach, spleen or bowel and or superficial liver metastases. If the PI score was ≥ 8 or ≤ 12, this was considered to be a high tumour load, related to lower chances of optimal cytoreduction and worse prognosis. The PI scoring system was based upon earlier work by the same group (Fagotti 2006; Fagotti 2013; Vizzielli 2014).

The initial phase of the Fagotti 2016 study identified 280 women, of whom 14.3% (40) were excluded: seven due to refusal to participate; 15 due to PS score > 2; and 18 due to age > 75 years. A further 15 women (6.25%) had an unsuccessful attempt at a staging laparoscopy, leaving 225 women who underwent a successful staging laparoscopy. Of those 225 women, a further 115 (51.1%) were excluded following staging laparoscopy: 69 due to a PI score < 8; 31 due to mesenteric retraction or PI score > 12; and 15 had non-EOC histology. This left 110 women, with 55 allocated to each arm of the study. These complexities in trial design introduce potential sources of bias and may limit the applicability to the general advanced ovarian cancer population.

The risk of selection bias in the Kehoe 2015 study was deemed to be low risk as the randomisation was performed centrally using a minimisation method based on randomising centre, largest radiological tumour size, clinical FIGO stage, and prespecified chemotherapy regimen.

The Onda 2016 study was deemed to be at low risk of selection bias. The Japan Clinical Oncology Group (JCOG) data centre randomly assigned treatment to each woman via a minimisation method based on institution, stage (III versus IV), performance status (0 to 1 versus 2 to 3) and age (< 60 versus > 60).

In Vergote 2010, randomisation and allocation were performed centrally and the study appeared to be at low risk of allocation bias, although details of the process of randomisation method and concealment were lacking in published data.

Blinding

The five included studies were open-label studies and outcome assessment were not blinded. This is probably not an issue for primary outcomes (i.e. survival); however, it may lead to detection bias with regard to other outcomes or subgroups (e.g. extent of debulking achieved). The importance of blinding of outcome assessment in ovarian cancer studies had been raised in a Gynecologic Cancer InterGroup (GCIG) consensus statement (Thigpen 2011). Data for such outcomes are thus to be interpreted with caution and all studies were deemed to be at high risk of bias.

Incomplete outcome data

Chekman 2015 was at unclear risk of attrition bias due to lack of reported details.

Fagotti 2016 was judged to be at high risk of attrition bias. After the recruitment of 110 women was achieved for the analysis of the first co-primary end point of major perioperative morbidity, further women were recruited to attain statistical power on PFS. The final trial cohort consisted of 171 women, with 84 randomised to PDS and 87 randomised to NACT. Information was not available for two patients who were lost during treatment, one for each arm. The initial published data reported QoL outcomes and short-term surgical outcomes. There were substantial missing data for QoL outcomes, but relative results (hazard ratios (HRs)) for survival (OS and PFS) were adequately reported and analysed. Of the women included in the analysis, 82/84 women in the PDS arm required upper abdominal surgical procedures compared to 28/74 women who underwent IDS (42.3%). Median duration of entire treatment from randomisation to completion of medical treatment was also longer in the PDS arm (38 weeks versus 28 weeks). This was due to an almost two-week difference in time to start post-surgery chemotherapy (median time post-PDS 40 days; median time post-IDS 27 days; P = 0.0001).

Kehoe 2015 and Onda 2016 were deemed to be at low risk of attrition bias, as all trial participants were accounted for and the results were analysed on an intention-to-treat basis.

In the Vergote 2010 study, data from 48 women from Argentina were excluded owing to "potential authorisation irregularities"; however, the investigators stated that their results were similar when these excluded data were included. The exclusions appeared erroneously as pre-randomisation exclusions on the published study-flow diagram. The study was, therefore, at unclear risk of attrition bias.

Selective reporting

Chekman 2015 was at unclear risk of reporting bias due to lack of detail.

Fagotti 2016 was at unclear risk of reporting bias due to the differences in numbers reported for different outcomes, as described above, and lack of quality of life data to date.

In Kehoe 2015, the risk of reporting bias was unclear. All prespecified outcome measures have been reported in some capacity, but QoL data were provided only in the form of a global score at baseline, six months and 12 months post-treatment.

The potential for reporting bias in the Onda 2016 study is now deemed to be low risk; surgical morbidities were reported in the initial publication and survival outcomes have now been published .

There was an unclear risk of selective reporting bias for QoL data in the Vergote 2010 study. Vergote 2010 (including Greimel and colleagues) subsequently published the QoL data from the Vergote 2010 study (see additional reference under Vergote 2010). They reported that compliance for all women was too restrictive and changes to the protocol-defined analysis plan were made. The dataset for QoL data was then restricted to institutions with the best compliance. The authors stated that the sample size of the Vergote 2010 was overpowered to detect clinically meaningful differences in QoL between the two study arms and they therefore decreased the sample size for QoL data to 400 participants. They further restricted QoL data collection to institutions that had 50% compliance at baseline and at least 35% on further follow-up over all enrolled women. Twenty-seven institutions out of 59 contributed 404 women (60.3% of the total 670 trial participants). The participants in institutions that were included in the QoL data had statistically significant differences compared to those participants not included: they had larger tumours (P < 0.01) and optimal debulking rates were 20% higher (P = 0.001). Those participants in institutions selected for inclusion in QoL data analysis had a greater median OS (nine months longer; P = 0.001) and a greater median progression-free survival (PFS) (2.4 months longer; P < 0.001) than the participants in the institutions that were

not included in the QoL data collection. In addition, as well as selecting institutions with the highest compliance with QoL data, the overall compliance from those institutions was still relatively poor over time. Compliance rates were 83.4% at baseline, 58.7% at chemotherapy cycle 3, 74% at chemotherapy cycle 6, 59.4% at six-month follow-up and 45.7% at 12-month follow-up.

The authors concluded that there were no differences in the QoL functioning or symptoms scales, other than for pain and dyspnoea.. At baseline, the PDS group had higher pain scores (P = 0.046; PDS mean 36.7; NACT mean 29.9) and lower dyspnoea scores (P = 0.049; PDS mean 22.9; NACT mean 27.9). As the difference between the groups was less than 10 points, the authors concluded that this did not represent a "clinically relevant difference".

There was, therefore, unclear risk of reporting bias for the QoL data, given the differences in disease that those participants selected for measurement of this outcome had in comparison with participants in the institutions not selected.

Other potential sources of bias

Due to lack of detail, Chekman 2015 was judged to be at unclear risk of other potential sources of bias.

The complexity of the inclusion criteria in Fagotti 2016, as described above, mean that we were unclear about other potential sources of bias and the study design limits the applicability of the study to a wider, less selected, cohort of women with ovarian cancer.

Supplementary data in Kehoe 2015 table 7 show that hysterectomy/bilateral salpingo-oophorectomy (BSO) and omentectomy were performed in varying proportions in the different arms. It is unclear what effect this might have on outcomes and this could be a potential source of bias.

In the Onda 2016 study, 14 women (one in PDS and 13 in NACT) underwent some type of additional surgery (off-protocol treatment). These off-protocol operations were not included as PDS or IDS in the analysis. There appeared to be more off-protocol surgery in the NACT group. No intention-to-treat analysis was performed. These issues could be another potential source of bias.

Effects of interventions

See: **Summary of findings 1** Neoadjuvant chemotherapy prior to interval surgery (NACT) compared to surgery followed by chemotherapy (PDS) for initial treatment in advanced ovarian epithelial cancer

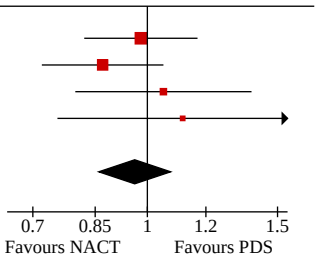
Overall survival (OS) (Analyses 1.1 to 1.4)

Meta-analysis of four studies (Fagotti 2016; Kehoe 2015; Onda 2016; Vergote 2010), assessing 1692 participants, demonstrated little or no difference in OS between neoadjuvant chemotherapy (NACT) and primary debulking surgery (PDS) for initial treatment in advanced ovarian cancer (hazard ratio (HR) = 0.96, 95% CI 0.86 to 1.08; high-certainty evidence); Analysis 1.1; Figure 3 and Figure 4).

Figure 3. Forest plot of comparison: 1 NACT vs PDS, outcome: 1.1 Overall survival.

Study or Subgroup	log[Hazard Ratio]	SE	Favours NACT Total	PDS Total	Weight	Hazard Ratio IV, Random, 95% CI
Vergote 2010 (1)	-0.0202	0.09	334	336	40.2%	0.98 [0.82 , 1.17]
Kehoe 2015	-0.1393	0.0966	274	276	34.9%	0.87 [0.72 , 1.05]
Onda 2016	0.05	0.14	152	149	16.6%	1.05 [0.80 , 1.38]
Fagotti 2016	0.11	0.199	87	84	8.2%	1.12 [0.76 , 1.65]
Total (95% CI)			**847**	**845**	**100.0%**	**0.96 [0.86 , 1.08]**

Heterogeneity: Tau² = 0.00; Chi² = 2.09, df = 3 (P = 0.55); I² = 0%
Test for overall effect: Z = 0.69 (P = 0.49)
Test for subgroup differences: Not applicable



Footnotes
(1) We have applied 95% CIs (investigators reported 90% CIs).

Figure 4. In the PDS group 757 people out of 1000 had died over 4 years compared to 743 (95% CI 704 to 783) out of 1000 for the NACT group. Green = alive at 4 years with PDS/chemo; yellow = additional people alive at 4 years with NACT/IDS; red = people who had died by 4 years with either NACT/PDS or PDS/chemo.

The results were also robust (i.e. no meaningful difference between subgroups) in terms of OS when three trials (Fagotti 2016; Kehoe 2015; Vergote 2010) were subgrouped by age (< 50, 50 to 70 and 70+ years) (Analysis 1.2), and extent of residual disease in two studies (Kehoe 2015; Vergote 2010) (up to 0.5 mm, 0.5-1 cm, > 1 cm) (Analysis 1.3). The results were also robust when three trials (Kehoe 2015; Onda 2016; Vergote 2010) were subgrouped by stage (III and IV) (Analysis 1.4). Survival data by stage were not yet available for one study (Fagotti 2016).

We were not able to extract time-to-event data for OS from the Chekman 2015 study. However, in total, 24 women died during the study period; 15 women (62.5%) in the PDS arm compared to nine women (37.5%) in the NACT arm.

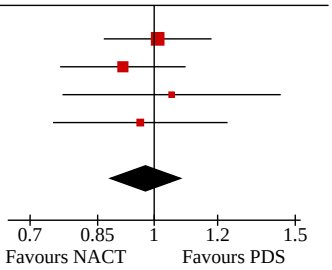
Progression-free survival (PFS) (Analysis 1.5)

Meta-analysis of four studies (Fagotti 2016; Kehoe 2015; Onda 2016; Vergote 2010), assessing 1692 participants, found there is probably little or no difference in risk of disease progression between NACT and PDS for initial treatment in advanced ovarian cancer (HR = 0.98, 95% CI 0.88 to 1.08; I^2 = 0%; moderate-certainty evidence) (Analysis 1.5; Figure 5 and Figure 6).

Figure 5. Forest plot of comparison: 1 NACT vs PDS, outcome: 1.4 Progression-free survival.

Study or Subgroup	log[Hazard Ratio]	SE	NACT Total	PDS Total	Weight	Hazard Ratio IV, Random, 95% CI
Vergote 2010 (1)	0.01	0.079	334	336	42.3%	1.01 [0.87 , 1.18]
Kehoe 2015 (2)	-0.09	0.092	274	276	31.2%	0.91 [0.76 , 1.09]
Fagotti 2016	0.05	0.16	87	84	10.3%	1.05 [0.77 , 1.44]
Onda 2016	-0.04	0.128	152	149	16.1%	0.96 [0.75 , 1.23]
Total (95% CI)			**847**	**845**	**100.0%**	**0.98 [0.88 , 1.08]**

Heterogeneity: Tau² = 0.00; Chi² = 0.93, df = 3 (P = 0.82); I² = 0%
Test for overall effect: Z = 0.49 (P = 0.62)
Test for subgroup differences: Not applicable



Footnotes
(1) We have applied 95% CIs (Investigators used 90% CIs)
(2) 0.09

Figure 6. In the PDS group 858 people out of 1000 had ovarian cancer that had recurred by 2 years compared to 852 (95% CI 821 to 879) out of 1000 for the NACT group. Green = not had recurrent disease by 2 years with PDS/chemo; yellow = additional people without recurrent disease by 2 years with NACT/IDS; red = peopel with recurrent disease by 2 years with either NACT/PDS or PDS/chemo.

From the Chekman 2015 study, we were not able to extract time-to event data for PFS. However, there were 36 recurrences (44%); 20 participants with progressive disease (55.5%) in the control arm (PDS) and 16 (44.5%) in the experimental (NACT) arm.

Of the 12 women in Chekman 2015 who were still alive with confirmed recurrence, five (41.6%) were in the PDS arm and seven (58.3%) were in the NACT arm. Peritoneal recurrence was reported to be most common. Further details about recurrence are given in the table Characteristics of included studies.

Extent of residual disease

In Kehoe 2015, 79/219 women (36%) and 39/255 women (15%) had no macroscopic residual disease in the NACT and PDS arms, respectively; 68/219 (31%) and 57/255 (22%) had 'optimal debulking' (defined as 0.1 cm to 1 cm residual disease) in the NACT and PDS arms, respectively; and 54/219 (25%) and 137/255 (54%) had suboptimal debulking (defined as > 1 cm) in the NACT and PDS arms, respectively. Overall, 147/219 (67%) women and 96/255 (38%) women in the NACT and PDS arms, respectively, had < 1 cm residual disease. Data on degree of resection were missing for 18 women in the NACT group and 22 in the PDS group.

In the NACT arm, 55/274 (20%) women did not have debulking surgery. In the PDS arm, 251 women had PDS and another four had surgery after NACT, so 21 of the 276 allocated to PDS women did not have debulking surgery (7.6%).

In Vergote 2010, of those who had debulking surgery, 151/295 women (51.2%) and 61/315 women (19.4%) had no macroscopic residual disease in the NACT and PDS arms, respectively; 87/295 (29.5%) and 70/315 (22.2%) had 1 mm to 10 mm residual disease in the NACT and PDS arms, respectively; and 52/295 (17.6%) and 167/315 (53%) had suboptimal debulking (> 1 cm residual disease) in the NACT and PDS arms, respectively. Data on debulking status were stated as missing for five (1.7%) women in the NACT group and 17 (5.4%) women in the PDS group. See Characteristics of included studies table for further details. Therefore, of those who had NACT and interval debulking surgery (IDS), 238 women (80.7%) had debulking to < 1 cm residual disease compared to 131 women (41.6%) who had PDS.

Of those assigned to NACT, 326/334 (98%) started chemotherapy and 295/334 (88%) went on to have IDS. In the PDS group, 315 (94.3%) had PDS and 88.4% started chemotherapy.

In Fagotti 2016, 57/74 women (77%) and 40/84 women (47.6%) had no macroscopic residual disease in the NACT and PDS arms, respectively; 16/74 (21.6%) and 38/84 (45.2%) had residual disease 0.1 cm to 1 cm in the NACT and PDS arms, respectively. Therefore, debulking to < 1 cm was achieved for 73/74 (98.6%) and 78/84 (92.8%) in the NACT and PDS arms, respectively; 1/74 (1.4%) and 6/84 (7.2%) had suboptimal debulking (residual disease > 1 cm) in the NACT and PDS arms, respectively (13 participants in the NACT arm did not undergo IDS). This is despite extensive pre-assessment and intraoperative exclusion (laparoscopic assessment), which differed significantly from the Kehoe 2015 and Vergote 2010 studies.

In Onda 2016, 83/150 women (55%) and 45/147 women (31%) had no macroscopic residual disease in the NACT and PDS arms, respectively; 24/150 (16%) and 47/147 (32%) had residual disease 0.1 cm to 1 cm in the NACT and PDS arms, respectively; and 23/150 (15%) and 55/147 (37%) had residual disease > 1 cm in the NACT and PDS arms, respectively. Overall, 107/150 women (71%) and 92/147 women (63%) had optimal debulking (defined as debulking to no residual disease > 1 cm) in the NACT and PDS arms, respectively. Higher optimal debulking rates than Kehoe 2015 and Vergote 2010 may be due to lower initial disease burden, since the entry criteria included all stage III disease, not just bulky stage IIIc, and 9 (6%) in the PDS and 10 (6.6%) in the NACT groups had no measurable disease (presumably by RECIST criteria (Eisenhauer 2009) but not stated) at outset.

Severe adverse effects (SAEs) (Analyses 1.6)

The trial of Fagotti 2016 reported major perioperative morbidity, initially when the trial had randomised 110 participants. Some level of granularity in adverse events was not given in the follow-up publication, so some analyses were based on the initial cohort (n = 110), whereas analyses included in the follow-up publication included all 171 women.

Some studies reported all SAEs during the study period (Kehoe 2015; Onda 2016; Vergote 2010), whereas some reported surgically-related SAEs only (Chekman 2015; Fagotti 2016). The following grade 3/4 (CTCAE 2017) SAEs were reported (Analysis 1.6):

Haemorrhage and blood transfusion requirements (Analyses 1.6.1 and 1.6.2)

Meta-analysis of three studies (Fagotti 2016; Kehoe 2015; Vergote 2010), assessing 1264 participants, found there may be little of no difference in risk of haemorrhage between NACT and PDS for initial treatment in advanced ovarian cancer (RR = 0.93, 95% CI 0.50 to 1.74; I^2 = 69%; low-certainty evidence).

In the Kehoe 2015 and Vergote 2010 studies, the need for blood transfusions and average blood loss were not reported in the published versions of the studies. However, Vergote 2010 provided unpublished data with respect to the number of women who received blood transfusions in the NACT and PDS groups. Meta-analysis of four trials (Chekman 2015; Fagotti 2016; Onda 2016; Vergote 2010), assessing 1085 participants, suggested NACT and IDS likely resulted in a slight reduction in needing a blood transfusion after surgery compared to PDS (RR 0.80, 95% CI 0.65 to 0.99; participants = 1085; I^2 = 50%; moderate-certainty evidence).

Venous thromboembolism (Analysis 1.6.3)

Meta-analysis of data from four studies (Fagotti 2016; Kehoe 2015; Onda 2016; Vergote 2010) suggested that there may be a reduction in the risk of venous thromboembolism in the NACT arm versus the PDS arm, although this was based on a low number of events (n = 27), so should be interpreted with caution (RR 0.28, 95% CI 0.09 to 0.90; participants = 1490; I^2 = 15%; low-certainty evidence).

Infection (Analysis 1.6.4)

Meta-analysis of data from four studies (Fagotti 2016; Kehoe 2015; Onda 2016; Vergote 2010) found women in the NACT arm probably had less risk of infection than in the PDS arm (RR 0.30; 95% CI 0.16 to 0.56; participants = 1490; I^2 = 0%, moderate-certainty evidence).

Gastrointestinal (GI) fistulae (Analysis 1.6.5)

Meta-analysis of data from four studies (Fagotti 2016; Kehoe 2015; Onda 2016; Vergote 2010), found that NACT may be associated with lower risk of severe gastrointestinal fistulae than PDS, although the overall event rate was very low (n = 17) (RR 0.30; 95% CI 0.09 to 0.97: 1490 participants; I^2= 0%; low-certainty evidence).

Other SAEs (Analyses 1.6.6 to 1.6.14 and 1.6.17)

Overall postoperative G3+ SAEs from two studies (Fagotti 2016; Onda 2016) found that the number of patients who had a G3+ SAE in the postoperative period was probably lower in the NACT group (RR 0.22, 95% CI 0.13 to 0.38; participants = 435; studies = 2; I^2 = 0%; moderate-certainty evidence) (see Analysis 1.6.18 and Figure 7).

Figure 7. In the PDS group 29 people out of 100 had G3+ post op serious adverse events (SAE) compared to 6 (95% CI 4 to 20) out of 100 for the NACT group. Green = no post op G3+ SAE with PDS/chemo; yellow = additional people who were better with NACT/IDS; red = people with G3+ SAEs with either NACT/PDS or PDS/chemo.

The proportion of remaining SAEs that were assessed was low. There was probably little or no difference between arms for risk of urinary/vaginal fistula, nausea, vomiting, diarrhoea, neutropenia, neurotoxicity, thrombocytopenia, anaemia, febrile neutropenia and renal toxicity (see analyses 1.6.6 to 1.6.14; 1.6.17; all low-certainty evidence). IDS may be associated with less risk of stoma formation, bowel resection, and postoperative grade 3+ events than PDS.

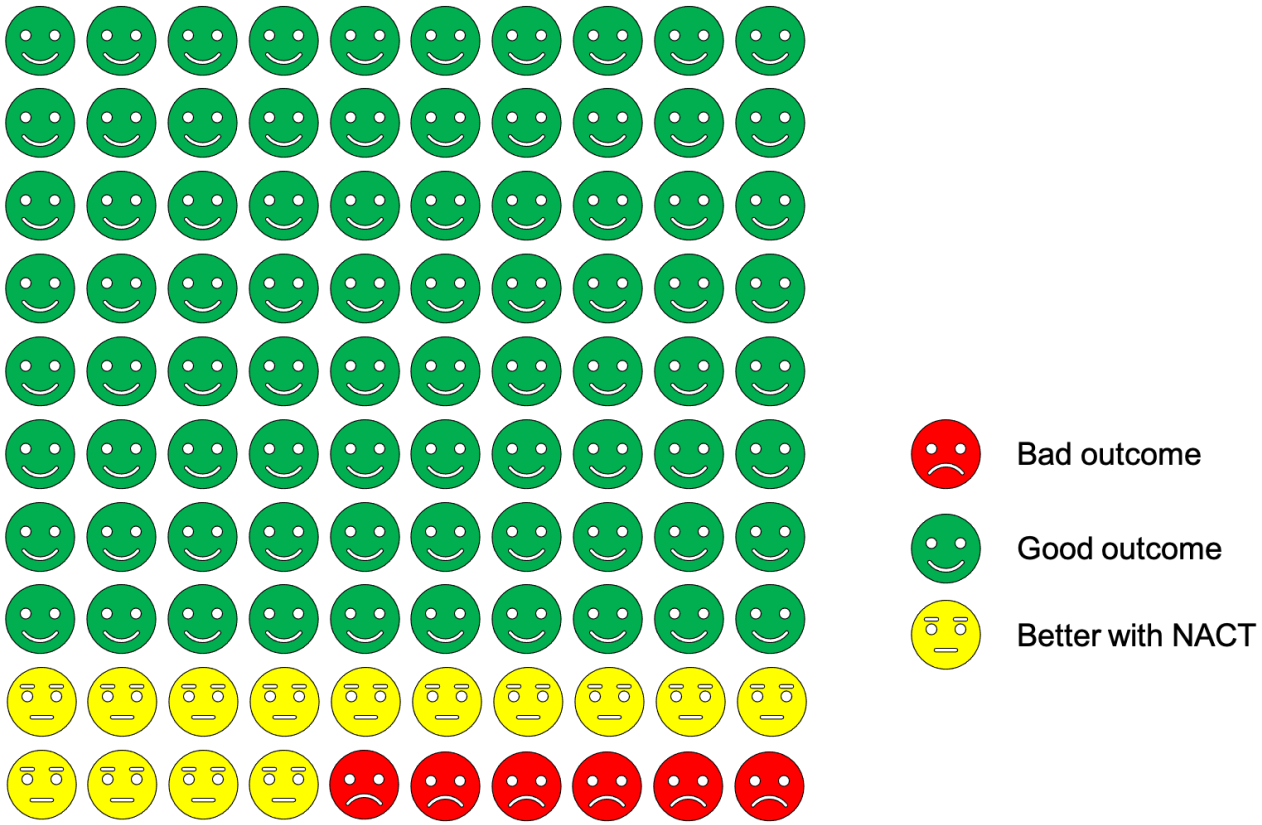
In the Chekman 2015 study, there were a total of 17 complications: 12/41 women in the PDS arm; 5/41 women in the NACT-IDS arm (intraoperative incidents). We were careful not to over interpret this result from a trial of low numbers in each arm, with issues regarding imprecision and unclear risk of bias.

The authors reported that eight re-operations (9.8%) were performed, mainly for abdominal and vascular complications; six (7.3%) in the PDS arm and two (2.4%) in the NACT-IDS arm.

Stoma formation (Analysis 1.6.15)

Women were less likely to require formation of a stoma (colostomy or ileostomy) in the NACT arm versus the PDS arm, although data were only presented in two of the studies (Fagotti 2016; Kehoe 2015) (RR 0.43, 95% CI 0.26 to 0.72; participants = 581; studies = 2; I^2 = 0%; moderate-certainty evidence) (Analysis 1.6.15 and Figure 8).

Figure 8. In the control group 20 people out of 100 had stoma formation following initial surgery compared to 6 (95% CI 2 to 15) out of 100 for the active treatment group. Green = no stoma with PDS/chemo; yellow = additional people who didn't require a stoma with NACT/IDS; red = people who required a stoma with either NACT/PDS or PDS/chemo.



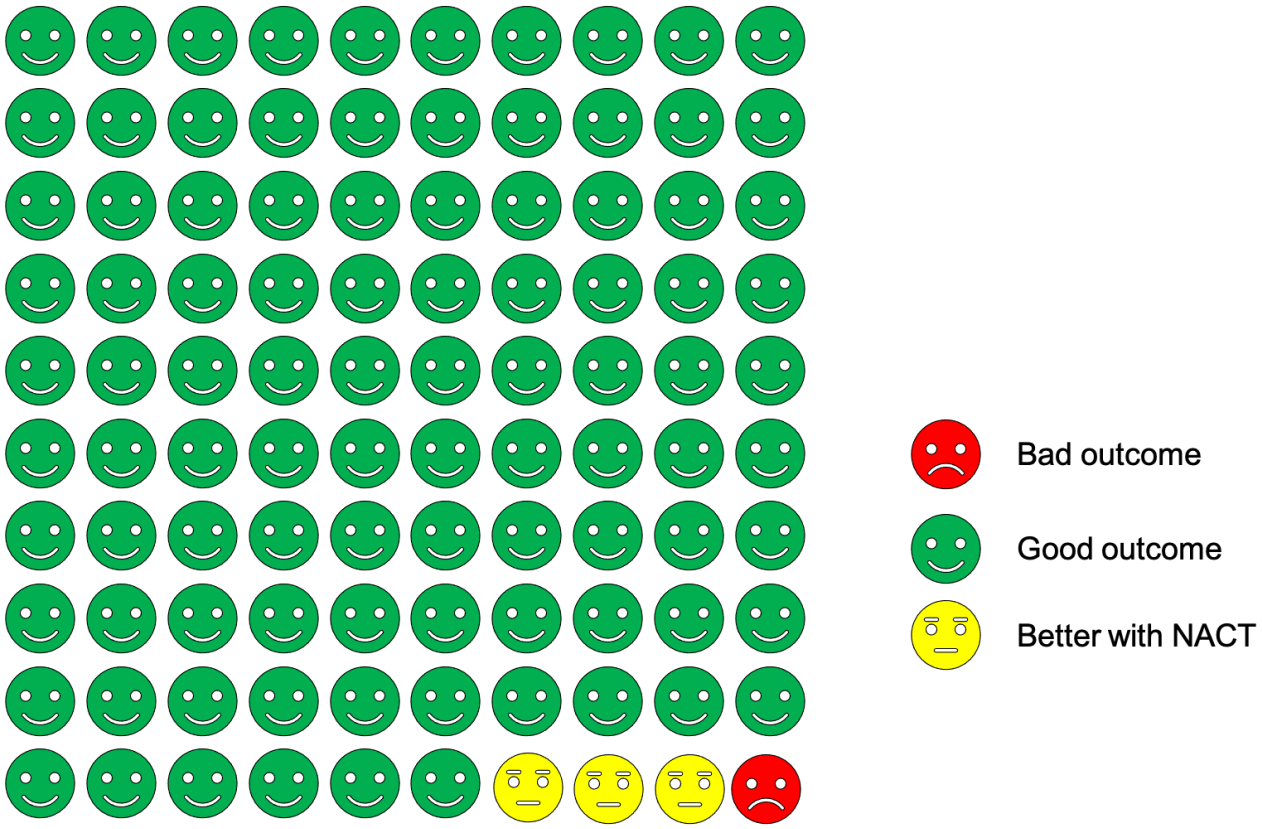
Bowel resection (Analysis 1.6.16)

Women were probably less likely to require a bowel resection (large and small bowel data combined) in the NACT arm versus the PDS arm from data in four studies (Fagotti 2016; Kehoe 2015; Onda 2016; Vergote 2010) (RR 0.49, 95% CI 0.30 to 0.79; participants = 1565; studies = 4; I^2 = 79%; moderate-certainty evidence) (Analysis 1.6.16).

Perioperative/postoperative mortality (Analysis 1.7)

Meta-analysis of five studies (Chekman 2015; Fagotti 2016; Kehoe 2015; Onda 2016; Vergote 2010), assessing the 1625 participants who had surgery, found women in the NACT arm had less risk of perioperative/postoperative mortality than in the PDS arm (Analysis 1.7; RR 0.16, 95% CI 0.06 to 0.46; participants = 1623; studies = 5; I^2 = 0%; high-certainty evidence). Three out of 787 (0.4%) women died within a month of surgery in the NACT arm compared to 26 out of 836 (3.1%) deaths in the PDS arm and, overall, 30/836 (3.6%) due to postoperative complications. There were an additional four deaths in Fagotti 2016 due to postoperative complications in women who survived more than 30 days after surgery, although these deaths were directly related to surgery. Overall, postoperative mortality was therefore 3.6% (30/836) in the PDS group versus 0.4% (3/787) in the NACT group (Analysis 1.7 and Figure 9).

Figure 9. In the PDS group 36 people out of 1000 died in the post-operative period compared to 6 (95% CI 2 to 17) out of 1000 for the NACT group. Green = alive at the end of the post-operative period with PDS/chemo; yellow = additional people who were alive with NACT/IDS; red = people who died in the post-operative period with either NACT/PDS or PDS/chemo.



In Chekman 2015, no deaths were recorded postoperatively (0 to 30 days), but one death was recorded after a second course of neoadjuvant chemotherapy (prior to surgery).

Chemotherapy-related toxicity (Analysis 1.8)

Chemotherapy-specific-related toxicity was not specifically reported in Vergote 2010 as all SAEs were reported together. However, median time to re-start chemotherapy after surgery was 18 days (range 5 to 55) and 19 days (range 0 to 84) in the NACT and PDS groups, respectively. In Fagotti 2016, the median time to start chemotherapy following surgery was lower in the NACT group (NACT = 27 days (range 16 to 37 days) versus PDS = 40 days (range 17 to 120 days); P < 0.0001)) for the initial 110 patient cohort.

Two trials (Kehoe 2015; Onda 2016), assessing 768 participants, found that there may be little or no difference in chemotherapy-related SAEs between arms, although we have low certainty in these results (Analysis 1.8; OR 0.88, 95% CI 0.57 to 1.36, I^2 = 54%; low-certainty evidence).

Quality of life (QoL) (Analyses 1.9 to 1.10)

Three studies (Kehoe 2015; Fagotti 2016; Vergote 2010), assessing 524 participants, reported on QoL at six months using the EORTC QLQ-C30 questionnaire. In two studies, individual symptoms were reported (Fagotti 2016; Vergote 2010). We did not interpret pooled results for individual symptoms due to heterogeneity in results and the summary effects are merely displayed in forest plots to demonstrate the heterogeneity. Results were either inconsistent or there did not appear to be any differences in QoL measures in individual domains between arms. The global health domain was the only domain to demonstrate a numerically significant difference between arms, but the magnitude of the difference was so small, it would be very unlikely to be clinically meaningful. Vergote 2010 and Kehoe 2015 also reported QoL at 12 months with similar results, but due to the high dropout rate, especially by 12 months, these results were of very low-certainty and should be interpreted with caution (Analysis 1.9; Analysis 1.10). Previously, the Kehoe 2015 results were reported separately, due to uncertainty in which QoL data were reported in the original paper. However, following clarification, we have been able to amalgamate these data. Further data from Kehoe 2015 for individual QoL parameters are awaited and it may be possible to combine further QoL data in future updates.

Duration of operation

Mean operating times in Chekman 2015 were 233 minutes (range 69 minutes to 360 minutes) and 273 minutes (range 144 minutes to 480 minutes) in the NACT and PDS groups, respectively. Mean operating times in the Fagotti 2016 study for IDS after NACT and PDS were 253.2 minutes (SD = 101.4) and 460.6 minutes (SD = 102.6), respectively.

In Vergote 2010, the median operating times were 180 minutes (range 30 minutes to 560 minutes) and 165 minutes (range 10 minutes to 720 minutes) in the IDS and PDS arms, respectively. Kehoe 2015 reported that the median operation time was 120 minutes in both groups (interquartiles ranges were 80 to 161 and 90 to 155 in the PDS and NACT arms, respectively; the overall range was 12 to 450 mins). Onda 2016 found that median operating time, when accounting for the main procedure only (not counting an additional debulking procedure in the PDS group) was 302 minutes in the NACT group and 240 minutes in the PDS group (P < 0.001). However, if the subsequent operative procedures were accounted for in both groups, median operating times were 270 minutes and 347 minutes in the NACT and PDS groups, respectively (P < 0.001). Due to disparities in the data collected, we are not able to combine these in a meta-analysis.

Length of stay following surgery

Fagotti 2016 reported mean length of hospital stay; in the NACT group, the mean was 6.7 days (SD = 3.9 days) and 14.8 days (SD = 11.3) in the PDS group (P < 0.001). In Kehoe 2015, length of stay was provided as follows: "fewer women were discharged from hospital within 14 days after surgery in the primary-surgery group compared with primary chemotherapy (198/249, 80% versus 197/211, 93%, P < 0·0001)". Data were not amenable to meta-analysis. These data were not available for Chekman 2015, Onda 2016 or Vergote 2010.

DISCUSSION

Summary of main results

We found five studies that met the inclusion criteria, including a total of 1774 randomised participants. One trial (Chekman 2015) was only available in abstract form (further details were provided by the trial author on request) and contributed to less than 5% of all participants included in the review. We found little or no difference in survival outcomes in women with stage IIIc/IV ovarian cancer who were treated with neoadjuvant chemotherapy (NACT) plus interval debulking surgery (IDS) compared with primary debulking surgery (PDS) plus chemotherapy. Surgically-related morbidity (grade 3/4) was probably higher in the PDS group (such as haemorrhagic, infective and thromboembolic adverse effects). NACT prior to surgery reduces postoperative deaths and the need for stoma formation by two-thirds and probably reduces the need for bowel resection by half. Quality of life (QoL) outcomes were poorly and incompletely reported and results were inconsistent in trials that reported this outcome. Choice of surgical treatment is likely to be dictated by clinical factors in and preferences of the patient, clinician training and surgeon preference.

Overall completeness and applicability of evidence

In a previous version of this review, the evidence for the non-inferiority of NACT versus PDS for advanced ovarian cancer was not widely applicable, as only participants with stage IIIc/IV ovarian tumours (extra-pelvic disease larger than 2 cm) were included in Vergote 2010, and the majority of participants had extensive disease (metastatic lesions larger than 10 cm were present in 61.6% of women)(Morrison 2012). In the subgroup of women with preoperative extra-pelvic tumour of less than 5 cm in diameter (189 women), PDS significantly improved OS compared with NACT (HR 0.64; 95% CI 0.44 to 0.93) (Vergote 2010 Supplementary appendix). Furthermore, when sub-grouped by FIGO stage, women with stage IV disease may have a survival advantage with NACT than with PDS although due to inconsistency between studies this should be interpreted with caution (HR 0.88, 95% CI 0.69 to 1.14; participants = 391; studies = 3)).

This update, with the addition of overall survival data, includes data from four studies with differing patient inclusion criteria, so the evidence for non-inferiority of NACT-IDS is more widely applicable.

Meta-analysis of four studies (Fagotti 2016; Kehoe 2015; Onda 2016; Vergote 2010), assessing 1692 participants, produced a hazard ratio of (HR 0.96, 95% CI 0.86 to 1.08), therefore there is high-certainty evidence for little or no difference in OS between NACT and PDS for initial treatment in advanced ovarian cancer, based on the relatively heterogeneous populations included in these studies.

Meta-analysis of four trials found moderate-certainty evidence for little or no difference in risk of disease progression between NACT and PDS for initial treatment in advanced ovarian cancer (HR 0.98, 95% CI 0.88 to 1.08; participants = 1692; studies = 4).

The QoL data analysis of variance, adjusted for baseline scores, showed that there may or may not be a difference in scores between NACT and PDS at six months (MD -0.29, 95% CI -2.77 to 2.20; participants = 524; studies = 3; I^2 = 81%)). However, we are very uncertain of these data and it is unlikely that there is a clinically meaningful difference. By 12 months we are even less certain of the data, due to high numbers of women dropping out, most likely due to disease progression.

The smaller studies of Onda 2016 (301 women) and Fagotti 2016 (171 women) published perioperative morbidity data initially. Updated survival data were published for both of these studies in 2020, including a larger cohort in Fagotti 2016 than in the initial cohort.

Heterogeneity of disease burden and treatments between studies

One of the criticisms of the Vergote 2010 and Kehoe 2015 studies has been that the macroscopic cytoreduction rates for both arms were lower than those reported in retrospective cohort studies. However, Vergote 2010 and Kehoe 2015 both included women with extensive disease: ~70% of women in each arm with metastatic deposits measuring > 5 cm, and a quarter of all participants had stage IV disease (Vergote 2010 specifically excluded stage IIIc disease based on para-aortic or pelvic lymph node metastases unless para-aortic lymph nodes larger than 2 cm). In Vergote 2010, 61% in the PDS arm had individual metastatic deposits larger than 10 cm (74% larger than 5 cm). Ten women in the PDS arm and 19 in the NAC/IDS arm were unable to receive either study treatment in Kehoe 2015 due to disease burden. This is similar to Onda 2016 where almost a third of women had stage IV disease. This is likely to represent the surgical equipoise at that time, so women with more bulky disease, thought to be less likely to be optimally debulked, were entered into the studies and women with disease thought amenable to surgery were not enrolled. This contrasts with Fagotti 2016 where much fewer women had stage IV disease (13 women (15.5%) women in the PDS arm versus eight women (9.2%) in the NACT/IDS arm). Additionally, in Fagotti 2016 women were only included, if they were deemed optimally debulkable (residual tumour < 1 cm) at laparoscopy, resulting in 130 of 240 women who underwent a staging laparoscopy being excluded from randomisation in the initial cohort of 110 patients (15 procedures

aborted due to too extensive disease for laparoscopy; 69 were excluded because of a PI score <8; 31 excluded due to a PI score > 12 or 31 who had presence of mesenteric retraction; and a further 15 found not to have epithelial ovarian/fallopian/peritoneal cancer). Women in Fagotti 2016 were also younger than those in the other three studies (PDS arm mean age 54.8 years (n = 84; SD = 9.7) versus 56.2 years (n = 87; SD = 10.7) in NACT arm). This study is therefore not representative of the many women with ovarian cancer, which limits its applicability when examined in isolation.

In the Japanese multi-centre Onda 2016 study, of 147 women who underwent PDS, optimal debulking was achieved in 37%. More than a third of women in the PDS arm underwent an additional attempt at cytoreductive surgery (additional debulking surgery (ADS)), despite maximal surgical effort at initial surgery, taking the total optimal debulking proportion (<1 cm residual disease) to 63% in the PDS arm (PDS + ADS after four cycles of chemotherapy). This is a significant amount of additional treatment in the PDS arm compared to the NACT/IDS arm and puts the study at high risk of performance bias, since these women received additional treatment compared to those in the NACT arm, which was selectively delivered, as the study participants and personnel were not blinded. A proportion of women in the Onda 2016 and Vergote 2010 studies underwent PDS and ADS (37% and 17%, respectively) (after four cycles of chemotherapy in Onda 2016 and six cycles in Vergote 2010). Kehoe 2015 also allowed for ADS after PDS, if incompletely debulked at PDS, but we have been unable to determine if any in the PDS arm underwent further ADS, and it would appear that none did. It would be expected that women in the PDS arm who underwent primary and ADS, to leave a lower volume of residual disease, should have superior outcomes to those women who had NACT-IDS, if surgical effort is the only determinant of survival; this does not seem to be the case from these RCT-level data.

The Fagotti 2016 trial was a mono-centric trial which only randomised women to the trial once they had undergone a staging laparoscopy that produced a predictive index score of disease burden of between ≥ 8 or ≤ 12, predictive of achieving optimal cytoreduction (Vizzielli 2014). If women were deemed as not able to have optimal cytoreduction, they were not eligible for randomisation. Not surprisingly, the macroscopic debulking rates achieved in the Fagotti 2016 study were higher than those of the other studies in the review; 90.9% of women in the PDS arm achieved optimal debulking to < 1 cm of residual disease (45.5% macroscopically debulked) compared with 90.4% in the NACT-IDS arm (57.7% macroscopically debulked). The improved median overall survival of up to 43 months in Fagotti 2016, in comparison with 27 months from the individual patient meta-analysis of the EORTC and CHORUS trial (Vergote 2018) represents differences in age, disease burden and additional chemotherapy agents (notably bevacizumab and Poly (ADP-ribose) polymerase (PARP) inhibitors). This is pertinent as the Vergote 2010 study, in further analyses (Van Meurs 2013) found that NACT particularly benefited women with stage IV disease with individual metastatic deposits of ≥ 4.5 cm, whereas PDS may be preferable for those with stage IIIc disease and individual metastatic deposits <4.5 cm. In women with either stage IIIc disease with larger metastatic deposits (≥4.5 cm) and those with stage IV disease and smaller volume metastatic disease (<4.5 cm) PDS and NACT were similarly effective. The more general applicability of the Fagotti 2016 trial is therefore compromised by selecting only those who are deemed as having the potential for optimal debulking rather than all-comers. Additionally, although complete debulking to no residual disease is associated with a survival advantage, given that, to date, there has been no RCT comparing PDS or NACT followed by IDS to chemotherapy alone, by not attempting any surgical treatment on the subset of women who had very bulky disease it is unclear if any differences in OS or PFS would have been apparent, if they had been included in the trial. Excluding women with a predictive index (PI) score of ≥12 therefore may have prevented those women who may have most benefited from NACT-IDS from inclusion in the study. It is therefore interesting that, despite differences in patient selection and subsequent treatment between the studies, findings were largely similar between the studies. This adds to the applicability of these findings.

Quality of the evidence

We consider the current evidence for primary outcomes of overall and progression-free survival to be of high to moderate-certainty. Further research may have an impact on our confidence in the estimates of effects and may change the estimates, overall and/or for subgroups of women with advanced ovarian cancer. We consider the evidence with regard to surgical morbidity and adverse events to be of high to low-certainty, downgraded due to risk of bias and a small number of events and further research may change these estimates. QoL outcomes provided very low-certainty evidence, mainly due to inconsistency, imprecision and substantial attrition.

Potential biases in the review process

To our knowledge there are no biases in the review process, other than a potential for bias due to the introduction of subgroup analyses (i.e. stage, age and residual disease) in the last update of the review that were not specified in the original protocol. At the stage this decision was made (first update), there was only one included study. The decision for subgroup analyses was therefore made prior to inclusion of the majority of studies in this version of the review. Specifically, the one author of previous versions of this review who was involved in a study included in this update had no role in screening, decisions about inclusion/exclusion, data extraction or analysis.

We still hope to include data from the Kumar 2009 trial. However, at the time of writing, the investigators had not published their final analyses, despite the trial being scheduled to be completed by 2012. We made the decision to discuss the interim data from this trial in Agreements and disagreements with other studies or reviews rather than as an included trial with incomplete outcomes to avoid potentially biasing the results. Once these data are published along with the results of the other ongoing trials (Mahner 2017; NCT04257786; NCT04515602; SUNNY), we plan to update the review.

Agreements and disagreements with other studies or reviews

Other studies

Investigators of the ongoing study Kumar 2009, have presented interim results (at the ACSO conferences in 2006 and 2007) despite the trial being scheduled for completion in 2012. Preliminary data from Kumar 2009 appear to corroborate the findings of the other included studies in this review. In the 2009 abstract, the

investigators reported no significant differences in OS and PFS with HRs for OS and PFS of 0.94 (95% CI 0.56 to 1.56) and 1.1 (95% CI 0.71 to 1.86), respectively (PDS versus NACT). Blood loss, perioperative mortality, postoperative infections and length of hospital stay were all reduced in the NACT group; in addition, QoL scores were significantly better in the NACT group "at the end of treatment" (P < 0.001). We understand from correspondence with Professor Kumar (from Sept 2011 to January 2012 and again in January 2019) that this trial is now closed, that new analyses are being undertaken and that data will be presented in manuscript form soon. Owing to insufficient data in the 2009 report and discrepancies in some of the reported findings over time, we took the decision to await the final statistical analyses before including the interim data in meta-analyses (see Characteristics of ongoing studies).

The study by Kumari 2020 was a prospective pilot RCT conducted in India (Jan 2012-Dec 2013) comparing early IDS after three cycles of NACT (control arm) with late IDS after six cycles of NACT (experimental arm). The study recruited 30 women with advanced ovarian epithelial cancer, the hypothesis being that late IDS would improve optimal cytoreduction rates. Optimal cytoreduction (defined as <1 cm deposits residual disease) was achieved more frequently in the late IDS arm (60%) compared to the early IDS arm (23%) (Odds ratio 10.5; P=0.01). Delivering six cycles of NACT before IDS increased the likelihood of achieving optimal cytoreduction, by a factor of 10, compared to early IDS. No other factor was associated with cytoreduction rate (CA125 / tumour size / age / performance status). However, women in the late IDS arm had a median of nine cycles of chemotherapy compared to a median of 6 cycles in the early IDS arm (P=0.0041), due to women in the late IDS arm having further chemotherapy following surgery.. Although at major risk of performance bias, this is a useful study, especially in the context of the COVID-19 pandemic when surgery has been delayed due to COVID-19 infection risk and limited access to operating theatres and ITU beds for major debulking procedures. It suggests there is still value in offering IDS to women who haven't been able to have surgery after 3 cycles.

Per-protocol pooled analysis of individual women data from two of the included studies

One study pooled longer-term survival data from women in the Kehoe 2015 and Vergote 2010 studies (Vergote 2018). We included this study as an additional reference to both of the studies from whom women were included. This was a pre-planned analysis prior to the launch of the Kehoe 2015 study. A total of 1220 women were included in the per-protocol pooled analysis (670 from Vergote 2010 and 550 from the Kehoe 2015), of whom 612 women received PDS and 608 NACT. Median follow-up was 7·6 years. When women from both studies were combined there was little or no difference in OS between the NACT and PDS groups (HR 0·97, 95% CI 0·86 to 1·09; P = 0·586). However, women with stage IV disease may have better OS and PFS outcomes with NACT versus PDS (OS HR 0·76, 95% CI 0·58 to 1·00; P = 0·048; PFS HR 0·77, 95% CI 0·59 to 1·00; P = 0·049). They concluded that when choosing between treatment strategies with women at diagnosis "one should account not only for the risk of perioperative morbidity and the possibility of debulking the women's disease to zero residual tumour, but also for FIGO stage and the extent of metastatic disease at presentation." They concluded that NACT, followed by IDS, should be standard of care in women with stage IV disease, with PDS reserved for "exceptional circumstances with easily resectable disease".

Systematic reviews

Systematic reviews of RCTs

A meta-analysis by Dai-yuan 2013 examining the role of IDS in ovarian cancer, combined the RCTs of Vergote 2010 and Rose 2004. However, the Rose 2004 study randomised women who had undergone PDS and three cycles of chemotherapy to undergo a further interval debulking surgery prior to completing three further cycles of chemotherapy or to complete three further cycles of chemotherapy without further IDS. Therefore, this meta-analysis did not compare the timing of chemotherapy in relation to surgery alone. There may also be some irregularities in the data extraction, as the authors state they were extracting data on atrial fibrillation duration, left ventricular size, ejection fraction and sinus rhythm maintenance without anti-arrhythmic drugs (which were not in the original study). The meta-analysis produced similar HRs to this review, despite using a fixed-effect model, as opposed to the random-effects model used in this review. HR for OS 0.98 (95% CI 0.85 to 1.14) and HR for PFS 1.03 (95% CI 0.91 to 1.16).

A systematic review by Yang 2017 included the same four studies (Fagotti 2016; Kehoe 2015; Onda 2016; Vergote 2010) as this review in their meta-analysis of serious adverse event and QoL data, but not survival data. They showed that the NACT group had lower risks of grade 3/4 infections (RR 0.30 95% CI 0.16 to 0.56), gastrointestinal (GI) fistulae (RR 0.24 95% CI 0.06 to 0.95) risk of any grade 3 or 4 event (RR 0.29 95% CI 0.11 to 0.78), and a lower rate of death within 28 days (RR 0.14 95% CI 0.04 to 0.49), although with a similar risk of blood transfusion (RR 0.60 95% CI 0.28 to 1.29). These findings are very similar to this review. Yang 2017 also found that the QoL data favoured the NACT group at the six months follow-up point. The likelihood of achieving a macroscopic debulk was higher in the NACT group (macroscopic debulk RR 1.95 95% CI 1.33-2.87; optimal debulk (< 1 cm) = RR 1.61 95% CI 1.05 to 2.47).

Systematic reviews of RCTs and non-randomised studies

A systematic review and meta-analysis by Xiao 2018 combined Vergote 2010 with nine cohort studies and two case-control studies. They calculated a median OS of 32 months with NACT and 37 months with PDS and a median PFS of 15 months with NACT and 15 months with PDS. Given the inclusion of observational studies in this review, there is likely to be critical risk of selection bias in the NACT group, as the NACT group contained older women with more co-morbidities, poorer performance status, higher CA125 at presentation and later FIGO stage, compared to the PDS group. This review also supported a higher optimal debulking rate achieved with NACT compared to PDS (despite more advanced disease in the NACT group) but, unsurprisingly given the imbalance between the groups, no survival benefit was conferred. The odds ratios produced for serious adverse events were in favour of NACT, although only major infection rates, wound complications and vascular events reached statistical significance.

A meta-analysis by Qin 2018 combined Kehoe 2015 and Vergote 2010 with 22 observational studies: 21 retrospective cohorts and one case-control study. The fixed-effect meta-analysis combining Kehoe 2015 and Vergote 2010 produced an HR for OS of 0.93 (95% CI 0.81 to 1.06) and an HR for PFS of 0.97 (95% CI 0.86 to 1.09), suggesting little or no difference between the two groups, similar to the findings of this review. Further, in keeping with the findings of this review, the risks of some serious adverse events (venous thromboembolism (VTE) , infection and GI events) were

lower in the NACT group. In addition, NACT was associated with a shorter stay in the intensive therapy unit (ITU) and overall shorter hospital stay compared to PDS. There was no difference found in risk of haemorrhage between the two groups. They included data from a trial by Melis 2016 , but this study has subsequently been withdrawn from publication calling into question its validity. As with our review and the reviews discussed below, the rates of optimal debulking were higher in the NACT group, but did not confer a survival advantage.

A meta-analysis by Zeng 2016 combined four RCTs, but like Daiyuan 2013 included different treatment strategies in the NACT/IDS arm: PDS versus NACT/IDS followed by completion chemotherapy (Kehoe 2015; Vergote 2010); PDS followed by chemotherapy with randomisation to either further cytoreductive surgery (ADS) (if progressive disease ruled out) and completion chemotherapy or completion chemotherapy alone (Rose 2004 and Van der Burg 1995). This meta-analysis produced HR for OS 0.94 (95% CI 0.81 to 1.08) and HR for PFS 0.89 (95% CI 0.77 to 1.03). As one would expect, there were high levels of heterogeneity between the studies included. This review also found that NACT favoured being able to achieve optimal cytoreduction (RR = 1.76 (95% CI 1.59 to 1.98)), but again did not confer a survival benefit.

Economic analyses

We did not specifically perform a search for articles examining the health economic effect of PDS versus NAC. However, our search found five studies which compared the approaches in a variety of settings. We will therefore discuss their results as a brief economic commentary and consider a formal economic analysis in future updates of this review.

Cost-effectiveness analyses based on non-randomised cohorts

Poonawalla 2015 identified a cohort of elderly women 65 years of age from the Surveillance, Epidemiology and End-results (SEER) Medicare-linked database in the USA from January 2000 to December 2009. These data are therefore not based on clinically equivalent groups in an RCT-setting, although propensity score was used to correct for differences in baseline characteristics. Costs of care from diagnosis to death or last Medicare claim were estimated, using the phase of care approach, and compared to years of survival to calculate the incremental cost-effectiveness-ratio (ICER). The authors calculated that the average life-time costs of NACT was $17,417 based on 2010 costs (estimated 2021 equivalent values of $21,007/€17,629/£15,109) more than PDS, and that the ICER was $174,173 (estimated 2021 equivalent values of $210,083/€176313/£151,101) due to the 0.1 incremental life-year gained from the NACT approach. Stratifying the women between high and low risk, the ICER for high-risk women was $42,988 per life-year saved (estimated 2021 equivalent values of $51,851/€43,516/£37,299), which met their threshold for cost-effectiveness. High-risk participants were those women known to have worse postoperative outcomes (those >75 years of age with stage 4 disease or those >75 years of age with stage 3 disease and co-morbidity score >/=1) and it was in this group that NACT was deemed cost-effective.

In another study, also from the SEER-Medicare database (1992 to 2009) Forde 2015 estimated the seven-month cost of care following PDS and NCACT for advanced ovarian cancer in women > 65 years of age. Of 4506 women, 82.4% received PDS and 17.6% NACT. Women with stage IV disease were more likely to have NACT. The authors found little or no difference in costs of care for women with stage IIIC disease between PDS and NACT. However, costs for those with stage IV disease were higher in those who had PDS (12% difference; $63,131 for PDS versus $55,302 for NACT; P < 0.0001. Costs were based on 2010 data and this difference of $7828 has an estimated 2021 values of $9441/€7925/£6791. Five-year OS in this non-randomised population was lower in the NACT group for both stage IIIC and IV (stage IIIC HR = 1.27, 95% CI 1.10 to 1.47; stage IV HR = 1.19, 95% CI 1.03 to 1.37).

Cost-effectiveness analyses modelled from RCT data

Rowland 2015 evaluated the cost implications of NACT versus PDS, limiting their analysis to those over 65 years of age. The authors modelled their analyses based on subgroup analyses, based on age, from Vergote 2010. They concluded that NACT was cost-saving compared to PDS in women over 65 years of age and that, assuming equal survival, NACT produced cost savings of $5616 based on 2010 USA Medicare reimbursement rates at that time (calculated as equivalent to $6773/€5685/£4871 in 2021).

A later cost-effectiveness study (Tran 2018) used data from all four studies included in our meta-analysis (Fagotti 2016; Kehoe 2015; Onda 2016; Vergote 2010) to model costs of NACT versus PDS, based on a hypothetical cohort of women aged 65 years with advanced epithelial ovarian cancer (EOC) of median baseline characteristics for women in the USA. They based costs on 2015 providers' fees for Medicare and Medicaid Services, taking into account both surgical and chemotherapy adverse events. They estimated that NACT costs $20,762 per woman compared with $27,796 for PDS, saving $7,034 per woman in the seven-month post-treatment time horizon (calculated as equivalent to $7805/€6549/£5613 in 2021). However, these data are affected by the relatively low macroscopic and optimal (< 1 cm residual disease) debulking rates in the RCTs used for the model.

The same team (Cole 2018) modelled costs of NACT and PDS based on the more aggressive surgical paradigm employed in Fagotti 2016. They based their model on a hypothetical annual cohort of 15,000 women in the USA with advanced ovarian cancer over a one-year time horizon based on US Medicare fee schedules and Hospital Cost and Utilization Project inflation adjusted to 2015. The authors based their calculations on the event rates in those randomised within Fagotti 2016 (not including those who underwent laparoscopy but were excluded from the study), thereby representing a cohort with less bulky disease than the other three studies (Kehoe 2015; Onda 2016; Vergote 2010). They found that NACT was associated with an estimated $142 million costs savings (calculated as equivalent to $157.6 million/€132 million/£113 million in 2021) based on the 15,000 women cohort. There were estimated to be 1098 fewer ovarian cancer related deaths, 1355 additional life-years and 1715 additional quality-adjusted life years (QALYs). NACT was associated with a predicted cost saving of $9452 per woman (calculated as equivalent to $10488/€8796/£7537 in 2021) and a 7.3% lower risk of postoperative death. These data may change now that OS data are available from Fagotti 2016, but have not been updated as yet.

Higher surgical complexity and higher optimal debulking rates are, as demonstrated, likely to widen the difference in costs, since those in the PDS arm require more complex surgery to achieve debulking, from the published RCT data. Re-calculating the costs and cost-effectiveness/QALY now that there are OS data

from Fagotti 2016 and the ongoing/unpublished studies, with higher macroscopic debulking rates and complexity, will be of great interest.

Other reviews

Many review articles and non-randomised cohort studies have been published on this subject, many representing single-institution cohorts and including criticisms of the studies included in this review. Many of these studies are at critical risk of selection bias, especially as many do not examine all patients within a population, including those not fit for surgery initially, and so are likely to over-estimate the benefits of upfront surgery (e.g. Mueller 2016). This emphasises the importance of to focusing on what is known from randomised data, where attempts have been made to limit these significant risks of bias. The reader is referred to the literature, since an in-depth narrative review of non-randomised studies is outside of the scope of this review.

Vergote 2010 performed post hoc multivariate analyses on their data. Achievement of macroscopic debulking was the strongest independent predictor of prolonged survival (P = 0.001), followed by stage IIIc disease (P = 0.001), small tumour size before randomisation (P = 0.001), endometrioid histological type (P = 0.005), and younger age (P = 0.005). This is in keeping with findings of a review by Du Bois 2009 and other non-randomised studies.

Vergote 2011b went on to review the results of their Vergote 2010 study, discussing their results in context with other studies (including Rose 2004 and Van der Burg 1995) and their implications for practice. They recommended selection criteria for utilising NACT in stage IIIc/IV disease. These are the Leuven selection criteria for women when considering NACT and IDS in stage IIIc/IV ovarian cancer include the following:

- tumours greater than 2 cm around the superior mesenteric artery or behind the porta hepatis; or
- intrahepatic metastases or extra-abdominal metastases (excluding resectable inguinal or supraclavicular lymph nodes); or
- poor general condition (e.g. over 80 years of age); or
- extensive serosal invasion necessitating bowel resections of greater than 1.5 m; or
- women who cannot be easily debulked to no residual tumour (e.g. more than one bowel resection, expected operating time greater than four hours).

According to Vergote 2011b, these criteria include ~50% of women with stage IIIc and IV disease in an otherwise unselected population. While agreeing that surgical skills are important, the authors stressed that radicality of surgery should be tailored to the general condition and extent of disease of the women, in order to decrease postoperative morbidity and mortality.

A non-systematic review/opinion piece by Schorge 2014 (interestingly entitled "Primary debulking surgery for advanced ovarian cancer: are you a believer or a dissenter?") argued that the decision about when to operate involves finely balancing an appropriately aggressive surgical technique to achieve macroscopic debulking whilst trying to avoid unnecessary morbidity. They state that data show that women benefit from a single maximal debulking effort, but the timing of that effort remains controversial. As the greatest survival benefit is associated with no macroscopic residual disease after surgery, the ability to assess preoperatively which women are most likely to by effectively cytoreduced, by triaging to either PDS or NACT-IDS, involves many complex factors. These factors include the woman's existing co-morbidities, her current physical condition, the surgical team, preoperative imaging and discussion and decision making between the multi-disciplinary team (MDT) and the woman.

The authors conclude that women who appear to benefit the most from PDS are those with stage IIIA or IIIB disease (excluded from the largest studies of Kehoe 2015 and Vergote 2010), those with stage IIIC and a Fagotti laparoscopic predicative index (PI) score of < 8 (Fagotti 2006; Fagotti 2013; Vizzielli 2014), or those with stage IIIC with promising MDT imaging review at an 'expert' centre routinely able to incorporate ultra-radical procedures. In contrast those women who appear to benefit the most from NACT-IDS are women with stage IIIC disease that is too extensive to be optimally debulked, based on imaging and/or laparoscopic scoring, women with stage IV disease, women with a performance status too poor to undergo an attempt at PDS or women without access to an experienced ovarian cancer surgical team, or elderly or morbidly obese women when ultra-radical procedures appear necessary.

A recent study (Havrilesky 2019) investigated patient preferences for attributes of PDS versus NACT for treatment of newly diagnosed ovarian cancer using a survey, educational video and discrete choice experiment activities. Overall the 101 participants preferred better clinical outcomes, less extensive surgery, lower surgical mortality risks, lower risks of readmission and longer PFS and OS. OS ranked the most important factor for consideration, followed by complications requiring readmission, PFS, surgical mortality, extent of surgery and lastly treatment order. Participants would tolerate higher risks of operative morbidity and mortality to gain more substantial survival outcomes (6 months). Conversely, participants were also willing to accept a reduction in survival outcomes (a 11-month reduction in PFS(95% CI 5 to 19 months) and a 7-month reduction in OS (95% CI 2 to 12 months)) to achieve a reduction in risk of surgical mortality. Limitations of this study were that 95% of participants had already received chemotherapy, a third were currently receiving chemotherapy and a third of all participants had recurrent disease. As the participants were not treatment naïve their previous experiences may have impacted on their perception of and tolerance for treatment risks versus survival advantages gained.

A review by Sato 2014 argues that there may be a difference in the assessment of the degree of macroscopic debulking achieved following PDS or NACT-IDS. As NACT-IDS is associated with tissue fibrosis and adhesions induced by chemotherapy, interpretation of tumour spread within the peritoneal cavity may be compromised. Incomplete tumour resection after NACT-IDS may occur, if perioperative evaluation of tumour spread is incorrect and therefore incomplete resection of potentially resectable areas may occur. The authors argue that microscopically carcinomatous areas have a benign appearance more often after NACT than at primary surgery. The authors highlighted that at present the optimal number of chemotherapy cycles in the NACT-IDS setting is unknown.

Based on the currently available data there has been a shift to offering NACT in some treatment settings. A retrospective national cohort study by Wright 2014 reviewed US SEER data from 1991 to 2007 for women with stage II-IV ovarian cancer. Using regression

analysis to adjust for effects of confounding variables on outcome and propensity score analysis to estimate the probability that a woman would undergo a given intervention, they performed a stratified analysis on women who lived longer than six months and underwent both surgery and chemotherapy in 'high volume' centres. This was defined as a hospital referral region that had more than 25 women attend for cancer-directed therapy, either surgery or chemotherapy over the 16-year period. In the initial observational analysis of 5345 (55.8%) women underwent PDS and 2238 (23.8%) underwent NACT, the remainder had no treatment.

The percentage of women undergoing NACT-IDS increased from 19.7% in 1991 to 31.8% in 2007, with a concomitant decrease in PDS from 63.2% in 1991 to 49.5% in 2007. Women most likely to receive NACT-IDS were older, recently diagnosed (i.e. in the 2000s not 1990s), have serous histology, live in metropolitan areas, have stage III or IV disease and have a Charlson co-morbidity score of 1. The substantial imbalance between treatment groups suggests strong selection bias in the cohort and there were strong associations between area of residence in the USA and primary treatment received. An instrumental variable analysis was performed to assess for geographic variation in treatment pattern (the difference in the expected rates of NACT use and the observed rates of NACT use). Once this instrumental variable analysis was performed, the primary treatment chosen had minimal effect on cancer-specific survival (HR 0.94, 95% CI 0.58 to 1.52) or OS (HR 1.04, % CI 0.67 to 1.60). When the observational cohort and propensity-scored cohort survival data were calculated this favoured PDS (HR 1.27 (95% CI 1.19 to 1.35) and HR 1.24 (95% CI 1.1.5 to 1.34), respectively). The authors concluded that in the subset of women who have both surgery and chemotherapy (regardless of total cycles completed), there is no evidence of a difference in survival regardless of timing of surgery. The median OS in the propensity-scored cohort was 27.2 months in the PDS group and 21 months in the NACT-IDS group, not hugely dissimilar to Vergote 2010 data of 30 months in the NACT-IDS group and 29 months in the PDS group, emphasising the applicability of the RCT data included in this review. The authors acknowledge that excluding women who survived less than six months from the analysis may have biased survival estimates.

A retrospective cohort Rauh-Hain 2017 of women less than 70 years of age without co-morbidities from the National Cancer Database in the USA found 22,962 women had been treated for stage III or IV ovarian cancer between 2003 to 2011. Of these, 3126 women had undergone NACT, with or without subsequent IDS. Using propensity scoring, the authors matched each woman in the NACT group with a woman in the PDS group, controlling for age, year at diagnosis, race, ethnicity, treating facility type, insurance status, stage, histological subtype and grade. The authors compared OS in 2935 matched pairs from the retrospective cohort. Once matched they calculated an OS HR of 1.18 (95% CI 1.11 to 1.26), an 18% higher hazard of death (all-cause mortality) in the NACT group. Although the authors compared the matched pairs on an intention-to-treat basis (women who underwent PDS but never received chemotherapy and women who underwent NACT but never underwent IDS were included) 26% of the NACT group never received surgery implying that either they were not fit enough to undergo surgery or their disease progressed on chemotherapy. As with any observational cohort data there is selection bias in the NACT cohort, as we do not know why treatment decision were made. Prior to the propensity scoring, the NACT group were known to be significantly older and less likely to have stage III disease in comparison with the PDS group. They noted that on sensitivity analysis, "lower survival in women who received NACT could be explained by a higher prevalence of limited performance status in women undergoing NACT". Propensity scoring attempts to reduce selection bias in observational studies, but there may well be other unidentified confounding variables that are present in the NACT group to account for the lower survival figures.

A Korean retrospective (2006 to 2014) cohort review of 435 consecutive women operated on in one centre looked at morbidity and survival differences after a paradigm shift in practice in 2010 to utilise more NACT-IDS (Lee 2018). The authors split the cohort into two groups. Group 1 were women operated on between 2006 to 2010. In this group 181 women (83.3%) underwent PDS and 35 women underwent NACT-IDS (16.2%). Group 2 consisted of women who were operated on between 2011 to 2014 during which time 112 women (51.1%) underwent PDS and 107 (48.9%) underwent NACT-IDS. The paradigm shift involved women being treated with NACT-IDS if they fulfilled one of three considerations: (1) pulmonary or liver parenchymal metastases visible on preoperative imaging; (2) medically inoperable due to co-morbidities; (3) optimal cytoreduction was deemed infeasible due to high tumour burden, as defined by a Fagotti PI score of > 8 at diagnostic laparoscopy. This is in contrast to the Fagotti 2016 study, which included women if the PI score was between 8 and 12. The two groups differed substantially in their baseline characteristics. Group 2 contained significantly more women with stage IV disease, ASA score 2, 3 and 4, higher median CA 125 levels and underwent > six cycles of chemotherapy. Intra-peritoneal chemotherapy was utilised in 13% of group 1 women but none of the women in group 2. The progression-free survival in group 2 compared to group 1 was HR 1.01 (95% CI 0.75 to 1.37) and overall survival HR 0.93 (95% CI 0.63 to 1.36) with no differences in survival despite the increased use of NACT in group 2. The shift to increased use of NACT was also associated with increased rates of achieving a macroscopic debulk (G1 = 10.2%; G2 = 21.5%) without increasing perioperative morbidity and mortality. The rates of performing more complex surgical procedures also increased in group 2 (G1 = 35.6%; G2 = 57.5%) with no change in perioperative morbidity between the two groups. The authors conclude that the use of NACT did not improve the survival rate, however, there were no survival differences between the groups after increased use of NACT, despite the women in group 2 having more stage IV disease, more co-morbidities and more extensive surgery than those women in group 1.

Melamed 2018 conducted a quasi-experimental fuzzy regression discontinuity design (Fuzzy RDD) and cross-sectional analysis comparing five regions in the USA. Two regions (New England and East South Central - 95 hospitals) had rapidly increased their use of NACT in 2011 to 2012 by 27.3% and 23.3%, respectively. These regions were compared to three control regions (South Atlantic, West North Central and East North Central - 378 hospitals) where rates of NACT use in 2011 to 2012 only increased by 2%. They compared survival outcomes, censored at three years after diagnosis, for 6034 women; 1156 women in the increased NACT regions and 4878 women in the control regions. The natural experiment compared the different regions and a cross-sectional analysis compared the year and percentage of NACT use on survival. In 2013, two out of the three control regions increased their use of NACT, which allowed for further comparison between control regions. All-cause mortality in the increased NACT regions decreased HR 0.81(95% CI 0.71 to 0.94) compared to the control

regions, which saw no change in all cause mortality (HR 1.02, 95% CI 0.93 to 1.12). Death rates within 30- and 90-days of surgery also decreased in the regions that had increased NACT (30-day mortality from 3.1% to 1.8% and 90-day mortality from 7.0% to 4.0%), which also differed from the control regions (30-day mortality from 1.9% to 2.2%; and 90-day mortality from 5.0% to 4.3%). The two control regions that went on to increase their use of NACT in 2013 also saw a reduction in mortality hazard compared to the control region that did not increase the use of NACT. The authors concluded that survival increased in the regions with increased use of NACT because NACT decreased surgical morbidity and mortality and that this reduction is greater in clinical practice than that seen in RCTs. They postulated whether PDS might be more extensive in the USA than in countries that have been involved in RCTs comparing PDS and NACT, which might explain the increased survival benefits in their cohort. The authors acknowledged that survival benefits may attenuate after three years, the time point at which their data were censored, compared to RCT data, which censored follow-up at five years. They concluded that not all women will benefit from NACT and that the survival benefit seen has been from increased adoption of NACT, occurring selectively in those women with stage IV disease and older women. They also highlight that the regions that increased their use of NACT had higher baseline perioperative mortality than control regions and speculated whether, in those regions with better than average surgical outcomes, increased use of NACT might not achieve the same increase in survival benefits.

AUTHORS' CONCLUSIONS

Implications for practice

It is of note that the role of NACT versus PDS remains an area of controversy in the gynaecological oncology community, despite four well-conducted studies, with differing inclusion criteria, demonstrating little or no difference in survival outcomes and reduced severe adverse events in those who had NACT. It is an area which often suffers from a distinct lack of equipoise. This is most often directed as criticism of the results of the included studies, largely based on concerns regarding low rates of optional/ macroscopic debulking achieved in Kehoe 2015 and Vergote 2010, especially. Further studies have been set up to specifically address some of these concerns, although it should be noted that the Fagotti 2016 study achieved excellent debulking rates, although with the exclusion of higher risk women, both in terms of age and disease status. This limits the applicability of the Fagotti 2016 data on its own to the wider population of women with advanced ovarian cancer, but strengthens the outcomes and applicability within the context of the meta-analysis.

Current evidence is that a combination of chemotherapy and debulking surgery with maximal tolerable effort, is standard treatment for women with advanced ovarian cancer. The order of these treatment modalities appears to have little or no difference on survival outcomes for the overall population. These data support the role of PDS as treatment for advanced (stage IIIc/ IV) ovarian cancer where achieving a macroscopic debulk can be reasonably expected. NACT may be a reasonable (or preferred) alternative for women with stage IV disease, poor performance status or co-morbidities. Compared to PDS, NACT may increase the rate of macroscopic cytoreduction, but this does not appear to translate into an increase in OS. e know from another RCT that removal of microscopic lymph node disease does not improve survival (Harter 2019). The authors of Fagotti 2016 in their discussion noted that those with macroscopically debulked disease and those with residual disease <1 cm at PDS "have superimposable median progression-free survival". These data suggest that small volume, chemotherapy-sensitive disease deposits are effectively treated by neoadjuvant chemotherapy.

The existing quality of evidence is of high to moderate certainty for survival outcomes and high to low certainty for adverse events and very-low certainty for quality of life (QoL) outcomes. One important outcome for women to consider is that, from these data, NACT reduces the risk by around two-thirds of needing a stoma following the operation (one stoma saved for every seven women who have NACT compared to PDS; number needed to treat for an additional beneficial outcome' (NNTB) = 6.89), which may or may not be reversible later, depending on indication and subsequent response to treatment. NACT also reduces the risk of dying after surgery (3 fewer postoperative death for every 100 women having NACT compared to PDS; NNTB = 30.3); these outcomes were of high certainty.

The Leuven selection criteria (Vergote 2011b; Vergote 2016) may offer a reasonable guide to women selection for PDS versus NACT, although it would be important to validate these criteria in a clinical trial setting.

As far as we are aware, there is, to date, no study that compares NACT/ interval debulking surgery (IDS) with NACT alone, although this review did not specifically search for studies in this area. These data are therefore limited to those patients in whom the intention was to perform IDS after NACT at the outset; we have not examined the role of IDS versus no IDS. However, those with disease refractory to chemotherapy have a very poor prognosis and QoL should be the primary concern in this situation, as they are unlikely to benefit from major surgery. The other patient cohort not addressed by these studies are those who may not have been fit enough to be considered surgical candidates at the outset, but whose performance status may be sufficiently improved by chemotherapy to be considered for IDS.

Interestingly, it would appear that some have misinterpreted retrospective data, which show an association between survival and degree of surgical debulking, as evidence that surgery is not indicated, if a macroscopic debulk is not thought achievable. This has not been tested in an RCT setting and cannot be extrapolated from the available data. A recent non-randomised study (NRS), comparing centres with a different surgical ethos, demonstrates that those who have chemotherapy alone, with no attempt at debulking surgery, do poorly (Hall 2019). A recent audit of ovarian cancer care in England demonstrated significant differences in rates of treatment for ovarian cancer between regions, including rates of surgery and combination of surgery and chemotherapy (http://www.ncin.org.uk/cancer_type_and_topic_specific_work/cancer_type_specific_work/gynaecological_cancer/gynaecological_cancer_hub/ovarian_cancer_audit_feasibility_pilot_outputs).

Importantly, data from the studies included in this review do not support or refute an ultra-radical approach to surgery, as patients in both arms had maximal surgical effort.

Cost-benefit analyses based on models derived from RCT data, suggest that a NACT strategy offers improved cost-effectiveness over a one-year time horizon following initial treatment, although these data will require updating now that OS data are available from all of the included studies in this review.

Implications for research

There are currently four ongoing studies (Mahner 2017; NCT04257786; NCT04515602; SUNNY) and one unpublished RCT (Kumar 2009). Mahner 2017 aims to address the role of ultra-radical primary debulking surgery (to achieve higher rates of macroscopic resection) versus NACT/IDS. The results of these studies will hopefully address questions raised by studies with lower optimal and macroscopic debulking rates.

Collection of QoL data is an important patient-centred outcome in advanced ovarian disease, especially if there is minimal difference in survival between treatment options. These were poorly and/or incompletely reported across included studies in this review. Data on rates of stoma formation should also be provided, since women worry about this prior to surgery and it is an important outcome for them.

This review does not address the role of NACT/IDS versus chemotherapy only, without IDS (NACT by definition is followed by other treatment). It can be extrapolated from other studies (e.g. Rose 2004; Van der Burg 1995), that NACT/IDS compared to chemotherapy alone is very likely to improve OS in first-line treatment. A Cochrane Review (Tangjitgamol 2010) demonstrated improved survival for women who had IDS following PDS, but only where there was no previous maximal debulking attempt by a gynaecological oncologist. In addition, results from the studies included in this review show a strong association between achievement of optimal debulking and an improved prognosis. However, studies of secondary debulking surgery in a recurrent disease setting have not been so clear cut and demonstrate improved survival outcomes only in women when macroscopic debulking can be achieved, in one study (Du Bois 2017; Du Bois 2020), but not in another (Coleman 2018). An RCT would be needed to address the value of adding IDS to first-line chemotherapy treatment versus chemotherapy alone, but is very unlikely to be thought to be ethical, as non-randomised data strongly support debulking surgery in a primary setting in women who are fit enough to be considered for major surgery (e.g. Hall 2019).

The Leuven selection criteria (Vergote 2011b; Vergote 2016) or similar triage tools to determine which women would be better served by PDS or NACT as first treatment for advanced ovarian cancer need to be validated in a clinical trial setting and prognostic selection criteria examined in a prognostic methods review.

An interesting article from one of our excluded studies (Wenzel 2017), examined the role of a women decision-making tool to help women come to an individual decision regarding intraperitoneal chemotherapy in ovarian cancer. A similar tool to aid shared decision-making for timing of primary surgery in advanced ovarian cancer would be extremely valuable.

As yet there has never been a randomised study to address the role of ultra-radical surgery in ovarian cancer. Data used to support this approach are based on retrospective review of data, often highly selected and at critical risk of bias. It would not be acceptable in a chemotherapy study to demonstrate survival curves divided retrospectively into groups based on initial response to treatment, yet this routinely happens in surgical studies. Furthermore, the argument for well-conducted prospective randomised trials to confirm or refute doctrine in ovarian cancer debulking is supported by the results of the recent LIONS study (Harter 2019). This was an area where a large number of non-randomised studies, including retrospective series, population studies, and re-analysis of prospective trials, reported an improved survival with systematic lymphadenectomy, as discussed in Eisenhauer 2019, which is similar to the evidence used to support ultra-radical surgery. Harter 2019 performed a well-conducted RCT that compared systematic removal of intra-abdominal lymph nodes with removal of clinically enlarged nodes only. Women were required to have had otherwise macroscopic debulking achieved and were randomised once this had been achieved, during surgery, to systematic lymphadenectomy or debulking of enlarged nodes. They demonstrated no survival benefit from the additional surgery (hazard ratio (HR) for death 1.06; 95% confidence interval (CI), 0.83 to 1.34; P = 0.65), and those who had systematic lymphadenectomy had clinically meaningful increases in serious postoperative complications, including repeat laparotomy (12.4% versus. 6.5%; P = 0.01) and higher death rates within 60 days of surgery (3.1% versus. 0.9%; P = 0.049). This study adds weight to the need for well-balanced RCTs to examine the role of surgery. It would be important to include details of all women not included and/or operated on within the study, so that we can compare outcomes at a population level, ascertain how selective the inclusion criteria are for involvement in the study, and how applicable their findings might be to the general population of women with advanced ovarian cancer. Interestingly, data from a cohort study where ultra-radical surgery was introduced at a population level, did not demonstrate improved outcomes (Falconer 2020). The shift to an ultra-radical surgical approach led to an reduction in the proportion of women who had surgery as part of their treatment (10% fewer), presumably because more women were not thought fit enough for an ultra-radical approach. The lead author, Dr. Salehi, Director of ovarian cancer surgery at Karolinska University Hospital, in Stockholm, Sweden, in a podcast discussing the paper (https://soundcloud.com/bmjpodcasts/salehi-outcomes-of-ultra-radical-surgery-in-ovarian-cancerwav) emphasised the need for studies on survival outcomes of ovarian cancer surgery to publish the outcomes including those who have and do not have debulking surgery within a defined population, since otherwise there is a significant risk of over-estimating the benefits of ultra-radical surgery by selecting out those who do less well. Other questions that remain in first-line treatment of advanced ovarian cancer include optimal treatment options in more elderly women, since few women over 70 years of age were included in any of the studies included in this review. This population is ill-served by clinical trials generally and, with an increasingly elderly population in many countries, this is an ever-expanding cohort of women for who we have little evidence to support recommendations for treatment.

ACKNOWLEDGEMENTS

We would like to thank the following people:

- Alexander Swanton, Krishnayan Haldar, Sally Collins, Tom Lyons and Richard Goodall who contributed to the searching,

evaluation of papers and writing for the original review and previous up-dates.
- Tess Lawrie: sifted updated search results, assessed new papers, performed data extraction and co-wrote the first review update.
- Sean Kehoe: initial idea, supervisor and approval of final version of first version of the review and previous updates. As an author on an included study, SK played no role in study selection, data collection or analysis to prevent undue bias in the review process, as per Cochrane conflict of interest guidelines.
- Ignace Vergote and Corneel Coens for making available unpublished data from Vergote 2010 for the purposes of this review.
- Sean Kehoe and Matthew Nankivell for making available unpublished data from Kehoe 2015 for the purposes of this review.
- Professor Chekman for making available unpublished data from Chekman 2015 for the purposes of this review.
- Anna Fagotti for providing additional data for Fagotti 2016.
- Takashi Onda for providing additional data for Onda 2016.
- J.P.H. Tam for his invaluable translation and for assistance with communication with an author team in China.
- Jane Hayes for designing the original search strategy and Jo Platt for running top-up searches (and for her endless patience with us).
- Gail Quinn and Clare Jess for their contribution to the editorial process.
- The Library Team (especially Natalie Parsley, Cate Newell and Denise Manning (now retired)) of Somerset NHS Foundation Trust. Their continued support and endless enthusiasm for evidence-based medicine and quality improvement is invaluable and we would like to offer our sincere thanks to them.
- Mojtiba Nouhi for obtaining the abstract list for the Iranian meeting found in the search strategy.

This project was supported by the National Institute for Health Research, via Cochrane infrastructure funding to the Cochrane Gynaecological, Neuro-oncology and Orphan Cancer Group. The views and opinions expressed therein are those of the review authors and do not necessarily reflect those of the Systematic Reviews Programme, NIHR, NHS, or the Department of Health.

The authors and Cochrane Gynaecological, Neuro-oncology and Orphan Cancers Team, are grateful to the following peer reviewers for their time and helpful comments: Lars Henning, Alexander Melamed and Monique Spillman.

The authors would like to sincerely thank Professor Paul Pharoah for his feedback on 29/1/21 regarding an error in the October 2019 version of this review.

REFERENCES

References to studies included in this review

Chekman 2015 *{published and unpublished data}*

Chekman C, Layoune R, Hocine O, Raissi N, Ferhat HA, Ali Khodja H, et al. An open prospective randomized trial comparing primary complete cytoreduction surgery to debulking surgery after chemotherapy in advanced stage (FIGO's IIIC) ovarian carcinoma. In: 19th International Meeting of the European Society of Gynaecological Oncology, ESGO 2015; 2015 Oct 24-27; Nice France. 2015:1316.

Fagotti 2016 *{published data only}*

* Fagotti A, Ferrandina G, Vizzielli G, Fanfani F, Gallotta V, Chiantera V, et al. Phase III randomised clinical trial comparing primary surgery versus neoadjuvant chemotherapy in advanced epithelial ovarian cancer with high tumour load (SCORPION trial): final analysis of peri-operative outcome. European Journal of Cancer 2016;**59**:22-33.

Fagotti A, Ferrandina MG, Vizzielli G, Pasciuto T, Fanfani F, Gallotta V, et al. Randomized trial of primary debulking surgery versus neoadjuvant chemotherapy for advanced epithelial ovarian cancer (SCORPION-NCT01461850). *International Journal of Gynecological Cancer* 2020;**30**(11):1657-64. [DOI: 10.1136/ijgc-2020-001640]

Fagotti A, Vizzielli G, Ferrandina G, Fanfani F, Gallotta V, Chiantera V, et al. Survival analyses from a randomized trial of primary debulking surgery versus neoadjuvant chemotherapy for advanced epithelial ovarian cancer with high tumor load (SCORPION trial). *Journal of Clinical Oncology* 2018;**36**(15 Suppl):5516.

Kehoe 2015 *{published and unpublished data}*

* Kehoe S, Hook J, Nankivell M, Jayson GC, Kitchener H, Lopes T, et al. Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer (CHORUS): an open-label, randomised, controlled, non-inferiority trial. *Lancet* 2015;**386**(9990):249-57.

Kehoe S, Hook J, Nankivell M. Chemotherapy or upfront surgery for newly diagnosed advanced ovarian cancer: results from the MRC CHORUS trial. *Journal of Clinical Oncology* 2013;**31 Suppl**(15):Abstract 5500.

Kehoe S, Wheeler S. CHORUS (Chemotherapy or Upfront Surgery). A randomised feasibility trial to determine the impact of timing of surgery and chemotherapy in newly diagnosed patients with advanced epithelial ovarian, primary peritoneal or fallopian tube carcinoma. www.ctu.mrc.ac.uk/plugins/StudyDisplay/protocols/CHORUS protocol Version 2.0 - 05 June 2008.pdf; and http://www.ctu.mrc.ac.uk/research_areas/study_details.aspx?s=9 (accessed 18 June 2012).

Kehoe S. Chemotherapy or upfront surgery for newly diagnosed advanced ovarian cancer: results from the MRC chorus trial. *International Journal of Gynecological Cancer (18th International Meeting of the European Society of Gynaecological Oncology, ESGO; 2013 Oct 19-22; Liverpool, United Kingdom)* 2013;**31**:17.

Law K, Murray C, Kehoe S. CHORUS - a randomised study to determine the impact of timing of surgery and chemotherapy in newly diagnosed patients with advanced epithelial ovarian, primary peritoneal or fallopian tube carcinoma. In: Annual Meeting of the British Gynaecological Cancer Society; 2006: Nov 30-Dec 1; Manchester, UK. 90.

Vergote I, Coens C, Nankivell M, Kristensen GB, Parmar MK, Ehlen T, et al. Neoadjuvant chemotherapy versus debulking surgery in advanced tubo-ovarian cancers: pooled analysis of individual patient data from the EORTC 55971 and CHORUS trials. *Lancet Oncology* 2018;**19**:1680-7.

Onda 2016 *{published data only}*

Onda T, Matsumoto K, Shibata T, Sato A, Fukuda H, Konishi I, et al. Phase III trial of upfront debulking surgery versus neoadjuvant chemotherapy for stage III/IV ovarian, tubal and peritoneal cancers: Japan Clinical Oncology Group Study JCOG0602. *Japanese Journal of Clinical Oncology* 2008;**38**(1):74-7.

Onda T, Satoh T, Ogawa G, Saito T, Kasamatsu T, Nakanishi T, et al, Japan Clinical Oncology Group. Comparison of survival between primary debulking surgery and neoadjuvant chemotherapy for stage III/IV ovarian, tubal and peritoneal cancers in phase III randomised trial. *European Journal of Cancer* 2020;**130**:114-25. [17193357]

* Onda T, Satoh T, Saito T, Kasamatsu T, Nakanishi T, Nakamura K, et al. Comparison of treatment invasiveness between upfront debulking surgery versus interval debulking surgery following neoadjuvant chemotherapy for stage III/IV ovarian, tubal, and peritoneal cancers in a phase III randomised trial: Japan Clinical Oncology Group Study JCOG0602. *European Journal of Cancer* 2016;**64**:22-31.

Onda T, Satoh T, Saito T, Kasamatsu T, Nakanishi T, Takehara K, et al. Comparison of survival between upfront primary debulking surgery versus neoadjuvant chemotherapy for stage III/IV ovarian, tubal and peritoneal cancers in phase III randomized trial: JCOG0602. *Journal of Clinical Oncology* 2018;**36**:15 Suppl.

Vergote 2010 *{published and unpublished data}*

EORTC-55971. Randomized phase III study comparing upfront debulking surgery versus neo-adjuvant chemotherapy in patients with Stage IIIc or IV epithelial ovarian carcinoma. Intergroup Study (EORTC 55971/NCIC OV13). www.cancer.gov/clinicaltrials/EORTC-55971 2003 (accessed 17 June 2012).

Greimel E, Kristensen G, Vergote I, Hoskins P, Van der Burg ME, Casado Herraez A, et al. Quality of life in advanced ovarian cancer patients: a randomized phase III study comparing upfront debulking surgery versus neo-adjuvant chemotherapy. *International Journal of Gynaecological Cancer* 2011;**21**:S620.

Greimel E, Kristensen GB, Van der Burg ME, Coronado P, Rustin G, Del Rio AS, et al. Quality of life of advanced ovarian cancer patients in the randomized phase III study comparing primary debulking surgery versus neo-adjuvant chemotherapy. *Gynecologic Oncology* 2013;**131**(2):437-44.

Vergote I, Amant F, Kristensen G, Ehlen T, Reed NS, Casado A. Primary surgery or neoadjuvant chemotherapy followed by interval debulking surgery in advanced ovarian cancer. *European Journal of Cancer* 2011;**47**(Suppl 3):S88-91.

Vergote I, Coens C, Nankivell M, Kristensen GB, Parmar MK, Ehlen T, et al. Neoadjuvant chemotherapy versus debulking surgery in advanced tubo-ovarian cancers: pooled analysis of individual patient data from the EORTC 55971 and CHORUS trials. *Lancet Oncology* 2018;**19**:1680-7.

Vergote I, Tropé CG, Amant F, Ehlen T, Reed NS, Casado A. Neoadjuvant chemotherapy is the better treatment option in some patients with stage IIIc to IV ovarian cancer. *Journal of Clinical Oncology* 2011;**29**(31):4076-8.

* Vergote I, Trope CG, Amant F, Kristensen GB, Ehlen T, Johnson N, et al. Neoadjuvant chemotherapy or primary surgery in stage IIIC or IV ovarian cancer. *New England Journal of Medicine* 2010;**363**(10):943-53. [Incl. Supplementary Appendix and Protocol]

Verleye L, Ottevanger PB, Kristensen GB, Ehlen T, Johnson N, Van der Burg ME, et al. Quality of pathology reports for advanced ovarian cancer: are we missing essential information? An audit of 479 pathology reports from the EORTC-GCG 55971/NCIC-CTG OV13 neoadjuvant trial. *European Journal of Cancer* 2011;**47**(1):57-64.

References to studies excluded from this review

Ansquer 2001 *{published data only}*

Ansquer Y, Leblanc E, Clough K, Morice P, Dauplat J, Mathevet P, et al. Neoadjuvant chemotherapy for unresectable ovarian carcinoma: a French multicenter study. *Cancer* 2001;**91**(12):2329-34.

Baekelandt 2003 *{published data only}*

Baekelandt M. The potential role of neoadjuvant chemotherapy in advanced ovarian cancer. *International Journal of Gynecological Cancer* 2003;**13 Suppl 2**:163-8.

Bertelsen 1990 *{published data only}*

Bertelsen K. Tumor reduction surgery and long-term survival in advanced ovarian cancer: a DACOVA study. *Gynecologic Oncology* 1990;**38**(2):203-9.

Bidzinski 2005 *{published data only}*

Bidzinski M, Danska-Bidzinska A, Ziólkowska-Seta I, Derlatka P, Sobiczewski P, Raczynski P. Analysis of the treatment of ovarian cancer patients with neo-adjuvant chemotherapy - preliminary results. *European Journal of Gynaecological Oncology* 2005;**26**(4):423-6.

Bristow 2001 *{published data only}*

Bristow R, Tomacruz RS, Armstrong DK, Trimble EL, Montz FJ. Survival impact of maximum cytoreductive surgery for advanced ovarian carcinoma during the platinum-era: a meta-analysis of 6,848 patients. *Proceedings of the American Society of Clinical Oncology* 2001;**20**:(Abstract 807) 202a.

Chambers 1990 *{published data only}*

Chambers JT, Chambers SK, Voynick IM, Schwartz PE. Neoadjuvant chemotherapy in stage X ovarian carcinoma. *Gynecologic Oncology* 1990;**37**(3):327-31.

Chan 2003 *{published data only}*

Chan YM, Ng TY, Ngan HY, Wong LC. Quality of life in women treated with neoadjuvant chemotherapy for advanced ovarian cancer: a prospective longitudinal study. *Gynecologic Oncology* 2003;**88**(1):9-16.

Chan 2017 *{published data only}*

Chan JK, Brady MF, Penson RT, Monk BJ, Kapp DS, Birrer MJ, et al. Neoadjuvant chemotherapy for advanced ovarian, fallopian tube and peritoneal cancer: an ancillary study of GOG 262. Gynecologic Oncology 2017;**145**(1):68.

Chi 2012 *{published data only}*

Chi DS, Musa F, Dao F, Zivanovic O, Sonoda Y, Leitao MM, et al. An analysis of patients with bulky advanced stage ovarian, tubal, and peritoneal carcinoma treated with primary debulking surgery (PDS) during an identical time period as the randomized EORTC-NCIC trial of PDS vs neoadjuvant chemotherapy (NACT). *Gynecologic Oncology* 2012;**124**:10-4.

Cole 2018 *{published data only}*

Cole AL, Barber EL, Gogate A, Tran A, Wheeler SB. Economic analysis of neoadjuvant chemotherapy versus primary debulking surgery for advanced epithelial ovarian cancer using an aggressive surgical paradigm. *International Journal of Gynecologic Cancer* 2018;**28**(6):1077-84.

Colombo 2009 *{published data only}*

Colombo PE, Mourregot A, Fabbro M, Gutowski M, Saint-Aubert B, Quenet F, et al. Aggressive surgical strategies in advanced ovarian cancer: a monocentric study of 203 stage IIIC and IV patients. *Journal of Cancer Surgery* 2009;**35**:135-43.

Cowan 2017 *{published data only}*

Cowan RA, Chi DS, Fagotti A, Scambia G. Point/counterpoint: primary debulking surgery versus neoadjuvant chemotherapy for newly diagnosed advanced ovarian cancer. *Oncology (Williston Park)* 2017;**31**(6):453-8 POINT and 460-1 COUNTERPOINT.

Da Costa 2014 *{published data only}*

Da Costa AA, Valadares CV, Saito A, Ribeiro AR, Tariki M, Guimaraes AP, et al. Primary versus interval debulking surgery and the risk to induce platinum resistance. *Journal of Clinical Oncology* 2014;**32**(15 Suppl):5588.

Dai-yuan 2013 *{published data only}*

Dai-yuan M, Bang-xian T, Xian-fu L, Ye-qin Z, Hong-Wei C. A meta-analysis: neoadjuvant chemotherapy versus primary surgery in ovarian carcinoma FIGO stage III and IV. *World Journal of Surgical Oncology* 2013;**11**:267-71.

Daniele 2017 *{published data only}*

Daniele G, Lorusso D, Scambia G, Cecere SC, Nicoletto MO, Breda E, et al. Feasibility and outcome of interval debulking surgery (IDS) after carboplatin-paclitaxel-bevacizumab (CPB): a

subgroup analysis of the MITO-16A-MaNGO OV2A phase 4 trial. Gynecologic Oncology 2017;**144**(2):256-9.

Deval 2003 *{published data only}*

Deval BP, Platini C, Combe M, Boiron C, Mignot L, Geay J, et al. Surgery: an option for patients with FIGO stage IV ovarian cancer treated by platinum-paclitaxel-based regimen? A GINECO study. *Proceedings of the American Society of Clinical Oncology* 2003;**22**:452; Abstract 1817.

Dutta 2005 *{published data only}*

Dutta T, Sharma H, Kumar L, Dinda AK, Kumar S, Bhatla N, et al. Neoadjuvant chemotherapy for epithelial ovarian cancer - role of apoptosis. *Cancer Chemotherapy and Pharmacology* 2005;**56**(4):427-35.

ESGO 2013 *{published data only}*

European Society of Gynaecological Oncology. European Society of Gynaecological Oncology, ESGO, 18th International Meeting; 2013. *International Journal of Gynecological Cancer* 2013;(Meeting abstracts).

Evdokimova 1982 *{published data only}*

Evdokimova NI, Grigorova TM. Comparative study of 2 combined treatment regimens in stage-III to -IV ovarian cancer. *Voprosy Onkologii* 1982;**28**(7):28-34.

Everett 2006 *{published data only}*

Everett EN, French AE, Stone RL, Pastore LM, Jazaeri AA, Andersen WA. Initial chemotherapy followed by surgical cytoreduction for the treatment of stage III/IV epithelial ovarian cancer. *American Journal of Obstetrics and Gynecology* 2006;**195**(2):574-6.

Fagö-Olsen 2014 *{published data only}*

Fagö-Olsen CL, Ottesen B, Kehlet H, Antonsen SL, Christensen IJ, Markauskas A, et al. Differences in regional diagnostic strategies and in intended versus actual first-line treatment of patients with advanced ovarian cancer in Denmark. *International Journal of Gynecological Cancer* 2014;**24**(7):1195-205.

Fagotti 2018 *{published data only}*

Fagotti A, Scambia G. Neoadjuvant chemotherapy versus upfront debulking surgery in advanced tubo-ovarian cancer. *Lancet Oncology* 2018;**19**(12):1558-60.

Fanfani 2003 *{published data only}*

Fanfani F, Ferrandina G, Corrado G, Fagotti A, Zakut HV, Mancuso S, et al. Impact of interval debulking surgery on clinical outcome in primary unresectable FIGO stage IIIc ovarian cancer patients. *Oncology* 2003;**65**(4):316-22.

Feng 1998 *{published data only}*

Feng Y, Sun T. Short-term effects of chemotherapy-surgery-chemotherapy regimen on clinically inoperable advanced ovarian cancer. *Chinese Medical Journal* 1998;**111**(8):722-5.

Forde 2015 *{published data only}*

Forde GK, Chang J, Ziogas A, Tewari KS, Bristow RE. Primary debulking surgery and neo-adjuvant chemotherapy in the Medicare population: an analysis of cost of care. *Gynecologic Oncology* 2015;**137**:109-10.

Fujiwara 2013 *{published data only}*

Fujiwara K, Kurosaki A, Hasegawa K. Clinical trials of neoadjuvant chemotherapy for ovarian cancer: what do we gain after an EORTC trial and after two additional ongoing trials are completed? *Current Oncology Reports* 2013;**15**(3):197-200.

Ghaemmaghami 2008 *{published data only}*

Ghaemmaghami F, Karimi-Zarchi M, Modares-Gilani M, Mousavi A, Behtash N. Clinical outcome of Iranian patients with advanced ovarian cancer with neoadjuvant chemotherapy versus primary debulking surgery. *Asia Pacific Journal of Cancer Prevention* 2008;**9**(4):719-24.

Giannopoulos 2006 *{published data only}*

Giannopoulos T, Butler-Manuel S. Clinical outcomes of neoadjuvant chemotherapy and primary debulking surgery in advanced ovarian carcinoma. *European Journal of Gynaecological Oncology* 2006;**27**(1):25-8.

Grosso 2013 *{published data only}*

Grosso LG, Lotti M, Rossetti D, Ansaloni L, Frigerio L. Cytoreduction and hipec vs only cytoreduction surgery after neoajuvant chemotherapy for treatment of ovarian cancer naive patients: a phase III multi-center randomized ongoing trial. *International Journal of Gynecological Cancer (18th International Meeting of the European Society of Gynaecological Oncology, ESGO, 2013 Oct 19-22; Liverpool, United Kingdom* 2013:897.

Hanker 2010 *{published data only}*

Hanker LC. Complete surgical debulking in advanced ovarian carcinoma improves prognosis in any FIGO stage: analysis of 3,126 prospectively randomized patients in AGO-OVAR/GINECO phase 3 trials. *Archives of Gynecology and Obstetrics (58th Congress of the German Society for Gynecology and Obstetrics (Deutsche Gesellschaft fur Gynakologie und Geburtshilfe, DGGG); 2010 Oct)* 2010;**282**.

Hegazy 2005 *{published data only}*

Hegazy MA, Hegazi RA, Elshafei MA, Setit AE, Elshamy MR, Eltatoongy M, et al. Neoadjuvant chemotherapy versus primary surgery in advanced ovarian carcinoma. *World Journal of Surgical Oncology* 2005;**3**:57.

Hou 2007 *{published data only}*

Hou JY, Kelly MG, Yu H, McAlpine JN, Azodi M, Rutherford TJ, et al. Neoadjuvant chemotherapy lessens surgical morbidity in advanced ovarian cancer and leads to improved survival in stage IV disease. *Gynecological Oncology* 2007;**105**(1):211-7.

Inciura 2006 *{published data only}*

Inciura A, Simavicius A, Juozaityte E, Kurtinaitis J, Nadisauskiene R, Svedas E, et al. Comparison of adjuvant and neoadjuvant chemotherapy in the management of advanced ovarian cancer: a retrospective study of 574 patients. *BMC Cancer* 2006;**6**:153.

Iranian Society Reproductive Medicine Conference *{published data only}*

Iranian Society for Reproductive Medicine. Iranian Society for Reproductive Medicine 2012 3rd International and 18th National Congress. *Iranian Journal of Reproductive Medicine* 2012;(Meeting abstracts).

Jacob 1991 *{published data only}*

Jacob JH, Gershenson DM, Morris M, Copeland LJ, Burke TW, Wharton JT. Neoadjuvant chemotherapy and interval debulking for advanced epithelial ovarian cancer. *Gynecologic Oncology* 1991;**42**(2):146-50.

Kayikcioglu 2000 *{published data only}*

Kayikcioglu F, Kose MF, Boran N, Ozdas E, Ozgul N, Tulunay G. Neoadjuvant chemotherapy in advanced stage ovarian carcinoma. In: VIII Meeting of the International Gynecologic Cancer Society; 2000; Buenos Aires, Argentina. 2000. [Abstract 48]

Kayikcioglu 2001 *{published data only}*

Kayikcioglu F, Kose MF, Boran N, Caliskan E, Tulunay G. Neoadjuvant chemotherapy or primary surgery in advanced epithelial ovarian carcinoma. *International Journal of Gynecological Cancer* 2001;**11**(6):466-70.

Kehoe 2011 *{published data only}*

Kehoe S, Nankivell M, Cairns J, Qian W, Swart AM. Problems recruiting to surgical trials: examples from the MRC/NRCI Chorus randomised clinical trial. *International Journal of Gynecological Cancer* 2011;**21**:S678.

Kuhn 2001 *{published data only}*

Kuhn W, Rutke S, Spathe K, Schmalfeldt B, Florack G, Von Hundelshausen B, et al. Neoadjuvant chemotherapy followed by tumor debulking prolongs survival for patients with poor prognosis in International Federation of Gynecology and Obstetrics Stage IIIC ovarian carcinoma. *Cancer* 2001;**92**(10):2585-91.

Kumar 2015 *{published data only}*

Kumar L, Pramanik R, Kumar S, Bhatla N, Malik S. Neoadjuvant chemotherapy in gynaecological cancers - implications for staging. *Best Practice & Research. Clinical Obstetrics & Gynaecology* 2015;**29**(6):790-801.

Lawton 1989 *{published data only}*

Lawton FG, Redman CW, Luesley DM, Chan KK, Blackledge G. Neoadjuvant (cytoreductive) chemotherapy combined with intervention debulking surgery in advanced, unresected epithelial ovarian cancer. *Obstetrics and Gynecology* 1989;**73**(1):61-5.

Lee 2006 *{published data only}*

Lee SJ, Kim BG, Lee JW, Park CS, Lee JH, Bae DS. Preliminary results of neoadjuvant chemotherapy with paclitaxel and cisplatin in patients with advanced epithelial ovarian cancer who are inadequate for optimum primary surgery. *Journal of Obstetric and Gynaecological Research* 2006;**32**(1):99-106.

Lee 2018 *{published data only}*

Lee YJ, Chung YS, Lee JY, Nam EJ, Kim SW, Kim S, et al. Impact of increased utilization of neoadjuvant chemotherapy on survival in patients with advanced ovarian cancer: experience from a comprehensive cancer center. *Journal of Gynecologic Oncology* 2018;**29**(4):e63.

Lim 1993 *{published data only}*

Lim JT, Green JA. Neoadjuvant carboplatin and ifosfamide chemotherapy for inoperable FIGO stage III and IV ovarian carcinoma. *Clinical Oncology (Royal College of Radiologists (Great Britain))* 1993;**5**(4):198-202.

Liu 1995 *{published data only}*

Liu S, Jiang D, Xu G. Advanced ovarian cancer: combination chemotherapy and cytoreductive surgery. *Acta Academiae Medicinae Hubei* 1995;**16**(4):343-4.

Liu 2004 *{published data only}*

Liu EL, Mi RR. Neoadjuvant intraarterial chemotherapy and embolization in treatment of advanced ovarian epithelial carcinoma. *Chinese Medical Journal* 2004;**117**(10):1547-51.

Liu 2015 *{published data only}*

Liu J. Should neoadjuvant chemotherapy be preferred to an alternative treatment for advanced ovarian cancer: comparison of neoadjuvant chemotherapy followed by interval debulking surgery and primary debulking surgery in patients with advanced ovarian cancer. *Gynecologic Oncology (46th Annual Meeting on Women's Cancer of the Society of Gynecologic Oncology, SGO, 2015 March 28-31; Chicago (IL) United States)* 2015:176.

Liu 2017 *{published data only}*

Liu EL, Mi RR, Wang DH, Wang LQ, Zhang YM, Chen WM. Application of combined intraperitoneal and intravenous neoadjuvant chemotherapy in senile patients with advanced ovarian cancer and massive ascites. *European Journal of Gynaecological Oncology* 2017;**38**(2):209-13.

Loizzi 2005 *{published data only}*

Loizzi V, Cormio G, Resta L, Rossi CA, Di Gilio AR, Cuccovillo A, et al. Neoadjuvant chemotherapy in advanced ovarian cancer: a case-control study. *International Journal of Gynecological Cancer* 2005;**15**(2):217-23.

Lotze 1987 *{published data only}*

Lotze W, Richter P, Sarembe B. Intra-arterial chemotherapy in advanced ovarian cancers 2. Therapeutic results in relation to prognostic factors. *Zentralblatt für Gynäkologie* 1987;**109**(9):578-85.

Lyngstadaas 2005 *{published data only}*

Lyngstadaas A, Ekanger R, Hagen B, Himmelmann A, Iversen OE, Iversen T, et al. Primary treatment of ovarian cancer. *Tidsskrift for den Norske Laegeforening* 2005;**125**(3):278-81.

Mackay 2011 *{published data only}*

Mackay HJ. Phase II/III study of intraperitoneal chemotherapy after neoadjuvant chemotherapy for ovarian cancer: NCIC CTG OV.21. *Current Oncology* 2011;**18**(2):84-90.

Mahner 2006 *{published data only}*

Mahner S, Park TW, Ortmann O, Hilfrich J, Breitbach GP, Höss C, et al. Neoadjuvant chemotherapy with carboplatin and docetaxel in advanced ovarian cancer. A randomized multicenter phase II study (PRIMOVAR). *International Journal of Gynecological Cancer* 2006;**16**(S3):659.

Mahner 2014 *{published data only}*

Mahner S, Harter P, Hilpert F, Pfisterer J, Du Bois A, Chi D. Neoadjuvant or postoperative therapy for advanced ovarian cancer. *Oncology Research and Treatment (Jahrestagung der Deutschen, Osterreichischen und Schweizerischen Gesellschaften fur Hamatologie und Medizinische Onkologie; 2014 Oct 10-14; Hamburg, Germany)* 2014:116.

Makar 2016 *{published data only}*

Makar AP, Trope CG, Tummers P, Denys H, Vandecasteele K. Advanced ovarian cancer: primary or interval debulking? Five categories of patients in view of the results of randomized trials and tumor biology: primary debulking surgery and interval debulking surgery for advanced ovarian cancer. *Oncologist* 2016;**21**(6):745-54.

Malzoni 1993 *{published data only}*

Malzoni M, Palagiano A, Palmese A. Neo-adjuvant chemotherapy in ovarian carcinoma (case report). *Rassegna Internazionale di Clinica e Terapia* 1993;**73**(7):309-12.

Mazzeo 2003 *{published data only}*

Mazzeo F, Berliere M, Kerger J, Squifflet J, Duck L, D'Hondt V. Neoadjuvant chemotherapy followed by surgery and adjuvant chemotherapy in patients with primarily unresectable, advanced-stage ovarian cancer. *Gynecologic Oncology* 2003;**90**(1):163-9.

Melamed 2018 *{published data only}*

Melamed A, Fink G, Wright AA, Keating NL, Gockley AA, Del Carmen MG, et al. Effect of adoption of neoadjuvant chemotherapy for advanced ovarian cancer on all cause mortality: quasi-experimental study. *BMJ* 2018;**360**:5463.

Morice 2003 *{published data only}*

Morice P, Dubernard G, Rey A, Atallah D, Pautier P, Pomel C, et al. Results of interval debulking surgery compared with primary debulking surgery in advanced stage ovarian cancer. *Journal of the American College of Surgeons* 2003;**197**(6):955-63.

Negretti 1988 *{published data only}*

Negretti E, Zambetti M, Luciani L, Gianni L. Timing of surgery and the role of cytoreductive chemotherapy in patients with advanced ovarian carcinoma. *Tumori* 1988;**74**(5):567-72.

Nick 2015 *{published data only}*

Nick AM, Coleman RL, Ramirez PT, Schmeler KM, Soliman PT, Lu KH, et al. Personalized surgical therapy for advanced ovarian cancer. *Gynecologic Oncology* 2015;**137**:10.

Oe 2011 *{published data only}*

Oe S, Hasegawa K, Ichikawa R, Torii Y, Kato R, Komiyama S, et al. Treatment outcomes for advanced ovarian cancers with peritoneal dissemination. *Gan to Kagaku Ryoho [Japanese Journal of Cancer & Chemotherapy]* 2011;**38**(4):591-7.

Onda 2009 *{published data only}*

Onda T, Kobayashi H, Nakanishi T, Hatae M, Iwasaka T, Konishi I, et al. Feasibility study of neoadjuvant chemotherapy followed by interval debulking surgery for stage III/IV ovarian, tubal, and peritoneal cancers: Japan Clinical Oncology Group Study JCOG0206. *Gynecologic Oncology* 2009;**113**(1):57-62.

Onnis 1996 *{published data only}*

Onnis A, Marchetti M, Padovan P, Castellan L. Neoadjuvant chemotherapy in advanced ovarian cancer. *European Journal of Gynaecologic Oncology* 1996;**17**(5):393-6.

Polcher 2009 *{published data only}*

Polcher M, Mahner S, Ortmann O, Hilfrich J, Diedrich K, Breitbach GP, et al. Neoadjuvant chemotherapy with carboplatin and docetaxel in advanced ovarian cancer - a prospective multicenter phase II trial (PRIMOVAR). *Oncology Reports* 2009;**22**(3):605-13.

Poonawalla 2015 *{published data only}*

Poonawalla IB, Lairson DR, Chan W, Piller LB, Du XL. Cost-effectiveness of neoadjuvant chemotherapy versus primary surgery in elderly patients with advanced ovarian cancer. *Value in Health* 2015;**18**:387-95.

Prescott 2016 *{published data only}*

Prescott LS, Vergote IB, Coens C, Sun CC, Munsell MF, Casado A, et al. Effect of perioperative blood transfusion on quality of life, progression-free and overall survival in primary treatment of advanced epithelial ovarian cancer: an EORTC ancillary study. *Gynecologic Oncology* 2016;**141**:198.

Qin 2018 *{published data only}*

Qin M, Jin Y, Ma L, Zhang Y, Pan L. The role of neoadjuvant chemotherapy followed by interval debulking surgery in advanced ovarian cancer: a systematic review and meta-analysis of randomized controlled trials and observational studies. *Oncotarget* 2018;**9**(9):8614-28.

Querleu 2013 *{published data only}*

Querleu D, Rafii A, Colombo PE, Ferron G, Rouanet P, Martinez A. Randomized study of aggressive surgery for advanced ovarian cancer. *International Journal of Gynecological Cancer* 2013;**23**(7):1170.

Rafii 2007 *{published data only}*

Rafii A, Deval B, Geay JF, Chopin N, Paoletti X, Paraiso D, et al. Treatment of FIGO stage IV ovarian carcinoma: results of primary surgery or interval surgery after neoadjuvant chemotherapy: a retrospective study. *International Journal of Gynecological Cancer* 2007;**17**(4):777-83.

Rauh-Hain 2017 *{published data only}*

Rauh-Hain JA, Melamed A, Wright A, Gockley A, Clemmer JT, Schorge JO, et al. Overall survival following neoadjuvant chemotherapy vs primary cytoreductive surgery in women with epithelial ovarian cancer: analysis of the National Cancer Database. *JAMA Oncology* 2017;**3**(1):76-82.

Recchia 2001 *{published data only}*

Recchia F, De Filippis S, Rosselli M, Saggio G, Carta G, Rea S. Primary chemotherapy in stage IV ovarian cancer. a prospective phase II study. *European Journal of Gynaecological Oncology* 2001;**22**(4):287-91.

Redman 1994 *{published data only}*

Redman CW, Warwick J, Luesley DM, Varma R, Lawton FG, Blackledge GR. Intervention debulking surgery in advanced epithelial ovarian cancer. *British Journal of Obstetrics and Gynaecology* 1994;**101**(2):142-6.

Robova 2003 *{published data only}*

Robova H, Rob L, Pluta M, Kacirek J, Strnad P, Schlegerova D. Neoadjuvant chemotherapy in patients with primary unresectable ovarian cancer. *International Journal of Gynecological Cancer* 2003;**13 Suppl 1**:44.

Rowland 2013 *{published data only}*

Rowland M, Farris C, Lesnock J, Krivak T. Neoadjuvant chemotherapy is less costly than primary debulking surgery for treatment of advanced stage ovarian cancer in patients > 65 years old. *Gynecologic Oncology (Annual Meeting of the Western Association of Gynecologic Oncologists; 2013 June 26-29; Seattle (WA) United States* 2013:278-9.

Rowland 2015 *{published data only}*

Rowland MR, Lesnock JL, Farris C, Kelley JL, Krivak TC. Cost-utility comparison of neoadjuvant chemotherapy versus primary debulking surgery for treatment of advanced-stage ovarian cancer in patients 65 years old or older. *American Journal of Obstetrics and Gynecology* 2015;**212**(6):763.e1-8.

Rutten 2012 *{published data only}*

Rutten MJ, Gaarenstroom KN, Van Gorp T, Van Meurs HS, Arts HJ, Bossuyt PM, et al. Laparoscopy to predict the result of primary cytoreductive surgery in advanced ovarian cancer patients (LapOvCa-trial): a multicentre randomized controlled study. *BMC Cancer* 2012;**12**:31.

Salzer 1990 *{published data only}*

Salzer H, Genger H, Gober S, Barrada M, Vavra N, Sevelda P. Surgery in the treatment concept of epithelial ovarian cancer. *Gynäkologische Rundschau* 1990;**30 Suppl 1**:26-9.

Sato 2014 *{published data only}*

Sato S, Itamochi H. Neoadjuvant chemotherapy in advanced ovarian cancer: latest results and place in therapy. *Therapeutic Advances in Medical Oncology* 2014;**6**(6):293-304.

Sayyah-Melli 2013 *{published data only}*

Sayyah-Melli M, Zonoozi GK, Hashemzadeh S, Esfahani A, Ouladehsahebmadarek E, Shobeiry MJ, et al. Comparison of platinum-based neoadjuvant chemotherapy and primary debulking surgery in patients with advanced ovarian cancer. *Journal of Obstetrics and Gynaecology of India* 2013;**63**(6):405-9.

Schorge 2014 *{published data only}*

Schorge JO, Clark RM, Lee SI, Penson RT. Primary debulking surgery for advanced ovarian cancer: are you a believer or a dissenter? *Gynecologic Oncology* 2014;**135**(3):595-605.

Schwartz 1994 *{published data only}*

Schwartz PE, Chambers JT, Makuch R. Neoadjuvant chemotherapy for advanced ovarian cancer. *Gynecologic Oncology* 1994;**53**(1):33-7.

Schwartz 1999 *{published data only}*

Schwartz PE, Rutherford TJ, Chambers JT, Kohorn EI, Thiel RP. Neoadjuvant chemotherapy for advanced ovarian cancer: long-term survival. *Gynecologic Oncology* 1999;**72**(1):93-9.

Shibata 2003 *{published data only}*

Shibata K, Kikkawa F, Mika M, Suzuki Y, Kajiyama H, Ino K, et al. Neoadjuvant chemotherapy for FIGO stage III or IV ovarian cancer: survival benefit and prognostic factors. *International Journal of Gynecological Cancer* 2003;**13**(5):587-92.

Shimizu 1993 *{published data only}*

Shimizu Y, Hasumi K. Treatment of stage III and IV ovarian cancer - is neoadjuvant chemotherapy effective? *Nippon Sanka Fujinka Gakkai Zasshi* 1993;**45**(9):1007-14.

Steed 2006 *{published data only}*

Steed H, Oza AM, Murphy J, Laframboise S, Lockwood G, Petrillo D, et al. A retrospective analysis of neoadjuvant platinum-based chemotherapy versus up-front surgery in advanced ovarian cancer. *International Journal of Gynecological Cancer* 2006;**16**(Suppl 1):47-53.

Sun 2000 *{published data only}*

Sun T, Feng Y, Zhu Y, Zheng Y. Therapeutic strategy in the management of stage II - IV epithelial ovarian carcinoma. *Chinese Medical Journal* 2000;**113**(7):625-7.

Surwit 1999 *{published data only}*

Surwit E, Childers J. Cytoreductive surgery in advanced ovarian cancer with or without neoadjuvant chemotherapy. *Gynecologic Oncology* 1999;**72**(3):468.

Taskin 2013 *{published data only}*

Taskin S, Gungor M, Ortac F, Oztuna D. Neoadjuvant chemotherapy equalizes the optimal cytoreduction rate to primary surgery without improving survival in advanced ovarian cancer. *Archives of Gynecology and Obstetrics* 2013;**288**(6):1399-403.

Taylor 2015 *{published data only}*

Taylor SE, Berger J, Johnson K, Boisen MM, Courtney-Brooks MB, Sukumvanich P, et al. Neoadjuvant chemotherapy reduces operative morbidity without effecting time to recurrence in advanced stage epithelial ovarian cancer. *Gynecologic Oncology* 2015;**137**:117-8.

Tran 2018 *{published data only}*

Tran AQ, Erim DO, Sullivan SA, Cole AL, Barber EL, Kim KH, et al. Cost effectiveness of neoadjuvant chemotherapy followed by interval cytoreductive surgery versus primary cytoreductive surgery for patients with advanced stage ovarian cancer during the initial treatment phase. *Gynecologic Oncology* 2018;**148**(2):329-35.

Trope 1997 *{published data only}*

Trope C. Primary debulking surgery is not an independent prognostic factor in advanced stage IIIC ovarian carcinoma. *Gynecologic Oncology* 1997;**64**(2):357.

Ushijima 2002 *{published data only}*

Ushijima K, Ota S, Komai K, Matsuo G, Motoshima S, Honda S, et al. Clinical assessment of neoadjuvant chemotherapy and interval cytoreductive surgery for unresectable advanced ovarian cancer. *International Surgery* 2002;**87**(3):185-90.

Van der Burg 1995 *{published data only}*

Van der Burg ME, Van Lent M, Buyse M, Kobierska A, Colombo N, Favalli G, et al. The effect of debulking surgery after induction chemotherapy on the prognosis in advanced epithelial ovarian cancer. Gynecological Cancer Cooperative Group of the European Organization for Research and Treatment of Cancer. *New England Journal of Medicine* 1995;**332**(10):629-34.

Van Meurs 2013 *{published data only}*

Van Meurs HS, Tajik P, Hof MH, Vergote I, Kenter GG, Mol BW, et al. Which patients benefit most from primary surgery or neoadjuvant chemotherapy in stage IIIC or IV ovarian cancer? An exploratory analysis of the European Organisation for Research and Treatment of Cancer 55971 randomised trial. *European Journal of Cancer* 2013;**49**(15):3191-201.

Varma 1990 *{published data only}*

Varma R, Blackledge G, Redman C, Luesley D, Chan KK, Mould J. A randomised trial of intervention debulking surgery and the duration of cis-platinum combination chemotherapy in advanced epithelial ovarian cancer (EOC). *Annals of Oncology (ESMO Congress)* 1990.;**1**(Suppl 9):4.

Vergote 1998 *{published data only}*

Vergote I, De Wever IW, Tjalma W, Van Gramberen M, Decloedt J, Van Dam P. Neoadjuvant chemotherapy or primary debulking surgery in advanced ovarian carcinoma: a retrospective analysis of 285 patients. *Gynecologic Oncology* 1998;**71**(3):431-6.

Vergote 2000 *{published data only}*

Vergote IB, De Wever I, Decloedt J, Tjalma W, Van Gramberen M, Van Dam P. Neoadjuvant chemotherapy versus primary debulking surgery in advanced ovarian cancer. *Seminars in Oncology* 2000;**27**(3 Suppl 7):31-6.

Vergote 2018 *{published data only}*

Vergote I, Coens C, Nankivell M, Kristensen GB, Parmar MKB, Ehlen T, et al. Neoadjuvant chemotherapy versus debulking surgery in advanced tubo-ovarian cancers: pooled analysis of individual patient data from the EORTC 55971 and CHORUS trials. *Lancet Oncology* 2018;**19**(12):1680-7.

Vergote 2019 *{published data only}*

Vergote I, Coens C, Nankivell M, Kristensen GB, Parmar MKB, Ehlen T, et al. Neoadjuvant chemotherapy versus debulking surgery in advanced tubo-ovarian cancers: pooled analysis of individual patient data from the EORTC 55971 and CHORUS trials. *Obstetrical & Gynecological Survey* 2019;**74**(3):156-8.

Vrscaj 2002 *{published data only}*

Vrscaj MU, Rakar S. Neoadjuvant chemotherapy for advanced epithelial ovarian carcinoma: a retrospective case-control study. *European Journal of Gynaecological Oncology* 2002;**23**(5):405-10.

Wenzel 2017 *{published data only}*

Wenzel L, Mukamel D, Osann K, Havrilesky L, Sparks L, Lipscomb J, et al. Rationale and study protocol for the Patient-Centered Outcome Aid (PCOA) randomized controlled trial: a personalized decision tool for newly diagnosed ovarian cancer patients. *Contemporary Clinical Trials* 2017;**57**:29-36.

Wright 2013 *{published data only}*

Wright J, Ananth C, Herzog T, Burke W, Lu Y, Lewin S, et al. Comparative effectiveness of upfront treatment strategies for advanced-stage ovarian cancer. *Gynecologic Oncology (44th Annual Meeting of the Society of Gynecologic Oncology; 2013 Mar 9-12; Los Angeles (CA) United States)* 2013:e117.

* Wright J, Ananth C, Tsui J, Glied S, Burke W, Lu Y, et al. Comparative effectiveness of upfront treatment strategies in elderly women with ovarian cancer. *Cancer* 2014;**120**(8):1246-54.

Wu 2012 *{published data only}*

Wu XY. Efficacy of neoadjuvant chemotherapy in advanced ovarian cancer. *Journal of Practical Oncology* 2012;**27**(6):650-2.

Xiao 2018 *{published data only}*

Xiao Y, Xie S, Zhang N, Wang J, Lv C, Guo J, et al. Platinum-based neoadjuvant chemotherapy versus primary surgery in ovarian carcinoma International Federation of Gynecology and Obstetrics Stages IIIc and IV: a systematic review and meta-analysis. *Gynecologic and Obstetric Investigation* 2017;**83**(3):209-18.

Yang 2017 *{published data only}*

Yang L, Zhang B, Xing G, Du J, Yang B, Yuan Q, et al. Neoadjuvant chemotherapy versus primary debulking surgery in advanced epithelial ovarian cancer: a meta-analysis of peri-operative outcome. *PLOS One* 2017;**12**:10.

Zamagni 2014 *{published data only}*

Zamagni C, Perrone M, Mandato VD, Bologna A, Rubino D, Zucchini G, et al. Randomized phase II study of 3 versus 6 courses of neoadjuvant carboplatin-paclitaxel chemotherapy in stage IIIC or IV epithelial ovarian cancer. *Journal of Clinical Oncology* 2014;**32**(15):5624.

Zeng 2016 *{published data only}*

Zeng LJ, Xiang CL, Gong YZ, Kuang Y, Lu FF, Yi SY, et al. Neoadjuvant chemotherapy for patients with advanced epithelial ovarian cancer: a meta-analysis. *Scientific Reports* 2016;**6**:35914.

References to studies awaiting assessment

Jiang 2018 *{published data only}*

Jiang Y, He W, Yang H, Su Z, Sun L. Analysis of clinical effects of neoadjuvant chemotherapy in advanced epithelial ovarian cance. *Journal of the Balkan Union of Oncology* 2018;**23**:3.

References to ongoing studies

Kumar 2009 *{published and unpublished data}*

Janga D, Kumar L, Kumar S, Shukla NK, Thulkar S, Singh R. Neoadjuvant chemotherapy (CT) followed by debulking surgery vs upfront surgery followed by chemotherapy in advanced epithelial ovarian carcinoma (EOC): a prospective, randomized study. *Proceedings of the American Society of Clinical Oncology* 2003;**22**:487.

Kumar L, Hariprasad R, Kumar S, Bhatla N, Shukla N, Thulkar S, et al. Neoadjuvant chemotherapy in advanced epithelial ovarian cancer (EOC): a phase III randomized study. *Journal of Clinical Oncology (ASCO Annual Meeting Proceedings Part I 2006 (June 20 Supplement))* 2006;**24**(Suppl):18.

* Kumar L, Hariprasad R, Kumar S, Bhatla N, Thulkar S, Shukla NJ. Neo-adjuvant chemotherapy in advanced epithelial ovarian cancer (EOC): a prospective, randomized study. *Indian Journal of Medical and Paediatric Oncology* 2009;**30**(1):15.

Kumar L, Hariprasad R, Kumar S, Bhatla N, Thulkar S, Vijayaraghavan M, et al. Neoadjuvant chemotherapy followed by interval debulking surgery versus upfront surgery followed by chemotherapy in advanced epithelial ovarian carcinoma: a prospective randomized study - interim results. *Journal of Clinical Oncology (ASCO Annual Meeting Proceedings Part I. 2007)* 2007;**25**(18S Suppl):5531.

Kumar L, Hariprasad R, Kumar S, Bhatla N, Thulkar S, Vijayaraghavan M. Neoadjuvant chemotherapy followed by interval debulking surgery versus upfront surgery followed by chemotherapy in advanced epithelial ovarian carcinoma: a prospective randomized study - interim results. *Journal of Clinical Oncology (ASCO Annual Meeting Proceedings Part I)* 2007;**25**(18 Suppl):5531.

Kumar L, Janga D, Berge S, Gupta S, Kumar S, Bhatla N, et al. Neoadjuvant chemotherapy in stage III & IV epithelial ovarian carcinoma (EOC). *Journal International Medical Sciences Academy* 2003;**16**(2):89-92.

Mahner 2017 *{published data only}*

Mahner S, Heitz F, Burges A, Reuss A, Kraemer B, Schmalfeldt B, et al. TRUST: trial of radical upfront surgical therapy in advanced ovarian cancer (ENGOT ov33/AGO-OVAR OP7). *Journal of Clinical Oncology (Annual Meeting of the American Society of Clinical Oncology, ASCO, 2017; United States 912 0139)* 2017;**35 (15 Supplement 1)**(15 Suppl 1):(no pagination).

Mahner S, Heitz F, Burges A, Reuss A, Kramer B, Schmalfeldt B, et al. Role of neoadjuvant chemotherapy in advanced ovarian cancer: TRUST-trial of radical upfront surgical therapy in advanced ovarian cancer (ENGOT ov33/AGO-OVAR OP7). In: Oncology Research and Treatment Conference (Jahrestagung der Deutschen, Osterreichischen und Schweizerischen); Annual Meeting of German, Austrian and Swiss Societies for Hematology and Medical Oncology 2017; Germany. Vol. 40 (Suppl 3). Berlin: Karger AG 0142 825, 2017.

Reuss A, Du Bois A, Harter P, Fotopoulou C, Sehouli J, Aletti G, et al. TRUST: Trial of Radical Upfront Surgical Therapy in advanced ovarian cancer (ENGOT ov33/AGO-OVAR OP7). *International Journal of Gynecological Cancer* 2019;**29**(8):1327-31.

NCT04257786 *{published data only}*

NCT04257786. Primary cyto-reductive surgery vs neo-adjuvant chemotherapy (NAC) in epithelial ovarian cancer. Clinicaltrials.gov.

NCT04515602 *{published data only}*

NCT04515602. Stratified evaluation of PDS and NACT-IDS in ovarian cancer. Clinicaltrials.gov.

SUNNY *{published data only}*

NCT02859038. Study of upfront surgery versus neoadjuvant chemotherapy in patients with advanced ovarian cancer (SUNNY). ClinicalTrials.gov.

Additional references

Allen 1995

Allen DG, Heintz AP, Touw FW. A meta-analysis of residual disease and survival in stage III and IV carcinoma of the ovary. *European Journal of Gynaecologic Oncology* 1995;**16**(5):349-56.

Bogani 2017

Bogani G, Matteucci L, Tamberi S, Arcangeli V, Ditto A, Maltese G, et al. The impact of number of cycles of neoadjuvant chemotherapy on survival of patients undergoing interval debulking surgery for stage IIIC–IV unresectable ovarian cancer: results from a multi-institutional study. *International Journal of Gynecologic Cancer* 2017;**27**(9):1856-62.

Bristow 2002

Bristow RE, Tomacruz RS, Armstrong DK, Trimble EL, Montz FJ. Survival effect of maximal cytoreductive surgery for advanced ovarian carcinoma during the platinum era: a meta-analysis. *Journal of Clinical Oncology* 2002;**20**(5):1248-59.

Bristow 2006

Bristow RE, Chi DS. Platinum-based neoadjuvant chemotherapy and interval surgical cytoreduction for advanced ovarian cancer: a meta-analysis. *Gynecologic Oncology* 2006;**103**(3):1070-6.

Bristow 2007

Bristow RE, Eisenhauer EL, Santillan A, Chi DS. Delaying the primary surgical effort for advanced ovarian cancer: a systematic review of neoadjuvant chemotherapy and interval cytoreduction. *Gynecologic Oncology* 2007;**104**:480-90.

Burghardt 1991

Burghardt E, Girardi F, Lahousen M, Tamussino K, Stettner H. Patterns of pelvic and paraaortic lymph node involvement

in ovarian cancer. *Gynecologic Oncology* 1991;**40**(2):103-6. [MEDLINE: 91184633]

Chi 2010

Chi D. An analysis of patients with bulky stage IIIC/IV ovarian, tubal and peritoneal carcinoma treated with primary debulking surgery (PDS) during the same period as the randomized EORTC-NCIC trial of PDS versus neoadjuvant chemotherapy. In: Gynecologic Oncology Conference: 41st Annual Meeting of the Society of Gynecologic Oncologists; 2010 Mar 14-17. San Francisco, CA, 2010.

Chi 2011

Chi D, Bristow RE, Armstrong DK, Karlan BY. Is the easier way ever the better way? *Journal of Clinical Oncology* 2011;**29**(31):4073-5.

Coleman 2018

Coleman RL, Enserro D, Spirtos N, Herzog TJ, Sabbatini P, Armstrong DK, et al. A phase III randomized controlled trial of secondary surgical cytoreduction (SSC) followed by platinum-based combination chemotherapy (PBC), with or without bevacizumab (B) in platinum-sensitive, recurrent ovarian cancer (PSOC): a NRG Oncology/Gynecologic Oncology Group (GOG) study. *Journal of Clinical Oncology* 2018;**36**(Suppl):abstr 5501.

CRUK 2018

Cancer Research UK . Cancer Research UK ovarian cancer fact sheet. www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer (accessed 8 April 2018).

CTCAE 2017

CTCAE. Common terminology criteria for adverse events. ctep.cancer.gov/protocolDevelopment/electronic_applications/docs/CTCAE_v5_Quick_Reference_8.5x11.pdf) (accessed prior to 22 June 2021);**v5.0**.

Deeks 2001

Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, Altman DG, editors(s). Systematic Reviews in Health Care: Meta-Analysis in Context. 2nd edition. London: BMJ Publication Group, 2001.

DerSimonian 1986

DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986;**7**:177-88.

Dewdney 2010

Dewdney SB, Rimel BJ, Reinhart AJ, Kizer NT, Brooks RA, Massad LS, et al. The role of neoadjuvant chemotherapy in the management of patients with advanced stage ovarian cancer: survey results from members of the Society of Gynecologic Oncologists. *Gynecologic Oncology* 2010;**119**(1):18-21.

Du Bois 2009

Du Bois A, Reuss A, Pujade-Lauraine E, Harter P, Ray-Coquard I, Pfisterer J. Role of surgical outcome as prognostic factor in advanced epithelial ovarian cancer: a combined exploratory analysis of 3 prospectively randomized Phase 3 multicenter trials: by the Arbeitsgemeinschaft Gynaekologische Onkologie Studiengruppe Ovarialkarzinom (AGO-OVAR) and the Groupe d'Investigateurs Nationaux Pour les Etudes des Cancers de l'Ovaire(GINECO). *Cancer* 2009;**115**(6):1234-44.

Du Bois 2011

Du Bois A, Marth C, Pfisterer J, Harter P, Hilpert F, Zeimet AG, et al. Neoadjuvant chemotherapy cannot be regarded as adequate routine therapy strategy of advanced ovarian cancer. *International Journal of Gynecological Cancer* 2011;**21**(6):1165-8.

Du Bois 2017

Du Bois A, Vergote I, Ferron G, Reuss A, Meier W, Greggi S, et al. Randomized controlled phase III study evaluating the impact of secondary cytoreductive surgery in recurrent ovarian cancer: AGO DESKTOP III/ENGOT ov20. *Journal of Clinical Oncology* 2017;**35**(15 Suppl):5501.

Du Bois 2020

Du Bois A , Sehouli J , Vergote I , Ferron G , Reuss A, Meier W, et al. Randomized phase III study to evaluate the impact of secondary cytoreductive surgery in recurrent ovarian cancer: final analysis of AGO DESKTOP III/ENGOT-ov20. Journal of Clinical Oncology 2020;**38**(15 Supp):6000-6000.

Eisenhauer 2009

Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Sargent R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *European Journal of Cancer* 2009;**45**:228-47.

Eisenhauer 2019

Eisenhauer EL, Chi DS. Ovarian cancer surgery — heed this LION's roar. *New England Journal of Medicine* 2019;**380**(9):871-3. [DOI: 10.1056/NEJMe1900044]

EUROCARE 2015

Rossi S, Baili P, Capocaccia R, Caldora M, Carrani E, Minicozzi P, et al, EUROCARE-5 Working Group. The EUROCARE-5 study on cancer survival in Europe 1999–2007: database, quality checks and statistical analysis methods. *European Journal of Cancer* 2015;**51**:2104-19.

Fagotti 2006

Fagotti A, Ferrandina G, Fanfani F, Ercoli A, Lorusso D, Rossi M, et al. A laparoscopy-based score to predict surgical outcome in patients with advanced ovarian carcinoma: a pilot study. *Annal of Surgical Oncology* 2006;**13**(8):1156-61.

Fagotti 2013

Fagotti A, Vizzielli G, De Laco P, Surico D, Buda A, Mandato VD, et al. A multicentric trial (Olympia-MITO 13) on the accuracy of laparoscopy to assess peritoneal spread in ovarian cancer. *American Journal of Obstetrics & Gynecology* 2013;**209**(5):e1-462.e11.

Falconer 2020

Falconer H, Joneborg U, Krawiec K, Palsdottir K, Bottai M, Salehi S. Ultra-radical upfront surgery does not improve survival in women with advanced epithelial ovarian cancer: a natural

experiment in a complete population. *Gynecologic Oncology* 2020;**159**(1):58-65.

FIGO 2009

FIGO Committee in Gynecologic Oncology. Current FIGO staging for cancer of the vagina, fallopian tube, ovary, and gestational trophoblastic neoplasia. *International Journal of Gynecology and Obstetrics* 2009;**105**:3-4.

GLOBOCAN 2020

Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* 2021. [DOI: 10.3322/caac.21660]

GRADEpro GDT [Computer program]

McMaster University (developed by Evidence Prime, Inc) GRADEpro GDT (GRADEpro Guideline Development Tool (Software)). Version (accessed prior to 22 June 2021). Hamilton (ON): McMaster University (developed by Evidence Prime, Inc), 2020. Available from gradepro.org. [gradepro.org]

Greimel 2003

Greimel E, Bottomley A, Cull A, Waldenstromd AC, Arrarase J, Chauvenet L, et al, EORTC Quality of Life Group and the Quality of Life Unit. An international field study of the reliability and validity of a disease specific questionnaire module (the QLQ-OV28) in assessing the quality of life of patients with ovarian cancer. *European Journal of Cancer* 2003;**39**:1402-8.

Griffiths 1975

Griffiths CT. Surgical resection of tumor bulk in the primary treatment of ovarian carcinoma. *National Cancer Institute Monograph* 1975;**42**:101-4. [MEDLINE: 77056303]

Hall 2019

Hall M, Savvatis K, Nixon K, Kyrgiou M, Hariharan K, Padwick M, et al. Maximal-effort cytoreductive surgery for ovarian cancer patients with a high tumor burden: variations in practice and impact on outcome. *Annals of Surgical Oncology* 2019;**26**(9):2943-51. [DOI: 10.1245/s10434-019-07516-3]

Harter 2019

Harter P, Sehouli J, Lorusso D, Reuss A, Vergote I, Marth C, et al. A randomized trial of lymphadenectomy in patients with advanced ovarian neoplasms. *New England Journal of Medicine* 2019;**380**(9):822-32. [DOI: 10.1056/NEJMoa1808424]

Havrilesky 2019

Havrilesky LJ, Yang J-C, Lee PS, Secord AA, Ehrisman JA, Davidson B, et al. Patient preferences for attributes of primary surgical debulking versus neoadjuvant chemotherapy for treatment of newly diagnosed ovarian cancer. *Cancer* 2019;**125**(24):4399-406.

Higgins 2003

Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**:557-60.

Higgins 2011

Higgins JP, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [March 2011]. The Cochrane Collaboration, 2011. Available from training.cochrane.org/handbook/archive/v5.1.

Hoskins 1992

Hoskins WJ, Bundy BN, Thigpen JT, Omura GA. The influence of cytoreductive surgery on recurrence-free interval and survival in small-volume stage III epithelial ovarian cancer: a Gynecologic Oncology Group study. *Gynecologic Oncology* 1992;**47**(2):159-66. [MEDLINE: 93106491]

Hunter 1992

Hunter RW, Alexander ND, Soutter WP. Meta-analysis of surgery in advanced ovarian carcinoma: is maximum cytoreductive surgery an independent determinant of prognosis? *American Journal of Obstetrics and Gynecology* 1992;**166**(2):504-11. [MEDLINE: 92160883]

Kang 2009

Kang S, Nam BH. Does neoadjuvant chemotherapy increase optimal cytoreduction rate in advanced ovarian cancer? Meta-analysis of 21 studies. *Annals of Surgical Oncology* 2009;**16**(8):2315-20.

Kehoe 1994

Kehoe S, Powell J, Wilson S, Woodman C. The influence of the operating surgeon's specialisation on patient survival in ovarian carcinoma. *British Journal of Cancer* 1994;**70**(5):1014-7.

Kumari 2020

Kumari A, Thakur M, Saha SC, Suri V, Prasad GRV, Patel FD, et al. To compare the optimal cytoreduction rate in advanced epithelial ovarian cancer stage III/IV after 3 versus 6 cycles of neoadjuvant chemotherapy. *Journal of Obstetrics and Gynaecology* 2020 [Epub ahead of print]:1-5. [DOI: 10.1080/01443615.2020.1787967]

Langendam 2013

Langendam MW, Akl EA, Dahm P, Glasziou, P, Guyatt G, Schünemann HJ. Assessing and presenting summaries of evidence in Cochrane Reviews. *Systematic Reviews* 2013;**2**:81-90.

Lawrie 2015

Lawrie TA, Winter-Roach BA, Heus P, Kitchener HC. Adjuvant (post-surgery) chemotherapy for early stage epithelial ovarian cancer. *Cochrane Database of Systematic Reviews* 2015, Issue 12. Art. No: CD004706. [DOI: 10.1002/14651858.CD004706.pub5]

Meader 2014

Meader N, King K, Llewellyn A, Norman G, Brown J, Rodgers M, et al. A checklist designed to aid consistency and reproducibility of GRADE assessments: development and pilot validation. *Systematic Reviews* 2014;**3**:82.

Melis 2016

Melis MH, Elagwany AMS. WITHDRAWN: Adjuvant chemotherapy followed by interval debulking surgery versus upfront surgery followed by chemotherapy in advanced epithelial ovarian

carcinoma. *Hematology/Oncology and Stem Cell Therapy* 2016 [Epub ahead of print]:(study subsequently withdrawn from publication).

Mueller 2016

Mueller JJ, Zhou QC, Iasonos A, O'Cearbhaill RE, Alvi FA, El Haraki A, et al. Neoadjuvant chemotherapy and primary debulking surgery utilization for advanced-stage ovarian cancer at a comprehensive cancer center. *Gynecologic Oncology* 2016;**140**(3):436-42. [DOI: 10.1016/j.ygyno.2016.01.008] [PMID: 26777991]

Onda 2010

Onda T, Yoshikawa H, Yasugi T, Matsumoto K, Taketani Y. The optimal debulking after neoadjuvant chemotherapy in ovarian cancer: proposal based on interval look during upfront surgery setting treatment. *Japanese Journal of Clinical Oncology* 2010;**40**(1):36-41.

Onda 2011

Onda T, Yoshikawa H. Neoadjuvant chemotherapy for advanced ovarian cancer: overview of outcomes and unanswered questions. *Expert Reviews in Anticancer Therapy* 2011;**11**(7):1053-67.

Osaba 1994

Osoba D, Zee B, Pater J, Warr D, Kaizer L, Latreille J. Psychometric properties and responsiveness of the EORTC QLQ-C30 in patients with breast, ovarian and lung cancer. *Quality of Life Research* 1994;**3**(5):353-64.

Parmar 1998

Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Statistics in Medicine* 1998;**17**(24):2815-34. [MEDLINE: 99120172]

Petru 1991

Petru E, Pickel H, Tamussino K, Lahousen M, Heydarfadai M, Posawetz W, et al. Pretherapeutic scalene lymph node biopsy in ovarian cancer. *Gynecologic Oncology* 1991;**43**(3):262-4. [MEDLINE: 92090796]

RevMan 2014 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration Review Manager (RevMan). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Rose 2004

Rose PG, Nerenstone S, Brady MF, Clarke-Pearson D, Olt G, Rubin SC, et al, Gynecologic Oncology Group. Secondary surgical cytoreduction for advanced ovarian carcinoma. *New England Journal of Medicine* 2004;**351**:2489-97.

Schünemann 2020

Schünemann HJ, Higgins JP, Vist GE, Glasziou P, Akl EA, Skoetz N, et al. Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.1 (updated September 2020). Cochrane, 2020. Available from training.cochrane.org/handbook/current/chapter-14. [www.cochrane-handbook.org]

Siegel 2018

Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA: A Cancer Journal for Clinicians* 2018;**68**:7–30.

Swart 2009

Swart PE. Contemporary considerations for neoadjuvant chemotherapy in primary ovarian cancer. *Current Oncology Reports* 2009;**11**:457-65.

Tangjitgamol 2010

Tangjitgamol S, Manusirivithaya S, Laopaiboon M, Lumbiganon P, Bryant A. Interval debulking surgery for advanced epithelial ovarian cancer. *Cochrane Database of Systematic Reviews* 2016, Issue 1. Art. No: CD006014. [DOI: 10.1002/14651858.CD006014.pub7]

Thigpen 2011

Thigpen T, DuBois A, McAlpine J, DiSaia P, Fujiwara K, Hoskins W, et al. First-line therapy in ovarian cancer trials. *International Journal of Gynecological Cancer* 2011;**21**(4):756-62.

Tierney 2007

Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 2007;**8**:16. [DOI: 10.1186/1745-6215-8-16]

Vergote 2011a

Vergote I, Tropé CG, Amant F, Ehlen T, Reed NS, Casado A. Neoadjuvant chemotherapy is the better treatment option in some patients with stage IIIc to IV ovarian cancer. *Journal of Clinical Oncology* 2011;**29**(31):4076-8.

Vergote 2011b

Vergote I, Amant F, Kristensen G, Ehlen T, Reed NS, Casado A. Primary surgery or neoadjuvant chemotherapy followed by interval debulking surgery in advanced ovarian cancer. *European Journal of Cancer* 2011;**47**(Suppl 3):S88-91.

Vergote 2016

Vergote I, Van Nieuwenhuysen E, Vanderstichele A. How to select neoadjuvant chemotherapy or primary debulking surgery in patients with stage IIIC or IV ovarian carcinoma. *Journal of Clinical Oncology* 2016;**34**(32):3827-8.

Vizzielli 2014

Vizzielli G, Costantini B, Tortorella L, Petrillo M, Fanfani F, Chiantera V, et al. Influence of intraperitoneal dissemination assessed by laparoscopy on prognosis of advanced ovarian cancer: an exploratory analysis of a single-institution experience. *Annals of Surgical Oncology* 2014;**21**(12):3970. [10.1245/s10434-014-3783-6]

Wright 2014

Wright JD, Ananth CV, Tsui J, Glied SA, Burke WM, Lu YS, et al. Comparative effectiveness of upfront treatment strategies in elderly women with ovarian cancer. *Cancer* 2014;**120**(8):1246–54.

References to other published versions of this review

Coleridge 2019

Coleridge SL, Bryant A, Lyons TJ, Goodall RJ, Kehoe S, Morrison J. Chemotherapy versus surgery for initial treatment in advanced ovarian epithelial cancer.. *Cochrane Database of Systematic Reviews* 2019, Issue 2. Art. No: CD005343. [DOI: 10.1002/14651858.CD005343.pub5]

Coleridge 2021

Coleridge SL, Bryant A, Kehoe S, Morrison J. Chemotherapy versus surgery for initial treatment in advanced ovarian epithelial cancer. *Cochrane Database of Systematic Reviews* 2021, Issue 2. Art. No: CD005343. [DOI: 10.1002/14651858.CD005343.pub5]

Morrison 2005

Morrison J, Swanton A, Kehoe S. Chemotherapy versus surgery for initial treatment in advanced ovarian epithelial cancer. *Cochrane Database of Systematic Reviews* 2005, Issue 3. Art. No: CD005343. [DOI: 10.1002/14651858.CD005343]

Morrison 2007

Morrison J, Swanton A, Collins S, Kehoe S. Chemotherapy versus surgery for initial treatment in advanced ovarian epithelial cancer. *Cochrane Database of Systematic Reviews* 2007, Issue 4. Art. No: CD005343. [DOI: 10.1002/14651858.CD005343.pub2]

Morrison 2012

Morrison J, Haldar K, Kehoe S, Lawrie TA. Chemotherapy versus surgery for initial treatment in advanced ovarian epithelial cancer. *Cochrane Database of Systematic Reviews* 2012, Issue 8. Art. No: CD005343. [DOI: 10.1002/14651858.CD005343.pub3]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Chekman 2015

Study characteristics	
Methods	Randomised trial, conducted in Algeria between 1 June 2008 and 31 April 2014 Single-centre study; single surgeon operated on all women in both groups.
Participants	90 women with FIGO stage IIIc ovarian carcinoma enrolled and underwent surgery. 82 women randomised, 41 to PDS and 41 to IDS The diagnosis of stage IIIC ovarian carcinoma was confirmed by laparoscopy (78 cases) or laparotomy (3 cases) A thoraco-abdomino-pelvic scan and tumour markers CA-125 and CA-19.9
Interventions	Primary complete cytoreduction surgery followed by chemotherapy (G1) or NACT chemotherapy followed by debulking surgery then further chemotherapy (G2) Chemotherapy regimen used was carboplatin ([AUC] 5) + paclitaxel 175 mg/m^2, every 3 weeks 44% of women in the IDS arm had 6 cycles of chemotherapy prior to debulking surgery, 10% had 4 cycles and 15% had 3 cycles. In the PDS arm, 78% of women had 6 cycles of chemotherapy after their surgery.
Outcomes	Rate of debulking to residual disease to nodules <1 cm or complete resection, OS, recurrence-free survival (RFS), morbidity and rate of lumboaortic lymphadenectomy
Notes	The trial was in abstract form only but Professor Chekman kindly provided us with the following information on request: The mean operating time was 254.2 min with (range 69 min to 480 min) PDS (G1); mean operating time 273 min; (range 144 min to 480 min) IDS (G2); mean operating time 233 min; (range 69 min to 360 min)

Chekman 2015 *(Continued)*

Average blood loss:

24 women (29%) were transfused; 13 women (16%) were transfused 1 unit; 9 women (11%) were transfused 2 units; 2 women (2.4%), were transfused 3 units

PDS group: 15 women underwent blood transfusion (18%) versus IDS (G2): 9 women underwent blood transfusion (11%).

There were no postoperative deaths (0 to 30 days)

1 death recorded after the second cycle of NACT

They performed 8 re-operations (9.8%) mainly for abdominal and vascular complications: PDS group (G1) six (7.3%); and IDS group (G2) two (2.4%)

Macroscopic resection was achieved in 30 women: 16 in PDS group (G1); and 14 in IDS group *(G2).*

There were 36 recurrences:

20 women in the PDS group (G1); and 16 women in the IDS group (G2)

Another frequently recurring recurrence was abdominal-pelvic lymph node recurrence with 19.4% of women with evidence of abdomino-pelvic nodal relapse in the total population. This was similar in both groups. The other recurrences were localised, in order of frequency, in the hepatic (n = 6), pulmonary (n = 2), cerebral (n = 1) and inguinal (n = 2) levels (it should be noted that one or more sites may be affected by tumour recurrence).

Isolated biological recurrences (increase in CA-125 without associated radiological evidence) were not recorded.

In this trial, 22% of women had recurred before the first year, 38% between the first and second year, 25% between the second and third year and 13.8% beyond the third year. Thus, most recurrences (86%) were recorded during the first three years and 15% after the third year (time of occurrence of recurrence (P = 0.49)).

There were 24 deaths:

15 in the PDS group (G1); and 9 in the IDS group (G2)

Of the 12 remaining women who had a recurrence and remained alive, 5 were in the PDS group (G1) and 7 were in the IDS group (G2).

The mean PFS was 13.15 months (95% CI 9.19 to 17.10).

In the PDS group (G1), mean PFS was 27.92 months [range 7 to 64] and in the IDS group (G2) mean PFS was 24.72 months [range 11 to 52].

Surgical management of recurrence occurred in 19.4% of cases.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	The randomisation was performed in the operating room by random draw by someone other than the surgeon, once verification of inclusion criteria and resectability under laparoscopy or laparotomy had been confirmed. Histological confirmation of carcinomatosis of ovarian origin was by extemporaneous examination.
Allocation concealment (selection bias)	Unclear risk	Information lacking about the concealment process

Chekman 2015 *(Continued)*

Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded study and so some outcomes at high risk of bias
Blinding of outcome assessment (detection bias) All outcomes	High risk	Open-label study and no details of independent blinded assessment
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Minimal data provided regarding outcomes; only percentages provided for OS and PFS, no raw numbers, no confidence intervals or statistical calculations provided. Morbidity rate provided but unclear as to what specific morbidities this rate referred to
Selective reporting (reporting bias)	Unclear risk	No information regarding why lumboaortic lymphadenectomy chosen as an outcome. No information regarding what constituted morbidity data
Other bias	Unclear risk	Insufficient information to permit judgement

Fagotti 2016

Study characteristics	
Methods	Single institution (Italy) randomised phase III clinical trial, superiority trial (SCORPION) enrolled 280 women
Participants	Women aged 18 to 75 years with FIGO stage IIIc or IV ovarian, fallopian tube, or primary peritoneal cancer and histological confirmation of diagnosis. Histological sample obtained through staging laparoscopy and high tumour load calculated through laparoscopic predictive index (PI). PI between 8 and 12 without evidence of mesenteric retraction became inclusion criteria to go onto randomisation into the trial arms (110 randomised initially and presented in 2016 reference; additional 61 patients in 2020 update with OS data; total 171 participants).
Interventions	PDS + systemic adjuvant chemotherapy (arm A, standard) or to NACT + ITS (arm B, experimental)
Outcomes	Co-primary outcome measures were PFS and perioperative outcomes (early and late postoperative complications). Secondary outcomes were OS and QoL. 171 patients were randomly assigned to primary debulking surgery (PDS) (n = 84) versus neoadjuvant chemotherapy (NACT) (n = 87). **Mean age (SD);** PDS = 54.8 (9.7); NACT = 56.2 (10.7) **ECOG performance status:** PS = 0: PDS = 40 (47.6%); NACT = 39 (44.8%) PS = 1: PDS = 35 (41.7%); NACT = 41 (47.1%) PS = 2: PDS = 9 (10.7%); NACT = 7 (8%) **FIGO Stage:** Stage IIIc: PDS = 71 = (84.5%); NACT = 79 (90.8%) Stage IV: PDS = 13 (15.5%); NACT = 8 (9.2%) **Median follow-up:** 59 months (95% CI 53 to 64 months)

Fagotti 2016 *(Continued)*

Median overall survival:

PDS = 41 months for patients; NACT = 43 months (HR 1.12, 95% CI 0.76 to 1.65; P = 0.56)

Median PFS:

PDS = 15 months; NACT = 14 months (HR 1.05, 95% CI 0.77 to 1.44; P = 0.73)

Median number of chemotherapy cycles = 6 in both groups; range 0 to 6 cycles in PDS arm and 3 to 6 in NACT arm

Women in the NACT arm received a median number of four cycles prior to IDS.

3 women in the PDS arm progressed and did not receive chemotherapy. Chemotherapy schedule was as follows:

- 3-weekly carboplatin-paclitaxel: 31 (60.8%) PDS arm versus 29 (55.8%) NACT arm (P = 0.691);

- 3-weekly carboplatin-paclitaxel-bevacizumab: 14 (27.4%) PDS arm versus 20 (38.5%) NACT arm (P = 0.296);

- weekly carboplatin-paclitaxel: 5 (9.8%) PDS arm versus 3 (5.7%) NACT arm (P = 0.444);

- weekly carboplatin: 1 (1.9%) PDS arm versus 0 (0%) NACT arm (P = 0.310).

Median duration of treatment (randomisation to completion): 38 weeks for PDS (range 17 to 45 weeks) and 28 weeks for NACT arm (range 16 to 34 weeks). This was largely due to increased time to start/restart chemotherapy after surgery: median time after PDS was 40 days (range 17 to 120 days) versus 27 days after IDS (range 16 to 37 days) (P = 0.001).

Operative time (mins), mean (SD): PDS = 460.6 mins (102.6); NACT = 253.2 mins (101.4); P < 0.0001

Surgical complexity score (SCS): (P < 0.0001)

SCS 1: PDS = 0 (0%); NACT = 43 (58.1%)

SCS 2: PDS = 9 (10.7%); NACT = 20 (27.0%)

SCS 3: PDS = 75 (89.3%); NACT = 11 (14.9%)

Size of residual disease (P = 0.001)

No macroscopic disease: PDS = 40 (47.6%); NACT = 57 (77.0%)

0.1-1 cm: PDS = 38 (45.2%); NACT = 16 (21.6%)

> 1 cm: PDS = 6 (7.1%); NACT = 1 (1.4%)

Patients with postoperative major complications (G3+ SAEs)

Early (≤ 30 days): PDS = 39 (46.4%); NACT = 7 (9.5%); P < 0.0001

Late (1-6 months): PDS = (11.9%); NACT = 1 (1.4%); P = 0.009

Notes	Trial registered on ClinicalTrials.gov (No. NCT01461850) We are very grateful to Professor Fagotti for providing additional information for this study. We understand that further information will be published.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	A centrally performed, computer-generated list for block randomisation (1:1 ratio) was used. Women randomly (max allowable percentage deviation =

Fagotti 2016 *(Continued)*

		10%) allocated to PDS + systemic adjuvant chemotherapy (arm A, standard) or to NACT + IDS (arm B, experimental)
Allocation concealment (selection bias)	Unclear risk	Randomisation was done centrally by an independent DMC (CUSH-CTC), however there was no mention of whether the sequence was protected prior to assignment.
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible to blind participants or personnel to interventions in the trial. It was unclear what impact this would have in terms of bias, although it did carry a high risk.
Blinding of outcome assessment (detection bias) All outcomes	High risk	Open-label study and no indication of independent blinded assessment
Incomplete outcome data (attrition bias) All outcomes	High risk	Substantial missing data for QoL outcomes. Unable to provide chemotherapy SAE data due to missing data. Postop SAEs more fully presented for initial 110 cohort. Additional unpublished data provided by author
Selective reporting (reporting bias)	Unclear risk	Data for OS and PFS for entire cohort provided in subsequent publication and unpublished data from author for entire cohort. SASs during chemotherapy not reported and only partial QoL outcomes reported due to missing data (see above)
Other bias	Unclear risk	The authors stated that the types of surgery performed on women in each arm of the study were significantly different. In women in the PDS arm, upper abdominal surgical procedures were performed in all women compared to 42.3% of women in the IDS arm. This is likely due to the beneficial effect of chemotherapy reducing the volume of disease but as the study was not blinded, there is potential for high risk of bias. Median duration of entire treatment from randomisation to completion of medical treatment was also longer in the PDS arm (38 weeks versus 28 weeks). This was due to a statistically significant difference in time to start post-surgery chemotherapy (median time post-PDS 40 days, median time post-IDS 27 days). This was likely due to the greater extent of surgery required for those with higher volume disease in the PDS group, but due to lack of blinding risk of bias was unclear. No conflict of interest declared

Kehoe 2015

Study characteristics	
Methods	Multicentre international RCT non-inferiority trial (CHORUS)
Participants	552 women with stage IIIc/IV EOC enrolled in the UK and New Zealand
Interventions	Primary surgery then 6 cycles of platinum-based chemotherapy or 3 cycles of platinum-based chemotherapy, surgery, then a further 3 cycles of platinum-based chemotherapy
Outcomes	OS, PFS, QoL Median follow-up of surviving women = 4.4 years (IQR 3.5–6.1) Surgery scheduled after 3 cycles of chemotherapy in NACT group

Kehoe 2015 *(Continued)*

Chemotherapy details:

Single-agent carboplatin: NACT = 63 (23%); PDS = 66 (24%);

Carboplatin paclitaxel: NACT = 210 (77%); PDS = 207 (75%);

Carboplatin plus other chemotherapy agent: NACT = 1 (< 1%); PDS = 3 (1%).

Dose modification required: NACT = 100 (39%); PDS = 87 (38%)

PDS group: 251 (91%) of 276 women started treatment as allocated; 212 (77%) had adjuvant chemotherapy.

- 15 had primary chemotherapy:
  - 11 unfit for surgery;
  - 3 clinician's choice;
  - 1 because of women's choice.
- Of the 15 who had primary chemotherapy:
  - 4 had surgery after chemotherapy (2 after four cycles);
    - 3 had more chemotherapy after surgery (2 had two cycles);
    - 1 did not have more chemotherapy after surgery.
- 11 did not have surgery after chemotherapy (7 had six cycles):
  - 5 unfit;
  - 3 disease progression;
  - 2 had a complete response to chemotherapy;
  - 1 through woman's choice.
- 10 did not have surgery or chemotherapy:
  - 3 died before treatment;
  - 3 unfit;
  - 2 withdrew from trial;
  - 1 disease progression;
  - 1 no malignancy.
- 10 did not have surgery or chemotherapy:
  - 3 died before treatment;
  - 3 unfit;
  - 2 withdrew from trial;
  - 1 disease progression;
  - 1 no malignancy.

NACT group: 253 (92%) of 274 women started treatment as allocated and 217 (79%) had IDS.

Median duration of treatment was 22 weeks in both groups (NACT interquartile range (IQR) 19 to 24 weeks; PDS IQR 17 to 24 weeks).

- 2 had primary surgery:
  - 1 unfit for primary chemotherapy, but then had six cycles after surgery;
  - 1 had benign disease.
- 19 did not have chemotherapy or surgery:
  - 6 ineligible malignancy;
  - 5 died before treatment;
  - 3 no malignancy;
  - 2 deemed inoperable;
  - 3 withdrew from the trial.

Kehoe 2015 *(Continued)*

- 16 did not have more chemotherapy after surgery:
  - 6 died;
  - 3 did not have ovarian cancer;
  - 3 had surgery after the full six cycles of chemotherapy;
  - 3 because of women's choice;
  - 1 progressive disease.

Notes	www.ctu.mrc.ac.uk/plugins/StudyDisplay/protocols/CHORUS protocol Version 2.0 - 05 June 2008.pdf Additional age and survival details: < 50 years: OS 22.8 months (18.5 to 34.4); PFS 13.2 months (9.9 to 17.1) 50 to 70 years: OS 24.1 (20.6 to 28.4); PFS 11.4 (10.5 to 12.5) > 70 years: OS 20.8 (14.7 to 25.8); PFS 10.4 (8.8 to 12.0) We are very grateful to Professor Kehoe and his team for providing additional information for this study.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random assignment centrally at the Medical Research Council Clinical Trials Unit by telephone using a minimisation method with a random element. Women stratified according to randomising centre, largest radiological tumour size, clinical FIGO stage, and prespecified chemotherapy regimen with equal probability of assignment to each treatment arm 2 women who had been randomised were subsequently excluded. One woman had been randomised by mistake as an administrative error and one woman was found not to have the capacity to consent and was therefore ineligible for the trial.
Allocation concealment (selection bias)	Low risk	Central randomisation by the Medical Research Council Clinical Trials Unit by telephone
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded, therefore high risk for some outcomes assessed by investigators involved with patient care (e.g. optimal debulking)
Blinding of outcome assessment (detection bias) All outcomes	High risk	No report of blinded central assessment
Incomplete outcome data (attrition bias) All outcomes	Low risk	All women accounted for and analysed by ITT analysis
Selective reporting (reporting bias)	Unclear risk	All pertinent outcomes appeared to have been reported in some capacity. Prespecified outcomes as per clinicaltrials.gov protocol for OS; PFS and QoL - see outcomes section in methods and clinical trials.gov website. Only global QoL outcomes reported at baseline, 6 months and 12 months
Other bias	Unclear risk	64 centres: surgery performed by specialist gynaecological oncologists; further 23 registered centres: only non-surgical management provided. Supplementary data in table 7 showed that hysterectomy/bilateral salpingo-oophorecto-

Kehoe 2015 *(Continued)*

	my (BSO) and omentectomy performed in varying proportions. Unclear what effect this might have on outcomes

Onda 2016

Study characteristics	
Methods	Randomised phase III non-inferiority study (JCOG0602) conducted in 34 institutions in Japan
Participants	301 women aged 20 to 75 years enrolled with stage III or IV ovarian, tubal and peritoneal cancers diagnosed by clinical findings, imaging studies (CT, MRI and CXR) and cytology of ascites, pleural effusions or tumour centesis
Interventions	PDS followed by 8 cycles of chemotherapy +/- additional IDS if not completely debulked prior to commencing chemotherapy compared to 4 cycles of NACT followed by IDS and a further 4 cycles of chemotherapy
Outcomes	Primary outcomes of OS and PFS Planned follow-up initially 5 years, extended to 6 years Secondary outcomes of adverse events, frequency and duration of surgery, amount of blood loss and frequency of blood, plasma and albumin transfusions, postoperative mortality within 30 days of surgery **Median age** (range): PDS = 62 (25-86); NACT = 63 (33-81) **Stage:** PDS: stage 3 = 257 (77%); stage 4 = 77 (23%); other = 2 (0.6%); NACT: stage 3 = 253 (76%); stage 4 = 81 (24%) **Performance status (PS):** PS 0-1: PDS = 130 (87.2%); NACT = 131 (86.2%) PS 2-3: PDS = 19 (12.8%); NACT = 21 (13.8%) **Median cycles of chemotherapy:** NACT = 8 (IQR 7 to 8); PDS = 8 (IQR 6 to 8) **Chemotherapy schedule:** Carboplatin (AUC6) and paclitaxel 175 mg/m^2 given 3-weekly for a total of 8 cycles with IDS scheduled after 4 cycles **Overall survival:** HR for death with NACT compared with PDS was 1.052 (90.8% CI, 0.835 to 1.326; P = 0.24 for non-inferiority calculated using the Cox proportional hazard model stratified by FIGO stage, PS and age) **Progression-free survival:** HR for progression with NACT compared with PDS was 0.96 (95% CI, 0.75 to 1.23 calculated by the Cox proportional hazard model stratified by the FIGO stage, PS and age) **Optimal debulking at first surgical effort (0 cm & < 1 cm):** PDS = 55/147 (37%); NACT = 107/130 (82%) **Postoperative G3+ events after initial surgical effort:** PDS = 15.0% (n = 22/147); NACT = 4.6% (n = 6/130)

Onda 2016 *(Continued)*

	Operation time: PDS = 341 min; NACT = 273 min; P < 0.001 **Postoperative any G3+ SAEs:** PDS = 15.6%; NACT = 4.6%; P = 0.003 **Chemotherapy-related non-haematological G3-4 SAEs:** First 4 cycles chemotherapy: PDS = 28/138 (20·3%); NACT = 27 (18·0%); P = 0.65 Second 4 cycles chemotherapy: PDS = 11 (8·8%); NACT = 15 (11·9%); P = 0·54 **Completion of treatment:** PDS = 99 (66.4%); NACT = 103 (67.8%); P = 0.90
Notes	49 women randomised to primary debulking arm underwent additional interval debulking surgery. We are very grateful to Professor Onda for providing additional data for this meta-analysis.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	The JCOG Data Centre randomly assigned treatment to each women via a minimisation method with equal probability of assignment to each treatment arm. Balancing factors were institution, stage (III versus IV), performance status (0 to 1 versus 2 to 3) and age (< 60 versus > 60).
Allocation concealment (selection bias)	Low risk	The JCOG Data Centre randomly assigned treatment to each women via a minimisation method with equal probability of assignment to each treatment arm.
Blinding of participants and personnel (performance bias) All outcomes	High risk	Women and treating physicians were not masked to assigned treatment.
Blinding of outcome assessment (detection bias) All outcomes	High risk	Individuals assessing outcomes and analysing data were not masked to assigned treatment.
Incomplete outcome data (attrition bias) All outcomes	Low risk	OS and PFS analysed using appropriate statistical methods. All women accounted for and similar numbers completed treatment in each arm
Selective reporting (reporting bias)	Low risk	Study recognised that QoL may contribute to measures of treatment invasiveness, but scope was on survival outcomes. Study protocol published alongside paper as supplementary information
Other bias	High risk	Fourteen women (one in PDS and 13 in NACT) underwent some type of surgery (off-protocol treatment). These off-protocol surgeries were not included as PDS or IDS in the analysis. Appeared to be significantly more in NACT group No ITT analysis carried out

Vergote 2010

Study characteristics

Methods	EORTC-GCG 55971 Multicentre non-inferiority RCT; 59 institutions in Belgium, Canada, the UK, Sweden, the Netherlands, Italy, Norway, Spain, Austria, Portugal, Ireland and Argentina

Vergote 2010 *(Continued)*

	Recruitment period: 1998 to 2006 Median follow-up: 56.4 months
Participants	718 women enrolled, 48 excluded post-randomisation owing to authorisation irregularities at the Argentinian centre leaving 670 women Inclusion criteria: evidence of stage IIIc/IV EOC, primary peritoneal cancer or fallopian tube cancer by intraperitoneal biopsy or FNA plus presence of extra-pelvic tumour of at least 2 cm (excluding ovaries) on laparoscopy or CT scan; WHO performance status of 0 to 2; no other serious disabling diseases contraindicating PDS or NACT; no prior primary malignancies; no brain metastases; adequate haematological, renal and hepatic function; absence of other factors that could affect compliance; CA-125:CEA ratio higher than 25 Treatment had to start within 3 weeks of initial biopsy/FNA.
Interventions	Experimental: NACT (334 women) - 3 cycles of platinum-based NACT, followed by IDS within 6 weeks of third cycle, then at least 3 more cycles of NACT Control: PDS (336 women) plus at least 6 cycles of platinum-based chemotherapy ± IDS All surgery was performed by gynaecological oncologists.

Outcomes	OS, PFS, QoL (QLQ-C30 and QLQ-Ov28), surgical morbidity and mortality, toxicity, optimal debulking
	
	Median follow-up = 4.7 years
	
	Chemotherapy details:
	
	Platinum-taxane: NACT = 283(87.9%); PDS = 243 (78.4%)
	
	Platinum only: NACT = 20 (6.2%); PDS = 25 (8.1%)
	
	Other: NACT = 19 (5.9%); PDS 21 (6.8%)
	
	No chemotherapy: NACT = 0 (0%); PDS = 21 (6.8%)
	
	Median time to re-start chemotherapy after surgery in days (range):
	
	NACT = 18 days (5 to 55) versus PDS 19 days (0 to 84)
	
	• 336 were assigned to PDS
	• 315 received assigned intervention
	• 21 did not receive assigned intervention
	  * 8 (38%) were withdrawn by physician
	  * 3 (14%) declined to participate
	  * 3 (14%) had different histologic diagnosis
	  * 1 (5%) died
	  * 2 (10%) had unresectable tumour
	  * 3 (14%) had logistic or administrative problem
	  * 1 (5%) had unknown reason
	• 315 (94%) underwent primary debulking
	  * 297 (88%) started chemotherapy
	  * 57 (17%) underwent interval debulking
	  * 11 (3%) underwent second-look procedure
	• 334 were assigned to NACT

Vergote 2010 *(Continued)*

- 326 received assigned intervention
  - 8 did not receive assigned intervention
  - 3 (38%) were withdrawn by physician
  - 2 (25%) declined to participate
  - 1 (13%) had different histologic diagnosis
  - 1 (13%) died
  - 1 (13%) had logistic or administrative problem
  - 2 (1%) underwent primary debulking
- 326 (98%) started NACT
- 295 (88%) underwent interval debulking
- 6 (2%) underwent second-look procedure

Notes

Baseline characteristics were similar: stage IIIc (75.7% versus 76.5%) or stage IV (22.9% versus 24.3%); mean age 63 years (NACT) versus 62 years (PDS); at least 6 cycles received by 276/322 (85.8%) of NACT group and 253/310 (81.6%) of PDS group.

The number of women with metastases > 5 cm at the time of surgery in the NACT group was half that of the PDS group (37.2% versus 74.5%) suggesting NACT-related tumour shrinkage. Optimal debulking (80.6% versus 41.6%) and complete debulking were achieved more often in NACT group, but this did not translate into improved survival, even though complete debulking was a prognostic indicator for OS.

Median OS was 30 versus 29 months (NACT versus PDS) and median PFS was 12 months for both groups.

Intervention effects on OS differed significantly between participating countries.

A per-protocol analysis of those who underwent surgery (322/334 in NACT arm and 310/336 in PDS arm) was performed. However, 295 women in the NACT group underwent IDS and 315 women underwent PDS. Data from the published supplementary data differed from those in Figure 2 of the published paper. These data were from the supplementary data, although we noted that the percentages are calculated from the 295 and 315 denominators of women who actually had NACT/IDS and PDS, respectively, rather than the per-protocol analysis, as the table suggested. After debulking surgery, 7 women assigned to NACT and 11 women assigned to PDS were subsequently found on final histology not to have EOC.

QoL data reported in separate publication (Greimel and et al. 2013 see additional reference under Vergote 2010 **)**

Only 404 women included in QoL analysis. QoL was limited to data from institutions with the best compliance. Over 50% baseline compliance rate and 35% at follow-up chosen as pragmatic cut-off

Women in the QoL study subset differed from the entire population.

Only institutions with good QoL compliance were included in the QoL substudy. The institutions with good QoL compliance differed from those studies excluded from the QoL analysis and compared to institutions with poor QoL compliance had:

- better OS (median 32.30 versus 23.29 months; P = 0.0006);
- PFS (median 12.35 versus 9.92 months; P = 0.0002);
- 39.9% optimal debulking surgery compared to 19.9% in excluded institutions (P = 0.0011);
- more women with biopsy-proven EOC (90.3% versus 79.3%; P = 0.0050);
- more women with larger tumours (P = 0.0034);
- laparoscopy used more frequently (40.3% versus 21.4%) and FNA cytology used less frequently (36.1% versus 56.0%) for biopsy in the selected centres (P = 0.0002);
- fewer women with unknown tumour grade (35.6% versus 48.5%; P = 0.0009);
- No differences were found in terms of age, WHO performance status and FIGO stage between institutions.

Vergote 2010 *(Continued)*

Quote: "No differences between the treatment arms in the QoL functioning or symptoms scales, except for pain and dyspnoea. At baseline women treated with PDS had significantly higher pain scores (P = 0.046; PDS mean 36.7; NACT mean 29.9) and significantly lower dyspnoea scores (P = 0.049; PDS mean 22.9; NACT mean 27.9) compared to women treated with NACT. However, the difference was below 10 points indicating no clinically relevant difference."

We are very grateful to Professor Vergote for providing additional information for this study.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Randomisation done centrally by computer-generated randomisation, but detail of methods lacking in published data. Minimisation used to stratify for institution, biopsy method, tumour stage and largest preoperative tumour size. QoL outcomes were based on a selected number of institutions selected for their QoL data compliance.
Allocation concealment (selection bias)	Unclear risk	Central allocation but detail of methods lacking and data from 48 women from Argentina were excluded after randomisation owing to "potential authorisation irregularities"
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded, therefore high risk for some outcomes assessed by investigators involved with patient care (e.g. optimal debulking)
Blinding of outcome assessment (detection bias) All outcomes	High risk	Unblinded study and no mention of central independent blinded assessment
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	3/336 versus 5/334 lost to follow-up but substantial proportion were missing for QoL outcome; overall outcomes were complete.
Selective reporting (reporting bias)	Unclear risk	All prespecified outcomes reported. Analysis by ITT and per-protocol However, QoL outcome was based on a selected number of institutions with better QoL compliance. While the trial authors offered justification for their approach, several differences were found when comparing the outcomes of the 404 selected women (of which, only 212 were assessed in QoL domains) to the overall populations of 670 women. Women from the selected institutions had significantly better OS and PFS when compared to women treated in institutions which were excluded because of poor compliance rates.
Other bias	Unclear risk	48 post-randomisation exclusions from the Argentinian centre owing to quote: "authorisation irregularities" were indicated erroneously as pre-randomisation exclusions on the study-flow diagram. The investigators stated that "The results of the study were similar whether the 48 patients....were included or excluded".

BSO: bilateral salpingo oophorectomy
CEA: carcinoembryonic antigen
CT: computer tomography
EOC: epithelial ovarian cancer
FIGO: Federation of International Gynaecologists and Obstetricians

FNA: fine needle aspiration; HR: hazard ratio; IDS: interval debulking surgery; ITT: intention to treat; IQR: interquartile range; MRI: magnetic resonance imaging; NACT: neoadjuvant chemotherapy; OS: overall survival; PDS: primary debulking surgery; PFS: progression-free survival; QoL: quality of life; RCT: randomised controlled trial; WHO: World Health Organization.

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Ansquer 2001	Retrospective study of 54 women with unresectable disease at primary laparotomy
Baekelandt 2003	Review article
Bertelsen 1990	RCT of chemotherapy (cisplatin versus cisplatin, cyclophosphamide, doxorubicin) no surgery randomisation
Bidzinski 2005	Retrospective study
Bristow 2001	Meta-analysis of the impact of optimal debulking. no surgical randomisation in any trial included
Chambers 1990	Retrospective case series of 17 women
Chan 2003	Prospective case control series of 17 women
Chan 2017	Wrong intervention, participants randomised to either weekly with 3-weekly paclitaxel. No surgical randomisation
Chi 2012	Wrong study design, retrospective review, no randomisation
Cole 2018	Wrong study design; economic analysis comparing treatment strategies but no randomisation
Colombo 2009	Not an RCT. Retrospective review of 203 women with stage IIIc/IV EOC; 142 received PDS and 61 received NACT. Overall median survival was 35 months. Concludes that PDS is management of choice. NACT is indicated in non-operable tumours or in women with poor performance status
Cowan 2017	Editorial article, not an RCT
Da Costa 2014	Wrong study design, retrospective cohort.
Dai-yuan 2013	Wrong study design, meta-analysis
Daniele 2017	Wrong Intervention. Evalution of adding Bevacizumab to NACT prior to IDS. Not an RCT
Deval 2003	RCT of different chemotherapy regimens. No surgical randomisation. 102 women with stage IV ovarian cancer. 53% primary surgery, 15% secondary surgery, 32% no surgery. No significant differences in survival
Dutta 2005	RCT, but comparing surgery after 3 or 6 cycles of chemotherapy, with no up-front surgery arm. Small study (24 women). No details of how women were randomised. No assessment of survival outcomes
ESGO 2013	Wrong study design, conference proceedings. No studies identified that had not already been found.
Evdokimova 1982	RCT of NACT then surgery versus surgery then chemotherapy. Chemotherapy - alternating cycles of cyclophosphamide/5-fluorouracil and cyclophosphamide hexamethylmelamine, therefore non-platinum based. Survival advantage for up-front surgery

Study	Reason for exclusion
Everett 2006	Not an RCT. Retrospective study in which 200 women with advanced ovarian cancer received NACT (98 women) or PDS (102 women). Optimal cytoreduction achieved more frequently in the NACT group. Optimal cytoreduction was associated with better survival
Fagö-Olsen 2014	Wrong study design, prospective cohort
Fagotti 2018	Commentary in response to per protocol joint analysis of Kehoe 2015 and Vergote 2010 studies
Fanfani 2003	Retrospective case-control series of 73 women with unresectable disease receiving NACT compared with 184 women with resectable disease undergoing conventional treatment
Feng 1998	Retrospective case series of 18 women with advanced ovarian cancer treated with NACT
Forde 2015	Wrong study design, cost analysis
Fujiwara 2013	Wrong study design, review article
Ghaemmaghami 2008	Not an RCT. Retrospective study of 92 women with advanced ovarian cancer. Compared 24 women with unresectable disease and NACT/IDS with 68 women with PDS and chemotherapy. PDS was associated with longer survival. Extent of residual tumour associated with poorer prognosis
Giannopoulos 2006	Not an RCT. Prospective cohort study of 64 women with stage IIIc/IV ovarian cancer. 35 women were considered unresectable and received NACT with IDS and 29 received PDS. Concluded that there was less morbidity in the IDS group. Optimal cytoreduction higher in NACT group (NS)
Grosso 2013	Wrong intervention, no randomisation
Hanker 2010	Not an RCT. Exploratory meta-analysis on the impact of surgical debulking, using individual patient data from 3 RCTs that investigated platinum/taxane-based regimens after primary surgery for advanced ovarian cancer. Concluded that the goal of 'optimal debulking' in PDS should be complete resection
Hegazy 2005	Not an RCT. Prospective study of 59 women with advanced ovarian cancer who received NACT if optimal cytoreduction was not feasible (27 women) or PDS (32 women) if it was feasible
Hou 2007	Not an RCT. Retrospective study of 172 women with advanced ovarian cancer: 109 received PDS and 63 received NACT. NACT was associated with less perioperative morbidity, more 'optimal cytoreduction' and less need for further aggressive surgery
Inciura 2006	Not an RCT. Retrospective study of 574 women; 213 received NACT and 361 received PDS. No significant differences in survival rates or 'optimal cytoreduction' rates
Iranian Society Reproductive Medicine Conference	Wrong study design, conference proceedings no RCTs identified
Jacob 1991	Retrospective case-control series
Kayikcioglu 2000	Retrospective series of 189 women. No randomisation
Kayikcioglu 2001	Retrospective series of 205 women. No randomisation
Kehoe 2011	Wrong study design, recruitment to CHORUS trial poster
Kuhn 2001	Prospective NRS of 31 women treated with NACT vs 32 women with conventional treatment
Kumar 2015	Wrong study design, review article.

Study	Reason for exclusion
Lawton 1989	Prospective case series of 23 women with suboptimally debulked disease at primary surgery
Lee 2006	Not an RCT. Prospective study of 40 women with advanced EOC. Compared 18 women who received NACT with 22 who received PDS. No significant survival differences between groups
Lee 2018	Wrong study design - non RCT - experience from a single cancer centre
Lim 1993	Non-randomised prospective case series of 30 women with untreated FIGO stage III and IV ovarian carcinoma given carboplatin (400 mg/m^2) and ifosfamide (5 g/m^2) with mesna. No surgical randomisation
Liu 1995	Retrospective case series
Liu 2004	Randomised 85 women with advanced ovarian cancer to NACT plus ovarian artery embolisation or PDS. 42 women received 1 cycle of neoadjuvant platinum-based chemotherapy (cisplatin, doxorubicin and cyclophosphamide) directly into the ovarian artery, followed by ovarian artery embolisation. These women then had debulking surgery followed by 7 cycles of intravenous platinum-based chemotherapy. The 43 women in the control arm underwent debulking surgery and then received 8 cycles of intravenous platinum-based chemotherapy. The results may have been attributable to the chemotherapy, embolisation or the combination
Liu 2015	Wrong study design, retrospective cohort study
Liu 2017	Trial comparing intra-peritoneal chemotherapy timing rather than timing of surgery in relation to chemotherapy administration.
Loizzi 2005	Retrospective case-control study of 30 women
Lotze 1987	RCT of intra-arterial chemotherapy, not surgery
Lyngstadaas 2005	Systematic review. No RCTs identified for NACT
Mackay 2011	Ongoing RCT of intravenous NACT versus intraperitoneal NACT (NCIC CTG OV.21 protocol)
Mahner 2006	Conference presentation of Polcher 2009
Mahner 2014	Review article
Makar 2016	Review article
Malzoni 1993	Case report
Mazzeo 2003	Retrospective case series of 45 women
Melamed 2018	Wrong study design: quasi-experimental fuzzy regression discontinuity design and cross-sectional analysis.
Morice 2003	Retrospective study of 57 women with unresectable disease undergoing chemotherapy then surgery with 28 women with resectable disease following surgery then chemotherapy
Negretti 1988	Retrospective case series of 27 women
Nick 2015	Wrong study design, case series
Oe 2011	Not an RCT but methods not clear. More details requested from authors

Study	Reason for exclusion
Onda 2009	Not an RCT. A cohort of 56 women with advanced mullerian tumours underwent a diagnostic laparoscopy, NACT and IDS. The aim of the study was to determine whether diagnostic laparoscopy was necessary before NACT. Clinical diagnosis plus cytology/histology yielded a positive predictive value > 95% for advanced mullerian tumours. Concluded that diagnostic laparoscopy not necessary before giving NACT
Onnis 1996	Retrospective case series of 88 women with NACT then surgery
Polcher 2009	Phase II RCT comparing 2 NACT treatment schedules, namely 3/6 cycles (40 women) or 2/6 cycles (43 women) of carboplatin/docetaxel followed by optimal debulking surgery. Primary outcome was pre-operative reduction in ascites volume. Secondary outcomes were residual tumour, perioperative morbidity and mortality. Concluded that 2 NACT cycles is a reasonable option. Any residual disease associated with survival rates
Poonawalla 2015	Non RCT - cost-effectiveness study comparing NACT and PDS in elderly patients
Prescott 2016	Wrong study design: retrospective study on effect of blood transfusion in Vergote 2010 study
Qin 2018	Systematic review of RCTS and observational studies
Querleu 2013	Wrong study design, letter
Rafii 2007	Not an RCT. Retrospective study on the benefit of debulking surgery in Stage IV ovarian cancer using data from GINECO randomised studies of platinum/taxane regimens
Rauh-Hain 2017	Wrong study design; population level comparison of OS outcomes of NACT versus PDS
Recchia 2001	Prospective non-randomised Phase II study of primary chemotherapy in 34 women with stage IV ovarian cancer. No surgical randomisation
Redman 1994	RCT comparing IDS versus no further surgery in women suboptimally debulked at primary surgery
Robova 2003	Not an RCT. Treated 87 women with inoperable EOC with NACT. Conference abstract only
Rowland 2013	Wrong study design, cost analysis (abstract)
Rowland 2015	Wrong study design, cost analysis (paper)
Rutten 2012	Wrong intervention, randomisation to laparoscopy or not prior to PDS
Salzer 1990	Prospective non-randomised cohort study of different chemotherapy regimens and IDS
Sato 2014	Wrong study design, review
Sayyah-Melli 2013	Wrong study design, prospective cohort
Schorge 2014	Wrong study design, review
Schwartz 1994	Retrospective case-control study of 11 women treated with NACT followed by surgery
Schwartz 1999	Retrospective case-control study of 59 women treated with NACT followed by surgery. Included long-term follow-up of 28 women from 2 other studies (Schwartz 1994 and Chambers 1990)
Shibata 2003	Retrospective, NRS

Study	Reason for exclusion
Shimizu 1993	Retrospective case series of 138 women with ovarian cancer. 77 women had conventional treatment, 82 had exploratory laparotomy alone with 74 then receiving chemotherapy
Steed 2006	Not an RCT. Retrospective analysis of 116 women with advanced ovarian cancer who received NACT (50 women) or primary surgery (66 women)
Sun 2000	Retrospective study. 95 women managed by traditional surgery-chemotherapy (76 women) or chemotherapy-surgery-chemotherapy (17 women)
Surwit 1999	Retrospective case series of 39 women receiving NACT prior to surgery
Taskin 2013	Wrong study design, not randomised, retrospective cohort study.
Taylor 2015	Wrong study design, retrospective case series.
Tran 2018	Wrong study design: cost-effectiveness study comparing different treatment approaches
Trope 1997	RCT study of chemotherapy regimens. No randomisation arm for surgery
Ushijima 2002	Retrospective case-control study of 65 women with unresectable ovarian cancer treated with NACT and surgery
Van der Burg 1995	RCT of IDS following suboptimal primary surgery (319 women)
Van Meurs 2013	Wrong study design, biomarker analysis
Varma 1990	Abstract of the later full Trial by Redman 1994, comparing secondary debulking surgery or chemotherapy after all women had initially undergone primary debulking surgery
Vergote 1998	Retrospective longitudinal study of 285 women: 112 in first cohort all underwent surgery; of second cohort (173 women) 43% received primary chemotherapy and 57% received PDS
Vergote 2000	Retrospective analysis of 338 women, including longer-term follow-up of those in Vergote 1998 paper
Vergote 2018	Pooled analysis of individual patient data from the EORTC 55971(Vergote 2010) and Kehoe 2015 trials. Data already included in review.
Vergote 2019	Pooled analysis of individual patient data from the EORTC 55971(Vergote 2010) and Kehoe 2015 trials. Data already included in review.
Vrscaj 2002	Retrospective case-control study of 75 women with advanced ovarian cancer
Wenzel 2017	Wrong Intervention. RCT trialling a patient decision making tool around IV or IP chemotherapy versus standard care. No surgical randomisation.
Wright 2013	Wrong study design, retrospective study
Wu 2012	Wrong study design, retrospective study
Xiao 2018	Systematic review and meta-analysis
Yang 2017	Meta-analysis of perioperative outcomes
Zamagni 2014	Wrong study design, comparison of 3 versus 6 cycles of chemotherapy

Study	Reason for exclusion
Zeng 2016	Wrong study design, systematic review of surgery in primary treatment of ovarian cancer

EOC: epithelial ovarian cancer; FIGO: Federation of International Gynaecologists and Obstetricians; GINECO: Group d'Investigateurs Nationaux pour l'Etude des Cancers Ovariens; IDS: interval debulking surgery; NACT: neoadjuvant chemotherapy; NCIC CTG: NCIC Clinical Trial Group; NRS: non-randomised study; NS: not significant; PDS: primary debulking surgery; RCT: randomised controlled trial

Characteristics of studies awaiting classification *[ordered by study ID]*

Jiang 2018

Methods	To investigate the role and significance of neoadjuvant chemotherapy in advanced ovarian cancer.
Participants	128 patients clinically diagnosed with stage IIC-IV advanced epithelial ovarian cancer (EOC)
Interventions	Neoadjuvant chemotherapy (NACT) combined with interval cytoreductive surgery (ICS) group (n=66) and primary cytoreductive surgery (PCS) group (n=62). Chemotherapy in the PCS group was administered after cytoreductive surgery.
Outcomes	Progression-free survival (PFS) and overall survival (OS) Secondary outcomes include operative time, bleeding, optimal debulking surgery, rate of clinical remission. Longer operating time in PDS group (mean 275.94mins +/- 70.84) versus NACT (mean 215.65mins +/- 68.48) P < 0.05. Higher blood loss in PDS group (mean 794.94mls +/- 250.16) versus NACT (mean 467.84mls +/-220.14) P < 0.05. Lower optimal debulking rate in PDS group (38.7%) versus NACT (60.6%) P < 0.05. Mean follow up time 61.3 months. 28 deaths in NACT group (42.4%) and 32 deaths in PDS group (51.6%) not significantly different. Mean PFS NACT 18.5 months versus 17.9 PDS not significantly different. Mean OS NACT 47.5 months versus 46.3 months PDS not significantly different.
Notes	Study describes itself as a retrospective cross sectional study although women were 'randomised' into NACT or PCS groups. Author contacted for clarification of study design and further data.

EOC: epithelial ovarian cancer; ICS: interval cytoreductive surgery; IDS: interval debulking surgery; NACT: neoadjuvant chemotherapy; NCIC CTG: NCIC Clinical Trial Group; NRS: non-randomised study; NS: not significant; OS: overall survival; PCS: primary cytoreductive surgery; PDS: primary debulking surgery; PFS; Progression-free survival; RCT: randomised controlled trial

Characteristics of ongoing studies *[ordered by study ID]*

Kumar 2009

Study name	Kumar
Methods	RCT; open-label
Participants	180 women Included if: age 20 to 65 years; EOC stage IIIc & IV (pleural effusion only); ECOG PS 0-2; cytology/biopsy-positive women; good compliance; previously untreated women

Kumar 2009 *(Continued)*

	Excluded if: any medical contraindication to surgery; psychiatric illness; cardiac, liver or renal dysfunction
Interventions	Upfront surgery followed by 6 cycles of paclitaxel + carboplatin (chemotherapy) (arm A) or upfront chemotherapy - 3 cycles chemotherapy followed by surgery then 3 more cycles of chemotherapy
Outcomes	Optimal debulking rate (≤ 1 cm), OS, PFS, clinical CR, QoL, operating time, blood loss, stay in ICU, duration of hospital stay, infections, chemo-toxicity
Starting date	
Contact information	lalitaiims@yahoo.com
Notes	Clinical Trials Register: NCT00715286 Interim results presented at 2007 ASCO meeting: 113/139 women evaluable, 20% optimally debulked in PDS group versus 85% in the NACT group. NACT group also experienced less blood loss (P = 0.01), shorter hospital stay (P = 0.04), less postoperative infection (2 cases versus 7 cases; P = 0.06) and less operative mortality (1 deaths versus 5 deaths; P = 0.08). Median OS was 29 months in PDS group versus 41 months in NACT group. Interim results presented in Kumar 2009: 128/133 women evaluable, 62 in PDS group, 66 in NACT group. Optimum debulking was achieved in 22.6% and 86.2% (P < 0.0001), respectively. The NACT group experienced less blood loss (413 mL versus 600 mL; P < 0.0001), reduced postoperative infections (1.54% versus 14.5%; P < 0.025), reduced operating time (75.4 minutes versus 89.2 minutes; P < 0.001) and shorter hospital stay (7.6 days versus 11.5 days; P < 0.001). Median follow-up at 42 months found similar OS of 42 months and 41 months in the PDS and NACT group, respectively (the 2007 results presented showed significantly better OS in the NACT group). HR for OS (PDS versus NACT) was 0.94; 95% CI 0.56 to 1.56. HR for PFS (PDS versus NACT) was 1.1; 95% CI 0.71 to 1.86. QoL score was significantly better in the NACT group 'at the end of treatment' (P < 0.001) There are some discrepancies in these data when compared with the 2007 interim results (e.g. OS data). Furthermore, the denominators used to create these data were not stated in Kumar 2009, and continuous data were presented without standard deviations. The authors stated that complete results will be published soon.

Mahner 2017

Study name	Role of neoadjuvant chemotherapy in advanced ovarian cancer: TRUST-trial of radical upfront surgical therapy in advanced ovarian cancer (ENGOT ov33 / AGO-OVAR OP7)
Methods	Multi-centre international randomised controlled trial comparing primary debulking surgery (maximally debulked - complete gross resection) followed by 6 cycles of chemotherapy (control arm) with 3 cycles of neoadjuvant chemotherapy followed by interval debulking surgery (maximally debulked - complete gross resection) and another 3 cycles of chemotherapy (experimental arm). There are 3 parts to the trial the first 2 parts were conducted in Germany alone. The 3rd part is the multi-centre international trial including centres in the UK (1), USA (1), France (3), Germany (8), Italy (3), Denmark (1), Austria (1) and Sweden (2). All are actively recruiting at present except Austria. The trial aims to recruit 686 participants
Participants	Suspected or histologically-confirmed, newly diagnosed invasive epithelial ovarian cancer FIGO stage IIIB-IV (IV only if resectable metastasis) Females aged ≥ 18 years Women who have given their written informed consent Good performance status (ECOG 0/1) Good ASA score (1/2)

Mahner 2017 *(Continued)*

	Preoperative CA 125/CEA ratio ≥ 25 (if CA-125 is elevated)* If < 25 and/or biopsy with non-serous, non-endometrioid histology, esophago-gastro-duodenoscopy (EGD) and colonoscopy mandatory to exclude gastrointestinal primary cancer Assessment of an experienced surgeon, that is based on all available information, the women can undergo the procedure and the tumour can potentially be completely resected Adequate bone marrow function: Absolute neutrophil count (ANC) ≥ 1.5 x 109/L. This ANC cannot have been induced or supported by granulocyte colony stimulating factors. Platelet count ≥ 100 x 109/L. Renal function: Serum-Creatinine ≤ 1.5 x institutional upper limit normal (ULN). Hepatic function: Bilirubin ≤ 1.5 x ULN. SGOT ≤ 3 x ULN Alkaline phosphatase ≤ 2.5 x ULN. Neurologic function: Neuropathy (sensory and motor) less than or equal to CTCAE Grade 1
Interventions	Primary debulking surgery followed by 6 cycles of chemotherapy (control arm) or 3 cycles of neoadjuvant chemotherapy followed by interval debulking surgery and a further 3 cycles of chemotherapy (experimental arm)
Outcomes	Primary outcome measure is OS (Women will be followed up for a minimum of 5 years after registration/randomisation or until death) Secondary outcome measures are: Progression-free survival (PFS) (Women will be followed up for a minimum of 5 years after registration/randomisation or until death) Progression-free survival time is calculated from the date of randomisation until the date of first progressive disease or death, whichever occurs first or date of last contact (censored observation). Progressive disease is defined as clinical or imaging-detected tumour progression or death in cases without prior documented tumour progression. Progression-free survival 2 (PFS2) (Women will be followed up for a minimum of 5 years after registration/randomisation or until death) PFS2 time is calculated from the date of randomisation until the date of second progressive disease or death, whichever occurs first or date of last contact (censored observation). Time to first subsequent anticancer therapy or death (TFST) (Time Frame: Women will be followed up for a minimum of 5 years after registration/randomisation or until death) Time to first subsequent anticancer therapy is calculated from the date of randomisation until the starting date of the first subsequent anticancer therapy or death, whichever occurs first or date of last contact (censored observation). Maintenance treatments following a cytostatic treatment are not considered separate treatment lines. Time to second subsequent anticancer therapy or death (TSST) (Time frame: Women will be followed up for a minimum of 5 years after registration/randomisation or until death) Time to second subsequent anticancer therapy is calculated from the date of randomisation until the starting date of the second subsequent anticancer therapy or death, whichever occurs first or date of last contact (censored observation). Maintenance treatments following a cytostatic treatment are not considered separate treatment lines. QoL (Time frame: women will be followed up for a minimum of 5 years after registration/randomisation or until death) QoL as measured by EORTC QLQ-C30 (Version 3), EORTC QLQ-OV28, EQ-5D-3L

Mahner 2017 *(Continued)*

	Documentation of surgical complications (Time frame: women will be followed up for 1 year after surgery or until death) Assessment of safety: documentation of surgical complications 28 days after surgery and 1 year after surgery.
Starting date	Recruitment commenced in July 2016 and is expected to close in April 2023.
Contact information	office-wiesbaden@ago-ovar.de
Notes	

NCT04257786

Study name	NCT04257786
Methods	Randomised open label study
Participants	80 participants. Females aged 18 years to 80 years with advanced epithelial ovarian cancer. Stage 2D or more ; Performance status (PS) according to Eastern Cooperative Oncology Group (ECOG) ≤ 2 No contra-indication to bevacizumab.
Interventions	Primary surgery then chemotherapy versus Neoadjuvant chemotherapy (NACT) followed by surgery
Outcomes	Primary outcome: Percentage of patient where complete resection of the tumour can be achieved
Starting date	1/3/2020
Contact information	Ali Hussien Ali Sayed, Specialist, Assiut University, Egypt
Notes	

NCT04515602

Study name	FOCUS (NCT04515602)
Methods	Randomised phase III open label multicenter study
Participants	410 female participants with pathologically confirmed stage IIIC and IV epithelial ovarian cancer, fallopian tube cancer or primary peritoneal carcinoma; **Part 1**

NCT04515602 *(Continued)*

- Females aged ≥ 18 years and cPCI score ≤ 8;
- Performance status (ECOG 0-2);
- Good ASA (1/2);
- Adequate bone marrow, renal and hepatic function to receive chemotherapy and subsequent surgery.

Part 2

- Females aged ≥ 18 years, and < 70 years with cPCI score ≥ 10;
- For FIGO IVB patients, abdominal lesions should be confined to one lobe of liver parenchyma metastasis or splenic metastasis. All extra-abdominal metastases should be resectable, such as inguinal lymph nodes, solitary supraclavicular, retrocrural or paracardial nodes;
- Good performance status (ECOG 0-1);
- Good ASA score (1/2);
- Adequate bone marrow, renal and hepatic function to receive chemotherapy and subsequent surgery

Interventions	**Part 1, Arm I** **(low/medium tumour burden)** **PDS:** Primary debulking surgery with a maximum cytoreduction, then followed by 6 cycles of Paclitaxel 175mg/m^2 or Docetaxel 60-75 mg/m^2 plus Carboplatin AUC (area under the curve) 5. For patients with gBRCA/sBRCA mutation and CR/PR after first-line chemotherapy, maintenance therapy of PARP inhibitors. **Part 1 Arm II (low/medium tumour burden)** **NACT:** 3 cycles of Paclitaxel 175mg/m^2 or Docetaxel 60-75 mg/m^2 plus Carboplatin AUC (area under the curve) 5, Interval debulking surgery with a maximal cytoreduction of complete gross resection, then followed by another 3 cycles of chemotherapy. For patients with gBRCA/sBRCA mutation and CR/PR after first-line chemotherapy, maintenance therapy of PARP inhibitors. **Part 2 Arm I (high tumour burden)** **PDS:** Primary debulking surgery with a maximum cytoreduction, then followed by 6 cycles of Paclitaxel 175mg/m^2 or Docetaxel 60-75 mg/m^2 plus Carboplatin AUC (area under the curve) 5. For patients with gBRCA/sBRCA mutation and CR/PR after first-line chemotherapy, maintenance therapy of PARP inhibitors. **Part 2 Arm II (high tumour burden)** **NACT:** 3 cycles of Paclitaxel 175mg/m^2 or Docetaxel 60-75 mg/m^2 plus Carboplatin AUC (area under the curve) 5, Interval debulking surgery with a maximal cytoreduction of complete gross resection, then followed by another 3 cycles of chemotherapy. For patients with gBRCA/sBRCA mutation and CR/PR after first-line chemotherapy, maintenance therapy of PARP inhibitors.
Outcomes	**Primary:** • Overall survival

NCT04515602 *(Continued)*

	Secondary: • Progression-free survival; • Postoperative complications evaluated at 30-day, 60-day, 90-day after upfront cytoreductive surgery or interval debulking surgery; • Quality of life (Qol) as measured by QOQ-C30; • Quality of life (Qol) as measured by FACT-O; • The overall survival time minus the total treatment time of surgery and chemotherapy after randomisation, regardless of the targeted therapy; • Time to first subsequent anticancer therapy; • Time to secondary subsequent anticancer therapy; • Progression-free survival 2.
Starting date	estimated start date January 2021
Contact information	Lina Shen (shen.lina@zs-hospital.sh.cn); Tingyu Luan (luan.yuting@zs-hospital.sh.cn)
Notes	**Sponsors and Collaborators:** Shanghai Gynecologic Oncology Group, Obstetrics & Gynecology Hospital of Fudan University, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine Shanghai First Maternity and Infant Hospital Estimated completion date January 2028.

SUNNY

Study name	Study of upfront surgery versus neoadjuvant chemotherapy in patients with advanced ovarian cancer (SUNNY) in China and Korea
Methods	To compare the efficacy and safety in women with FIGO (2014) stage IIIC or IV epithelial ovarian cancer, fallopian tube cancer, or peritoneal carcinoma treated with neoadjuvant chemotherapy followed by interval debulking surgery versus upfront surgery. A randomised phase III multi-centre study
Participants	A total of 456 women will be accrued for this study within 5 years. Inclusion criteria • Age ≥ 18 years. • Pathologic confirmed stage IIIC and IV epithelial ovarian cancer, fallopian tube cancer or primary peritoneal carcinoma (diagnosis by biopsy or fine needle aspiration*). Laparoscopic biopsy with pictures is recommended. * If fine needle aspiration showing an adenocarcinoma, women should satisfy the following conditions: a. the patient has a pelvic mass, and b. omental cake or other metastasis larger than 2 cm in the upper abdomen, or pathologic confirmed extra-abdominal metastasis, and c. serum CA-125/CEA ratio>25. If serum CA-125/CEA ratio<25 or malignancies of other origins, such as breasts and digestive tract, are suspected from symptoms, physical examinations or imaging diagnosis, endoscopy or ultrasonography should be done to exclusive metastasis ovarian cancer. • ECOG performance status of 0 to 2 • ASA score of 1 to 2 • Adequate bone marrow, liver and renal function to receive chemotherapy and subsequently to undergo surgery • White blood cells >3,000/μL, absolute neutrophil count ≥1,500/μL, platelets ≥100,000/μL, haemoglobin ≥9 g/dL

SUNNY *(Continued)*

	• Serum creatinine <1.25 x upper limit of normal (ULN) or creatinine clearance ≥60 mL/min according to Cockroft-Gault formula or to local lab measurement • Serum bilirubin <1.25 x ULN, AST(SGOT) and ALT(SGPT) < 2.5 x ULN • Comply with the study protocol and follow-up • Written informed consent Exclusion Criteria • Women with non-epithelial tumours as well as borderline tumours • Mucinous ovarian cancer • Low-grade ovarian cancer • Synchronous or metachronous (within 5 years) malignancy other than carcinoma in situ • Any other concurrent medical conditions contraindicating surgery or chemotherapy that could compromise the adherence to the protocol • Other conditions, such as religious, psychological and other factors, that could interfere with provision of informed consent, compliance to study procedures, or follow-up
Interventions	Women will receive upfront maximal cytoreductive surgery followed by at least 6 cycles of adjuvant chemotherapy or 3 cycles of neoadjuvant chemotherapy followed by interval debulking surgery, and then at least 3 cycles of adjuvant chemotherapy. Women are followed every 3 months within the first 5 years, and then every 6 months.
Outcomes	Primary outcome measure • OS Secondary outcome measures • PFS • Postoperative complications - the surgical complications will be evaluated at 30-day after upfront cytoreductive surgery or interval debulking surgery • QoL assessments using QOQ-C30 questionnaire
Starting date	December 2015
Contact information	Rong Jiang, MD - jiang.rong@zs-hospital.sh.cn Yuting Luan, RN - yutingluan@163.com
Notes	Estimated study completion date December 2022

ALT: alanine aminotransferase; ASCO: American Society of Clinical Oncology; ASA; American Society of Anesthesiology; AST: aspartate aminotransferase; AUC: area under the curve; BRCA: Breast cancer susceptibility protein (g = germline; s = somatic); CI: confidence interval; cPCI: clinical peritoneal cancer index; CR: complete response; ECOG PS: Eastern Cooperative Oncology Group Performance Scale; EOC: epithelial ovarian carcinoma; HR: hazard ratio; ICU: intensive care unit; NACT: neoadjuvant chemotherapy; OS: overall survival; PDS: primary debulking surgery; PFS: progression-free survival; QoL: quality of life; PR: partial regression; RCT: randomised controlled trial; ULN: upper limit of normal.

DATA AND ANALYSES

Comparison 1. NACT vs PDS

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.1 Overall survival	4	1692	Hazard Ratio (IV, Random, 95% CI)	0.96 [0.86, 1.08]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.2 Overall survival by age	3	1391	Hazard Ratio (IV, Random, 95% CI)	0.94 [0.83, 1.06]
1.2.1 Age < 50 years	2	129	Hazard Ratio (IV, Random, 95% CI)	1.12 [0.64, 1.96]
1.2.2 Age <60 years	1	157	Hazard Ratio (IV, Random, 95% CI)	0.71 [0.50, 1.01]
1.2.3 Age 50-60 years	1	57	Hazard Ratio (IV, Random, 95% CI)	1.17 [0.59, 2.29]
1.2.4 Age 50-70 years	1	439	Hazard Ratio (IV, Random, 95% CI)	0.96 [0.77, 1.19]
1.2.5 Age 60-70 years	2	271	Hazard Ratio (IV, Random, 95% CI)	0.93 [0.71, 1.22]
1.2.6 Age > 70 years	3	338	Hazard Ratio (IV, Random, 95% CI)	0.99 [0.78, 1.25]
1.3 Overall survival by residual disease	2	1173	Hazard Ratio (IV, Random, 95% CI)	0.93 [0.79, 1.11]
1.3.1 Residual disease up to 0.5cm	2	334	Hazard Ratio (IV, Random, 95% CI)	1.12 [0.58, 2.13]
1.3.2 0.5cm > Residual disease ≤ 1cm	2	399	Hazard Ratio (IV, Random, 95% CI)	0.86 [0.69, 1.08]
1.3.3 Residual tumour > 1 cm	1	172	Hazard Ratio (IV, Random, 95% CI)	0.89 [0.64, 1.24]
1.3.4 Residual disease 1-2cm	1	218	Hazard Ratio (IV, Random, 95% CI)	0.82 [0.61, 1.10]
1.3.5 Residual disease >2cm	1	50	Hazard Ratio (IV, Random, 95% CI)	1.08 [0.59, 1.99]
1.4 Overall survival by stage	3	1519	Hazard Ratio (IV, Random, 95% CI)	0.95 [0.84, 1.08]
1.4.1 Stage 3	3	1128	Hazard Ratio (IV, Random, 95% CI)	0.98 [0.85, 1.14]
1.4.2 Stage 4	3	391	Hazard Ratio (IV, Random, 95% CI)	0.88 [0.69, 1.14]
1.5 Progression-free survival	4	1692	Hazard Ratio (IV, Random, 95% CI)	0.98 [0.88, 1.08]
1.6 Surgically-related severe adverse effects (grade 3+)	5		Risk Ratio (IV, Random, 95% CI)	Subtotals only
1.6.1 Haemorrhage	3	1264	Risk Ratio (IV, Random, 95% CI)	0.93 [0.50, 1.74]
1.6.2 Need for blood transfusion	4	1085	Risk Ratio (IV, Random, 95% CI)	0.80 [0.65, 0.99]
1.6.3 Venous thromboembolism	4	1490	Risk Ratio (IV, Random, 95% CI)	0.28 [0.09, 0.90]
1.6.4 Infection	4	1490	Risk Ratio (IV, Random, 95% CI)	0.30 [0.16, 0.56]

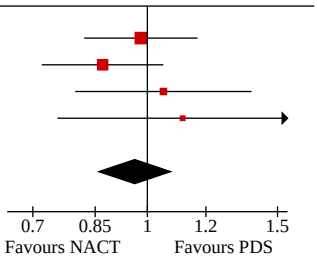
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.6.5 Gastrointestinal fistula	4	1541	Risk Ratio (IV, Random, 95% CI)	0.30 [0.09, 0.97]
1.6.6 Urinary/vaginal fistula	2	1106	Risk Ratio (IV, Random, 95% CI)	1.06 [0.15, 7.49]
1.6.7 Nausea	2	577	Risk Ratio (IV, Random, 95% CI)	0.42 [0.02, 8.23]
1.6.8 Vomiting	2	577	Risk Ratio (IV, Random, 95% CI)	0.41 [0.03, 6.03]
1.6.9 Diarrhoea	1	474	Risk Ratio (IV, Random, 95% CI)	0.58 [0.11, 3.15]
1.6.10 Neutropenia	1	103	Risk Ratio (IV, Random, 95% CI)	1.15 [0.48, 2.74]
1.6.11 Neutrotoxicity	1	103	Risk Ratio (IV, Random, 95% CI)	1.02 [0.15, 6.97]
1.6.12 Thrombocytopenia	1	103	Risk Ratio (IV, Random, 95% CI)	5.10 [0.25, 103.61]
1.6.13 Febrile neutropenia	1	103	Risk Ratio (IV, Random, 95% CI)	3.06 [0.13, 73.36]
1.6.14 Renal toxicity	1	103	Risk Ratio (IV, Random, 95% CI)	Not estimable
1.6.15 Stoma formation	2	632	Risk Ratio (IV, Random, 95% CI)	0.29 [0.12, 0.74]
1.6.16 Bowel resection	4	1565	Risk Ratio (IV, Random, 95% CI)	0.49 [0.30, 0.79]
1.6.17 Splenectomy	3	1067	Risk Ratio (IV, Random, 95% CI)	0.31 [0.08, 1.12]
1.6.18 Post- operative G3+ events	2	435	Risk Ratio (IV, Random, 95% CI)	0.22 [0.13, 0.38]
1.7 Postoperative mortality	5	1623	Risk Ratio (IV, Random, 95% CI)	0.16 [0.06, 0.46]
1.8 Chemotherapy-related SAEs (G3+)	2	768	Odds Ratio (IV, Random, 95% CI)	0.88 [0.57, 1.36]
1.9 EORTC QLQ-C30 QoL at 6 months	3		Mean Difference (IV, Random, 95% CI)	Subtotals only
1.9.1 Global health	3	524	Mean Difference (IV, Random, 95% CI)	-0.29 [-2.77, 2.20]
1.9.2 Fatigue	2	307	Mean Difference (IV, Random, 95% CI)	-0.55 [-6.02, 4.93]
1.9.3 Nausea	2	307	Mean Difference (IV, Random, 95% CI)	2.12 [-0.36, 4.61]
1.9.4 Pain	2	307	Mean Difference (IV, Random, 95% CI)	0.35 [-7.41, 8.12]
1.9.5 Constipation	2	307	Mean Difference (IV, Random, 95% CI)	-2.17 [-7.24, 2.89]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.9.6 Insomnia	2	307	Mean Difference (IV, Random, 95% CI)	0.30 [-0.86, 1.47]
1.9.7 Apetite loss	2	307	Mean Difference (IV, Random, 95% CI)	0.47 [-0.31, 1.24]
1.9.8 Dyspneoa	2	307	Mean Difference (IV, Random, 95% CI)	2.47 [-3.42, 8.36]
1.9.9 Diarrhoea	2	307	Mean Difference (IV, Random, 95% CI)	-0.77 [-12.69, 11.15]
1.9.10 Financial difficulties	2	307	Mean Difference (IV, Random, 95% CI)	2.46 [-5.33, 10.25]
1.10 EORTC QLQ-C30 QoL at 12 months	2		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.10.1 Global health	2		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.10.2 Fatigue	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.10.3 Nausea	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.10.4 Pain	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.10.5 Dyspneoa	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.10.6 Insomnia	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.10.7 Apetite loss	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.10.8 Constipation	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.10.9 Diarrhoea	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.10.10 Financial difficulties	1		Mean Difference (IV, Random, 95% CI)	Totals not selected

Analysis 1.1. Comparison 1: NACT vs PDS, Outcome 1: Overall survival

Study or Subgroup	log[Hazard Ratio]	SE	Favours NACT Total	PDS Total	Weight	Hazard Ratio IV, Random, 95% CI
Vergote 2010 (1)	-0.0202	0.09	334	336	40.2%	0.98 [0.82 , 1.17]
Kehoe 2015	-0.1393	0.0966	274	276	34.9%	0.87 [0.72 , 1.05]
Onda 2016	0.05	0.14	152	149	16.6%	1.05 [0.80 , 1.38]
Fagotti 2016	0.11	0.199	87	84	8.2%	1.12 [0.76 , 1.65]
Total (95% CI)			**847**	**845**	**100.0%**	**0.96 [0.86 , 1.08]**

Heterogeneity: Tau² = 0.00; Chi² = 2.09, df = 3 (P = 0.55); I² = 0%
Test for overall effect: Z = 0.69 (P = 0.49)
Test for subgroup differences: Not applicable

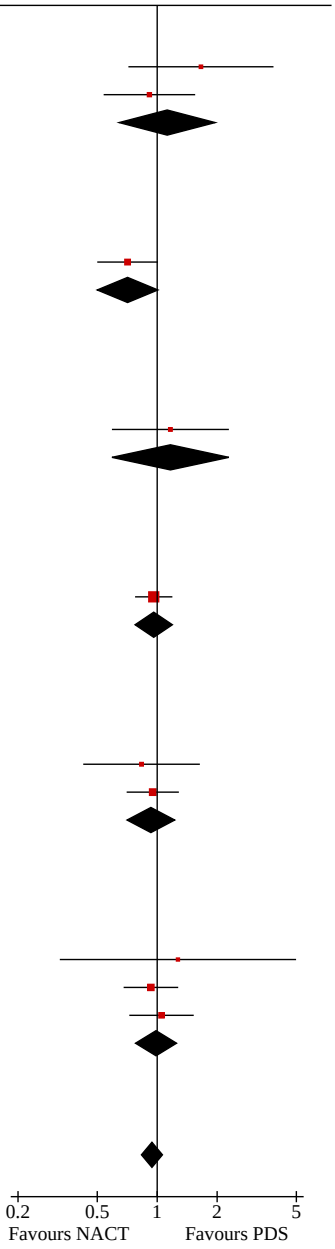


Footnotes

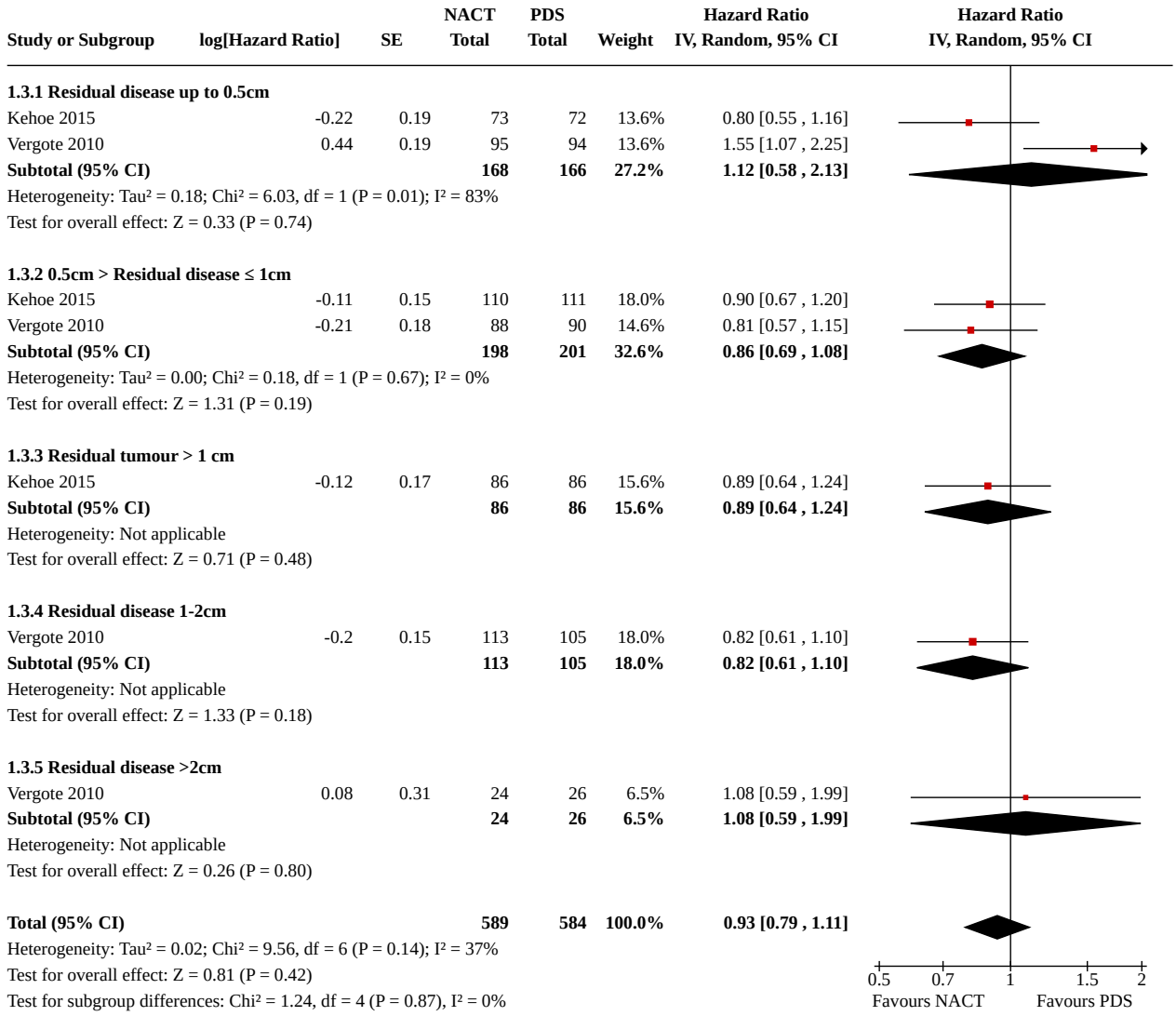
(1) We have applied 95% CIs (investigators reported 90% CIs).

Analysis 1.2. Comparison 1: NACT vs PDS, Outcome 2: Overall survival by age

Study or Subgroup	log[Hazard Ratio]	SE	NACT Total	PDS Total	Weight	Hazard Ratio IV, Random, 95% CI
1.2.1 Age < 50 years						
Fagotti 2016	0.506	0.428	23	22	2.1%	1.66 [0.72 , 3.84]
Vergote 2010	-0.09	0.27	47	37	5.3%	0.91 [0.54 , 1.55]
Subtotal (95% CI)			**70**	**59**	**7.4%**	**1.12 [0.64 , 1.96]**
Heterogeneity: Tau² = 0.05; Chi² = 1.39, df = 1 (P = 0.24); I² = 28%						
Test for overall effect: Z = 0.41 (P = 0.68)						
1.2.2 Age <60 years						
Kehoe 2015	-0.3425	0.1789	75	82	12.0%	0.71 [0.50 , 1.01]
Subtotal (95% CI)			**75**	**82**	**12.0%**	**0.71 [0.50 , 1.01]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.91 (P = 0.06)						
1.2.3 Age 50-60 years						
Fagotti 2016	0.1528	0.345	23	34	3.2%	1.17 [0.59 , 2.29]
Subtotal (95% CI)			**23**	**34**	**3.2%**	**1.17 [0.59 , 2.29]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.44 (P = 0.66)						
1.2.4 Age 50-70 years						
Vergote 2010	-0.04	0.11	210	229	31.7%	0.96 [0.77 , 1.19]
Subtotal (95% CI)			**210**	**229**	**31.7%**	**0.96 [0.77 , 1.19]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.36 (P = 0.72)						
1.2.5 Age 60-70 years						
Fagotti 2016	-0.181	0.344	33	23	3.2%	0.83 [0.43 , 1.64]
Kehoe 2015	-0.0513	0.154	116	99	16.2%	0.95 [0.70 , 1.28]
Subtotal (95% CI)			**149**	**122**	**19.4%**	**0.93 [0.71 , 1.22]**
Heterogeneity: Tau² = 0.00; Chi² = 0.12, df = 1 (P = 0.73); I² = 0%						
Test for overall effect: Z = 0.52 (P = 0.60)						
1.2.6 Age > 70 years						
Fagotti 2016	0.24	0.697	8	5	0.8%	1.27 [0.32 , 4.98]
Kehoe 2015	-0.0726	0.161	83	95	14.8%	0.93 [0.68 , 1.28]
Vergote 2010	0.05	0.19	77	70	10.6%	1.05 [0.72 , 1.53]
Subtotal (95% CI)			**168**	**170**	**26.2%**	**0.99 [0.78 , 1.25]**
Heterogeneity: Tau² = 0.00; Chi² = 0.38, df = 2 (P = 0.83); I² = 0%						
Test for overall effect: Z = 0.11 (P = 0.91)						
Total (95% CI)			**695**	**696**	**100.0%**	**0.94 [0.83 , 1.06]**
Heterogeneity: Tau² = 0.00; Chi² = 5.32, df = 9 (P = 0.81); I² = 0%						
Test for overall effect: Z = 0.98 (P = 0.33)						
Test for subgroup differences: Chi² = 3.45, df = 5 (P = 0.63), I² = 0%						

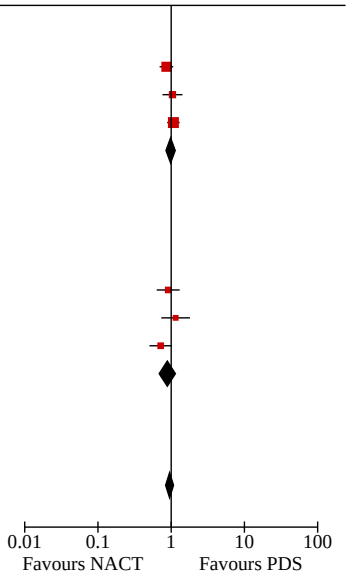


Analysis 1.3. Comparison 1: NACT vs PDS, Outcome 3: Overall survival by residual disease



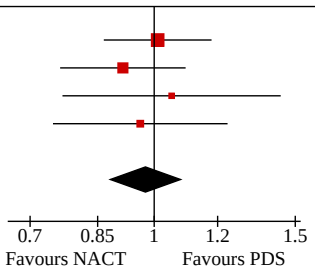
Analysis 1.4. Comparison 1: NACT vs PDS, Outcome 4: Overall survival by stage

Study or Subgroup	log[Hazard Ratio]	SE	NACT Total	PDS Total	Weight	Hazard Ratio IV, Random, 95% CI
1.4.1 Stage 3						
Kehoe 2015	-0.150823	0.109	206	206	26.5%	0.86 [0.69 , 1.06]
Onda 2016	0.04	0.16	105	100	13.9%	1.04 [0.76 , 1.42]
Vergote 2010	0.07	0.1	253	258	30.3%	1.07 [0.88 , 1.30]
Subtotal (95% CI)			**564**	**564**	**70.7%**	**0.98 [0.85 , 1.14]**
Heterogeneity: Tau² = 0.00; Chi² = 2.39, df = 2 (P = 0.30); I² = 16%						
Test for overall effect: Z = 0.25 (P = 0.80)						
1.4.2 Stage 4						
Kehoe 2015	-0.094311	0.1836	68	70	10.9%	0.91 [0.63 , 1.30]
Onda 2016	0.14	0.23	47	49	7.2%	1.15 [0.73 , 1.81]
Vergote 2010	-0.33	0.18	81	76	11.3%	0.72 [0.51 , 1.02]
Subtotal (95% CI)			**196**	**195**	**29.3%**	**0.88 [0.69 , 1.14]**
Heterogeneity: Tau² = 0.01; Chi² = 2.65, df = 2 (P = 0.27); I² = 25%						
Test for overall effect: Z = 0.94 (P = 0.35)						
Total (95% CI)			**760**	**759**	**100.0%**	**0.95 [0.84 , 1.08]**
Heterogeneity: Tau² = 0.00; Chi² = 5.77, df = 5 (P = 0.33); I² = 13%						
Test for overall effect: Z = 0.80 (P = 0.42)						
Test for subgroup differences: Chi² = 0.48, df = 1 (P = 0.49), I² = 0%						



Analysis 1.5. Comparison 1: NACT vs PDS, Outcome 5: Progression-free survival

Study or Subgroup	log[Hazard Ratio]	SE	NACT Total	PDS Total	Weight	Hazard Ratio IV, Random, 95% CI
Vergote 2010 (1)	0.01	0.079	334	336	42.3%	1.01 [0.87 , 1.18]
Kehoe 2015 (2)	-0.09	0.092	274	276	31.2%	0.91 [0.76 , 1.09]
Fagotti 2016	0.05	0.16	87	84	10.3%	1.05 [0.77 , 1.44]
Onda 2016	-0.04	0.128	152	149	16.1%	0.96 [0.75 , 1.23]
Total (95% CI)			**847**	**845**	**100.0%**	**0.98 [0.88 , 1.08]**
Heterogeneity: Tau² = 0.00; Chi² = 0.93, df = 3 (P = 0.82); I² = 0%						
Test for overall effect: Z = 0.49 (P = 0.62)						
Test for subgroup differences: Not applicable						



Footnotes

(1) We have applied 95% CIs (Investigators used 90% CIs)

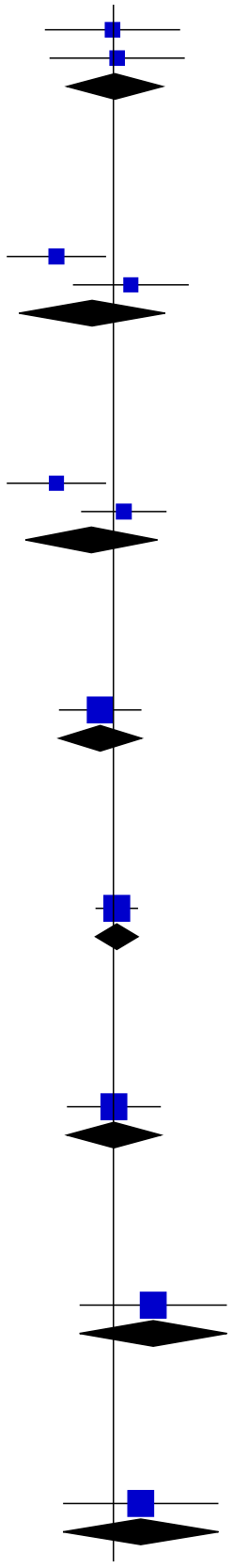
(2) 0.09

Analysis 1.6. Comparison 1: NACT vs PDS, Outcome 6: Surgically-related severe adverse effects (grade 3+)

Study or Subgroup	NACT Events	NACT Total	PDS Events	PDS Total	Weight	Risk Ratio IV, Random, 95% CI
1.6.1 Haemorrhage						
Vergote 2010 (1)	12	322	23	310	31.0%	0.50 [0.25 , 0.99]
Kehoe 2015	14	219	8	255	25.6%	2.04 [0.87 , 4.77]
Fagotti 2016 (2)	34	74	42	84	43.4%	0.92 [0.66 , 1.27]
Subtotal (95% CI)		**615**		**649**	**100.0%**	**0.93 [0.50 , 1.74]**
Total events:	60		73			
Heterogeneity: Tau² = 0.20; Chi² = 6.39, df = 2 (P = 0.04); I² = 69%						
Test for overall effect: Z = 0.21 (P = 0.83)						
1.6.2 Need for blood transfusion						
Vergote 2010	155	289	181	310	47.0%	0.92 [0.80 , 1.06]
Chekman 2015	9	41	15	41	7.9%	0.60 [0.30 , 1.21]
Fagotti 2016	5	52	15	55	4.7%	0.35 [0.14 , 0.90]
Onda 2016	79	150	98	147	40.4%	0.79 [0.65 , 0.96]
Subtotal (95% CI)		**532**		**553**	**100.0%**	**0.80 [0.65 , 0.99]**
Total events:	248		309			
Heterogeneity: Tau² = 0.02; Chi² = 6.03, df = 3 (P = 0.11); I² = 50%						
Test for overall effect: Z = 2.07 (P = 0.04)						
1.6.3 Venous thromboembolism						
Vergote 2010	0	322	8	310	15.0%	0.06 [0.00 , 0.98]
Kehoe 2015	0	219	5	255	14.6%	0.11 [0.01 , 1.90]
Fagotti 2016	0	52	3	55	14.1%	0.15 [0.01 , 2.85]
Onda 2016	4	130	7	147	56.3%	0.65 [0.19 , 2.16]
Subtotal (95% CI)		**723**		**767**	**100.0%**	**0.28 [0.09 , 0.90]**
Total events:	4		23			
Heterogeneity: Tau² = 0.25; Chi² = 3.53, df = 3 (P = 0.32); I² = 15%						
Test for overall effect: Z = 2.14 (P = 0.03)						
1.6.4 Infection						
Vergote 2010	5	322	25	310	43.8%	0.19 [0.07 , 0.50]
Kehoe 2015	6	219	16	255	46.4%	0.44 [0.17 , 1.10]
Fagotti 2016	0	52	4	55	4.7%	0.12 [0.01 , 2.13]
Onda 2016	1	130	1	147	5.2%	1.13 [0.07 , 17.90]
Subtotal (95% CI)		**723**		**767**	**100.0%**	**0.30 [0.16 , 0.56]**
Total events:	12		46			
Heterogeneity: Tau² = 0.00; Chi² = 2.77, df = 3 (P = 0.43); I² = 0%						
Test for overall effect: Z = 3.75 (P = 0.0002)						
1.6.5 Gastrointestinal fistula						
Vergote 2010	1	322	3	310	27.9%	0.32 [0.03 , 3.07]
Kehoe 2015	1	219	2	255	24.8%	0.58 [0.05 , 6.38]
Onda 2016	0	130	5	147	17.1%	0.10 [0.01 , 1.84]
Fagotti 2016 (3)	1	74	4	84	30.2%	0.28 [0.03 , 2.48]
Subtotal (95% CI)		**745**		**796**	**100.0%**	**0.30 [0.09 , 0.97]**
Total events:	3		14			
Heterogeneity: Tau² = 0.00; Chi² = 0.83, df = 3 (P = 0.84); I² = 0%						
Test for overall effect: Z = 2.01 (P = 0.04)						
1.6.6 Urinary/vaginal fistula						
Vergote 2010	1	322	1	310	50.0%	0.96 [0.06 , 15.32]
Kehoe 2015	1	219	1	255	50.0%	1.16 [0.07 , 18.51]

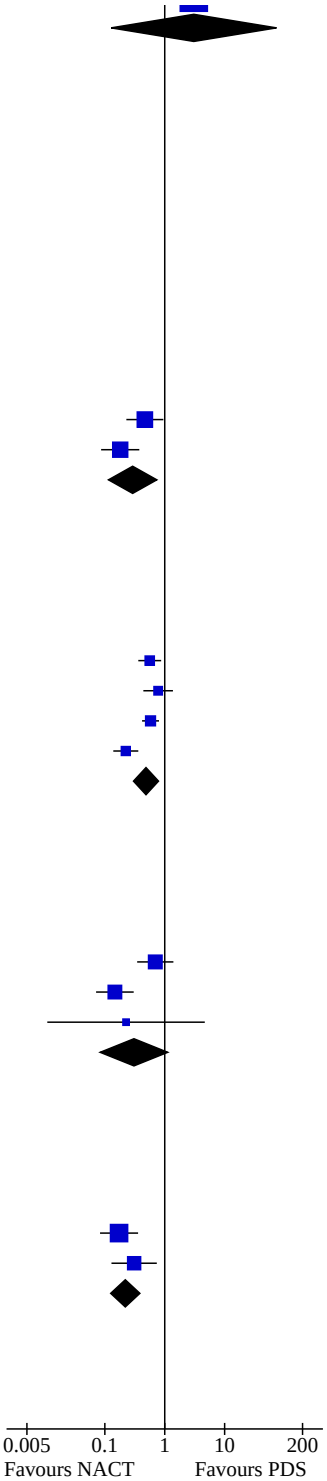
Analysis 1.6. (Continued)

Vergote 2010	1	322	1	310	50.0%	0.96 [0.06 , 15.32]
Kehoe 2015	1	219	1	255	50.0%	1.16 [0.07 , 18.51]
Subtotal (95% CI)		**541**		**565**	**100.0%**	**1.06 [0.15 , 7.49]**
Total events:	2		2			
Heterogeneity: Tau² = 0.00; Chi² = 0.01, df = 1 (P = 0.92); I² = 0%						
Test for overall effect: Z = 0.06 (P = 0.95)						
1.6.7 Nausea						
Kehoe 2015	1	219	12	255	52.1%	0.10 [0.01 , 0.74]
Fagotti 2016	2	51	1	52	47.9%	2.04 [0.19 , 21.80]
Subtotal (95% CI)		**270**		**307**	**100.0%**	**0.42 [0.02 , 8.23]**
Total events:	3		13			
Heterogeneity: Tau² = 3.37; Chi² = 3.66, df = 1 (P = 0.06); I² = 73%						
Test for overall effect: Z = 0.57 (P = 0.57)						
1.6.8 Vomiting						
Kehoe 2015	1	219	12	255	48.2%	0.10 [0.01 , 0.74]
Fagotti 2016	3	51	2	52	51.8%	1.53 [0.27 , 8.77]
Subtotal (95% CI)		**270**		**307**	**100.0%**	**0.41 [0.03 , 6.03]**
Total events:	4		14			
Heterogeneity: Tau² = 2.87; Chi² = 4.07, df = 1 (P = 0.04); I² = 75%						
Test for overall effect: Z = 0.66 (P = 0.51)						
1.6.9 Diarrhoea						
Kehoe 2015	2	219	4	255	100.0%	0.58 [0.11 , 3.15]
Subtotal (95% CI)		**219**		**255**	**100.0%**	**0.58 [0.11 , 3.15]**
Total events:	2		4			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.63 (P = 0.53)						
1.6.10 Neutropenia						
Fagotti 2016	9	51	8	52	100.0%	1.15 [0.48 , 2.74]
Subtotal (95% CI)		**51**		**52**	**100.0%**	**1.15 [0.48 , 2.74]**
Total events:	9		8			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.31 (P = 0.76)						
1.6.11 Neutrotoxicity						
Fagotti 2016	2	51	2	52	100.0%	1.02 [0.15 , 6.97]
Subtotal (95% CI)		**51**		**52**	**100.0%**	**1.02 [0.15 , 6.97]**
Total events:	2		2			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.02 (P = 0.98)						
1.6.12 Thrombocytopenia						
Fagotti 2016	2	51	0	52	100.0%	5.10 [0.25 , 103.61]
Subtotal (95% CI)		**51**		**52**	**100.0%**	**5.10 [0.25 , 103.61]**
Total events:	2		0			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.06 (P = 0.29)						
1.6.13 Febrile neutropenia						
Fagotti 2016	1	51	0	52	100.0%	3.06 [0.13 , 73.36]
Subtotal (95% CI)		**51**		**52**	**100.0%**	**3.06 [0.13 , 73.36]**
Total events:	1		0			



Analysis 1.6. (Continued)

Subtotal (95% CI)		**51**		**52**	**100.0%**	**3.06 [0.13 , 73.36]**
Total events:	1		0			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.69 (P = 0.49)						
1.6.14 Renal toxicity						
Fagotti 2016	0	51	0	52		Not estimable
Subtotal (95% CI)		**51**		**52**		**Not estimable**
Total events:	0		0			
Heterogeneity: Not applicable						
Test for overall effect: Not applicable						
1.6.15 Stoma formation						
Kehoe 2015	10	219	25	255	50.5%	0.47 [0.23 , 0.95]
Fagotti 2016	7	74	44	84	49.5%	0.18 [0.09 , 0.38]
Subtotal (95% CI)		**293**		**339**	**100.0%**	**0.29 [0.12 , 0.74]**
Total events:	17		69			
Heterogeneity: Tau² = 0.31; Chi² = 3.30, df = 1 (P = 0.07); I² = 70%						
Test for overall effect: Z = 2.60 (P = 0.009)						
1.6.16 Bowel resection						
Vergote 2010	28	322	48	310	25.4%	0.56 [0.36 , 0.87]
Kehoe 2015	18	219	27	255	22.2%	0.78 [0.44 , 1.37]
Onda 2016	39	152	66	149	28.1%	0.58 [0.42 , 0.80]
Fagotti 2016 (4)	14	74	71	84	24.4%	0.22 [0.14 , 0.36]
Subtotal (95% CI)		**767**		**798**	**100.0%**	**0.49 [0.30 , 0.79]**
Total events:	99		212			
Heterogeneity: Tau² = 0.19; Chi² = 14.10, df = 3 (P = 0.003); I² = 79%						
Test for overall effect: Z = 2.94 (P = 0.003)						
1.6.17 Splenectomy						
Vergote 2010	13	322	18	310	43.5%	0.70 [0.35 , 1.39]
Fagotti 2016	7	74	54	84	43.1%	0.15 [0.07 , 0.30]
Onda 2016	0	130	2	147	13.4%	0.23 [0.01 , 4.66]
Subtotal (95% CI)		**526**		**541**	**100.0%**	**0.31 [0.08 , 1.12]**
Total events:	20		74			
Heterogeneity: Tau² = 0.88; Chi² = 9.26, df = 2 (P = 0.010); I² = 78%						
Test for overall effect: Z = 1.79 (P = 0.07)						
1.6.18 Post- operative G3+ events						
Fagotti 2016 (5)	7	74	46	84	58.7%	0.17 [0.08 , 0.36]
Onda 2016	6	130	22	147	41.3%	0.31 [0.13 , 0.74]
Subtotal (95% CI)		**204**		**231**	**100.0%**	**0.22 [0.13 , 0.38]**
Total events:	13		68			
Heterogeneity: Tau² = 0.00; Chi² = 1.00, df = 1 (P = 0.32); I² = 0%						
Test for overall effect: Z = 5.31 (P < 0.00001)						



Footnotes

(1) Results for all SAEs in this trial are per protocol, not ITT.

(2) Estimated Blood Loss >750 ml for those who had surgery

(3) three pancreatic fistulae and one biliary fistula

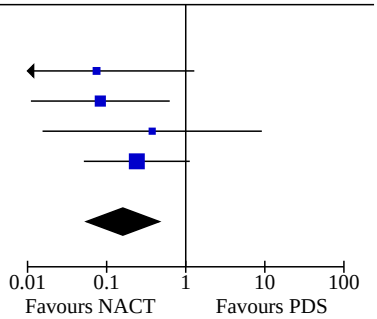
(4) Single bowel resection (NACT = 10 versus PDS = 52); multiple bowel resections (NACT= 4 versus PDS =19)

(5) within 30 days of surgery. Further post-op SAE > 30 days (NACT =1; PDS = 13)

Analysis 1.7. Comparison 1: NACT vs PDS, Outcome 7: Postoperative mortality

Study or Subgroup	NACT Events	NACT Total	PDS Events	PDS Total	Weight	Risk Ratio IV, Random, 95% CI
Chekman 2015	0	41	0	40		Not estimable
Fagotti 2016 (1)	0	75	7	84	13.9%	0.07 [0.00 , 1.28]
Kehoe 2015 (2)	1	219	14	255	27.6%	0.08 [0.01 , 0.63]
Onda 2016 (3)	0	130	1	147	11.1%	0.38 [0.02 , 9.16]
Vergote 2010 (2)	2	322	8	310	47.4%	0.24 [0.05 , 1.12]
Total (95% CI)		**787**		**836**	**100.0%**	**0.16 [0.06 , 0.46]**
Total events:	3		30			

Heterogeneity: Tau² = 0.00; Chi² = 1.22, df = 3 (P = 0.75); I² = 0%
Test for overall effect: Z = 3.38 (P = 0.0007)
Test for subgroup differences: Not applicable



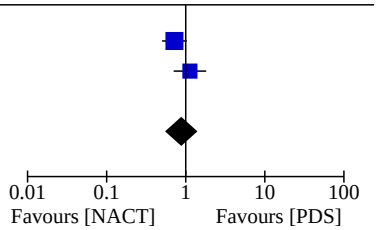
Footnotes

(1) Fagotti 2016 includes 3 post-op deaths within 30 days and a further 4 late post-op deaths, over 30 days, due to post-op complications.
(2) deaths within 28 days of surgery
(3) Defined as 'treatment-related deaths related to surgery' within 4 weeks of surgery

Analysis 1.8. Comparison 1: NACT vs PDS, Outcome 8: Chemotherapy-related SAEs (G3+)

Study or Subgroup	Experimental Events	Experimental Total	Control Events	Control Total	Weight	Odds Ratio IV, Random, 95% CI
Kehoe 2015	102	254	110	228	56.0%	0.72 [0.50 , 1.03]
Onda 2016 (1)	62	149	53	137	44.0%	1.13 [0.70 , 1.81]
Total (95% CI)		**403**		**365**	**100.0%**	**0.88 [0.57 , 1.36]**
Total events:	164		163			

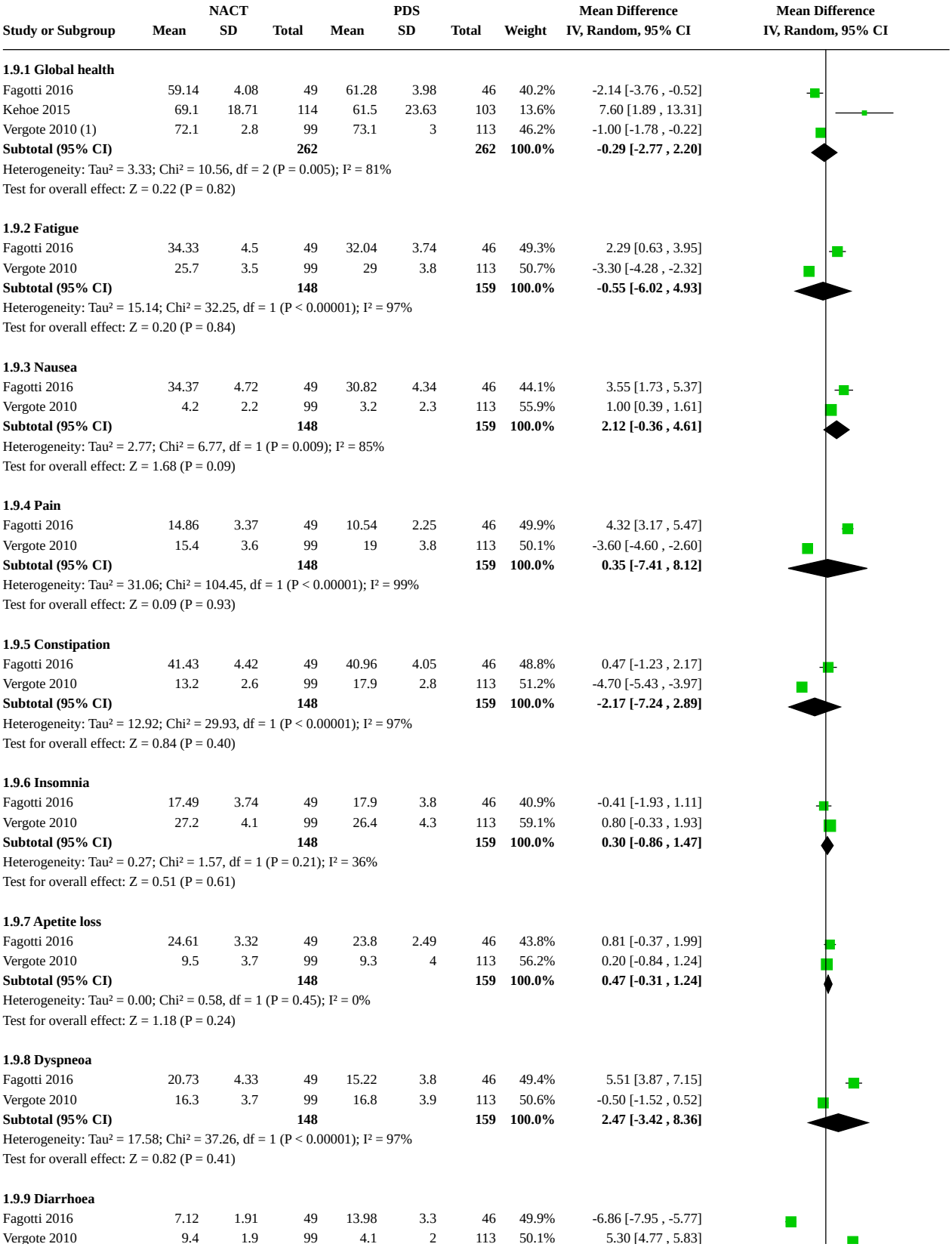
Heterogeneity: Tau² = 0.06; Chi² = 2.20, df = 1 (P = 0.14); I² = 54%
Test for overall effect: Z = 0.58 (P = 0.56)
Test for subgroup differences: Not applicable



Footnotes

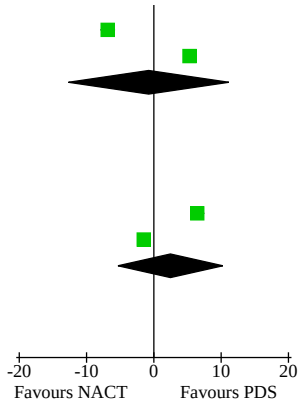
(1) Combination of SAEs during cycles 1-4 and 5-8 (All SAEs excluding bone marrow suppression)

Analysis 1.9. Comparison 1: NACT vs PDS, Outcome 9: EORTC QLQ-C30 QoL at 6 months



Analysis 1.9. (Continued)

Fagotti 2016	7.12	1.91	49	13.98	3.3	46	49.9%	-6.86 [-7.95 , -5.77]
Vergote 2010	9.4	1.9	99	4.1	2	113	50.1%	5.30 [4.77 , 5.83]
Subtotal (95% CI)			**148**			**159**	**100.0%**	**-0.77 [-12.69 , 11.15]**
Heterogeneity: Tau² = 73.74; Chi² = 386.02, df = 1 (P < 0.00001); I² = 100%								
Test for overall effect: Z = 0.13 (P = 0.90)								
1.9.10 Financial difficulties								
Fagotti 2016	39.47	2.56	49	33.02	2.66	46	49.8%	6.45 [5.40 , 7.50]
Vergote 2010	10.2	1.9	99	11.7	2	113	50.2%	-1.50 [-2.03 , -0.97]
Subtotal (95% CI)			**148**			**159**	**100.0%**	**2.46 [-5.33 , 10.25]**
Heterogeneity: Tau² = 31.42; Chi² = 175.84, df = 1 (P < 0.00001); I² = 99%								
Test for overall effect: Z = 0.62 (P = 0.54)								
Test for subgroup differences: Chi² = 4.05, df = 9 (P = 0.91), I² = 0%								

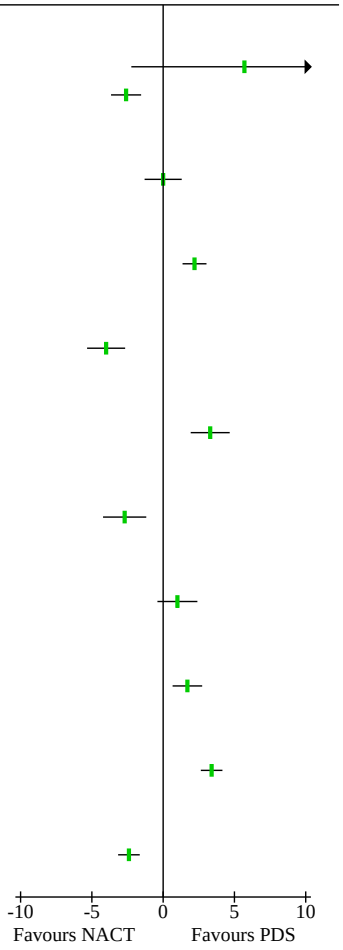


Footnotes

(1) Kehoe 2015 data now combined, as authors confirm Global QoL scores were on same EORTC QLQ-C30 scale as Vergote 2010 and Fagotti 2016 studies

Analysis 1.10. Comparison 1: NACT vs PDS, Outcome 10: EORTC QLQ-C30 QoL at 12 months

Study or Subgroup	NACT Mean	NACT SD	NACT Total	PDS Mean	PDS SD	PDS Total	Mean Difference IV, Random, 95% CI
1.10.1 Global health							
Kehoe 2015	67.5	22.38	69	61.8	24.16	64	5.70 [-2.23 , 13.63]
Vergote 2010	67.8	3.1	64	70.4	3.3	78	-2.60 [-3.66 , -1.54]
1.10.2 Fatigue							
Vergote 2010	29.1	3.8	64	29.1	4.1	78	0.00 [-1.30 , 1.30]
1.10.3 Nausea							
Vergote 2010	5.6	2.4	64	3.4	2.7	78	2.20 [1.36 , 3.04]
1.10.4 Pain							
Vergote 2010	15.1	3.9	64	19.1	4.2	78	-4.00 [-5.33 , -2.67]
1.10.5 Dyspneoa							
Vergote 2010	18.9	4	64	15.6	4.3	78	3.30 [1.93 , 4.67]
1.10.6 Insomnia							
Vergote 2010	22.1	4.4	64	24.8	4.8	78	-2.70 [-4.22 , -1.18]
1.10.7 Apetite loss							
Vergote 2010	10.6	4.1	64	9.6	4.4	78	1.00 [-0.40 , 2.40]
1.10.8 Constipation							
Vergote 2010	14.2	3	64	12.5	3.3	78	1.70 [0.66 , 2.74]
1.10.9 Diarrhoea							
Vergote 2010	8.1	2.2	64	4.7	2.4	78	3.40 [2.64 , 4.16]
1.10.10 Financial difficulties							
Vergote 2010	10	2.2	64	12.4	2.4	78	-2.40 [-3.16 , -1.64]



ADDITIONAL TABLES

Table 1. Carcinoma of the ovary: FIGO* nomenclature

Stage	Extent of tumour	Substage	Details
I	Limited to ovaries	Ia	Limited to 1 ovary, no tumour on surface or capsule rupture, no positive ascites
		Ib	Limited to both ovaries, no tumour on surface or capsule rupture, no positive ascites
		Ic	Stage Ia or Ib but with capsule ruptured, tumour on ovarian surface or positive peritoneal washings/ascites
II	Limited to 1 or both ovaries with pelvic extension	IIa	Extension, metastases to uterus, tubes, or a combination
		IIb	Extension to other pelvis tissues
		II c	Stage IIa or IIb with tumour on the surface of 1 or both ovaries, or with capsule ruptured, or with positive peritoneal washings/ascites
III	Limited to abdomen with histologically confirmed peritoneal implants outside the pelvis or positive nodes, or both, or extension to small bowel or omentum	IIIa	Tumour grossly limited to the true pelvis with negative regional lymph nodes, microscopic seeding of abdominal peritoneal surfaces or extension to small bowel or mesentery
		IIIb	Macroscopic metastases < 2 cm; negative regional lymph nodes
		IIIc	Macroscopic metastases > 2 cm or positive regional lymph nodes, or both
IV	Distant metastases		Growth outside the abdominal cavity (e.g. lung, liver parenchyma (superficial liver metastases is stage III))

FIGO: Federation of International Gynaecologists and Obstetricians
* From FIGO 2009 as all included studies used 2009 classification not 2018

APPENDICES

Appendix 1. Embase search strategy

Embase (R) 1980 to Sept 2006 via Ovid:
The search: (ovar*) and (cancer* or carcinoma* or malignan* or neoplas* or tumour* or tumor*) and (chemotherap*) and (surg*) and (rct or random* or study or studies or trial* or investigation*) and (advanced or stage III or stage IV)

Embase Sept 2006 to date:

1. exp ovary tumor/
2. (ovar* adj5 (neoplas* or tumor* or tumour* or cancer* or malignan* or carcinoma*)).mp.
3. 1 or 2
4. chemotherap*.mp.
5. dt.fs.
6. exp antineoplastic agent/
7. exp cancer chemotherapy/
8. adjuvant chemotherapy/

9. 4 or 5 or 6 or 7 or 8
10. surg*.mp.
11. su.fs.
12. exp surgery/
13. 10 or 11 or 12
14. 3 and 9 and 13
15. random*.ti,ab.
16. factorial*.ti,ab.
17. (crossover* or cross over* or cross-over*).ti,ab.
18. placebo*.ti,ab.
19. (doubl* adj blind*).ti,ab.
20. (singl* adj blind*).ti,ab.
21. assign*.ti,ab.
22. allocat*.ti,ab.
23. volunteer*.ti,ab.
24. crossover procedure/
25. double blind procedure/
26. randomised controlled trial/
27. single blind procedure/
28. 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27
29. 14 and 28

Appendix 2. MEDLINE search strategy

The full MEDLINE search strategy via Silver Platter, from 1966 to Sept 2006 was: (ovar*) and (cancer* or carcinoma* or malignan* or neoplas* or tumour* or tumor*) and (chemotherap*) and (surg*) and (rct or random* or study or studies or trial* or investigation*) and (advanced or stage III or stage IV)

It contained free text (including alternative spellings) and MeSH terms, and MeSH headings were exploded.

MEDLINE Sept 2006 to date:

1. exp Ovarian Neoplasms/
2. (ovar* adj5 (neoplas* or tumor* or tumour* or cancer* or malignan* or carcinoma*)).mp.
3. 1 or 2
4. chemotherap*.mp.
5. drug therapy.fs.
6. exp Antineoplastic Agents/
7. Antineoplastic Combined Chemotherapy Protocols/
8. Neoadjuvant Therapy/
9. 4 or 5 or 6 or 7 or 8
10. surg*.mp.
11. surgery.fs.
12. exp Surgical Procedures, Operative/
13. 10 or 11 or 12
14. 3 and 9 and 13
15. randomized controlled trial.pt.
16. controlled clinical trial.pt.
17. randomized.ab.
18. placebo.ab.
19. clinical trials as topic.sh.
20. randomly.ab.
21. trial.ti.
22. 15 or 16 or 17 or 18 or 19 or 20 or 21
23. 14 and 22

key:

mp=title, original title, abstract, name of substance word, subject heading word, unique identifier
fs=floating subheading
pt=publication type
ab=abstract

Appendix 3. CENTRAL search strategy

#1 MeSH descriptor Ovarian Neoplasms explode all trees
#2 ovar* near/5 (neoplas* or tumor* or tumour* or cancer* or malignan* or carcinoma*)
#3 (#1 OR #2)
#4 chemotherap*
#5 Any MeSH descriptor with qualifier: DT
#6 MeSH descriptor Antineoplastic Agents explode all trees
#7 MeSH descriptor Antineoplastic Combined Chemotherapy Protocols explode all trees
#8 MeSH descriptor Neoadjuvant Therapy explode all trees
#9 (#4 OR #5 OR #6 OR #7 OR #8)
#10 surg*
#11 Any MeSH descriptor with qualifier: SU
#12 MeSH descriptor Surgical Procedures, Operative explode all trees
#13 (#10 OR #11 OR #12)
#14 (#3 AND #9 AND #13)

Appendix 4. Assessing 'Risk of bias' of included studies

We assessed the risk of bias of included studies according to the following criteria.

(1) Random sequence generation (checking for possible selection bias)

We described for each included study the method used to generate the allocation sequence in sufficient detail to allow an assessment of whether it produced comparable groups. We assessed the method as:

- low risk of bias (any truly random process, e.g. random number table; computer random number generator);
- high risk of bias (any non-random process, e.g. odd or even date of birth; hospital or clinic record number);
- unclear risk of bias.

(2) Allocation concealment (checking for possible selection bias)

We described for each included study the method used to conceal allocation to interventions prior to assignment and assessed whether intervention allocation could have been foreseen in advance of, or during recruitment, or changed after assignment. We assessed the methods as:

- low risk of bias (e.g. telephone or central randomisation; consecutively numbered sealed opaque envelopes);
- high risk of bias (open random allocation; unsealed or non-opaque envelopes, alternation; date of birth);
- unclear risk of bias.

(3) Blinding of outcome assessment (checking for possible detection bias)

We described for each included study the methods used, if any, to blind outcome assessors from knowledge of which intervention a participant received. We assessed methods used to blind outcome assessment as:

- low, high or unclear risk of bias.

(4) Incomplete outcome data (checking for possible attrition bias owing to the amount, nature and handling of incomplete outcome data)

We described for each included study the completeness of data including attrition and exclusions from the analysis. We stated whether attrition and exclusions were reported and the numbers included in the analysis at each stage (compared with the total randomised participants), reasons for attrition or exclusions where reported, and whether missing data were balanced across groups or were related to outcomes. Where sufficient information was reported, or could be supplied by the trial authors, we re-included missing data in the analyses that we undertook. We assessed methods as:

- low risk of bias (e.g. no missing outcome data or missing data < 20%; missing outcome data balanced across groups);

- high risk of bias (e.g. numbers or reasons for missing data imbalanced across groups; 'as treated' analysis done with substantial departure of intervention received from that assigned at randomisation or <80% assessed at endpoint for at least the primary outcomes);
- unclear risk of bias.

(5) Selective reporting (checking for reporting bias)

We described for each included study how we investigated the possibility of selective outcome reporting bias and what we found. We assessed the methods as:

- low risk of bias (where it is clear that all of the study's pre-specified outcomes and all expected outcomes of interest to the review were reported);
- high risk of bias (where not all the study's pre-specified outcomes were reported; one or more reported primary outcomes were not pre-specified; outcomes of interest were reported incompletely and so could not be used; study failed to include results of a key outcome that would have been expected to have been reported);
- unclear risk of bias.

(6) Other bias (checking for bias owing to problems not covered by 1 to 5 above)

We described for each included study any important concerns we had about other possible sources of bias. We assessed each study as:

- low risk of other bias;
- high risk of other bias;
- unclear whether there is risk of other bias.

WHAT'S NEW

Date	Event	Description
7 April 2021	New citation required but conclusions have not changed	Review updated but conclusions not changed
7 April 2021	New search has been performed	New search to 9 October 2020 and data added from studies included in previous version

HISTORY

Protocol first published: Issue 3, 2005
Review first published: Issue 4, 2007

Date	Event	Description
1 February 2021	Amended	Correction to survival data for Kehoe 2015
1 February 2021	New citation required but conclusions have not changed	New citation required but conclusions have not changed. Correction to survival data for Kehoe 2015
29 May 2019	New search has been performed	Search updated 11 February 2019.
28 May 2019	New citation required but conclusions have not changed	Updated with inclusion of four new studies. Three ongoing unpublished studies identified.
27 March 2014	Amended	Contact details updated.

Date	Event	Description
21 June 2012	New search has been performed	Search updated; 26 newly identified reports added to studies awaiting classification, including five reports of three ongoing studies (CHORUS #a; Kumar #a; Onda #a).
21 June 2012	New citation required and conclusions have changed	One new trial (Vergote 2010) included. Conclusions changed.

CONTRIBUTIONS OF AUTHORS

- Sarah Coleridge: co-review author, sifted original search results, assessed papers, evaluated included papers, extracted data and co-wrote this review update.
- Andrew Bryant: assisted with data extraction, data analysis and writing of the final version of the review update.
- Sean Kehoe: original idea for review and approved final versions of the protocol, original review and previous updates.
- Jo Morrison: co-review author, wrote protocol, sifted search results, assessed papers, evaluated included papers, extracted data, contributed to analysis and co-wrote the review and its updates.

DECLARATIONS OF INTEREST

Sarah Coleridge: no conflict of interest
Andrew Bryant: no conflict of interest
Sean Kehoe: principle investigator of included study, therefore excluded from title screening, data extraction and all analyses/GRADE decisions
Jo Morrison: no conflict of interest

SOURCES OF SUPPORT

Internal sources

- New Source of support, UK

  The review update was performed without formal internal support.

External sources

- 10/4001/12 NIHR Cochrane Programme Grant Scheme, UK

  A previous up-date of the review received methodological and statistical support as part of the 10/4001/12 NIHR Cochrane Programme Grant Scheme - Optimising care, diagnosis and treatment pathways to ensure cost effectiveness and best practice in gynaecological cancer: improving evidence for the NHS. This most recent updates has been performed without specific funding. No external support was received for this update.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We have updated the methodology of this review to be consistent with the latest Cochrane guidelines, therefore the method of assessing and reporting the risk of bias of included studies has changed from the protocol.

We apply GRADE approach and have added a 'Summary of findings' table, which was not part of Cochrane methodology at the time the original protocol was published.

Although these were not in the original protocol, these were included in the previous update of this review and applied again to this latest update, so were pre-specified prior to this update.

On advice of a reviewer we have added bowel resection and stoma formation to the outcome measures and included these in the Summary of findings 1, as these are important outcomes for women and can have life-long effects. In this update we have also included post-operative death as a specific outcome in the Summary of findings 1, which although it is a grade 5 SAE of surgical morbidity, which was therefore one of the specified outcomes for collection, was not separately reported in previous versions of the review.

INDEX TERMS

Medical Subject Headings (MeSH)

Antineoplastic Agents [*therapeutic use]; Bias; *Carcinoma, Ovarian Epithelial [drug therapy] [mortality] [pathology] [surgery]; Chemotherapy, Adjuvant [methods] [mortality]; Cytoreduction Surgical Procedures [adverse effects] [*methods] [mortality]; Neoadjuvant Therapy [*methods]; *Ovarian Neoplasms [drug therapy] [mortality] [pathology] [surgery]; Postoperative Complications [epidemiology] [etiology]; Preoperative Care; Progression-Free Survival; Randomized Controlled Trials as Topic; Treatment Outcome

MeSH check words

Female; Humans

Appendix 4: Publication 2: Maximal effort debulking (ultraradical or more extensive) surgery versus standard surgery for the primary cytoreduction of advanced epithelial ovarian cancer

Cochrane Library

Cochrane Database of Systematic Reviews

Ultra-radical (extensive) surgery versus standard surgery for the primary cytoreduction of advanced epithelial ovarian cancer (Review)

Hiu S, Bryant A, Gajjar K, Kunonga PT, Naik R

Hiu S, Bryant A, Gajjar K, Kunonga PT, Naik R.
Ultra-radical (extensive) surgery versus standard surgery for the primary cytoreduction of advanced epithelial ovarian cancer.
Cochrane Database of Systematic Reviews 2022, Issue 8. Art. No.: CD007697.
DOI: 10.1002/14651858.CD007697.pub3.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Review]

Ultra-radical (extensive) surgery versus standard surgery for the primary cytoreduction of advanced epithelial ovarian cancer

Shaun Hiu[1a], Andrew Bryant[1a], Ketankumar Gajjar[2], Patience T Kunonga[1], Raj Naik[3]

[1]Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK. [2]Department of Gynaecological Oncology, 1st Floor Maternity Unit, City Hospital Campus, Nottingham, UK. [3]Queen Elizabeth Hospital, Northern Gynaecological Oncology Centre, Gateshead, UK

[a]Joint first author

Contact: Shaun Hiu, shaun.hiu@newcastle.ac.uk.

Editorial group: Cochrane Gynaecological, Neuro-oncology and Orphan Cancer Group.
Publication status and date: New search for studies and content updated (no change to conclusions), published in Issue 8, 2022.

Citation: Hiu S, Bryant A, Gajjar K, Kunonga PT, Naik R. Ultra-radical (extensive) surgery versus standard surgery for the primary cytoreduction of advanced epithelial ovarian cancer. *Cochrane Database of Systematic Reviews* 2022, Issue 8. Art. No.: CD007697. DOI: 10.1002/14651858.CD007697.pub3.

ABSTRACT

Background

Ovarian cancer is the seventh most common cancer among women and the leading cause of death in women with gynaecological malignancies. Opinions differ regarding the role of ultra-radical (extensive) cytoreductive surgery in ovarian cancer treatment.

Objectives

To evaluate the effectiveness and morbidity associated with ultra-radical/extensive surgery in the management of advanced-stage epithelial ovarian cancer.

Search methods

We searched CENTRAL (2021, Issue 11), MEDLINE Ovid and Embase Ovid up to November 2021. We also searched registers of clinical trials, abstracts of scientific meetings, reference lists of included studies and contacted experts in the field.

Selection criteria

Randomised controlled trials (RCTs) or non-randomised studies (NRS), analysed using multivariate methods, that compared ultra-radical/extensive and standard surgery in women with advanced primary epithelial ovarian cancer.

Data collection and analysis

Two review authors independently assessed whether potentially relevant studies met the inclusion criteria, abstracted data and assessed the risk of bias. We identified three NRS and conducted meta-analyses where possible.

Main results

We identified three retrospective observational studies for inclusion in the review. Two studies included women exclusively undergoing upfront primary debulking surgery (PDS) and the other study including both PDS and interval debulking surgical (IDS) procedures. All studies were at critical risk of bias due to retrospective and non-randomised study designs.

Meta-analysis of two studies, assessing 397 participants, found that women who underwent radical procedures, as part of PDS, may have a lower risk of mortality compared to women who underwent standard surgery (adjusted HR 0.60, 95% CI 0.43 to 0.82; I^2 = 0%; very low-certainty evidence), but the evidence is very uncertain. The results were robust to a sensitivity analysis including women with more-

extensive disease (carcinomatosis) (adjusted HR 0.61, 95% CI 0.44 to 0.85; I^2 = 0%; n = 283, very low-certainty evidence), but the evidence is very uncertain.

One study reported a comparison of radical versus standard surgical procedures associated with both PDS and IDS procedures, but a multivariate analysis was only undertaken for disease-free survival (DFS) and therefore the certainty of the evidence was not assessable for overall survival (OS) and remains very low. The lack of reporting of OS meant the study was at high risk of bias for selective reporting of outcomes.

One study, 203 participants, found that women who underwent radical procedures as part of PDS may have a lower risk of disease progression or death compared to women who underwent standard surgery (adjusted HR 0.62, 95% CI 0.42 to 0.92; very low-certainty evidence), but the evidence is very uncertain. The results were robust to a sensitivity analysis in one study including women with carcinomatosis (adjusted HR 0.52, 95% CI 0.33 to 0.82; n = 139; very low-certainty evidence), but the evidence is very uncertain.

A combined analysis in one study found that women who underwent radical procedures (using both PDS and IDS) may have an increased chance of disease progression or death than those who received standard surgery (adjusted HR 1.60, 95% CI 1.11 to 2.31; I^2 = 0%; n = 527; very low-certainty evidence), but the evidence is very uncertain. In absolute and unadjusted terms, the DFS was 19.3 months in the standard surgery group, 15.8 in the PDS group and 15.9 months in the IDS group.

All studies were at critical risk of bias and we only identified very low-certainty evidence for all outcomes reported in the review. Perioperative mortality, adverse events and quality of life (QoL) outcomes were either not reported or inadequately reported in the included studies. Two studies reported perioperative mortality (death within 30 days of surgery), but they did not use any statistical adjustment. In total, there were only four deaths within 30 days of surgery in both studies. All were observed in the standard surgery group, but we did not report a risk ratio (RR) to avoid potentially misleading results with so few deaths and very low-certainty evidence. Similarly, one study reported postoperative morbidity, but the authors did not use any statistical adjustment. Postoperative morbidity occurred more commonly in women who received ultra-radical surgery compared to standard surgery, but the certainty of the evidence was very low.

Authors' conclusions

We found only very low-certainty evidence comparing ultra-radical surgery and standard surgery in women with advanced ovarian cancer. The evidence was limited to retrospective, NRSs and so is at critical risk of bias. The results may suggest that ultra-radical surgery could result in improved OS, but results are based on very few women who were chosen to undergo each intervention, rather than a randomised study and intention-to-treat analysis, and so the evidence is very uncertain. Results for progression/DFS were inconsistent and evidence was sparse. QoL and morbidity was incompletely or not reported in the three included studies.

A separate prognostic review assessing residual disease as a prognostic factor in this area has been addressed elsewhere, which demonstrates the prognostic effect of macroscopic debulking to no macroscopic residual disease.

In order to aid existing guidelines, the role of ultra-radical surgery in the management of advanced-stage ovarian cancer could be addressed through the conduct of a sufficiently powered, RCT comparing ultra-radical and standard surgery, or well-designed NRSs, if this is not possible.

PLAIN LANGUAGE SUMMARY

Ultra-radical (extensive) surgery versus standard surgery to remove tumours in women with advanced ovarian cancer

Review question

What are the benefits and harms of ultra-radical (extensive) versus standard surgery in the management of ovarian cancer?

Background

The ovaries are small glands found on either side of the womb that produce and store eggs, and make hormones that control the menstrual cycle (periods). Ovarian cancer is the most common cause of death in women with a cancer of the reproductive system. Opinions differ about whether women with advanced ovarian cancer have better outcomes if they have 'ultra-radical' surgery, which is much more extensive than standard surgery, to remove tumours. Standard surgery in an advanced disease setting still has an element of radicality and comprises as a minimum many of the surgical procedures involved in more radical surgery. Ultra-radical (extensive) surgery is an extension of standard surgery and may include at least one additional extensive surgical procedure.

Review methods

We searched the scientific literature for studies comparing ultra-radical and standard surgery for women with advanced ovarian cancer. We looked for randomised controlled trials, which are regarded as the best type of study, and for non-randomised studies that were analysed using methods that allow for differences between the groups of women receiving different types of surgery.

Key results

We identified three non-randomised studies. The evidence is very limited and uncertain for all results since women were chosen to undergo each type of treatment, rather than randomly allocated, so there is a very high (critical) risk of bias in these types of studies.

In two studies (397 women), women who had radical surgery to remove the tumour may have 18% to 57% less chance of death compared to women who had standard surgery. The results were similar for women with more-extensive disease. There were very few deaths within 30 days of surgery. There may be less chance of disease progression with radical surgery.

One study compared radical versus standard surgery associated with both upfront primary (tumour removed before starting chemotherapy) and interval debulking (tumour removed between chemotherapy sessions) surgery on death, but the comparison was not fair and there was high risk of bias for reporting of outcomes.

One study (203 women) found that women who had radical procedures as part of upfront primary debulking surgery may have 8% to 58% less chance of disease progression or death compared to women who had standard surgery. The results were similar when including only the 139 women with more-extensive disease (where risk was 18% to 67% lower).

One analysis (527 women) merging radical surgery groups in one study found that women who underwent ultra-radical procedures (using both upfront primary and interval debulking surgical procedures) may be associated with 11% to 60% increased chance of disease progression or death than those who received standard surgery.

All studies were at very high (critical) risk of bias and we were very unsure about the evidence. We included relatively few women due to our stringent inclusion criteria. Studies either did not report or inadequately reported death, side effects or quality of life.

Main conclusions and certainty in the evidence

Although some of these results may suggest that survival may be better in women receiving upfront primary ultra-radical surgery rather than standard surgery, extreme caution is required with interpretation, as the studies were not well designed or analysed, and thus the effects could even be in the opposite direction.

We are unable to reach any definite conclusions about the relative benefits and harms of the two types of surgery. Better designed, large studies are needed.

Summary of findings 1. Ultra-radical (extensive) surgery compared to standard surgery in women with stage IIIc or IV ovarian cancer

Ultra-radical (extensive) surgery compared to standard (radical) surgery in women with stage IIIc or IV ovarian cancer

Patient or population: women with stage IIIc or IV ovarian cancer
Setting: –
Intervention: ultra-radical (extensive) surgery
Comparison: standard surgery

Outcomes	Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What does this mean?	Comments
Survival (overall and disease- specific) Follow-up: median 43–49 months Overall survival was listed as the desired primary outcome in the protocol and we note the potential issues of reporting disease-specific survival.	**HR 0.60** (0.43 to 0.82)	397 (2 studies)	⊕⊖⊖⊖ **Very low**[a,b,c]	Survival may be prolonged in woman who received ultra-radical surgery compared to standard surgery but the evidence was limited and very uncertain. More studies are needed.	We could not present illustrative absolute effects because a representative control group risk could not be ascertained from the studies. The HR estimates were adjusted for in multivariable analyses and this cannot be done in absolute terms so we made no attempt as numbers were likely to mislead. 1 study reported 5-year disease-specific survival rather than categorising deaths by any cause. We made an assumption that most women with advanced-stage disease would die of the disease rather than other comorbidities. The results were robust to a sensitivity analysis that included 2 studies assessing 283 women with more-extensive disease (carcinomatosis) (adjusted HR 0.61, 95% CI 0.44 to 0.85; I^2 = 0%; very low-certainty evidence).
Progression-free survival Follow-up: median 43	**HR 0.62** (0.42 to 0.92)	203 (1 observational study)	⊕⊖⊖⊖ **Very low**[a,b,c]	Disease progression may be delayed in woman who received ultra-radical surgery compared to standard surgery but the evidence was limited and very uncertain. More studies are needed.	Participants received upfront debulking surgery. The results were robust to a sensitivity analysis assessing a subset of 139 women with carcinomatosis (adjusted HR 0.52, 95% CI 0.33 to 0.82; very low-certainty evidence).

Disease-free survival Follow-up: median 49 months	**HR 1.60** (1.11 to 2.31)	527 (2 analyses from 1 observational study)	⊕⊖⊖⊖ **Very low**[a,b,c]	Disease may relapse earlier in woman who received ultra-radical surgery compared to standard surgery but the evidence was limited and very uncertain. More studies are needed.	Participants received upfront and interval debulking surgical procedures.
Rate of optimal cytoreduction	Although a secondary outcome, 'Optimal cytoreduction' was not reported in any multivariate analyses in any of the studies. We did not present any unadjusted results for this as it is likely that 'optimal cytoreduction' will be higher in ultraradical surgery and would not be a fair comparison.				
Recurrence rate	Not reported				
(Loco)regional control	Not reported				
Adverse event: **perioperative mortality** Follow-up: median 43–49 months	In total there were only 4 deaths within 30 days of surgery in both studies and none in the ultra-radical group. We did not report a RR as to not provide potentially misleading results with so few deaths.	397 (2 observational studies)	⊕⊖⊖⊖ **Very low**[a,b,c]	In total there were only 4 deaths within 30 days of surgery in both studies and none in the ultra-radical group. However, the evidence is limited and very uncertain and more studies are needed.	None of the studies reporting this serious adverse event used any statistical adjustment. Upfront debulking surgery In 1 study, there were 0 reported cases of perioperative mortality within 30 days in the ultra-radical surgery group versus 3 women died in the standard surgical group. In another study, perioperative death within 30 days occurred in 0/119 (0%) in the surgery group versus 1/84 (1.2%) in the standard group.
Adverse event: **serious postoperative morbidity** Follow-up: median 43	**RR 3.24** (1.84 to 5.68)	203 (1 observational study)	⊕⊖⊖⊖ **Very low**[a,b,c]	Significant postoperative morbidity occurred in 32/84 (38.1%) women in the in ultra-radical group versus 14/119 (11.8%) women in the standard surgery group. However, the evidence is limited and very uncertain and more studies are needed.	This study did not use any statistical adjustment for this adverse event. Women received upfront debulking surgery.
Quality of Life	Not reported				

CI: confidence interval; **HR:** hazard ratio; **RR:** risk ratio.

GRADE Working Group grades of evidence
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

---

[a]Downgraded one level for sparse data.
[b]Downgraded one level for high risk of bias concerns.
[c]Downgraded one level as outcomes were incompletely or inadequately (or both) reported.

BACKGROUND

Description of the condition

Ovarian cancer is the seventh most common cancer among women and the leading cause of death in women with gynaecological malignancies. Globally, there are over 200,000 new cases per year, with approximately 6.1 new cases per 100,000 women per year. A woman's cumulative risk of developing ovarian cancer by the age of 75 years is 0.7%: 0.5% in low-income countries and 1.0% in low- to middle-income countries (GLOBOCAN 2018). It is less common in women under the age of 40 years, and the incidence increases with age. In Europe, approximately 38% of women with ovarian cancer are alive five years after diagnosis (EUROCARE 2015), largely because the early stages of the disease often present with very few, if any, specific symptoms so most women present with advanced-stage disease (Bast 2020; Kirby 2020; Kurman 2008; Lancet 2007; Siegel 2020; Visintin 2008; Webb 2017). Symptoms include: abdominal distension, bloating, indigestion, urinary frequency, urinary urgency, early satiety, weight loss, reduced appetite, abdominal and pelvic pain, and, less commonly, vaginal bleeding (Rani 2018).

Cancers of the ovary are classified according to their cells of origin. Most ovarian cancers originate from the surface (epithelial) cells of the ovary/fallopian tubes and are termed epithelial tumours, although some cancers can also arise from the substance of the ovary, called stromal tumours, or from embryological differentiation (sex cord and germ cell tumours) (American Cancer Society 2020; CRUK 2018; Kurman 2014). The staging of ovarian cancer is based on the International Federation of Gynaecology and Obstetrics (FIGO) classification system (Berek 2018; PDQ Adult Treatment Editorial Board 2021; Prat 2014)). FIGO staging depends on the findings at the time of surgery. Stages I and II constitute early-stage disease, where stage I is limited to the ovaries and stage II tumours extend to the pelvis. Stages III and IV constitute advanced disease. In stage III, the tumour extends outside the pelvis, or involves lymph nodes within the pelvis, and stage IV is where the tumour has spread to distant sites such as the liver, lungs and lymph nodes in the neck (Berek 2018).

Description of the intervention

Treatment for women with epithelial ovarian cancer (EOC) is a combination of surgery and platinum- and taxane-based chemotherapy. Prognosis depends not only on the stage and histological type of the tumour, but also how much disease is left behind (residual disease) following surgery. Studies have shown that residual disease after initial surgery is a strong independent prognostic factor for survival, with improvements in both overall and progression-free survival (PFS) being greatest in women with no visible disease, also known as no macroscopic residual disease (NMRD) or minimal (less than 1 cm, currently termed near-optimal cytoreduction) visible residual disease at the end of surgery (Bryant 2021). Women who undergo more-extensive surgery may be more likely to have tumour deposits of 2 cm or less at the end of surgery (Bristow 2002; Crawford 2005; Horowitz 2015). Survival for women who have residual tumour deposits of more than 2 cm or up to 2 cm at the end of the surgery appears to be similar, further suggesting that optimal cytoreduction is associated with improved survival rates (Bristow 2002; Bryant 2021). However, the extent of surgical resection required to achieve optimal cytoreduction remains controversial. There appears to be a universally diverse practice with huge variations in achieving the NMRD rate of between 22% and 98% (Bryant 2021).

Although there is a lack of evidence demonstrating a benefit from performing a hysterectomy at the time of debulking surgery, this is accepted practice as it aids the diagnosis of a primary tumour site, for example, serous papillary cancers and carcinosarcomas may originate from both the uterus and ovaries. It also helps in excluding synchronous primary uterine tumours. While systematic lymphadenectomy of non-bulky nodes has been shown to worsen outcomes (Harter 2019), removing the uterus and cervix, both tubes and ovaries, the omentum and enlarged lymph nodes is part of standard surgery (Aletti 2006a; Norell 2020; PDQ Adult Treatment Editorial Board 2021; Todo 2003; Vergote 2016).

There has been a shift in recent years in some centres to attempts at achieving complete cytoreduction with use of more-extensive and radical procedures in performing cytoreductive surgery (Phillips 2019). To achieve NMRD, surgeons often have to perform radical and ultra-radical procedures with associated significant postoperative morbidity and mortality. There were Grade 3 and 4 complications in 19% of women after debulking surgery for advanced ovarian cancer (Benedetti Panici 2015). In one meta-analysis, there were no important differences in the quality of life (QoL) of women in three randomised controlled trials (RCTs) comparing primary surgery with improvements over baseline at six and 12 months. However, there was insufficient evidence on QoL outcomes of women undergoing extensive or ultra-radical surgery compared with those undergoing less-extensive surgery (Kumar 2019). However, the results of the SOCQER-2 (Surgery in Ovarian Cancer – Quality of Life Evaluation Research – 2) cohort study showed the global QoL of women undergoing low-, intermediate- and high-complex surgery (based on a surgical complexity score) improved at 12 months after surgery and was no worse in women undergoing extensive surgery (Sundar 2021). This is an interesting result, as, if there are no significant differences in QoL and general morbidity after more-extensive surgery, then centres may be more inclined to perform more-aggressive surgery more often. Postoperative mortality within 28 days following debulking surgery for ovarian cancer was reported in 2.5% of cases who underwent primary debulking surgery in the EORTC (European Organisation for Research and Treatment of Cancer) 55971 and 6% of cases in CHORUS (Chemotherapy Or Upfront Surgery) trials (Kehoe 2015; Vergote 2010).

Women with widespread disease, which involves the upper abdomen, affecting the diaphragm, liver, spleen and omentum, or widespread disease affecting the bowel, will need much more radical surgery in order to achieve NMRD or optimal cytoreduction. The complexity of the procedures required to achieve these outcomes undoubtedly increases. Radical surgery including bowel resection, splenectomy, liver resection and diaphragmatic stripping has been described in the literature as treatment for advanced ovarian cancer with low complication rates (Bristow 2003; Eisenkop 2001; Jaeger 2001; Merideth 2003; Montz 1989; Norell 2020; Pomel 2004; Vergote 2016). NICE (National Institute for Health and Care Excellence) has previously published guidance on ultra-radical (extensive) surgery for advanced ovarian cancer (NICE 2013). Standard surgery in an advanced disease setting still has an element of radicality and comprises as a minimum, total hysterectomy, bilateral adnexectomy with excision of the pelvic peritoneum, total omentectomy including the

supracolic omentum, removal of bulky pelvic and lumbo-aortic nodes, simple peritonectomies, localised colonic resection, or a combination of these. Procedures such as appendicectomy may have previously been considered part of standard surgery, but evidence now suggests that this could be unnecessary and may cause harm. Ultra-radical (extensive) surgery is an extension of standard surgery including at least one of the following: stripping of the peritoneum over the diaphragm, extensive stripping of the peritoneum, multiple resections of the bowel (excluding localised colonic resection), liver resection, partial gastrectomy, cholecystectomy and splenectomy (with or without resection of the tail of the pancreas) (NICE 2013).

How the intervention might work

It has been proposed that multiple factors, including tumour biology, determine the manner of disease progression, which in turn influences the likelihood of surgical cytoreduction (Colombo 2019; Eisenkop 2001; Hoskins 1992; Markman 2007). Supporters of less-radical surgery argue that the initial extent of advanced disease reflects the aggressiveness of the tumour, and ultimately dictates treatment success. Therefore, when radical surgery becomes necessary to achieve optimal cytoreduction, it may not improve survival, despite leaving minimal residual disease (Colombo 2019; Covens 2000). Furthermore, the role of surgery has been questioned because patients who undergo surgery to achieve NMRD often represent women who may be younger and fitter, and have relatively small preoperative tumour loads and, therefore, less biologically aggressive tumours, and that differences in tumour biology account for the survival benefits that are reported to be from surgery (Eisenkop 1998; Hoskins 1992; Norell 2020; Vergote 2016). Perhaps of greater concern is the patient morbidity that is incurred during such radical procedures, both in the perioperative and postoperative periods (Chen 1985; Sundar 2021; van Dam 1996; Venesmaa 1992).

Ultra-radical surgery is associated with a prolonged operating time and exposure to anaesthesia. This may increase the risk of hypothermia; respiratory complications such as atelectasis (lung collapse), infection, adult respiratory distress syndrome; blood loss; and intraoperative ureteric, bowel and bladder injury. In the postoperative period, these women may require a longer hospital stay and recovery time, with an increased risk of infection (chest, wound, urine), venous thromboembolic disease, poorer mobility and poorer nutritional status. The cost-effectiveness of such surgery would also require evaluation.

There is also a suggestion from one before-after study that a structured shift to an ultra-radical upfront primary surgical approach may not improve survival in surgically treated women (Falconer 2020). In this population-based cohort study, women with suspected advanced EOC near Stockholm in Sweden were included via the Swedish Quality Registry for Gynecologic Cancer (SQRGC) and the National Cancer Registry (NCR). Women were selected in two sets of three-year cohorts, based on the year of their diagnosis (a before cohort or an after cohort 2 change in surgical treatment algorithm) and were followed for at least three years. Five-year overall survival (OS) in non-surgically and surgically treated women was analysed. After a median follow-up of around 28 months in 752 women, the complete resection rate increased from 37% to 67% as well as proportion of non-surgically treated women (from 24% to 33%). This study also demonstrated that a shift to ultra-radical surgery increased the proportion of non-surgically treated women. However, this study was not a 'controlled' before-after study and as a consequence was prone to bias. The use of historical controls are known to overestimate the benefit of new treatments. Before-after studies also have a high risk of bias because there may be unidentified differences between the intervention and control groups that may affect changes in the outcome measure (Sterne 2022).

Why it is important to do this review

To our knowledge, there have been no comprehensive and rigorous systematic reviews on ultra-radical (extensive) surgery versus standard surgery. There is no consensus in clinical guidelines, and there is widespread variation in surgical practice globally with varying rates of survival (Norell 2020). Willingness to undertake more-extensive surgery was correlated with three-year survival by distant stage, and clinicians from higher performing countries appeared to be more likely than those from lower performing countries to be proponents of 'ultra-radical' surgery (Norell 2020). Guidelines from Belgium in 2016 supported the use of radical surgical techniques to obtain resection of all macroscopic tumour (Vergote 2016).

Given the differences in opinion regarding the role of extensive debulking surgery in ovarian cancer treatment, we aimed to systematically review the available evidence for ultra-radical surgery in ovarian cancer management.

OBJECTIVES

To evaluate the effectiveness and morbidity associated with ultra-radical/extensive surgery in the management of advanced-stage epithelial ovarian cancer.

METHODS

Criteria for considering studies for this review

Types of studies

- Randomised controlled trials (RCTs)

As we expected to find few, if any, RCTs of surgical interventions (Johnson 2008), we included the following types of non-randomised studies with concurrent comparison groups.

- Quasi-randomised trials, non-randomised studies, prospective and retrospective cohort studies, and case series of 100 or more participants

We excluded case-control studies, uncontrolled observational studies and case series of fewer than 100 participants.

In order to minimise selection bias, we decided to include only studies that used statistical adjustment for baseline case mix (e.g. age, performance status, grade, etc.) using multivariate analyses.

Types of participants

Women diagnosed with stage III and IV EOC. Women having ultra-radical surgery as part of upfront primary debulking surgery (PDS) or interval debulking surgery (IDS; surgery halfway through the course of chemotherapy) were included.

There is evidence that a high percentage of so-called 'ovarian' high-grade serous carcinomas arise in the fimbrial end of the

fallopian tube. Serous tubal intraepithelial carcinoma is considered a precursor lesion (Harley 2014).

Women with other concurrent malignancies women with recurrent disease were excluded.

Types of interventions

- Intervention: ultra-radical surgery defined as total abdominal hysterectomy, bilateral salpingo-oophorectomy, omentectomy, removal of enlarged lymph nodes (para-aortic, pelvic, obturator) and one or more of the following: upper abdominal surgery (splenectomy, diaphragmatic or peritoneal stripping, liver resection), bowel surgery or stoma formation (excluding localised colonic resection) or urinary tract surgery, peritonectomy (en bloc or excision of nodules, depending on disease involvement).
- Comparison: standard surgery defined as total abdominal hysterectomy, bilateral salpingo-oophorectomy, omentectomy either with or without removal of enlarged lymph nodes (para-aortic, pelvic, obturator), localised colonic resection and debulking of any other superficial tumour plaques.

The types of interventions defined above have been widely described in the literature and in the published NICE guidance on ultra-radical (extensive) surgery for advanced ovarian cancer (NICE 2013).

Two included studies in the review also included some elements of extensive surgery in the standard surgery group: segmental small bowel resection (Chang 2012a), and rectosigmoid resection and appendectomy (Luyckx 2012). It was decided to include these two studies, because these surgical additions are common practice in some countries such as Belgium (Vergote 2016).

Types of outcome measures

Primary outcomes

- Overall survival (OS): survival until death from all causes. Survival was assessed from the time when women were enrolled in the study. One study reported disease-specific survival rather than death from any cause (Aletti 2006a). We additionally added this as an outcome in the review.

Secondary outcomes

- Progression-free survival (PFS).
- Disease-free survival (DFS).
- Optimal cytoreduction, defined as residual tumour less than 1 cm, or complete cytoreduction.
- Death within 30 days of intervention.
- Adverse events classified according to CTCAE 2017:
  - direct surgical morbidity: for example, vascular injury, injury to bladder, ureter, small bowel or colon, presence and complications of adhesions, febrile morbidity, intestinal obstruction, anastomotic leak, haematoma, collection, local infection.
  - surgically related systemic morbidity, for example, chest/wound/urine infection, thromboembolic events (deep vein thrombosis and pulmonary embolism), cardiac events (cardiac ischaemia, myocardial infarction and cardiac failure), cerebrovascular accident, transfusion reaction, pulmonary oedema;
  - recovery: delayed discharge, unscheduled re-admission.
- Quality of life (QoL) measured using a scale that has been validated through reporting of norms in a peer-reviewed publication.

Search methods for identification of studies

We sought papers in all languages and carried out translations when necessary.

Electronic searches

See: Cochrane Gynaecological, Neuro-oncology and Orphan Cancer Group methods used in reviews.

For this review update, we searched the following electronic databases on 25 November 2021:

- the Cochrane Central Register of Controlled Trials (CENTRAL; 2021, Issue 11), in the Cochrane Library;
- MEDLINE via Ovid (November 2010 to November week 3 2021);
- Embase via Ovid (November 2010 to 2021 week 46).

The CENTRAL, MEDLINE and Embase search strategies are presented in Appendix 1, Appendix 2, and Appendix 3.

All relevant articles found were identified on PubMed and using the 'related articles' feature, a further search was carried out for newly published articles.

Searching other resources

Unpublished and grey literature

We searched metaRegister, Physicians Data Query, the ISRCTN Registry (www.controlled-trials.com/rct), ClinicalTrials.gov (www.clinicaltrials.gov), and the National Cancer Institute Register (www.cancer.gov/clinicaltrials) for ongoing trials. We used search terms derived from the main searches.

Reference lists

We searched reference lists of all included studies for additional studies.

Handsearching

We handsearched abstracts of meetings from the International Gynaecological Cancer Society (2000 to 2020), the British Gynaecological Cancer Society (2008 to 2021), European Society of Gynaecological Oncology (2003, 2005, 2009, 2015 and 2019) and the Society of Gynecologic Oncology (2009, 2010, 2015 and 2019) to identify unpublished studies.

Data collection and analysis

Selection of studies

We downloaded all titles and abstracts retrieved by electronic searching to the reference management database Endnote and removed duplicates. Three review authors (AB, PB, SH) independently examined the remaining references. We excluded those studies that clearly did not meet the inclusion criteria and obtained copies of the full text of potentially relevant references. Three review authors (AB, PB, SH) independently assessed the eligibility of retrieved papers. We resolved disagreements by discussion between the three review authors

and, when necessary, with fourth and fifth review authors (RN, KG). We documented reasons for exclusion in the Characteristics of excluded studies table.

Data extraction and management

For included studies, we recorded the following data.

- Author, year of publication and journal citation (including language).
- Country.
- Setting.
- Inclusion and exclusion criteria.
- Study design, methodology.
- Study population, abstracted by treatment arm if possible:
  - total number enrolled;
  - participant characteristics;
  - age;
  - ethnicity;
  - comorbidities;
  - response to neoadjuvant chemotherapy.
- Ovarian cancer details at diagnosis:
  - FIGO stage (III or IV);
  - histological cell type;
  - differentiation.
- Previous treatment (neoadjuvant chemotherapy subgroup analysis: responders versus non-responders).
- Surgical details:
  - type of surgeon (gynae-oncologist, gynaecologist, general surgeon);
  - type of surgery (ultra-radical (extensive) versus standard).
- Risk of bias in study (see below).
- Duration of follow-up.
- Outcomes (see above) – OS, PFS, QoL and adverse events.
  - For each outcome:
    - outcome definition (with diagnostic criteria if relevant);
    - unit of measurement (if relevant);
    - for scales: upper and lower limits, and whether high or low score is good.
  - For results: number of participants allocated to each intervention group.
  - For each outcome of interest: sample size; missing participants.

We extracted data on outcomes as follows.

- For time to event data (OS), we extracted the log of the hazard ratio (HR) and its standard error from trial reports; if these were not reported, we attempted to estimate the log (HR) and its standard error using the methods of Parmar 1998.

We reported the HR and its 95% confidence interval (CI). For adjusted statistics, we noted the variables used in adjustment. Where possible, all data extracted were those relevant to an intention-to-treat analysis, in which participants were analysed in groups to which they were assigned. We noted the time points at which outcomes were collected and reported.

Two review authors (PB, SH) independently extracted data onto a data abstraction form specially designed for the review. We resolved differences between review authors by discussion or appeal to a third review author (RN), when necessary.

Assessment of risk of bias in included studies

We used the Risk of Bias In Non-randomised Studies of Interventions (ROBINS-I) tool to assess bias in our included studies (Sterne 2016). According to the ROBINS-I, non-randomised studies of interventions (NRSI) aim to mimic a target trial (i.e. a hypothetical pragmatic RCT), which may not be feasible or ethical to conduct. Bias in this sense is defined as "systematic difference between the results of the NRSI and the results expected from the target trial". Given that the NRSIs included in our study were concerned with the effect of having undergone ultra-radical or standard surgery and not the effect of having been assigned to a surgery type, we may further specify that bias is the systematic difference between the results of NRSIs and the per-protocol effect of a target trial.

The ROBINS-I rates bias along seven domains (see Appendix 4):

- confounding;
- selection of participants into the study;
- classification of interventions;
- deviation from intended interventions;
- missing data;
- measurement of outcomes; and
- selection of reported result.

Responses to signalling questions lead to the formulation of domain-specific risk of bias ratings – no information, low, moderate, serious and critical risk of bias – which then guide the judgement for an overall risk of bias rating. We also added additional signalling questions to the ones in ROBINS-I domains in accordance with additional criteria for confounding and selection of women so that we were confident in our judgements (Taggart 2001). These additional criteria for confounding included an assessment of the comparability of treatment groups to see if there were no differences between the two groups or that differences had been controlled for, in particular with reference to age, FIGO stage, histological cell type, differentiation, previous treatment (neoadjuvant chemotherapy – responders versus non-responders) and type of surgeon (gynae-oncologist, gynaecologist, general surgeon). At least three of these characteristics were reported and any reported differences were controlled for. To aid signalling questions in selection of women into the study, we assessed whether relevant details of criteria for assignment of women to treatments was provided and whether the group of women who received each intervention were representative and were not selected by a subset of the population. If these additional signalling questions were questionable in any way, then the risk of bias judgement in that domain would be of serious or critical concern, which is above a 'high' risk of bias judgement.

Three review authors (AB, PB, SH) independently applied the risk of bias tool and resolved differences by discussion or by appeal to a fourth review author (RN, KG). We tabulated results and presented them in a risk of bias graph.

Measures of treatment effect

We used the following measures of the effect of treatment.

- For time to event data, we used the HR with 95% CI.

Unit of analysis issues

We did not expect or encounter any unit of analysis issues.

Dealing with missing data

We did not impute missing outcome data for any outcomes. For the primary outcome, if data were missing or only imputed outcome data were reported, we contacted study authors to request data on the outcomes among participants who were assessed.

Assessment of heterogeneity

We assessed heterogeneity between studies by visual inspection of forest plots, by estimation of the percentage heterogeneity between trials that could not be ascribed to sampling variation (Higgins 2003), by a formal statistical test of the significance of the heterogeneity (Deeks 2001), and, where possible, by subgroup analysis (see below). If there was evidence of substantial heterogeneity, we investigated and reported the possible reasons for this.

Assessment of reporting biases

We did not examine funnel plots corresponding to meta-analysis of the primary outcome to assess the potential for small-study effects due to an insufficient number of included studies. If there had been evidence of small-study effects, we would have considered publication bias as only one of a number of possible explanations. If these plots had suggested that treatment effects may not have been sampled from a symmetric distribution, as assumed by the random-effects model, we would have performed a sensitivity analysis using the fixed-effect model.

Data synthesis

If there were sufficient clinically similar studies available, we pooled their results in a meta-analysis. We used adjusted summary statistics as specified in Types of studies.

- For time-to-event data, we pooled HRs using the generic inverse variance facility of Review Manager 5 (Review Manager 2014).

We used random-effects models with inverse variance weighting for all meta-analyses (DerSimonian 1986).

Subgroup analysis and investigation of heterogeneity

We performed subgroup analysis, grouping the studies by:

- reporting of survival (overall and disease-specific; progression and disease-free);
- radicality of procedures in the ultra-radical groups.

We considered factors such as age, FIGO stage, type of surgery (upfront primary debulking surgery (PDS) or IDS), type of surgeon and length of follow-up in interpretation of any heterogeneity.

Sensitivity analysis

We planned to perform a sensitivity analysis excluding studies at high risk of bias, but all three studies were at a high risk of bias. However, we did perform sensitivity analyses including only women with more-extensive disease (with carcinomatosis).

Summary of findings and assessment of the certainty of the evidence

We presented the overall certainty of the evidence for each outcome according to the GRADE approach, which takes into account issues related to internal validity (risk of bias, inconsistency, imprecision, publication bias) and to external validity such as directness of results (see Summary of findings 1 based on the methods described the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2019). We downgraded the evidence from 'high' certainty by one level for serious (or by two for very serious) concerns for each limitation.

- High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
- Moderate certainty: we are moderately confident in the effect estimate. The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
- Low certainty: our confidence in the effect estimate is limited. The true effect may be substantially different from the estimate of the effect.
- Very low certainty: we have very little confidence in the effect estimate. The true effect is likely to be substantially different from the estimate of effect.

RESULTS

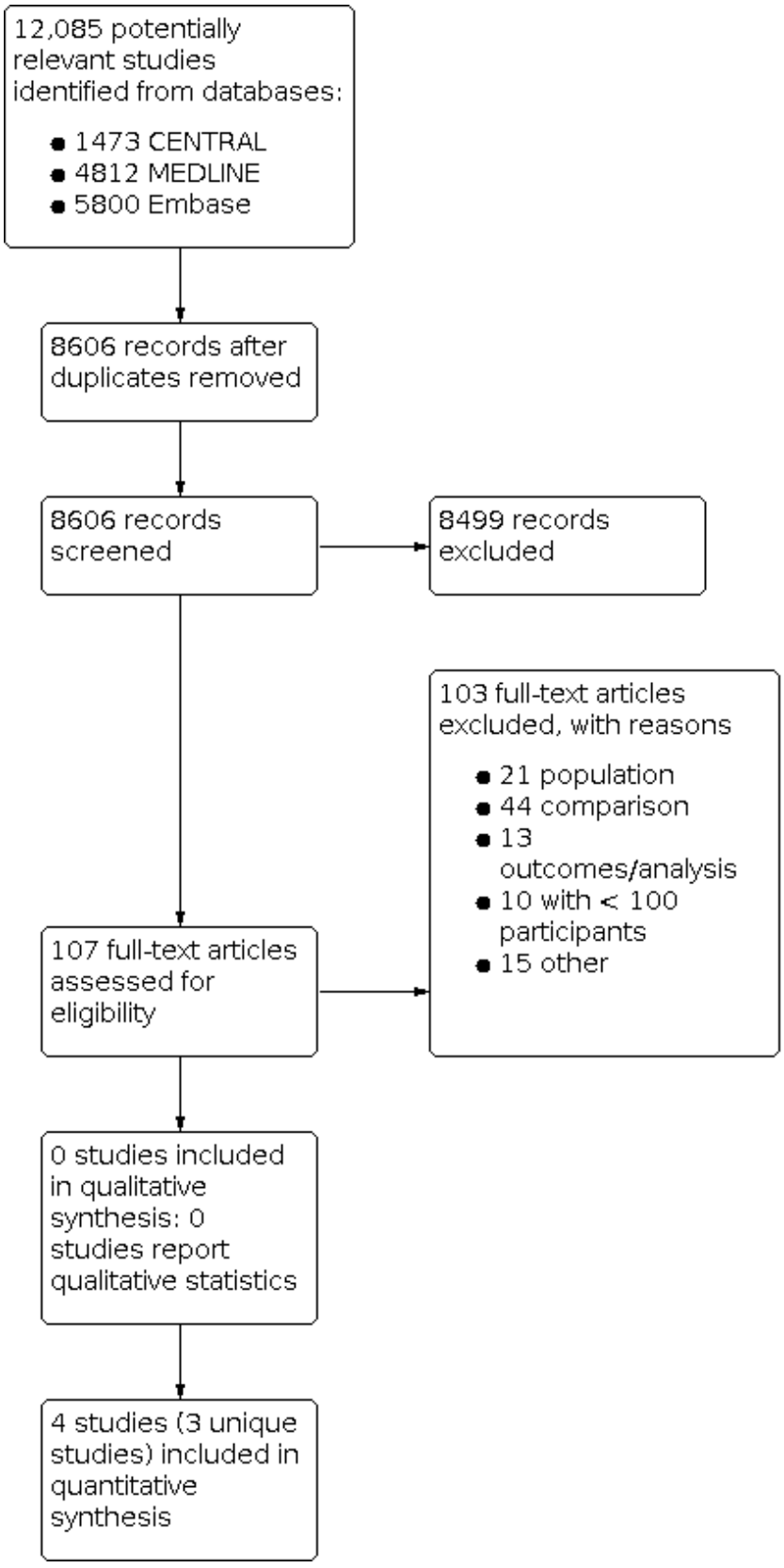
Description of studies

Results of the search

When the search results were merged into EndNote and duplicates were removed, there were 8606 unique references. The title and abstract screening identified 107 references as potentially eligible. The full-text screening excluded 103 of these, with the reasons for exclusion presented in the Characteristics of excluded studies table. We included four references reporting on three studies that met our inclusion criteria (Aletti 2006a; Chang 2012a; Luyckx 2012).

The PRISMA flow diagram of the search results is presented in Figure 1.

Figure 1. Study flow diagram up to 25 November 2021.



Searches of the grey literature did not identify any additional relevant studies.

The three included studies are described in the Characteristics of included studies table (Aletti 2006a; Chang 2012a; Luyckx 2012). While it did not meet the inclusion criteria, we also identified a study worthy of discussion that measured the introduction of ultra-radical surgery on a population level (Falconer 2020). The analysis had a before-after design but was excluded as it was not a controlled study. Details of this study are given in Agreements and disagreements with other studies or reviews.

Included studies

All three included studies (924 women) compared ultra-radical or extensive surgery with standard surgery. The two most recent studies also included some elements of extensive surgery in the standard surgery group: segmental small bowel resection (Chang 2012a), and rectosigmoid resection and appendectomy (Luyckx 2012). We decided to include these two studies, because these surgical additions are common practice in some countries such as Belgium (Vergote 2016).

All studies enrolled women who underwent primary surgery and adjusted their analyses in attempts to reduce selection bias in assignment of participants to surgical treatment. All studies were considered to have a high risk of bias. Despite each study reporting multivariate analyses, confounding by indication could not be excluded. In addition, in all cases, the adjusted HRs were derived from prognostic models, which seem to have been assessed based on significance testing and not on the inclusion of putative confounders in the analysis, irrespective of statistical significance.

Design

All three studies reported retrospective analyses of participants identified from surgical or medical records (Aletti 2006a; Chang 2012a; Luyckx 2012). Aletti 2006a reported a retrospective analysis of 194 women from the Mayo clinic in Minnesota (USA), Chang 2012a (203 women) was set in South Korea (Ajou University Hospital, Republic of Korea) and Luyckx 2012 (527 women) analysed data from seven gynaecological oncology centres in France.

Participants

The median age at diagnosis of advanced EOC ranged from 54 years in Chang 2012a to 64 years in Aletti 2006a (ages across studies ranged from 24 to 90 years). About 65% to 82% of participants had a serous histological tumour cell type. Most participants had Grade IIIC tumour (84% to 100%) and 93% of women had tumour Grade III in Aletti 2006a whereas the proportion with Grade III in the other two studies was lower (49 to 58% based on non-missing observations), with over a third having tumour Grade II in these studies. Ascites varied across studies with Aletti 2006a reporting mean ascites of 2076 mL and median of 1000 mL (range 0 mL to 12,000 mL). In Chang 2012a, 45% of women had ascites greater than 1000 mL, which was in contrast to Luyckx 2012 where median ascites was 50 mL (range 0 mL to 8000 mL). In terms of residual disease after primary surgery, Luyckx 2012 had most favourable outcome with 71% being cytoreduced to microscopic disease and 18.5% of remaining women having optimal cytoreduction (residual disease less than 1 cm). The other two studies were fairly similar with around two-thirds of women being optimally or completely cytoreduced with the remaining third or so having residual disease greater than 1 cm. Two studies reported American Society of Anesthesiologists (ASA) score at baseline: Aletti 2006a (48% had ASA scores 1 to 2 and 49% had ASA scores 3 to 4 with remaining scores unknown) and Chang 2012a (56% had ASA scores 1 to 2 and 39% had ASA scores 3 to 4, with remaining scores unknown). Two studies also reported extent of disease: 144 (74%) women had carcinomatosis in Aletti 2006a and 149 (73%) had carcinomatosis in Chang 2012a. Luyckx 2012 reported the extent of peritoneal carcinomatosis with a median peritoneal cancer index of 10. Approximately 39% had no upper abdominal lesions, 40% had abdominal lesions of 2.5 cm or less, and 21% had upper abdominal lesions greater than 2.5 cm.

Interventions

All three studies compared ultra-radical or extensive surgery with standard surgery. However, the two more recent studies additionally included some elements of extensive surgery in the standard surgery group: segmental small bowel resection (Chang 2012a), and rectosigmoid resection and appendectomy (Luyckx 2012). Aletti 2006a and Chang 2012a included only surgery from upfront primary debulking surgery whereas Luyckx 2012 included a mixture of PDS and IDS.

Aletti 2006a performed initial surgery for diagnosis, staging and surgical cytoreduction. Ultra-radical surgery was defined as having any diaphragmatic surgery, bowel resection, splenectomy or extensive abdominal peritoneal stripping or resection and was compared to standard surgery defined as hysterectomy, complete omentectomy, stripping of pelvic peritoneum or limited resection of peritoneal-based nodules. Participants were first classified by the extent of peritoneal dissemination. Those with tumour nodules diffusely covering most of the bowel serosal surfaces and the parietal peritoneum of the abdomen and pelvis were classified as having carcinomatosis. The centre's division of gynaecological surgery contained a mixed group of surgeons, some being more likely to carry out ultra-radical surgery but all sharing a uniform referral base with similar patient demographics, practising at a single institution where each surgeon had access to identical services and nursing support. The mean length of follow-up was 3.5 years and median was 2.7 years (range 0.02 to 10.5 years). For the overall cohort of 194 women, 83 (42%) received ultra-radical surgery and 111 (57%) received standard surgery. For the subset of 144 women with worse disease (carcinomatosis), 68 (47%) underwent ultra-radical surgery and 76 (53%) received standard surgery.

Radical cytoreductive procedures in Chang 2012a included radical oophorectomy with or without rectosigmoid colectomy, total omentectomy, multiple bowel resections, diaphragm peritonectomy or resection, liver resection, splenectomy, distal pancreatectomy and gastric resection. Simple surgery included total abdominal hysterectomy, bilateral salpingo-oophorectomy, peritoneal biopsies or excisions, infracolic omentectomy, pelvic lymphadenectomy, para-aortic lymphadenectomy and segmental resection of small bowel. After surgery, all participants received adjuvant platinum-based chemotherapy in combination with paclitaxel for six to nine cycles. The median length of follow-up was 43 months (range 1 to 124 months).

Luyckx 2012 defined ultra-radical surgery as involving standard surgery plus upper abdominal surgery such as stripping of the

diaphragmatic peritoneum and splenectomy alone (group 2A in the study), or a combination of digestive tract resections (right colon and caecum, total colectomy and others), organ resection (spleen, gallbladder, partial gastrectomy and others), coeliac lymph node dissection, and total abdominal peritoneum stripping in addition to standard surgery (group 2B in the study). The comparison group was standard surgery with hysterectomy, bilateral salpingo-oophorectomy, rectosigmoid resection, infragastric omentectomy, pelvic and aortic lymphadenectomy and, when applicable, appendectomy (group 1 in the study). The median length of follow-up was 49 months.

Outcomes reported

Survival

Two studies reported outcomes for survival. One study applied a multivariate analysis of OS adjusting the HR for surgery type, age (continuous), FIGO stage and residual disease (Chang 2012a). Although not reported in the original paper, Aletti 2006a provided estimates of the HR from a multivariable Cox model, comparing five-year disease-specific survival (DSS) (event being death from advanced ovarian cancer) in the ultra-radical surgery group with that in the standard surgery group for all 194 women and for the 144 women with carcinomatosis. The HR for DSS in Aletti 2006a including all 194 women was adjusted for: age, ASA score, carcinomatosis, mesenteric involvement, diaphragmatic involvement, ascites, residual disease and operative time. The HR for DSS in the subset of 144 women with carcinomatosis was adjusted for: age, ASA score, tumour grade, residual disease and operative time. Luyckx 2012 reported cox regression estimates for OS but did not include type of surgery in their multivariate model as it was not significant in univariate analysis.

Progression-free survival

Chang 2012a reported PFS and adjusted the HR for FIGO stage, tumour grade, residual disease and surgery type.

Disease-free survival

Luyckx 2012 reported an HR for disease-free survival (DFS) and adjusted for FIGO stage, tumour grade, presence of upper abdominal disease, amount of residual disease, and timing of surgery (primary or interval) and surgery type.

Death within 30 days of intervention

Aletti 2006a and Chang 2012a reported perioperative death within 30 days. In this review, we used 'death within 30 days' as a secondary outcome measure because this cut-off has been widely used in the literature and would include people who died of complications directly related to surgery that may only manifest one to two weeks after surgery.

Adverse events

Chang 2012a reported postoperative morbidity defined as infected lymphocyst, thromboembolism, intestinal obstruction, anastomostic leakage, ureteral injury, sepsis, intra-abdominal abscess, pneumothorax, postoperative death within 30 days, or a combination of these. Luyckx 2012 and Aletti 2006a did not report adverse events by type of surgery.

None of the studies reported recurrence rate, QoL or (loco)regional control.

For further details see the Characteristics of included studies table.

Excluded studies

We excluded 103 references after obtaining the full text for the following reasons (see Characteristics of excluded studies table).

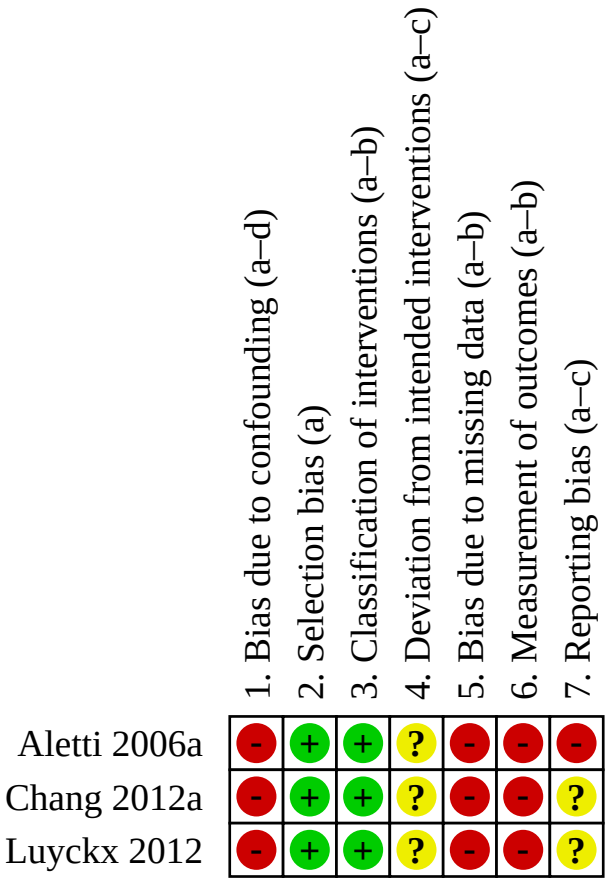
- In 12 studies, a comparison of ultra-radical and standard surgery was not possible (Aletti 2006b; Aletti 2009a; Bahra 2013; Bertelsen 1990; Eisenkop 2001; Eisenkop 2003; Grimm 2017; Laios 2019; Pelissier 2018; Vidal 2016; Wimberger 2007; Yildirim 2014).
- In 22 studies, the comparison was not of interest to our study (Chua 2011; Clark 2012; Clark 2014; Favero 2014; Ferrero 2014; Fotopoulou 2012; Gremeau 2014; Guyon 2014; Hamilton 2011; Hwang 2014; Hudry 2013; Janda 2014; Kato 2013a; Kehoe 2013; Li 2014; Perri 2013; Pushpalatha 2011; Qin 2012; Rouzier 2010; Sandadi 2014; Scalici 2014; Sehouli 2010).
- Participants in the comparison (standard surgery) group also had extensive bowel surgery (which is classified as ultra-radical) in 13 studies (Aletti 2006b; Canlorbe 2018; Chi 2004; Eisenhauer 2006; Eisenkop 1993; Eisenkop 1998; Elgamal 2019; Eoh 2017; Filippova 2019; Gockley 2019; Kommoss 2010; Kuhn 1998; Tozzi 2019), diaphragmatic stripping in two studies (Tsolakidis 2010a; Tsolakidis 2010b), ultra-radical with splenectomy in one study (Davies 2019), and extensive upper abdominal surgery in two studies (Chi 2009; Oseledchyk 2016).
- In three studies, the intervention was a specific form of ultra-radical surgery, but it was unclear whether those in the comparison group received a different form of ultra-radical surgery or standard surgery (Aletti 2006c; Cai 2007; Eisenkop 2006).
- In three studies, there was no ultra-radical surgery performed (Chang 2012b; Cormier 2012; Park 2011).
- Four studies included participants with recurrent disease (Bristow 1999; Kato 2013b; Kolev 2014; van de Laar 2014), whereas in one study it was unclear whether women with recurrent disease were included (von Hugo 1989).
- Ten studies analysed data by descriptive statistics, no multivariate analysis was performed (Barlin 2013; Chereau 2011; Eng 2018; Soo Hoo 2015; McCann 2011; Muallem 2018; Phillips 2018; Sagara 2019; Zamurovic 2013; Zapardiel 2012).
- Ten studies included fewer than 100 participants in their analyses (Angioli 2012; Butler 2012; Kim 2011; Liu 2013a; Pathiraja 2011; Pathiraja 2013; Ratnavelu 2014; Stefanović 2011; Sundar 2014; Wat 2012).
- Four studies included people with borderline tumours (Kristensen 2014), germ cell tumours (Liu 2013b), only stages pT1-2 (Oshita 2013), and where those with suboptimal debulking were excluded (Rodriguez 2013).
- Thirteen studies were conference abstracts and the full text was not available to make a decision (Campos 2014; Cummins 2019; Jiang 2013; Liberale 2019; Lee 2017; Martinez 2014; Rodriguez 2012; Sundar 2018; Sundar 2019; Suzuki 2008; Szczesny 2016; Wallace 2016; Wright 2012), one of which reported outcomes that were not of interest (Wright 2012).
- In Ren 2015 (Jiang 2013 in abstract form), the type of surgery was not included as a variable in a multivariable analysis of PFS.
- One study was an uncontrolled before-after study, but forms part of the discussion in Agreements and disagreements with other studies or reviews (Falconer 2020).

Risk of bias in included studies

We did not identify any RCTs, so we did not apply Cochrane's risk of bias tool (we planned to use ROB-1) for the assessment of these types of studies. Instead, we used the ROBINS-I tool to assess bias in our included studies (Sterne 2016).

The risk of bias assessments of the three included comparative observational studies are summarised in Table 1 and Figure 2. All studies had critical bias due to confounding because no known prognostic factors could be identified that would have the potential for confounding the effect on intervention. In Chang 2012a, adjusted HRs were derived from a prognostic model. No details were presented on how modelling was performed, but this seems to have been done based on significance testing (and not on including putative confounders in the analysis, irrespective of statistical significance). The adjusted HRs for Luyckx 2012 were derived from a prognostic model based on univariate significance testing ($P < 0.10$) and not on including putative confounders in the analysis, irrespective of statistical significance. In addition, the data were collected retrospectively.

Figure 2. Methodological quality summary: review authors' judgements about each methodological quality item for each included study.



All three studies were at low risk of bias in selection of participants into the study. In all three, the intervention and follow-up occurred simultaneously as outcomes would be observed immediately after the cytoreductive surgery. Additionally, there was no evidence of selection into the study due to variables measured after the intervention, since participants were included retrospectively.

In all three studies, bias in classification of interventions was low because the intervention statuses were clearly defined as either having: aggressive surgery or not (Aletti 2006a), simple versus radical surgical procedure (Chang 2012a), or standard surgery versus standard surgery plus relatively routine upper abdominal surgery versus ultra-radical surgery (Luyckx 2012).

All three studies were at unclear risk of bias due to deviations from intended interventions. There was no evidence of any deviations from interventions or usual practice but these may either be due to omission or that deviations did not happen.

In all three studies, there was no differential follow-up or missing data reported and no participants were reportedly omitted due to

missing data. Although there is no reason to believe there was a serious bias due to missing data, they were still not comparable to a randomised trial. Therefore, we rated bias due to missing data as moderate to high.

All three studies were at critical bias in measurement of outcomes. In Aletti 2006a, the multivariate analysis adjusted for variables that were measured after the time origin in some of the analyses, namely extent of residual disease and operative time (Altman 1995). Residual disease was also used as an adjustment prognostic factor in the other two included studies (Chang 2012a; Luyckx 2012). This is likely to distort the estimate of survival as this adjustment is made after surgery and is a key prognostic factor.

Only one study was at critical bias in selection of the reported result. In Aletti 2006a, the authors reported DSS rather than OS, which is a more appropriate and reliable outcome measure and did not report any QoL data, or state if there were any predefined outcome measures prior to data analysis. Therefore, it is possible that the outcomes may have been selectively reported. DSS is not a good outcome measure to use for several reasons, for example the coding of death certificates is notoriously prone to error. Also data were reported in a subset of the 144 women with carcinomatosis (more-extensive disease) only. There was unclear evidence of selective reporting in the other studies as no protocol was available (Chang 2012a; Luyckx 2012). However, all outcomes mentioned in the methods section seemed to have been reported in the results section.

Effects of interventions

See: **Summary of findings 1** Ultra-radical (extensive) surgery compared to standard surgery in women with stage IIIc or IV ovarian cancer

We only identified very low-certainty evidence for all outcomes reported in the review, but mainly due to relatively few women being included due to stringent inclusion criteria. A breakdown of adverse events was not adequately reported in two studies (Aletti 2006a; Luyckx 2012) and QoL was not reported in any of the three included studies. Although a secondary outcome, none of the studies reported 'Optimal cytoreduction' in any multivariate analyses. We did not present any unadjusted results for this as it is likely that 'optimal cytoreduction' will be higher in ultraradical surgery and would not be a fair comparison.

Survival (overall and disease-specific)

Upfront primary debulking surgery

Meta-analysis of two studies (397 women) found that women who underwent radical procedures as part of PDS had 40% less chance of mortality compared to women who underwent standard surgery (adjusted HR 0.60, 95% CI 0.43 to 0.82; I^2 = 0%; very low-certainty evidence; Analysis 1.1; Summary of findings 1) (Aletti 2006a; Chang 2012a). Aletti 2006a reported the five-year DSS rather than categorising deaths by any cause. In Chang 2012a, the median OS (unadjusted) was 66 months in the ultra-radical surgery group and 38 months in the standard surgery group. The five-year DSS rate (unadjusted) was 46% in the ultra-radical surgery group compared with 13% in the standard surgery group.

The results were robust to a sensitivity analysis which included two studies assessing 283 participants with more-extensive disease (carcinomatosis), which found that women who underwent radical procedures as part of PDS had 39% less chance of mortality compared to women who underwent standard surgery (adjusted HR 0.61, 95% CI 0.44 to 0.85; I^2 = 0%; very low-certainty evidence; Analysis 1.2; Summary of findings 1) (Aletti 2006a; Chang 2012a).

Upfront primary and interval debulking surgery

One study, which included women with stage IIIC and IV disease, reported a comparison of radical versus standard surgical procedures associated with both PDS and IDS procedures (Luyckx 2012). The study authors did not report the magnitude of effect in multivariate analyses and only included variables associated with P < 0.05 on univariate analysis in Cox regression model. The study found no difference in the risk of mortality between women undergoing radical surgery versus standard surgery (403 women) or ultra-radical versus standard surgery (424 women), in univariate analyses. Multivariate analyses were not reported and, therefore, the certainty of the evidence was not assessable and remained very low.

Progression-free survival

Upfront primary debulking surgery

Chang 2012a, which assessed 203 participants, found that women who underwent radical procedures as part of PDS had nearly 40% less chance of disease progression or death compared to women who underwent standard surgery (adjusted HR 0.62, 95% CI 0.42 to 0.92; very low-certainty evidence; Analysis 1.3; Summary of findings 1). The median PFS (unadjusted) was 18 in the ultra-radical surgery group and 11 months in the standard surgery group. The results were robust to a sensitivity analysis assessing a subset of 139 women with carcinomatosis, which found that women who underwent radical procedures as part of PDS had nearly 50% less chance of disease progression or death compared to women who underwent standard surgery (adjusted HR 0.52, 95% CI 0.33 to 0.82; very low-certainty evidence; Analysis 1.4; Summary of findings 1).

Disease-free survival

Upfront and interval debulking surgery

One study, which included women with stage IIIC and IV disease, reported DFS for a comparison of radical versus standard surgical procedures associated with both PDS and IDS procedures (Luyckx 2012). A combined analysis in one study (Luyckx 2012), assessing 527 women, found that those who underwent radical procedures were associated with increased chance of disease progression or death than those who received standard surgery (adjusted HR 1.60, 95% CI 1.11 to 2.31; I^2 = 0%; very low-certainty evidence; Analysis 1.5; Summary of findings 1). In absolute and unadjusted terms, the DFS was 19.3 months in the standard surgery group (group 1), 15.8 months in group 2A and 15.9 months in group 2B (the two ultra-radical surgery groups) (see Characteristics of included studies table for details of groups).

Death within 30 days of surgery

Upfront debulking surgery

None of the studies reporting death within 30 days of surgery used any statistical adjustment.

In Aletti 2006a, three women died within 30 days of their standard surgical procedure whereas there were no reported cases of perioperative mortality in the ultra-radical surgery group (very low-certainty evidence).

In Chang 2012a, perioperative death within 30 days occurred in 0/119 (0%) in the ultra-radical surgery group versus 1/84 (1.2%) in the standard surgery group (very low-certainty evidence).

In total there were only four deaths within 30 days of surgery in both studies and none in the ultra-radical group so we did not report a risk ratio (planned outcome of choice a priori) to avoid potentially misleading results with so few deaths (Summary of findings 1).

Adverse events

Upfront debulking surgery

Chang 2012a did not use any statistical adjustment for any adverse events.

In Chang 2012a, there was postoperative morbidity in 32/84 (38.1%) women in the ultra-radical surgery group versus 14/119 (11.8%) women in the standard surgery group (RR 3.24, 95% CI 1.84 to 5.68; very low-certainty evidence).

Women who underwent ultra-radical surgery had significantly larger median estimated blood loss (800 mL with ultra-radical surgery versus 500 mL with standard surgery; P = 0.03), were more likely to receive a intraoperative or postoperative blood transfusion (intraoperative: 25% with ultra-radical surgery versus 17.6% with standard surgery; postoperative: 39.3% with ultra-radical surgery versus 26.1% with standard surgery; P = 0.01), had longer median days in the intensive care unit (1.5 days with ultra-radical surgery versus 0.8 days with standard surgery; P < 0.01), and were more likely to experience postoperative morbidity (38% with ultra-radical surgery versus 11.8% with standard surgery; P < 0.01) than those who underwent standard surgery (very low-certainty evidence; Summary of findings 1).

Operative time

Upfront debulking surgery

Chang 2012a did not use any statistical adjustment for operative times between groups.

In Chang 2012a, women who underwent ultra-radical surgery had significantly longer median operative times than those who had standard surgery (307 with ultra-radical surgery versus 235 minutes with standard surgery; P < 0.01). This outcome was not specified in the summary of findings table.

DISCUSSION

Summary of main results

We found three studies that met our inclusion criteria (Aletti 2006a; Chang 2012a; Luyckx 2012). These studies reported retrospective data for 924 women with advanced EOC (stage III/IV) who underwent either ultra-radical or standard surgery. Two studies reported on women who exclusively received PDS (Aletti 2006a; Chang 2012a), whereas Luyckx 2012 included women who had received both PDS and IDS procedures.

Of the six outcomes examined, only survival (overall and disease-specific and progression/disease-free), and perioperative mortality were reported in more than one study. There is from two observational retrospective studies providing very low-certainty evidence that ultra-radical surgery compared to standard surgery was associated with better OS in multivariate analyses (Chang 2012a). However, the evidence for better OS was not corroborated in Luyckx 2012, as surgery type was not found to be associated with OS in a univariate analysis. In contrast, we also found evidence from Luyckx 2012 that ultra-radical surgery was associated with worse DFS compared to standard surgery. We found that ultra-radical surgery was no better than standard surgery regarding DSS in multivariate analysis (Aletti 2006a).

In women with advanced-stage ovarian cancer, a difference in perioperative mortality between ultra-radical surgery and standard surgery could neither be demonstrated nor refuted due to the low number of reported deaths within 30 days of surgery. We found that there is very low-certainty evidence that these participants who underwent ultra-radical surgery may be more likely to experience postoperative morbidity, have longer operative time, greater estimated blood loss, more likely to have intraoperative or postoperative blood transfusions, and longer stay in the intensive care unit compared to those who underwent standard surgery (Chang 2012a).

In summary, across the three studies, we found insufficient evidence in assessing ultra-radical surgery versus standard surgery.

We did not identify any RCTs or comparative observational studies that used statistical adjustment that addressed recurrence rate, QoL or (loco)regional control.

Overall completeness and applicability of evidence

The included studies did not adequately address the objectives of the review, with outcomes being incompletely reported or not reported at all (e.g. QoL). The settings of the three studies spread across three countries: USA, France and South Korea.

We assumed that the decision to perform standard or ultra-radical procedures in these three retrospective studies was determined by the surgeon's discretion unless the study authors explicitly stated the reasons. Furthermore, descriptive information on participant and disease characteristics were not reported by type of surgery. Thus, confounding by indication cannot be ruled out. Significance at univariate analyses was the primary method for variable selection in multivariate analyses for all three studies, highlighting the exploratory nature of these studies with regards to identifying potential confounders. Putative confounders, irrespective of statistical significance on their own, should always be reported in statistical models. Depending on the outcome, selected variables included a combination of age, FIGO stage, residual disease, ASA score, operative time, timing of surgery, tumour grade, or a combination of these in addition to surgery type. Prognostic factors that are commonly known (or could, in principle, be known) before the operation is performed (e.g. age, ASA score, carcinomatosis, mesenteric involvement, diaphragmatic involvement, ascites) are moderating variables and it is valid to adjust for them. It may even be necessary to adjust for them as they may confound the assignment to type of surgery.

The sensitivity analysis including women with carcinomatosis appeared to suggest that in women with more-extensive disease, there is more benefit of radical surgery. Despite the certainty of the evidence being very low, there was a suggestion that unless there is some indication of both the preprocedure extent of disease and the postprocedure residual disease, it is difficult to define the value of surgery. FIGO staging may simply be too crude to use to define extent of disease for most cases.

One included study reported DSS, which included deaths from ovarian cancer and deaths from surgical treatment (Aletti 2006a). Adjusted HRs for OS, which was this review's strict primary outcome, was only reported in Chang 2012a, although we combined these outcomes for the meta-analysis with the assumption that deaths from other causes would be minimal (potentially a dubious assumption but we did present the different outcomes as a subgroup for transparency).

One limitation observed was the inclusion of potential mediating variables in the multivariate models reported in the identified studies The extent of residual disease is likely to be a consequence of both the initial extent of disease (e.g. the pattern rather than the stage and bulk) and the type of surgery (e.g. the degree of surgical radicality). If residual disease is a putative risk factor for survival, then it is likely to be the case that residual disease is a potential mediator in the hypothesised causal pathway from surgery type to survival. As such, residual disease does not meet the criteria for a confounder (Kyriacou 2016), and its inclusion in multivariate models would reduce the effect of surgery type on survival.

Women with advanced ovarian cancer are generally in poor health and have a relatively short life expectancy. A good QoL after treatment is therefore an important issue in this group of women, but unfortunately this review was unable to assess this important outcome but the results of the recent SOCQER-2 study are discussed below under Agreements and disagreements with other studies or reviews (Sundar 2021).

Quality of the evidence

Overall, the certainty of the evidence was very low for all outcomes because the review found only three relevant NRSs, all of which were at high risk (critical) of bias (GRADE Working Group 2004). This severely limits any conclusions. The three included studies analysed 874 women, but not all could be included in the same pooled analyses. All three studies were at critical risk of bias, largely because they were retrospective in nature. Participant characteristics were not reported by surgical group so it was not possible to assess whether the groups receiving different types of surgery were similar prior to surgery. However, the univariate analysis showed which factors were important predictors of survival individually and analysis of the type of surgery that adjusted for these prognostic factors and generally gave similar effect estimates for survival estimates to the unadjusted results, suggesting that prognostic factors were likely balanced between surgical groups. However, it is possible that factors not significant in univariate analysis could influence the estimates of effect in the multivariate model. Furthermore, the dichotomy of several covariates is also questionable and variables that were not considered in the analysis, such as comorbidities and ethnicity, could also influence results.

There were also other contributing factors to downgrade the level of evidence to providing very low-certainty evidence. We had concerns that residual disease after surgery had been adjusted for in the Cox models for survival in all three included studies. When assessing the effect of ultra-radical versus standard surgery, the extent of residual disease is likely to be a consequence of whether ultra-radical or standard surgery was performed; therefore, adjusting for extent of residual disease is likely to dilute the estimate of the effect of the type of surgery. 'Extent of residual disease' is a mediating variable, on the causal pathway between type of surgery and outcome (Altman 1995). Likewise, operative time which was included in the Cox model in Aletti 2006a is also a mediating variable. Prognostic factors that are known (or could in principle be known) before the operation is performed (e.g. age, ASA score, carcinomatosis, mesenteric involvement, diaphragmatic involvement, ascites) are moderating variables and it is completely valid to adjust for them. Indeed, it is necessary to adjust for them and they formed part of our inclusion criteria because they are probably confounded with assignment to treatment group. A separate prognostic review assessing residual disease as a prognostic factor in this area has been conducted (Bryant 2021). This review shows the prognostic impact of achieving NMRD.

In the included studies, as well as many that were excluded, there appeared to be an over-interpretation of statistical significance. For example, in Aletti 2006a, the adjusted HR was 0.64 (95% CI 0.40 to 1.04). The HR in women with carcinomatosis was 0.64 (95% CI 0.41 to 0.98). The former was somewhat dismissed because it was "not statistically significant" but the authors were more convinced by the latter. This carries through into the conclusions where ultra-radical surgery is deemed to be beneficial for women with carcinomatosis, but not for others. However, the point estimates are in this instance identical. The reason the former is not significant and the latter is, could simply be because the former was adjusted for a large number of factors (Higgins 2019; Schisterman 2009). We reflected this in our certainty of the evidence judgements, although the power of meta-analyses did provide us as review authors increased scope to make slightly more generalised conclusions (than single study authors) as there was a suggestion that women with carcinomatosis may have benefited from more radical surgery.

There were also many other factors affecting assignment to the surgical groups. It may have been the case that surgeons were more likely to perform ultra-radical surgery if women are in better health or they are themselves more experienced. We suggest that because the adjusted and unadjusted HRs are similar, the prognostic factors may be well balanced at baseline. Nonetheless, we have been cautious and this is reflected in the certainty of evidence judgements. While the groups may be well balanced, it is possible that the ultra-radical group started off healthier and their apparently better survival in upfront surgery was an artefact. However, this is not possible to ascertain and more evidence of better certainty is needed.

Aletti 2006a reported disease-specific OS. We assumed this to be DSS, as DSS and OS are different. OS counts all deaths (from whatever cause) as an event; DSS counts only deaths from ovarian cancer as an event. This raises the question about how disease-specific survival counts deaths from other causes, where presumably such deaths are censored. DSS is not a non-ideal outcome measure to use due to the potentially poor and error-

prone coding of death certificates. Furthermore, if someone dies because of the treatment they receive, this may not be counted as a death from ovarian cancer, but it is just as important to the patient as a death from ovarian cancer. Thus the evaluation of the relative benefits of the treatments should include these deaths. DSS was not one of our prespecified outcomes, however we chose to subgroup by DSS in the meta-analysis.

Potential biases in the review process

We performed a comprehensive search, including a thorough search of the grey literature and at least two review authors independently sifted and extracted data for all studies. The review included NRSs and was not restricted to RCTs, which provide the strongest level of evidence available. We made every attempt to minimise bias in the review process. We anticipated that selection bias was likely to be a real problem due to the non-randomised assignment of women to surgery as it was likely that treatment allocation depended on the clinical indication and the level of surgical expertise available. We attempted to minimise this bias by only including RCTs or quasi-RCTs or NRSs of sufficient quality that adjusted for baseline differences between the groups receiving different types of surgery. Unfortunately, we were only able to include three studies of such quality that met the inclusion criteria.

A further threat to the validity of the review is likely to be the possibility of publication bias. Studies that did not find a statistically significant difference between treatments may not have been published. We were unable to assess this possibility as the analysis was restricted to just three included studies.

Agreements and disagreements with other studies or reviews

One before-after study suggested that a structured shift to an ultra-radical upfront primary surgical approach may not improve survival in surgically treated women (Falconer 2020). In this population-based cohort study, women with suspected advanced EOC near Stockholm in Sweden were included via the Swedish Quality Registry for Gynecologic Cancer (SQRGC) and the NCR. Women were selected in two sets of three-year cohorts, based on the year of their diagnosis (a 'before' cohort or an 'after' cohort with a change in surgical treatment algorithm) and were followed for at least three years. Five-year OS in non-surgically and surgically treated women was analysed. After a median follow-up of around 28 months in 752 women, the complete resection rate increased from 37% to 67% as well as the proportion of non-surgically treated women, from 24% to 33%. This study also demonstrated that a shift to ultra-radical surgery was associated with an increase in the proportion of non-surgically treated women. However, this study was not a 'controlled' before-after study and as a consequence was prone to bias (Goodacre 2015). The use of historical controls are known to overestimate the benefit of new treatments. Before-after studies also have a high risk of bias because there may be unidentified differences between the intervention and control groups that may affect changes in the outcome measure (Sterne 2022).

One of the excluded studies evaluated the impact of different prognostic factors for surgical outcome and evaluated the impact of surgical outcome on survival in women with advanced-stage ovarian cancer (Wimberger 2007). In this prospective study, 798 women with FIGO IIB-IV disease from 136 centres within Germany were operated on and then randomised to receive either cisplatin plus paclitaxel or carboplatin plus paclitaxel chemotherapy. Complete surgical data were obtained from 761 women and were analysed using multivariable logistic regression. Complete cytoreduction with no macroscopic residual tumour was achieved in 29.8% of women, with a significant improved OS compared to women with visible, including small, remaining disease (P < 0.0001). In women with FIGO stages IIIC and IV, complete cytoreduction was less likely in older women, those with a higher preoperative tumour load, worse performance status, and peritoneal carcinomatosis. FIGO stage was not an independent factor for complete cytoreduction in this group of women. The authors identified a subgroup of 71 centres (referred to as type A) which demonstrated the capability of performing ultra-radical surgery having carried out pelvic or para-aortic lymphadenectomy (or both) and peritoneal stripping in at least one of the enrolled participants in the study. This group included 534 (69.8%) women. The remaining 65 centres were identified as type B centres and treated 227 women. A higher percentage of women with worse performance status were treated in type A centres (53.9% in type A versus 43.6% in type B; P = 0.009). Type A centres more often achieved complete cytoreduction compared to type B centres (32.8% in type A versus 22.9% in type B; P = 0.007). Treatment in type A centres was associated with greater OS compared to treatment in type B centres (45.2 months in type A versus 35 months in type B; P = 0.045).

Their results suggest an advantage for aggressive primary surgery and complete cytoreduction in women with more advanced disease when operated on in experienced centres. Although this study was excluded from the review because the comparative groups were by treatment centres that contained a mixed case load of ultra-radical and standard surgery, it does provide some evidence that aggressive primary cytoreductive surgery can negate the effects of aggressive tumour biology in advanced ovarian cancer, with a subsequent improvement in OS.

In Aletti 2006a, the authors reported that radical procedures were performed at the same rate regardless of age (49% for age less than 65 years versus 51% for age greater than 65 years; P = 0.45) and that participants with better ASA scores (1 or 2 versus 3 or 4) were more likely to have aggressive procedures performed (59% with ASA 1 or 2 versus 36% with ASA 3 or 4; P = 0.005), which implies the overall medical condition of the participant at least partially influences the decision to perform aggressive surgery. However, the numbers of women in each surgical group were not reported. For further details, see the Characteristics of included studies table.

One recent review of guidelines showed clear international differences in ovarian cancer survival and these differences in treatment could be contributing to survival disparities (Norell 2020). The objective of the review by Norell and colleagues was to compare clinical practice guidelines and patterns of care across seven high-income countries. They included guidelines widely used in routine ovarian cancer treatment. The review also included an expert questionnaire component, which included questions on surgical practice and was validated and tested by an expert clinical working group. Guideline and survey results were crudely compared with three-year survival by 'distant' stage using Spearman's rank order correlation.

Norell 2020 compared 27 guidelines, and 119 clinicians completed the survey. Guideline-related measures varied between countries but did not correlate with survival internationally. Reported

patterns of surgical care varied internationally, including for rates of extensive/'ultra-radical' surgery, and perceived barriers to optimal cytoreduction. When surveyed, Norwegian and Australian clinicians either agreed or strongly agreed with ultra-radical surgery, whereas clinicians from Canada and the UK agreed with ultra-radical surgery to a lesser extent, with some respondents either disagreeing or strongly disagreeing with this approach. When crudely compared, willingness to undertake extensive/ultra-radical surgery correlated with three-year survival by distant stage (Spearman's rank correlation coefficient (r_s) = 0.94, P = 0.017). Norell 2020 reported that most guidelines that were identified did not explicitly recommend ultra-radical (extensive) surgery, but clinicians from higher-performing countries were more likely than those from lower-performing countries to be proponents of 'ultra-radical' surgery. Norwegian clinicians were least likely to perceive age as a barrier to achieving optimal cytoreduction and Norway demonstrated the highest survival in elderly women with distant-stage disease. In the UK, where clinicians perceived a lack of supportive care, survival for these women was lower. Women with advanced ovarian cancer are more likely to have severe comorbidities and higher mortality, and historically, elderly women were less likely to receive comprehensive surgical treatment. One Dutch study recently found that older participants and those with advanced disease were significantly less likely to receive any cancer-directed treatment (Zijlstra 2019). It was also noted by the authors of Norell 2020 that available resources and operating theatre time may influence a surgeons' ability to perform extensive surgery and could impact patient outcomes. They also added that it is this subcategory of elderly women with advanced disease where survival is lowest and where significant differences exist.

Guidelines from Belgium in 2016 provided recommendations based on scientific evidence for the diagnosis, treatment and follow-up of epithelial ovarian, fallopian tube and primary peritoneal cancer (Vergote 2016). The report stated that clinicians were encouraged to interpret their recommendations in the context of the individual patient situation and her own values and preferences. Furthermore, in the absence of good-quality evidence on optimal treatment options, patient participation in clinical trials was to be encouraged as much as possible. The guidelines reported by Vergote 2016 acknowledged that the evidence was limited, but suggested that it supports the use of radical surgical techniques (such as diaphragm resection, peritoneal stripping, splenectomy, etc.) to obtain complete resection of all macroscopic tumour. The guidelines showed the prognostic value of debulking to no macroscopic disease at the end of surgery and supporting evidence from the use of radical surgery. The guidelines formulated a strong recommendation (despite low level of evidence) that complete debulking should be the aim of cytoreductive surgery (PDS or IDS) and that the term optimal should no longer be used as old definitions of optimal surgery (residual disease less than 2 cm or less than 1 cm). We are more cautious in the interpretations in our systematic review than the guidelines we identified. While the results of the guidelines are compelled to make recommendations, our review is restricted to the inclusion of just three NRSs and we were bound by systematic review reporting guidance (Higgins 2019).

The guidelines reported by Norell 2020 also accounted for the experience of patient representatives. The influence of radical surgery on long-term QoL was reported not to be a major drawback, the survival benefit weighing more importantly in the overall balance. The SOCQER-2 study reported QoL as a primary outcome, which was a prospective, non-randomised multicentre observational study run across the UK, India and Australia (Sundar 2021). Women were eligible if they had suspected or confirmed EOC with radiological spread beyond pelvis and if primary or delayed debulking surgery was planned.

The SOCQER-2 study found that women with late-stage ovarian cancer had no important differences in European Organisation for Research and Treatment of Cancer Quality of Life of Cancer Patients – 30 (EORTC QLQ-C30) global scores measured across six weeks, six months and 12 months' postsurgery when undergoing surgery of varying complexity, despite a higher preoperative disease burden in people undergoing more radical surgical procedures (Sundar 2021). Across all groups of women receiving all forms of complex surgery (categorised by surgery complexity scores (SCS) and grouped into low, intermediate and high), global QoL showed a small but significant improvement by 12 months postoperatively. Women who underwent the most complex surgery (high-SCS group) had small-to-moderate detriments in EORTC QLQ-C30 physical function, role function and emotional function at six weeks postsurgery compared with women undergoing less-extensive surgery (intermediate- and low-SCS groups), but by six to 12 months' postsurgery, these functions were comparable across all SCS categories. Most women undergoing high-SCS surgery without disease progression experienced a positive change in QoL by 12 months' postsurgery. There were no clinically meaningful differences in QoL among women undergoing surgery of different complexities. The authors of the study concluded that women undergoing high-complexity surgery can be reassured that by 12 months' postsurgery most will have better QoL after than immediately before surgery (Sundar 2021).

The authors of SOCQER-2 found that women who underwent low-complexity surgery had higher rates of residual disease and lower survival compared with those with a similar disease burden undergoing surgery of intermediate complexity (Sundar 2021). However, there was no statistical adjustment performed in these analyses. Postoperative residual disease was associated with poorer OS, particularly in women undergoing low-complexity surgery, but again they made no statistical adjustment. Sundar 2021 acknowledged potential selection bias, but since research nurses carried out recruitment to the SOCQER-2 study, that systematic bias introduced by surgeons recruiting women whom they believed would recover well after extensive surgery was unlikely.

AUTHORS' CONCLUSIONS

Implications for practice

We found only very low-certainty evidence comparing ultra-radical and standard surgery in women with advanced ovarian cancer and also subgroups with carcinomatosis. The evidence suggested that ultra-radical surgery may result in better survival, but results are based on retrospective studies, at critical risk of bias, in relatively few women. Results for progression/disease-free survival were inconsistent and evidence was sparse. Quality of life (QoL) and morbidity was not reported in the two groups, but the results of the SOCQER-2 (Surgery in Ovarian Cancer – Quality of Life Evaluation Research – 2) study are promising for those undertaking high-complexity surgery (Sundar 2021). This study was the only one

to adequately investigate QoL and it concluded that there can be confidence in clinical practice that the use of high-complexity surgery in advanced ovarian cancer will not have a detrimental effect on global QoL compared with less-complex surgery.

While we were unable to reach definite conclusions about the relative benefits and adverse effects of the two types of surgery in our review (that applied stringent inclusion criteria), the guidelines that we identified are worthy of consideration (Norell 2020; Vergote 2016). These guidelines generally supported the use of radical surgical techniques to obtain no macroscopic residual disease in appropriate women.

Implications for research

To date, most studies of ultra-radical (extensive) surgery for advanced-stage ovarian cancer have assessed residual disease as an outcome rather than survival. Other studies that have assessed the role of ultra-radical surgery have not compared it with standard surgery and have included women with recurrent disease, making this a heterogeneous group of women and hence limiting the inferences that can be made about the role of ultra-radical surgery. In order to aid existing guidelines, the role of ultra-radical surgery in the management of advanced-stage ovarian cancer could be addressed through the conduct of a sufficiently powered randomised controlled trial comparing ultra-radical and standard surgery.

If randomised controlled trials are not feasible, high-quality non-randomised studies should be designed to add to the existing evidence base in the review. Such studies should include all women diagnosed within a fixed population and agree criteria for prognostic factors that will form the key adjustment in analyses. Population-level, multicentre studies are important in this area as what works or does not work in one institution may be very different from what works elsewhere. It would be important to test the effect of ultra-radical surgical adoption on the rates of surgery and the effects on those women who do not undergo surgery. Multivariable analysis should allow for baseline prognostic factors, but not for variables (such as extent of residual disease or operating time) that were recorded after women were assigned to surgical groups. The experience of the treating surgeon should also be factored in.

ACKNOWLEDGEMENTS

We thank Jo Morrison (Co-ordinating Editor for the Cochrane Gynaecological, Neuro-oncology and Orphan Cancers (GNOC)) for clinical and editorial advice, Jo Platt (GNOC) for running the searches for the update, and Gail Quinn and Clare Jess (GNOC) for their contribution to the editorial process.

This project was supported by the National Institute for Health Research (NIHR), via Cochrane Infrastructure funding to the Cochrane Gynaecological, Neuro-oncology and Orphan Cancer Group. The views and opinions expressed herein are those of the authors and do not necessarily reflect those of the Systematic Reviews Programme, NIHR, National Health Service or the Department of Health.

The authors and Cochrane GNOC Team are grateful to the following peer reviewers for their time and comments: Sean Kehoe, Mary Lunnen and Monique Spillman.

We would also like to acknowledge the Christine Ang, Karen KL Chan and Heather O Dickinson for their contributions to earlier versions of this review. We thank Professor Cilby and Dr Aletti for providing us with the additional information we required.

REFERENCES

References to studies included in this review

Aletti 2006a *{published data only}*

Aletti GD, Dowdy SC, Gostout BS, Jones MB, Stanhope CR, Wilson TO, et al. Aggressive surgical effort and improved survival in advanced-stage ovarian cancer. *Obstetrics and Gynecology* 2006;**107**(1):77-85.

Chang 2012a *{published data only}*

Chang SJ, Bristow RE, Ryu HS. Impact of complete cytoreduction leaving no gross residual disease associated with radical cytoreductive surgical procedures on survival in advanced ovarian cancer. *Annals of Surgical Oncology* 2012;**19**(13):4059-67.

Luyckx 2012 *{published data only}*

Luyckx M, Leblanc E, Filleron T, Morice P, Darai E, Classe JM, et al. Maximal cytoreduction in patients with FIGO stage IIIC to stage IV ovarian, fallopian, and peritoneal cancer in day-to-day practice: a Retrospective French Multicentric Study. *International Journal of Gynecological Cancer* 2012;**22**(8):1337-43.

Martinez A, Picaud L, Luyckx M, Pomel C, Leblanc E, Morice P, et al. Surgical approach and survival in elderly patients with advanced ovarian cancer. 15th Biennial Meeting of the International Gynecologic Cancer Society; 2014 Nov 8-11; Melbourne (VIC).

References to studies excluded from this review

Aletti 2006b *{published data only}*

Aletti GD, Podratz KC, Jones MB, Cliby WA. Role of rectosigmoidectomy and stripping of pelvic peritoneum in outcomes of patients with advanced ovarian cancer. *Journal of the American College of Surgeons* 2006;**203**(4):521-6.

Aletti 2006c *{published data only}*

Aletti GD, Dowdy SC, Podratz KC, Cliby WA. Surgical treatment of diaphragm disease correlates with improved survival in optimally debulked advanced stage ovarian cancer. *Gynecologic Oncology* 2006;**100**(2):283-7.

Aletti 2009a *{published data only}*

Aletti GD, Dowdy SC, Gostout BS, Jones MB, Stanhope RC, Wilson TO, et al. Quality improvement in the surgical approach to advanced ovarian cancer: the Mayo Clinic Experience. *Journal of the American College of Surgeons* 2009;**208**:614-20.

Aletti 2009b *{published data only}*

Aletti GD, Podratz KC, Moriarity JP, Cliby WA, Long KH. Aggressive and complex surgery for advanced ovarian cancer: an economic analysis. *Gynecologic Oncology* 2009;**112**:16-21.

Angioli 2012 *{published data only}*

Angioli R, Plotti F, Aloisi A, Capriglione S, Terranova C, Ricciardi R, et al. Does extensive upper abdomen surgery during primary cytoreduction impact on long-term quality of life? *International Journal Gynecologic Cancer* 2012;**2012**:E1014.

Bahra 2013 *{published data only}*

Bahra M, Fotopoulou C, Braicu EI, Kwee SL, Kuhberg M, Richter R, et al. Salvage surgery due to bowel obstruction in advanced or relapsed ovarian cancer resulting in short bowel syndrome and long-life total parenteral nutrition: surgical and clinical outcome. *International Journal of Gynecologic Cancer* 2013;**23**(8):1495-500.

Barlin 2013 *{published data only}*

Barlin JN, Long KC, Tanner EJ, Gardner GJ, Leitao MM, Levine DA, et al. Optimal. *Gynecologic Oncology* 2013;**130**(2):284-8.

Bartl 2018 *{published data only}*

Bartl T, Schwameis R, Stift A, Bachleitner-Hofmann T, Reinthaller A, Grimm C, et al. Predictive and prognostic implication of bowel resections during primary cytoreductive surgery in advanced epithelial ovarian cancer EMT. *International Journal of Gynecological Cancer* 2018;**28**(9):1664-71.

Bertelsen 1990 *{published data only}*

Bertelsen K. Tumor reduction surgery and long-term survival in advanced ovarian cancer: a DACOVA study. *Gynecologic Oncology* 1990;**38**(2):203-9.

Bristow 1999 *{published data only}*

Bristow RE, Montz FJ, Lagasse LD, Leuchter RS, Karlan BY. Survival impact of surgical cytoreduction in stage IV epithelial ovarian cancer. *Gynecologic Oncology* 1999;**72**(3):278-87.

Butler 2012 *{published data only}*

Butler J, Watt J, Brockbank E, Pomel C, Jeyarajah A, Reynolds K, et al. Upper abdominal resections in gynaecological malignancy – the Barts experience. *International Journal Gynecologic Cancer* 2012;**1**:E1014.

Cai 2007 *{published data only}*

Cai HB, Zhou YF, Chen HZ, Hou HY. The role of bowel surgery with cytoreduction for epithelial ovarian cancer. *Clinical Oncology (Royal College of Radiologists)* 2007;**19**(10):757-62.

Campos 2014 *{published data only}*

Campos Cascales P, Gil Jose, Parrilla Pascual. Morbidity and mortality outcomes of cytoreductive surgery and hyperthermic intraperitoneal chemotherapy in patients with primary and recurrent advanced ovarian cancer. *European Journal of Surgical Oncology* 2014;**40**(8):970-5.

Canlorbe 2018 *{published data only}*

Canlorbe G, Touboul C, Chargari C, Bentivegna E, Maulard A, Pautier P, et al. Transitory stoma at the time of complete cytoreductive surgery affects survival for patients with advanced-stage ovarian cancer. *Anticancer research* 2018;**38**(3):1517-23.

Chang 2012b *{published data only}*

Chang SJ, Bristow RE, Ryu HS. Prognostic significance of systematic lymphadenectomy as part of primary debulking

surgery in patients with advanced ovarian cancer. *Gynecologic Oncology* 2012;**126**(3):381-6.

Chereau 2011 *{published data only}*

Chereau E, Ballester M, Selle F, Rouzier R, Darai E. Ovarian cancer in the elderly: impact of surgery on morbidity and survival. *European Journal of Surgical Oncology* 2011;**37**(6):537-42.

Chi 2004 *{published data only}*

Chi DS, Franklin CC, Levine DA, Akselrod F, Sabbatini P, Jarnagin WR, et al. Improved optimal cytoreduction rates for stages IIIC and IV epithelial ovarian, fallopian tube, and primary peritoneal cancer: a change in surgical approach. *Gynecologic Oncology* 2004;**94**(3):650-4.

Chi 2009 *{published data only}*

Chi DS, Zivanovic O, Kolev V, Huh J, Joseph D, Leitao MM, et al. Incidence of major surgical complications after the performance of extensive upper abdominal surgical procedures during primary cytoreduction of advanced ovarian, tubal and peritoneal carcinomas. *Gynecologic Oncology* 2009;**112**(2):S2-3.

Chua 2011 *{published data only}*

Chua TC, Liauw W, Saxena A, Al-Mohaimeed K, Fransi S, Zhao J, et al. Evolution of locoregional treatment for peritoneal carcinomatosis: single-center experience of 308 procedures of cytoreductive surgery and perioperative intraperitoneal chemotherapy. *American Journal of Surgery* 2011;**201**(2):149-56.

Clark 2012 *{published data only}*

Clark RM, Growdon WB, Wiechert A, del Carmen MG, Goodman A, Boruta DM, et al. Hospital readmission after surgical cytoreduction for epithelial ovarian carcinoma: an assessment of risk. *Journal of Clinical Oncology* 2012;**30**:15.

Clark 2014 *{published data only}*

Clark RM, Clemmer JT, Melamed A, Rauh-Hain JA, Joseph N, Boruta DM, et al. Primary debulking surgery in stage IIIC and IV ovarian cancer results in improved survival compared to those undergoing neoadjuvant chemotherapy with interval cytoreduction. *Gynecologic Oncology* 2014;**133**:91-2.

Cormier 2012 *{published data only}*

Cormier B, Long K, Ducie J, Tanner E, Wadhawan I, Jewell E, et al. Do patients with complete gross resection of advanced stage ovarian cancer benefit from lymphadenectomy? *Gynecologic Oncology* 2012;**125**:S24.

Cummins 2019 *{published data only}*

Cummins C, Patrick H, Long J, Kumar S, Sundar S, Bramley G. Ultraradical ovarian cancer surgery comparative clinical effectiveness. *International Journal of Technology Assessment in Health Care* 2019;**35**:97.

Davies 2019 *{published data only}*

Davies J, Asher V, Bali A, Abdul S, Gomez D, Tou S, et al. Impact of splenectomy on survival in advanced ovarian cancer (AOC) in a propensity matched cohort. *International Journal of Gynecological Cancer* 2019;**29**:A508-9.

Eisenhauer 2006 *{published data only}*

Eisenhauer EL, Abu-Rustum NR, Sonoda Y, Levine DA, Poynor EA, Aghajanian C, et al. The addition of extensive upper abdominal surgery to achieve optimal cytoreduction improves survival in patients with stages IIIC–IV epithelial ovarian cancer. *Gynecologic Oncology* 2006;**103**(3):1083-90.

Eisenkop 1993 *{published data only}*

Eisenkop SM, Nalick RH, Wang HJ, Teng NN. Peritoneal implant elimination during cytoreductive surgery for ovarian cancer: impact on survival. *Gynecologic Oncology* 1993;**51**(2):224-9.

Eisenkop 1998 *{published data only}*

Eisenkop SM, Friedman RL, Wang HJ. Complete cytoreductive surgery is feasible and maximizes survival in patients with advanced epithelial ovarian cancer: a prospective study. *Gynecologic Oncology* 1998;**69**(2):103-8.

Eisenkop 2001 *{published data only}*

Eisenkop SM, Spirtos NM. Procedures required to accomplish complete cytoreduction of ovarian cancer: is there a correlation with "biological aggressiveness" and survival? *Gynecologic Oncology* 2001;**82**(3):435-41.

Eisenkop 2003 *{published data only}*

Eisenkop SM, Spirtos NM, Friedman RL, Lin WC, Pisani AL, Perticucci S. Relative influences of tumor volume before surgery and the cytoreductive outcome on survival for patients with advanced ovarian cancer: a prospective study. *Gynecologic Oncology* 2003;**90**(2):390-6.

Eisenkop 2006 *{published data only}*

Eisenkop SM, Spirtos NM, Lin WC. Splenectomy in the context of primary cytoreductive operations for advanced epithelial ovarian cancer. *Gynecologic Oncology* 2006;**100**(2):344-8.

Elgamal 2019 *{published data only}*

Elgamal M, Saha S, Cherry M, Saharan V, Buttar R, Wiese D, et al. Prognostic implications of lymph node metastasis in advanced ovarian cancer: analysis of the National Cancer Database 2006 to 2014. Journal of Clinical Oncology 2019;**37**(15 Suppl):e17042-e17042.

Eng 2018 *{published data only}*

Eng OS, Raoof M, Yu X, Lee SJ, Han ES, Wakabayashi MT, et al. Outcomes after upper abdominal debulking in ovarian malignancies. *Annals of Surgical Oncology* 2018;**12**(1):S138-9.

Eoh 2017 *{published data only}*

Eoh KJ, Lee JY, Yoon JW, Nam EJ, Kim S, Kim SW, et al. Role of systematic lymphadenectomy as part of primary debulking surgery for optimally cytoreduced advanced ovarian cancer: reappraisal in the era of radical surgery. *Oncotarget* 2017;**8**(23):37807-16.

Falconer 2020 *{published data only}*

Falconer H, Joneborg U, Krawiec K, Palsdottir K, Bottai M, Salehi S. Ultra-radical upfront surgery does not improve survival in women with advanced epithelial ovarian cancer; a natural experiment in a complete population. *Gynecologic Oncology* 2020;**159**(1):58-65.

Favero 2014 *{published data only}*

Favero G, Macerox N, Pfiffer T, Ribeiro A, Miranda VC, Diz M, et al. Laparoscopic versus laparotomic cytoreduction in patients with advanced ovarian cancer submitted to NACT: evaluation of oncologic safety. *Journal of Clinical Oncology* 2014;**32**:15.

Ferrero 2014 *{published data only}*

Ferrero A, Ditto A, Giorda G, Gadducci A, Greggi S, Daniele A, et al. Secondary cytoreductive surgery for isolated lymph node recurrence of epithelial ovarian cancer: a multicenter study. *European Journal of Surgical Oncology* 2014;**40**(7):891-8.

Filippova 2019 *{published data only}*

Filippova OT, Broach V, Gardner GJ, Sonoda Y, Zivanovic O, Chi DS, et al. Trends in specific procedures performed at the time of cytoreduction for ovarian cancer: is interval debulking surgery truly less radical? *Gynecologic Oncology* 2019;**154**:142-3.

Fotopoulou 2012 *{published data only}*

Fotopoulou C, Braicu I, Vergote IB, Cadron I, Amant F, Chekerov R, et al. Interval versus primary tumor debulking surgery in advanced ovarian cancer: analysis of the European OVCAD data. *Journal of Clinical Oncology* 2012;**30**:5071.

Gockley 2019 *{published data only}*

Gockley AA, Fiascone S, Hicks Courant K, Pepin K, Del Carmen M, Clark RM, et al. Clinical characteristics and outcomes after bowel surgery and ostomy formation at the time of debulking surgery for advanced-stage epithelial ovarian carcinoma. *International Journal of Gynecological Cancer* 2019;**29**(3):585-92.

Gremeau 2014 *{published data only}*

Gremeau AS, Bourdel N, Jardon K, Rabischong B, Mage G, Pouly JL, et al. Surgical management of non-epithelial ovarian malignancies: advantages and limitations of laparoscopy. *European Journal of Obstetrics & Gynecology and Reproductive Biology* 2014;**172**:106-10.

Grimm 2017 *{published data only}*

Grimm C, Harter P, Alesina PF, Prader S, Schneider S, Ataseven B, et al. The impact of type and number of bowel resections on anastomotic leakage risk in advanced ovarian cancer surgery. *Gynecologic Oncology* 2017;**146**(3):498-503.

Guyon 2014 *{published data only}*

Guyon F, Doublier M, Quincy P, Babin G, Floquet A. Surgical management of advanced ovarian cancer: what does the plasmajet brings? *International Journal of Gynecological Cancer* 2014;**24**:416.

Hamilton 2011 *{published data only}*

Hamilton CA, Miller A, Miller C, Krivak TC, Farley JH, Chernofsky MR, et al. The impact of disease distribution on survival in patients with stage III epithelial ovarian cancer cytoreduced to microscopic residual: a Gynecologic Oncology Group study. *Gynecologic Oncology* 2011;**122**(3):521-6.

Hudry 2013 *{published data only}*

Hudry D, Cannone F, Houvenaeghel G, Buttarelli M, Jauffret C, Chéreau E, et al. Comparison of single-port laparoscopy and conventional laparoscopy for extraperitoneal para-aortic lymphadenectomy. *Surgical Endoscopy* 2013;**27**(11):4319-24.

Hwang 2014 *{published data only}*

Hwang EC, Hwang I, Jung SI, Kang TW, Kwon DD, Heo SH, et al. Prognostic factors for recurrence-free and overall survival after adrenalectomy for metastatic carcinoma: a retrospective cohort pilot study. *BMC Urology* 2014;**14**(1):41.

Janda 2014 *{published data only}*

Janda M, Graves N, Bauer J, Baker J, Obermair A. Quality of life and nutritional status after early enteral feeding versus standard care after surgery for advanced epithelial ovarian cancer. *International Journal of Gynecological Cancer* 2014;**24**(9 Suppl 4):45.

Jiang 2013 *{published data only}*

Jiang R, Yin S, Liu D, Wu X, Wang H, Li Z, et al. Extensive upper abdominal surgery for bulky stage IIIC and IV ovarian cancer: is it just a "belief"? 18th International Meeting of the European Society of Gynaecological Oncology, ESGO; 2013 Oct 19-22; Liverpool (UK).

Kato 2013a *{published data only}*

Kato K, Tate S, Nishikimi K, Shozu M. Assessment of intraoperative tube thoracostomy after diaphragmatic resection as part of debulking surgery for primary advanced-stage Müllerian cancer. *Gynecologic Oncology* 2013;**131**(1):32-5.

Kato 2013b *{published data only}*

Kato K, Tate S, Nishikimi K, Shozu M. Bladder function after modified posterior exenteration for primary gynecological cancer. *Gynecologic Oncology* 2013;**129**(1):229-33.

Kehoe 2013 *{published data only}*

Kehoe S, Hook J, Nankivell M, Jayson GC, Kitchener HC, Lopes T, et al. Chemotherapy or upfront surgery for newly diagnosed advanced ovarian cancer: results from the MRC CHORUS trial. *Journal of Clinical Oncology* 2013;**31**:15.

Kim 2011 *{published data only}*

Kim HS, Kim EN, Jeong SY, Chung HH, Kim YB, Kim JW, et al. Comparison of the efficacy of low anterior resection with primary anastomosis and Hartmann's procedure in advanced primary or recurrent epithelial ovarian cancer. *European Journal of Obstetrics & Gynecology and Reproductive Biology* 2011;**156**(2):194-8.

Kolev 2014 *{published data only}*

Kolev V, Pereira EB, Schwartz M, Sarpel U, Roayaie S, Labow D, et al. The role of liver resection at the time of secondary cytoreduction in patients with recurrent ovarian cancer. *International Journal of Gynecologic Cancer* 2014;**24**(1):70-4.

Kommoss 2010 *{published data only}*

Kommoss S, Rochon J, Harter P, Heitz F, Grabowski JP, Ewald-Riegler N. Prognostic impact of additional extended surgical procedures in advanced-stage primary ovarian cancer. *Annals of Surgical Oncology* 2010;**17**:279-86.

Kristensen 2014 *{published data only}*

Kristensen GS, Schledermann D, Mogensen O, Jochumsen KM. The value of random biopsies, omentectomy, and hysterectomy in operations for borderline ovarian tumors. *International Journal of Gynecologic Cancer* 2014;**24**(5):874-9.

Kuhn 1998 *{published data only}*

Kuhn W, Florack G, Roder J, Schmalfeldt B, Pache L, Rust M, et al. The influence of upper abdominal surgery on perioperative morbidity and mortality in patients with advanced ovarian cancer FIGO III and FIGO IV. *International Journal of Gynecological Cancer* 1998;**8**(1):56-63.

Laios 2019 *{published data only}*

Laios A, Kaufmann A, Otify M, Broadhead T, Hutson R, Nugent D, et al. The prognostic role of age, timing of surgery and surgical radicability in patients with advanced stage epithelial ovarian cancer; a single institution experience. *International Journal of Gynecological Cancer* 2019;**119**:A485.

Lee 2017 *{published data only}*

Lee Y, Lu L, Xu W, Brown T, May T. Investigating the impact of interval from primary debulking surgery to initiation of adjuvant chemotherapy in advanced high-grade serous ovarian cancer. *Gynecologic Oncology* 2017;**145**:185.

Li 2014 *{published data only}*

Li X, Xing H, Li L, Huang Y, Zhou M, Liu Q, et al. Clinical significance of para-aortic lymph node dissection and prognosis in ovarian cancer. *Frontiers of medicine* 2014;**8**(1):96-100.

Liberale 2019 *{published data only}*

Liberale G, Pop CF, Polastro L, Kerger J, Moreau M, Chintinne M, et al. A radical approach to achieve complete cytoreductive surgery improve survival of patients with advanced ovarian cancer. *Journal of Visceral Surgery* 2019;**157**:79-86.

Liu 2013a *{published data only}*

Liu Q, Ding X, Yang J, Cao D, Shen K, Lang J, et al. The significance of comprehensive staging surgery in malignant ovarian germ cell tumors. *Gynecologic Oncology* 2013;**131**(3):551-4.

Liu 2013b *{published data only}*

Liu Q, Ding XL, Yang JX, Cao DY, Shen K, Lang JH, et al. Multicenter randomized controlled clinical study for the operative treatment of malignant ovarian germ cell tumors. *Zhonghua Fu Chan Ke za Zhi* 2013;**48**(3):188-92.

Martinez 2014 *{published data only}*

Martinez A, Picaud L, Luyckx M, Pomel C, Leblanc E, Morice P, et al. Surgical approach and survival in elderly patients with advanced ovarian cancer. *Journal of Ovarian Research* 2014;**1**:297-8.

McCann 2011 *{published data only}*

McCann CK, Growdon WB, Munro EG, Del Carmen MG, Boruta DM, Schorge JO, et al. Prognostic significance of splenectomy as part of initial cytoreductive surgery in ovarian cancer. *Annals of Surgical Oncology* 2011;**18**(10):2912-8.

Muallem 2018 *{published data only}*

Muallem MZ. Total retroperitoneal en bloc resection of peritoneal-multi visceral packet by advanced ovarian cancer. *International Journal of Gynecological Cancer* 2018;**30**:648-53.

Oseledchyk 2016 *{published data only}*

Oseledchyk A, Hunold LE, Mallmann MR, Domrose CM, Abramian A, Debald M, et al. Impact of extended primary surgery on suboptimally operable patients with advanced ovarian cancer. *International Journal of Gynecological Cancer* 2016;**26**(5):873-83.

Oshita 2013 *{published data only}*

Oshita T, Itamochi H, Nishimura R, Numa F, Takehara K, Hiura M, et al. Clinical impact of systematic pelvic and para-aortic lymphadenectomy for pT1 and pT2 ovarian cancer: a retrospective survey by the Sankai Gynecology Study Group. *International Journal of Clinical Oncology* 2013;**18**(6):1107-13.

Park 2011 *{published data only}*

Park NH, Kim MK, Kim JW, Song YS, Kang SB. Effect of lymphadenectomy on survival outcomes in advanced epithelial ovarian cancer. 17th International Meeting of the European Society of Gynaecological Oncology, ESGO; 2011 Sep 11-14; Milan (Italy).

Pathiraja 2011 *{published data only}*

Pathiraja PN, Eltz S, Garruto R, Spain G, Martinek I, Tozzi R. Evaluation surgical morbidity of diaphragmatic surgery with/without pleurectomy to achieve optimum cytoreduction in ovarian cancer. 17th International Meeting of the European Society of Gynaecological Oncology, ESGO, 11-14 September 2011. In: 17th International Meeting of the European Society of Gynaecological Oncology, ESGO; 2011 Sep 11-14; Milan (Italy).

Pathiraja 2013 *{published data only}*

Pathiraja PN, Garruto-Campanile R, Tozzi R. Diaphragmatic peritonectomy versus full thickness diaphragmatic resection and pleurectomy during cytoreduction in patients with ovarian cancer. *International Journal of Surgical Oncology* 2013;**2013**:6.

Pelissier 2018 *{published data only}*

Pelissier A, Franke O, Darai E, Houvenaeghel G, Chereau E, Rouzier R. Value of diaphragmatic surgery during interval debulking surgery. *Anticancer Research* 2018;**38**(1):411-6.

Perri 2013 *{published data only}*

Perri T, Ben-Baruch G, Kalfon S, Beiner ME, Helpman L, Hogen LB, et al. Abdominopelvic cytoreduction rates and recurrence sites in stage IV ovarian cancer: is there a case for thoracic cytoreduction? *Gynecologic Oncology* 2013;**131**(1):27-31.

Phillips 2018 *{published data only}*

Phillips A, Sundar S, Singh K, Pounds R, Nevin J, Kehoe S, et al. The NICE classification for 'Ultra-radical (extensive) surgery for advanced ovarian cancer' guidance does not meaningfully predict postoperative complications: a cohort study. *BJOG* 2018;**126**:96-104.

Pushpalatha 2011 *{published data only}*

Pushpalatha K, Kumar KS, Kumar L, Julka PK, Mathur S. Impact of pelvic lymphadenectomy during primary cytoreductive surgery on survival in epithelial ovarian cancer: a prospective study. *Journal of Clinical Oncology* 2011;**29**(15 Suppl):e15563.

Qin 2012 *{published data only}*

Qin JC, Yang ZJ, Xiong L, Li L. Systematic lymphadenectomy for overall survival in epithelial ovarian cancer: a meta-analysis. *Chinese Journal of Evidence-Based Medicine* 2012;**12**(2):224-30.

Ratnavelu 2014 *{published data only}*

Ratnavelu N, Biliatis I, Patel A, Founta C, Kucukmetin A, Naik R. Total colectomy to achieve cytoreduction in advanced ovarian cancer: a three-arm matched cohort study. *Gynecologic Oncology* 2014;**114**:450-1.

Ren 2015 *{published data only}*

Ren Y, Jiang R, Yin S, You C, Liu D, Cheng X, et al. Radical surgery versus standard surgery for primary cytoreduction of bulky stage IIIC and IV ovarian cancer: an observational study. *BMC Cancer* 2015;**15**(1):583.

Rodriguez 2012 *{published data only}*

Rodriguez N, Miller A, Richard S, Rungruang B, Hamilton C, Bookman M, et al. Upper abdominal procedures in advanced stage ovarian or primary peritoneal carcinoma patients with minimal or no gross residual disease: an analysis of GOG 182. 43rd Annual Meeting of the Society of Gynecologic Oncology; 2012 Mar 24-27; Austin (TX).

Rodriguez 2013 *{published data only}*

Rodriguez N, Miller A, Richard SD, Rungruang B, Hamilton CA, Bookman MA, et al. Upper abdominal procedures in advanced stage ovarian or primary peritoneal carcinoma patients with minimal or no gross residual disease: an analysis of Gynecologic Oncology Group (GOG) 182. *Gynecologic Oncology* 2013;**130**(3):487-92.

Rouzier 2010 *{published data only}*

Rouzier R, Bergzoll C, Brun JL, Dubernard G, Selle F, Uzan S, et al. The role of lymph node resection in ovarian cancer: analysis of the Surveillance, Epidemiology, and End Results (SEER) database. *BJOG* 2010;**117**(12):1451-8.

Sagara 2019 *{published data only}*

Sagara A, Motohara T, Iwagoi Y, Saito F, Takaishi K, Miyahara Y, et al. Survival impact of wide resection of the pelvic peritoneum in patients with epithelial ovarian cancer. *International Journal of Gynecological Cancer* 2019;**29**:A515.

Sandadi 2014 *{published data only}*

Sandadi S, Long K, Andikyan V, Vernon J, Zivanovic O, Eisenhauer EL, et al. Postoperative outcomes among patients undergoing thoracostomy tube placement at time of diaphragm peritonectomy or resection during primary cytoreductive surgery for ovarian cancer. *Gynecologic Oncology* 2014;**132**(2):299-302.

Scalici 2014 *{published data only}*

Scalici JM, DeCotis-Smith D, Wang B, Finan MA, Rocconi RP. A comparative analysis of the treatment strategies for advanced ovarian cancer. *Gynecologic Oncology* 2014;**133**:60-1.

Sehouli 2010 *{published data only}*

Sehouli J, Savvatis K, Braicu EI, Schmidt SC, Lichtenegger W, Fotopoulou C. Primary versus interval debulking surgery in advanced ovarian cancer: results from a systematic single-center analysis. *International Journal of Gynecologic Cancer* 2010;**20**(8):1331-40.

Soo Hoo 2015 *{published data only}*

Soo Hoo S, Marriott N, Houlton A, Nevin J, Balega J, Singh K, et al. Patient-reported outcomes after extensive (ultraradical) surgery for ovarian cancer: results from a prospective longitudinal feasibility study. *International Journal of Gynecological Cancer* 2015;**25**(9):1599-607.

Stefanović 2011 *{published data only}*

Stefanović A, Jeremić K, Kadija S, Milincić N, Mircić A, Petković S, et al. Intestinal surgery in treatment of advanced ovarian cancer review of our experience. *European Journal of Gynaecological Oncology* 2011;**32**(4):419-22.

Sundar 2014 *{published data only}*

Sundar S. Quality of life in patients undergoing extensive (radical/ ultraradical surgery) for ovarian cancer – the SOCQER-1 study. *International Journal of Gynecological Cancer* 2014;**9**:15.

Sundar 2018 *{published data only}*

Sundar S, Kumar S, Long J, Balega J, Fotopoulou C, Broadhead T, et al. Patient reported outcomes (PRO) after surgery in advanced ovarian cancer-initial results from the international, prospective, multicenter SOCQER 2 study. *International Journal of Gynecological Cancer* 2018;**24**:604-6.

Sundar 2019 *{published data only}*

Sundar S, Cummins C, Kumar S, Long J, Arora V, Balega J, et al. Quality of life after surgery of varying surgical complexity in advanced ovarian cancer: results from the international, prospective, multicenter cohort SOCQER2 study. *International Journal of Gynecological Cancer* 2019;**24**:A638-9.

Suzuki 2008 *{published data only}*

Suzuki S, Kajiyama H, Shibata K, Ino K, Nawa A, Sakakibara K, et al. Is there any association between retroperitoneal lymphadenectomy and survival benefit in ovarian clear cell carcinoma patients? *Annals of Oncology* 2008;**19**(7):1284-7.

Szczesny 2016 *{published data only}*

Szczesny W, Vistad I, Kaern J, Nakling J, Trope C, Paulsen T. Impact of hospital type and treatment on long-term survival among patients with FIGO Stage IIIC epithelial ovarian cancer: follow-up through two recurrences and three treatment lines in search for predictors for survival. *European Journal of Gynaecological Oncology* 2016;**37**(3):305-11.

Tozzi 2019 *{published data only}*

Tozzi R, Casarin J, Baysal A, Garruto-Campanile R, Majd HS, Kilic Y, et al. Morbidity of multiple bowel resection compared

to single bowel resection after debulking surgery for ovarian cancer. *BJOG* 2019;**126**:98.

Tsolakidis 2010a *{published data only}*

Tsolakidis D, Amant F, Van Gorp T, Leunen K, Neven P, Vergote I. The role of diaphragmatic surgery during interval debulking after neoadjuvant chemotherapy. *International Journal of Gynecological Cancer* 2010;**20**:542-51.

Tsolakidis 2010b *{published data only}*

Tsolakidis D, Amant F, van Gorp T, Leunen K, Neven P, Vergote I. Diaphragmatic surgery during primary debulking in 89 patients with stage IIIB-IV epithelial ovarian cancer. *Gynecologic Oncology* 2010;**116**:489-96.

van de Laar 2014 *{published data only}*

van de Laar R, Zusterzeel PL, van Gorp T, Buist MR, van Driel WJ, Gaarenstroom KN, et al. Cytoreductive surgery followed by chemotherapy versus chemotherapy alone for recurrent platinum-sensitive epithelial ovarian cancer (SOCceR trial): a multicenter randomised controlled study. *BMC Cancer* 2014;**14**(1):22.

Vidal 2016 *{published data only}*

Vidal F, Al Thani H, Haddad P, Luyckx M, Stoeckle E, Morice P, et al. Which surgical attitude to choose in the context of non-resectability of ovarian carcinomatosis: beyond gross residual disease considerations. *Annals of Surgical Oncology* 2016;**23**(2):434-42.

von Hugo 1989 *{published data only}*

von Hugo R, Holscher M, Janicke F. Morbidity, mortality and quality of life following radical surgical interventions in advanced ovarian cancer. *Archives of Gynecology & Obstetrics* 1989;**245**(1-4):625-7.

Wallace 2016 *{published data only}*

Wallace S, Kumar A, Mc Gree M, Weaver A, Cliby W. Residual disease after primary cytoreduction in stage IIIC ovarian cancer; is aggressive cytoreduction worth it? *Gynecologic Oncology* 2016;**143**(1):206.

Wat 2012 *{published data only}*

Wat J, Butler J, Brockbank E, Pomel C, Jeyarajah A, Rosenthal A, et al. Upper abdominal resections in gynaecological malignancy – the Barts experience. *Official Journal of the International Hepato Pancreato Biliary Association* 2012;**32**:E1014.

Wimberger 2007 *{published data only}*

Wimberger P, Lehmann N, Kimmig R, Burges A, Meier W, Du Bois A. Prognostic factors for complete debulking in advanced ovarian cancer and its impact on survival. An exploratory analysis of a prospectively randomized phase III study of the Arbeitsgemeinschaft Gynaekologische Onkologie Ovarian Cancer Study Group (AGO-OVAR). *Gynecologic Oncology* 2007;**106**(1):69-74.

Wright 2012 *{published data only}*

Wright JD, Herzog TJ, Neugut AI, Burke WM, Lu YS, Lewin SN, et al. Effect of radical cytoreductive surgery on omission and delay of chemotherapy for advanced-stage ovarian cancer. *Obstetrics & Gynecology* 2012;**120**(4):871-81.

Yildirim 2014 *{published data only}*

Yildirim Y, Ertas IE, Nayki U, Ulug P, Nayki C, Yilmaz I, et al. En-bloc pelvic resection with concomitant rectosigmoid colectomy and immediate anastomosis as part of primary cytoreductive surgery for patients with advanced ovarian cancer. *European Journal of Gynaecological Oncology* 2014;**35**(4):400-7.

Zamurovic 2013 *{published data only}*

Zamurovic M, Soldo V, Cutura N. Survival rate analysis of patients with advanced ovarian cancer treated according to different protocols. *Internet Journal of Oncology* 2013;**9**(1):629-36.

Zapardiel 2012 *{published data only}*

Zapardiel I, Peiretti M, Zanagnolo V, Biffi R, Bocciolone L, Landoni F, et al. Splenectomy as part of primary cytoreductive surgery for advanced ovarian cancer: a retrospective cohort study. *International Journal of Gynecological Cancer* 2012;**22**(6):968-73.

Additional references

Altman 1995

Altman DG, de Stavola BL, Love SB, Stepniewska KA. Review of survival analyses published in cancer journals. *British Journal of Cancer* 1995;**72**:511-8.

American Cancer Society 2020

American Cancer Society. Cancer facts & figures 2020. www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2020/cancer-facts-and-figures-2020.pdf (last accessed 5 August 2022).

Bast 2020

Bast RC Jr, Lu Z, Han CY, Lu KH, Anderson KS, Drescher CW, et al. Biomarkers and strategies for early detection of ovarian cancer. *Cancer Epidemiology, Biomarkers & Prevention* 2020;**29**(12):2504-12.

Benedetti Panici 2015

Benedetti Panici P, Donato Di V, Fischetti M, Casorelli A, Perniola G, Musella A, et al. Predictors of postoperative morbidity after cytoreduction for advanced ovarian cancer: analysis and management of complications in upper abdominal surgery. *Gynecologic Oncology* 2015;**137**(3):406-11.

Berek 2018

Berek JS, Kehoe ST, Kumar L, Friedlander M. Cancer of the ovary, fallopian tube, and peritoneum. *International Journal of Gynaecology and Obstetrics* 2018;**143**:59-78.

Bristow 2002

Bristow RE, Tomacruz RS, Armstrong DK, Trimble EL, Montz FJ. Survival effect of maximal cytoreductive surgery for advanced ovarian carcinoma during the platinum era: a meta-analysis. *Journal of Clinical Oncology* 2002;**20**:1248-59.

Bristow 2003

Bristow RE, del Carmen MG, Kaufman HS, Montz FJ. Radical oophrectomy with primary stapled colorectal anastomosis for resection of locally advanced epithelial ovarian cancer. *Journal of the American College of Surgeons* 2003;**197**:565-74.

Bryant 2021

Bryant A, Hiu S, Kunonga P, Gajjar K, Craig D, Vale L, et al. Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery. *Cochrane Database of Systematic Reviews* 2021, Issue 9. Art. No: CD015048. [DOI: 10.1002/14651858.CD015048]

Chen 1985

Chen SS, Bochner R. Assessment of morbidity and mortality in primary cytoreductive surgery for advanced ovarian carcinoma. *Gynecologic Oncology* 1985;**20**:190-5.

Colombo 2019

Colombo N, Sessa C, du Bois A, Ledermann J, McCluggage WG, McNeish I, et al. ESMO-ESGO consensus conference recommendations on ovarian cancer: pathology and molecular biology, early and advanced stages, borderline tumours and recurrent disease. *Annals of Oncology* 2019;**30**(5):672-705.

Covens 2000

Covens AL. A critique of surgical cytoreduction in advanced ovarian cancer. *Gynecologic Oncology* 2000;**78**:269-74.

Crawford 2005

Crawford SC, Vasey PA, Paul J, Hay A, Davis JA, Kaye SB. Does aggressive surgery only benefit patients with less advanced ovarian cancer? Results from an international comparison within the SCOTROC-1 Trial. *Journal of Clinical Oncology* 2005;**23**:8802-11.

CRUK 2018

Cancer Research UK. Types of ovarian cancer. www.cancerresearchuk.org/about-cancer/ovarian-cancer/types (last accessed 5 August 2022).

CTCAE 2017

CTCAE. Common terminology criteria for adverse events (CTCAE) v5.0, 2017. ctep.cancer.gov/protocolDevelopment/electronic_applications/docs/CTCAE_v5_Quick_Reference_5x7.pdf (last accessed 5 August 2017).

Deeks 2001

Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, Altman DG, editors(s). Systematic Reviews in Health Care: Meta-Analysis in Context (2nd edition). London (UK): BMJ Publication Group, 2001.

DerSimonian 1986

DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986;**7**:177-88.

EUROCARE 2015

Sant M, Chirlaque Lopez MD, Agresti R, Sánchez Pérez MJ, Holleczek B, Bielska-Lasota M, et al, EUROCARE-5 Working Group. Survival of women with cancers of breast and genital organs in Europe 1999–2007: results of the EUROCARE-5 study. *European Journal of Cancer* 2015;**51**(15):2191-205.

GLOBOCAN 2018

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a Cancer Journal for Clinicians* 2018;**68**(6):394-424.

Goodacre 2015

Goodacre S. Uncontrolled before-after studies: discouraged by Cochrane and the EMJ. *Emergency Medicine Journal* 2015;**32**:507-8.

GRADE Working Group 2004

GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;**328**:1490-4.

Harley 2014

Harley IJ, Quinn J, Beirne JP, McCluggage WG. The distal fallopian tube as the origin of non-uterine pelvic high-grade serous carcinomas; scientific paper No. 44, 2014. www.rcog.org.uk/media/s0nlcrlx/sip44hgscs.pdf (last accessed 5 August 2022).

Harter 2019

Harter P, Sehouli J, Lorusso D, Reuss A, Vergote I, Marth C, et al. A randomized trial of lymphadenectomy in patients with advanced ovarian neoplasms. *New England Journal of Medicine* 2019;**380**(9):822-32.

Higgins 2003

Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**:557-60.

Higgins 2019

Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Chapter 8: Assessing risk of bias in included studies. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 6.0 (updated August 2019). Cochrane, 2019. training.cochrane.org/handbook/archive/v6.

Horowitz 2015

Horowitz NS, Miller A, Rungruang B, Richard SD, Rodriguez N, Bookman MA, et al. Does aggressive surgery improve outcomes? Interaction between preoperative disease burden and complex surgery in patients with advanced-stage ovarian cancer: an analysis of GOG 182. *Journal of Clinical Oncology* 2015;**33**(8):937-43.

Hoskins 1992

Hoskins WJ, Bundy BN, Thigpen JT, Omura GA. The influence of cytoreductive surgery on recurrence-free survival in small-volume stage III epithelial ovarian cancer: a Gynecologic Oncology Group study. *Gynecological Oncology* 1992;**47**:159-66.

Jaeger 2001

Jaeger W, Ackermann S, Kessler H, Katalinic A, Lang N. The effect of bowel resection on survival in advanced epithelial ovarian cancer. *Gynecologic Oncology* 2001;**83**:286-91.

Johnson 2008

Johnson NP, Selman T, Zamora J, Khan KS. Gynaecologic surgery from uncertainty to science: evidence-based surgery is no passing fad. *Human Reproduction* 2008;**23**(4):832-9.

Kehoe 2015

Kehoe S, Hook J, Nankivell M, Jayson GC, Kitchener H, Lopes T, et al. Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer (CHORUS): an open-label, randomised, controlled, non-inferiority trial. *Lancet* 2015;**386**(9990):249-57.

Kirby 2020

Kirby T. Confronting a rare ovarian cancer during lockdown. *Lancet Respiratory Medicine* 2020;**8**(12):1176-8.

Kumar 2019

Kumar S, Long J, Kehoe S, Sundar S, Cummins C. Quality of life outcomes following surgery for advanced ovarian cancer: a systematic review and meta-analysis. *International Journal of Gynecological Cancer* 2019;**29**(8):1285-91.

Kurman 2008

Kurman RJ, Visvanathan K, Roden R, Wu TC, Shih IM. Early detection and treatment of ovarian cancer: shifting from early stage to minimal volume of disease based on a new model of carcinogenesis. *American Journal of Obstetrics and Gynecology* 2008;**198**(4):351-6.

Kurman 2014

Kurman RJ, Carcangiu ML, Herrington CS, Young RH. WHO Classification of Tumours of Female Reproductive Organ. Lyon (France): IARC Press, 2014.

Kyriacou 2016

Kyriacou DN, Lewis RJ. Confounding by indication in clinical research. *JAMA* 2016;**316**(17):1818-9.

Lancet 2007

Editorial. An experiment in earlier detection of ovarian cancer. *Lancet* 2007;**369**(9579):2051.

Markman 2007

Markman M. Concept of optimal cytoreduction in advanced ovarian cancer: a brief critique and a call for action. *Journal of Clinical Oncology* 2007;**25**(27):4168-70.

Merideth 2003

Merideth MA, Cliby WA, Keeney GL, Lesnick TG, Nagorney DM, Podratz KC. Hepatic resection for metachronous metastases from ovarian cancer. *Gynecologic Oncology* 2003;**89**:16-21.

Montz 1989

Montz FJ, Schlaerth JB, Berek JS. Resection of diaphragmatic peritoneum and muscle: role in cytoreductive surgery for ovarian cancer. *Gynecologic Oncology* 1989;**35**:338-40.

NICE 2013

National Institute for Health and Care Excellence. Ultra-radical (extensive) surgery for advanced ovarian cancer, 2013. www.nice.org.uk/guidance/ipg470 (last accessed 5 August 2022).

Norell 2020

Norell CH, Butler J, Farrell R, Altman A, Bentley J, Cabasag CJ, et al. Exploring international differences in ovarian cancer treatment: a comparison of clinical practice guidelines and patterns of care. *International Journal of Gynecological Cancer* 2020;**30**(11):1748-56.

Parmar 1998

Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Statistics in Medicine* 1998;**17**(24):2815-34.

PDQ Adult Treatment Editorial Board 2021

PDQ Adult Treatment Editorial Board. Ovarian epithelial, fallopian tube, and primary peritoneal cancer treatment (PDQ®), 2021. www.ncbi.nlm.nih.gov/books/NBK66007/ (last accessed 5 August 2022).

Phillips 2019

Phillips A, Sundar S, Singh K, Pounds R, Nevin J, Kehoe S, et al. The NICE classification for 'Ultra-radical (extensive) surgery for advanced ovarian cancer' guidance does not meaningfully predict postoperative complications: a cohort study. *BJOG* 2019;**126**:96-104.

Pomel 2004

Pomel C, Dauplat J. Management of malignant epithelial tumours of the ovary. *Journal de Chirurgie* 2004;**141**:277-84.

Prat 2014

Prat J, FIGO Committee on Gynecologic Oncology. Staging classification for cancer of the ovary, fallopian tube, and peritoneum. *International Journal of Gynecological Cancer* 2014;**124**(1):1-5.

Rani 2018

Rani G, Bandupadhyay S, Medhi AC, Shafi F. Ovarian cancer screening. *Journal of Medical Science and Clinical Research* 2018;**6**(5):1042-4.

Review Manager 2014 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration Review Manager 5 (RevMan 5). Version 5.4. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Schisterman 2009

Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* 2009;**20**(4):488-95.

Siegel 2020

Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA: a Cancer Journal for Clinicians* 2020;**70**(1):7-30.

Sterne 2016

Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;**355**:i4919.

Sterne 2022

Sterne JA, Hernán MA, McAleenan A, Reeves BC, Higgins JP. Chapter 25: Assessing risk of bias in a non-randomized study. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editor(s).Cochrane Handbook for Systematic Reviews of Interventions Version 6.3 (updated February 2022). Cochrane, 2022. Available from training.cochrane.org/handbook.

Sundar 2021

Sundar S, Cummins C, Kumar S, Long J, Arora V, Balega J, et al. Quality of life from cytoreductive surgery in advanced ovarian cancer: investigating the association between disease burden and surgical complexity in the international, prospective, SOCQER-2 cohort study. *International Journal of Obstetrics and Gynaecology* 2021;**129**:1122-32.

Taggart 2001

Taggart DP, D'Amico R, Altman DG. Effect of arterial revascularisation on survival: a systematic review of studies comparing bilateral and single internal mammary arteries. *Lancet* 2001;**358**(9285):870-5.

Todo 2003

Todo Y, Sakuragi N, Oikawa M, Negishi H, Yamamoto R, Yoshiaki K, et al. Cytoreductive surgery combined with organ resection for advanced ovarian cancer. *International Journal of Clinical Oncology* 2003;**8**:90-6.

van Dam 1996

van Dam PA, Tjalma W, Weyler J. Ultraradical debulking of epithelial ovarian cancer with the ultrasonic surgical aspirator: a prospective randomised trial. *Obstetrics and Gynecology* 1996;**174**:943-50.

Venesmaa 1992

Venesmaa P, Ylikorkala O. Morbidity and mortality associated with primary and repeat operations for ovarian cancer. *Obstetrics and Gynecology* 1992;**79**:168-72.

Vergote 2010

Vergote I, Trope CG, Amant F, Kristensen GB, Ehlen T, Johnson N, et al. Neoadjuvant chemotherapy or primary surgery in stage IIIC or IV ovarian cancer. *New England Journal of Medicine* 2010;**363**(10):943-53. [Incl. Supplementary Appendix and Protocol]

Vergote 2016

Vergote I, Vlayen J, Heus P, Hoogendam JP, Damen JA, van de Wetering F, et al. Ovarian cancer: diagnosis, treatment and follow-up, 2016. Good Clinical Practice (GCP) Brussels: Belgian Health Care Knowledge Centre (KCE). KCE Reports 268. D/2016/10.273/49.

Visintin 2008

Visintin I, Feng Z, Longton G, Ward DC, Alvero AB, Lai Y, et al. Diagnostic markers for early detection of ovarian cancer. *Clinical Cancer Research* 2008;**14**:1065-72.

Webb 2017

Webb PM, Jordan SJ. Epidemiology of epithelial ovarian cancer. *Best Practice & Research. Clinical Obstetrics & Gynaecology* 2017;**41**:3-14.

Zijlstra 2019

Zijlstra M, Timmermans M, Fransen H, van der Aa M, Reyners A, Raijmakers N, et al. Treatment patterns and associated factors in patients with advanced epithelial ovarian cancer: a population-based study. *International Journal of Gynecological Cancer* 2019;**29**:1032-7.

References to other published versions of this review

Ang 2009

Ang C, Chan KK, Naik R, Bryant A, Dickinson HO. Ultraradical surgery for the primary debulking of epithelial ovarian cancer. *Cochrane Database of Systematic Reviews* 2009, Issue 2. Art. No: CD007697. [DOI: 10.1002/14651858.CD007697]

Ang 2011

Ang C, Chan KK, Bryant A, Naik R, Dickinson HO. Ultra-radical (extensive) surgery versus standard surgery for the primary cytoreduction of advanced epithelial ovarian cancer. *Cochrane Database of Systematic Reviews* 2011, Issue 4. Art. No: CD007697. [DOI: 10.1002/14651858.CD007697.pub2]

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Aletti 2006a

Study characteristics	
Methods	Retrospective cohort study of consecutive participants identified from surgical records. Surgery carried out at Mayo Clinic, Minnesota (USA)
Participants	Women with FIGO stage IIIC ovarian cancer, where disease status was extracted from surgical exploration notes.

Aletti 2006a *(Continued)*

	Age at study entry: mean 64.4 years; median 64 years; range 24–87 years All women presented with FIGO stage IIIC: 194/194 (100%) Tumour cell type: serous 126 (64.9%), mucinous: 4 (2.1%), endometrioid: 18 (9.3%), clear cell: 7 (3.6%), mixed: 17 (8.8%), seroanaplastic: 17 (8.8%), Müllerian origin: 2 (1%) Tumour grade: 1: 1 (0.5%), 2: 13 (6.7%), 3: 180 (92.8%) ASA score: 1: 7 (3.6%), 2: 87 (44.8%), 3: 88 (45.4%), 4: 7 (3.6%), unknown: 5 (2.6%) Ascites: mean 2076 mL, median 1000 mL, range 0–12,000 mL Extent of disease: carcinomatosis: 144 (74.2%), diaphragm involvement: 137 (70.6%), mesentery: 138 (71.1%), cul-de-sac: 163 (84%), omentum: 168 (86.6%), ascites: 160 (82.5%) Residual disease: no gross visible: 46 (23.7%); 0–1 cm: 85 (43.8%); 1–2 cm: 22 (11.3%); > 2 cm: 41 (21.1%) Baseline details for 144 women with carcinomatosis were not reported. However, it is known that 68 (47.2%) women underwent ultra-radical surgery and 76 (52.8%) underwent standard surgery.
Interventions	Initial surgery performed for diagnosis, staging and surgical cytoreduction. **Intervention:** ultra-radical surgery: if any diaphragmatic surgery, bowel resection, splenectomy or extensive abdominal peritoneal stripping or resection. **Comparison:** standard surgery: hysterectomy, complete omentectomy, stripping of pelvic peritoneum or limited resection of peritoneal-based nodules.
Outcomes	• Disease-specific overall survival • Perioperative mortality Disease-specific survival: HR for death from advanced epithelial ovarian cancer (adjusted for age, ASA score, carcinomatosis, mesenteric involvement, diaphragmatic involvement, ascites, residual disease and operative time): 0.64 (95% CI 0.40 to 1.04). Provided through personnel communication with the study authors. Median disease-free survival: 15.9 with ultra-radical surgery; 19.3 months with standard surgery; significant; not adjusted
Notes	Follow-up: mean: 3.5 years; median: 2.7 years; range: 0.02–10.5 years Participants were first classified by the extent of peritoneal dissemination. Those with tumour nodules diffusely covering most of the bowel serosa surfaces and the parietal peritoneum of the abdomen and pelvis were classified as having carcinomatosis. In multivariate analysis, only residual disease and radical surgery were independent factors predicting participant survival (Table 4). Quote: "When examining the effect of radical surgery on all patients with carcinomatosis (n = 144), we observed an improved disease-specific overall survival rate (38% versus 9%; log-rank test, P=0.001) favouring patients who underwent radical procedures versus non-radical procedures (Fig. 3)". Quote: "Radical procedures were performed at the same rate regardless of age (49% for age 65 years versus 51% for age 65 years; P = 0.45). Patients with better ASA scores (1 or 2 versus 3 or 4) were more likely to have aggressive procedures performed (59% versus 36%, respectively; P = 0.005), which implies the overall medical condition of the patient at least partially influences the decision to perform aggressive surgery". Quote: "The 5-year disease-specific overall survival rate was 46% compared with 13% for patients with radical and non-radical surgeries, respectively (log-rank test, P = 0.001; Fig. 4A)". Quote: "The rate of optimal resection (residual disease 1 cm) was 84.5% compared with 51% on the basis of surgeon tendency to use radical procedures".

Aletti 2006a *(Continued)*

Quote: "Our division of gynaecologic surgery shares a uniform referral base with similar patient demographics, and we practice at a single institution where each surgeon has access to identical services and nursing support".

Risk of bias

Bias	Authors' judgement	Support for judgement
1. Bias due to confounding (a–d)	High risk	Domain had a critical risk of bias: no known prognostic factors that have potential for confounding of the effect on intervention. Information was collected retrospectively.
2. Selection bias (a)	Low risk	Intervention and follow-up start were simultaneous as a rule for cytoreductive surgery. No evidence of selection into the study due to variables measured after the intervention since participants were included retrospectively.
3. Classification of interventions (a–b)	Low risk	Well-defined surgical interventions based on aggressive surgery (yes versus no).
4. Deviation from intended interventions (a–c)	Unclear risk	No evidence of any deviations from interventions or usual practice but may be due to omission or that deviations did not happen.
5. Bias due to missing data (a–b)	High risk	Domain had moderate-to-high risk of bias: no differential follow-up or missing data reported; no participant selection due to missing data reported. In some respects, there was no reason to believe there was serious bias due to missing data as the study was sound for a non-randomised study with regard to this domain but cannot be considered comparable to a well-performed randomised trial. Therefore, it was sensible to judge the missing data domain at moderate-to-high risk of bias.
6. Measurement of outcomes (a–b)	High risk	Domain had a critical risk of bias: disease-specific survival is not a good outcome measure to use for several reasons, namely the coding of death certificates is notoriously error-prone. If someone dies because of the treatment they receive, this may not be counted as a death from ovarian cancer. But it is just as important to the patient as a death from ovarian cancer and the evaluation of the relative benefits of the treatments should include these deaths. The study authors would have had access to data for death from all causes.
7. Reporting bias (a–c)	High risk	Domain had a critical risk of bias: there is a serious problem in the multivariate analysis. It adjusted for variables that were measured after the time origin, namely extent of residual disease and operative time (Altman 1995). Also data were reported in a subset of the 144 women with carcinomatosis (more-extensive disease) only.

Chang 2012a

Study characteristics

Methods	Retrospective review of medical records. The decision to perform simple or radical procedures was determined by the surgeon.
Participants	Consecutive women with FIGO stage IIIC and IV primary epithelial ovarian, fallopian tube or peritoneal cancer who underwent primary cytoreductive surgery at Ajou University Hospital, Republic of Korea (enrolment 1 January 2000 to 31 December 2011). Age: median 54 years; range 30–78 years

Chang 2012a *(Continued)*

	FIGO stage IIIC: 189 (93.1%); IV: 14 (6.9%) Tumour cell type: serous 167 (82.3%), mucinous: 4 (2.0%), endometrioid: 5 (2.5%), clear cell: 9 (4.4%), mixed: 18 (8.9%) ASA score 1–2: 114 (56.2%); 3–4: 80 (39.4%); 9 not available Tumour grade 1: 26 (12.8%), 2: 72 (35.5%), 3: 100 (49.3%), unknown: 5 Ascites > 100 mL: 92 (54.7%) Peritoneal carcinomatosis 149 (73.4%) Residual disease: no gross visible: 63 (31.0%); 0–1 cm: 67 (37.9%); > 1 cm: 63 (31.0%) Median BMI: 23.3 (range 11.7–35.2) Carcinomatosis: 149 (73.4%) Baseline details not presented according to type of surgery.
Interventions	**Intervention:** radical cytoreductive procedures included radical oophorectomy with or without rectosigmoid colectomy, total omentectomy, multiple bowel resections, diaphragm peritonectomy or resection, liver resection, splenectomy, distal pancreatectomy, and gastric resection. **Comparison:** simple surgery included total abdominal hysterectomy, bilateral salpingo-oophorectomy, peritoneal biopsies or excisions, infracolic omentectomy, pelvic lymphadenectomy, para-aortic lymphadenectomy, and segmental resection of small bowel. After surgery, all participants received adjuvant platinum-based chemotherapy plus paclitaxel for 6–9 cycles.
Outcomes	Overall survival, progression-free survival and adverse events Ultra-radical versus standard surgery Median operative time (minutes): 307 versus 235; P < 0.01 Median estimated blood loss (mL): 800 versus 500; P = 0.03 Intra- or postoperative blood transfusion: 25% versus 17.6%; P = 0.01 Median stay in intensive care unit (days): 1.5 versus 0.8; P < 0.01 Postoperative mortality within 30 days: 1 versus 0 Any postoperative morbidity: 38% versus 11.8%; P < 0.01 Postoperative morbidity defined as infected lymphocyst, thromboembolism, intestinal obstruction, anastomotic leakage, ureteral injury, sepsis, intra-abdominal abscess, pneumothorax or postoperative death within 30 days.
Notes	Follow-up: median 43 months; range 1–124 months Retrospective non-randomised study. The decision to perform simple or radical procedures was determined by the surgeon. Confounding by indication could not be excluded. Participant and disease characteristics not reported per type of surgery. Blinding not reported. Adjusted HRs were derived from a prognostic model. No details on how modelling was performed, but this seems to have been done based on significance testing (and not on including putative confounders in the analysis, irrespective of statistical significance).
Risk of bias	

Chang 2012a *(Continued)*

Bias	Authors' judgement	Support for judgement
1. Bias due to confounding (a–d)	High risk	Domain had a critical risk of bias: (quote) "The decision to perform simple or radical procedures was determined by the surgeon's discretion". Confounding by indication could not be excluded. Also, no known prognostic factors that had potential for confounding of the effect on intervention. Information was collected retrospectively.
2. Selection bias (a)	Low risk	Intervention and follow-up start were simultaneous as a rule for cytoreductive surgery. No evidence of selection into the study due to variables measured after the intervention since participants were included retrospectively.
3. Classification of interventions (a–b)	Low risk	Well-defined surgical interventions based on type of surgical procedure (simple versus radical).
4. Deviation from intended interventions (a–c)	Unclear risk	No evidence of any deviations from interventions or usual practice – which may either be an error of omission or that deviations did not happen.
5. Bias due to missing data (a–b)	High risk	Domain had a moderate-to-high risk of bias: all selected participants seem to have been included in the analyses. No differential follow-up or missing data reported; no participant selection due to missing data reported. There was no reason to believe there was a serious bias due to missing data as the study was sound for a non-randomised study with regard to this domain but could not be considered comparable to a well-performed randomised trial. Therefore, it was sensible to judge the missing data domain at moderate-to-high risk of bias.
6. Measurement of outcomes (a–b)	High risk	Domain had a critical risk of bias: adjusted HRs were derived from a prognostic model. No details on how modelling was performed, but this seems to have been done based on significance testing (and not on including putative confounders in the analysis, irrespective of statistical significance). Also, adjustment were made for residual disease and this was likely to distort the estimate of survival as this adjustment was made after surgery and was a key prognostic factor.
7. Reporting bias (a–c)	Unclear risk	Difficult to judge. No protocol available. All outcomes mentioned in the methods section seemed to have been reported in the results section.

Luyckx 2012

Study characteristics	
Methods	Retrospective review of medical records of patients treated in 7 French gynaecological oncology and surgery centres.
Participants	Women with FIGO stage IIIC and IV (pleural invasion only) ovarian, tubal or peritoneal epithelial carcinoma who underwent either primary or interval debulking. All had ≥ 6 cycles of carboplatin plus paclitaxel (enrolment 1 January 2003 to 31 December 2007) Age: median 59 years; range 24–90 years FIGO stage IIIC: 441 (83.7%); IV: 86 (16.3%) Tumour cell type: serous papillary 382 (72.8%), mucinous: 11 (2.1%), endometrioid: 54 (10.3%), clear cell: 13 (2.5%), undifferentiated 54 (10.3%), other: 11 (2.1%) Tumour grade 1: 34 (8.3%), 2: 138 (33.8%), 3: 236 (57.8%), unknown: 119

Luyckx 2012 *(Continued)*

	Ascites: median 50 mL; range 0–8000 mL Residual disease: no gross visible: 374 (71.1%), 0–1 cm: 97 (18.5%), > 1 cm: 55 (10.5%) Peritoneal cancer index: median 10.0 Upper abdominal lesion: 0 mm: 175 (38.5%); 0–25 mm: 182 (40.0%); > 25 mm: 97 (21.4%) Baseline details not presented according to type of surgery.
Interventions	**Intervention 1:** ultra-radical surgery involving a combination of digestive tract resections (right colon and caecum, total colectomy, and others), organ resection (spleen, gallbladder, partial gastrectomy, and others), coeliac lymph node dissection, and total abdominal peritoneum stripping in addition to standard surgery (group 2B in the study). **Intervention 2:** standard surgery plus relatively routine upper abdominal surgery (group 2A in the study). **Comparison:** standard surgery with hysterectomy, bilateral salpingo-oophorectomy, rectosigmoid resection, infragastric omentectomy, pelvic and aortic lymphadenectomy, and, when applicable, appendectomy (group 1 in the study).
Outcomes	Overall survival, disease-free survival
Notes	Follow-up: median 49 months Retrospective non-randomised study. We assume that the decision to perform simple or radical procedures was determined by the surgeon. Confounding by indication could not be excluded. Participant and disease characteristics not reported per type of surgery. Blinding not reported (but may not be relevant to this research question). Sample also included a mixture of primary and interval debulking surgery. Adjusted HRs were derived from a prognostic model. Characteristics were selected based on statistical significance in the univariate analysis (P < 0.10) and not on including putative confounders in the analysis, irrespective of statistical significance.

Risk of bias

Bias	Authors' judgement	Support for judgement
1. Bias due to confounding (a–d)	High risk	Domain had a critical risk of bias: no known prognostic factors that have potential for confounding of the effect on intervention. Information collected retrospectively.
2. Selection bias (a)	Low risk	Intervention and follow-up start were simultaneous as a rule for cytoreductive surgery. No evidence of selection into the study due to variables measured after the intervention since participants were included retrospectively.
3. Classification of interventions (a–b)	Low risk	Well-defined surgical interventions based on type of surgical procedure: group 1: standard surgery; group 2A: standard surgery plus relatively routine upper abdominal surgery; group 2B: ultra-radical surgery
4. Deviation from intended interventions (a–c)	Unclear risk	No evidence of any deviations from interventions or usual practice but may be due to omission or that deviations did not happen.
5. Bias due to missing data (a–b)	High risk	Domain had a moderate-to-high risk of bias: all selected participants may have been included in the analyses but this could not be confirmed. Therefore, it was sensible to judge the missing data domain as being at moderate-to-high risk of bias.
6. Measurement of outcomes (a–b)	High risk	Domain had a critical risk of bias: adjusted HRs are derived from a prognostic model based on univariate significance testing (P < 0.10) and not on including putative confounders in the analysis, irrespective of statistical significance. Al-

Luyckx 2012 *(Continued)*

		so, adjustment were made for residual disease and this was likely to distort the estimate of survival as this adjustment was made after surgery and was a key prognostic factor.
7. Reporting bias (a–c)	Unclear risk	Difficult to judge. No protocol available. All outcomes mentioned in the methods section seemed to have been reported in the results section.

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Aletti 2006b	Participants in comparison standard surgery group also had extensive bowel surgery, which is ultra-radical. It was also unclear whether women with recurrent disease were included.
Aletti 2006c	Intervention was ultra-radical (removal of tumour from diaphragm), but unclear whether those in comparison group received different form of ultra-radical surgery.
Aletti 2009a	Comparison of ultra-radical versus standard surgery groups not possible – low complexity scores also included possible small bowel resection.
Aletti 2009b	Comparison of ultra-radical versus standard surgery groups not possible – low complexity scores also included possible small bowel resection.
Angioli 2012	< 100 participants in analysis.
Bahra 2013	Comparison not possible.
Barlin 2013	No multivariate analysis.
Bartl 2018	No comparator.
Bertelsen 1990	Comparison of ultra-radical versus standard surgery groups not possible. It was also unclear whether women with recurrent disease were included.
Bristow 1999	Women with recurrent disease also included.
Butler 2012	< 100 participants; conference abstract.
Cai 2007	Comparisons were made between bowel resection versus no bowel resection regardless of the nature of surgery, so those in the no bowel resection group may have still received a form of ultra-radical surgery. It was also unclear whether women with recurrent disease were included.
Campos 2014	Conference abstract.
Canlorbe 2018	Participants in comparison standard surgery group also had extensive bowel surgery, which is ultra-radical. It was also unclear whether women with recurrent disease were included.
Chang 2012b	No ultra-radical surgery.
Chereau 2011	Mixed FIGO stages; no multivariate analysis.
Chi 2004	Participants in comparison standard surgery group also had extensive bowel surgery, which is ultra-radical. It was also unclear whether women with recurrent disease were included.

Study	Reason for exclusion
Chi 2009	Comparison between standard surgery and ultra-radical surgery groups not possible as all women underwent extensive upper abdominal surgery.
Chua 2011	Comparison not of interest.
Clark 2012	Outcomes not of interest; conference abstract.
Clark 2014	Comparison not of interest.
Cormier 2012	No ultra-radical surgery; conference abstract.
Cummins 2019	Conference abstract.
Davies 2019	Participants in comparison standard surgery group also had extensive bowel surgery, which is ultra-radical. It was also unclear whether women with recurrent disease were included.
Eisenhauer 2006	Participants in comparison standard surgery group also had extensive bowel surgery, which is ultra-radical. Also unclear if women with recurrent disease included.
Eisenkop 1993	Participants in comparison standard surgery group also had extensive bowel surgery or diaphragmatic stripping (or both) which is ultra-radical. It was also unclear whether women with recurrent disease were included.
Eisenkop 1998	Participants in comparison standard surgery group also had extensive bowel surgery or diaphragmatic stripping (or both) which is ultra-radical. Women with recurrent disease were also included.
Eisenkop 2001	Comparison of ultra-radical versus standard surgery groups not possible.
Eisenkop 2003	Comparison of ultra-radical versus standard surgery groups not possible. Women with recurrent disease were also included.
Eisenkop 2006	Comparisons were made between splenectomy versus no splenectomy regardless of the nature of surgery, so those in the no splenectomy group may have still received a form of ultra-radical surgery. Women with recurrent disease were also included.
Elgamal 2019	Participants in comparison standard surgery group also had extensive bowel surgery, which is ultra-radical. It was also unclear whether women with recurrent disease were included.
Eng 2018	No multivariate analysis.
Eoh 2017	Participants in comparison standard surgery group also had extensive bowel surgery, which is ultra-radical. It was also unclear whether women with recurrent disease were included.
Falconer 2020	Uncontrolled before-after study.
Favero 2014	Comparison not of interest.
Ferrero 2014	Comparison not of interest.
Filippova 2019	Participants in comparison standard surgery group also had extensive bowel surgery, which is ultra-radical. It was also unclear whether women with recurrent disease were included.
Fotopoulou 2012	Comparison not of interest; conference abstract.
Gockley 2019	Participants in comparison standard surgery group also had extensive bowel surgery, which is ultra-radical. It was also unclear whether women with recurrent disease were included.

Study	Reason for exclusion
Gremeau 2014	Comparison not of interest.
Grimm 2017	No comparator.
Guyon 2014	Comparison not of interest.
Hamilton 2011	Comparison not of interest.
Hudry 2013	Comparison not of interest.
Hwang 2014	Mixed population; comparison not of interest.
Janda 2014	Comparison not of interest.
Jiang 2013	Abstract form only but appeared to be same study as Ren 2015 where multivariate analysis did not include surgery type.
Kato 2013a	Comparison not of interest.
Kato 2013b	Population not of interest.
Kehoe 2013	Comparison not of interest.
Kim 2011	Comparison not of interest.
Kolev 2014	Recurrent cancer.
Kommoss 2010	Comparison between groups not possible as both groups also included participants undergoing bowel resection.
Kristensen 2014	Borderline tumours.
Kuhn 1998	Participants in comparison standard surgery group also had extensive bowel surgery or diaphragmatic stripping (or both), which is ultra-radical. Women with recurrent disease were also included.
Laios 2019	No comparator.
Lee 2017	Conference abstracts.
Li 2014	Comparison not of interest.
Liberale 2019	Conference abstract.
Liu 2013a	Germ cell tumours; < 100 participants.
Liu 2013b	Germ cell tumours; article in Chinese.
Martinez 2014	Conference abstract.
McCann 2011	No multivariate analysis.
Muallem 2018	No multivariate analysis.
Oseledchyk 2016	Participants in comparison standard surgery group also had extensive bowel surgery, which is ultra-radical. It was also unclear whether women with recurrent disease were included.

Study	Reason for exclusion
Oshita 2013	Included only stages pT1-2.
Park 2011	No ultra-radical surgery; conference abstract.
Pathiraja 2011	< 100 participants; conference abstract.
Pathiraja 2013	< 100 participants; outcomes not of interest.
Pelissier 2018	No comparator.
Perri 2013	Comparison not of interest.
Phillips 2018	No multivariate analysis.
Pushpalatha 2011	Comparison not of interest.
Qin 2012	No meta-analysis; comparison not of interest.
Ratnavelu 2014	< 100 participants; conference abstract.
Ren 2015	Multivariate analysis did not include surgery type.
Rodriguez 2012	Women with suboptimal debulking were excluded (see also full publication Rodriguez 2013).
Rodriguez 2013	Women with suboptimal debulking were excluded.
Rouzier 2010	Mixed population; comparison not of interest.
Sagara 2019	No multivariate analysis.
Sandadi 2014	Comparison not of interest.
Scalici 2014	Comparison not of interest.
Sehouli 2010	Comparison not of interest.
Soo Hoo 2015	No multivariate analysis.
Stefanović 2011	< 100 participants.
Sundar 2014	< 100 participants; conference abstract.
Sundar 2018	Conference abstract.
Sundar 2019	Conference abstract.
Suzuki 2008	Conference abstract.
Szczesny 2016	Conference abstract.
Tozzi 2019	Participants in comparison standard surgery group also had extensive bowel surgery, which is ultra-radical. It was also unclear whether women with recurrent disease were included.
Tsolakidis 2010a	Comparison between 'standard surgery' and 'ultra-radical surgery' groups not possible as all women underwent diaphragmatic stripping.

Study	Reason for exclusion
Tsolakidis 2010b	Comparison between standard surgery and ultra-radical surgery groups not possible as all women underwent diaphragmatic stripping.
van de Laar 2014	Protocol for a new study; applies to recurrent cancer.
Vidal 2016	No comparator.
von Hugo 1989	Unclear if women with recurrent disease were included.
Wallace 2016	No comparator.
Wat 2012	< 100 participants; conference abstract.
Wimberger 2007	Comparison of ultra-radical versus standard surgery groups not possible, as the comparative groups include participants who had both types of surgery.
Wright 2012	Outcomes not of interest.
Yildirim 2014	No comparator.
Zamurovic 2013	No multivariate analysis.
Zapardiel 2012	No multivariate analysis.

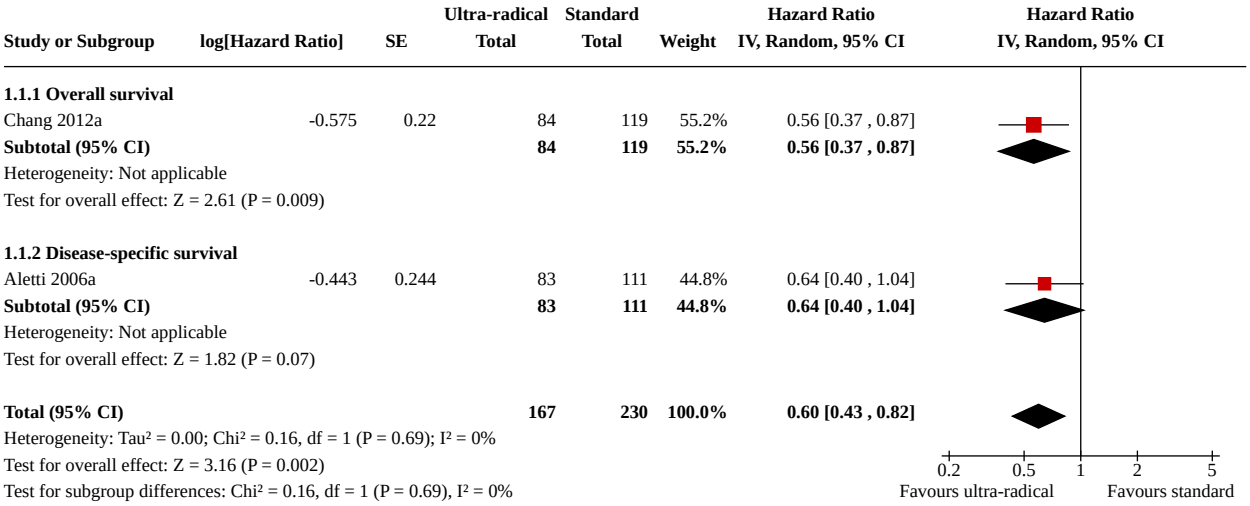
DATA AND ANALYSES

Comparison 1. Ultra-radical versus standard surgery (upfront surgery)

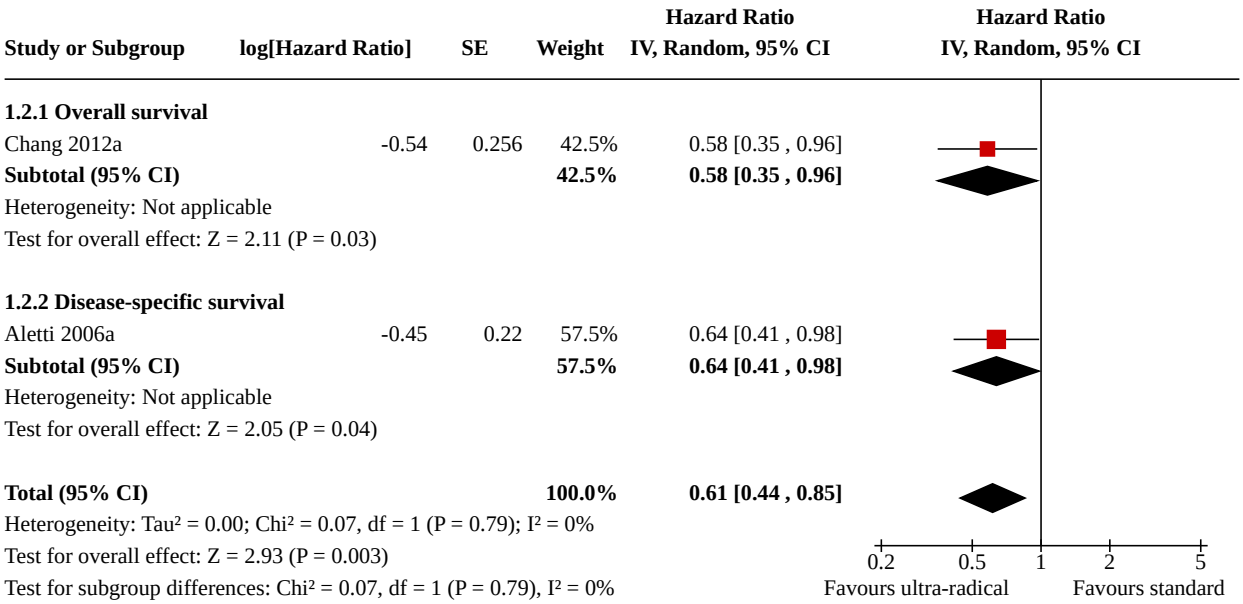
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.1 Survival	2	397	Hazard Ratio (IV, Random, 95% CI)	0.60 [0.43, 0.82]
1.1.1 Overall survival	1	203	Hazard Ratio (IV, Random, 95% CI)	0.56 [0.37, 0.87]
1.1.2 Disease-specific survival	1	194	Hazard Ratio (IV, Random, 95% CI)	0.64 [0.40, 1.04]
1.2 Survival: women with carcinomatosis (upfront surgery)	2		Hazard Ratio (IV, Random, 95% CI)	0.61 [0.44, 0.85]
1.2.1 Overall survival	1		Hazard Ratio (IV, Random, 95% CI)	0.58 [0.35, 0.96]
1.2.2 Disease-specific survival	1		Hazard Ratio (IV, Random, 95% CI)	0.64 [0.41, 0.98]
1.3 Progression-free survival	1		Hazard Ratio (IV, Random, 95% CI)	Totals not selected

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.3.1 Upfront primary debulking surgery	1		Hazard Ratio (IV, Random, 95% CI)	Totals not selected
1.4 Progression-free survival: women with carcinomatosis (upfront surgery)	1		Hazard Ratio (IV, Random, 95% CI)	Subtotals only
1.5 Disease-free survival	1	527	Hazard Ratio (IV, Random, 95% CI)	1.60 [1.11, 2.31]
1.5.1 Mix of upfront and interval debulking surgical procedures – including group 2A	1	258	Hazard Ratio (IV, Random, 95% CI)	1.54 [0.91, 2.60]
1.5.2 Mix of upfront and interval debulking surgical procedures – including group 2B	1	269	Hazard Ratio (IV, Random, 95% CI)	1.66 [1.00, 2.78]

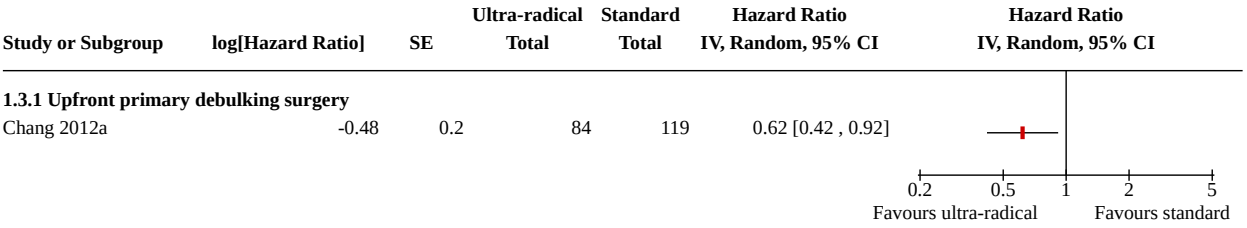
Analysis 1.1. Comparison 1: Ultra-radical versus standard surgery (upfront surgery), Outcome 1: Survival



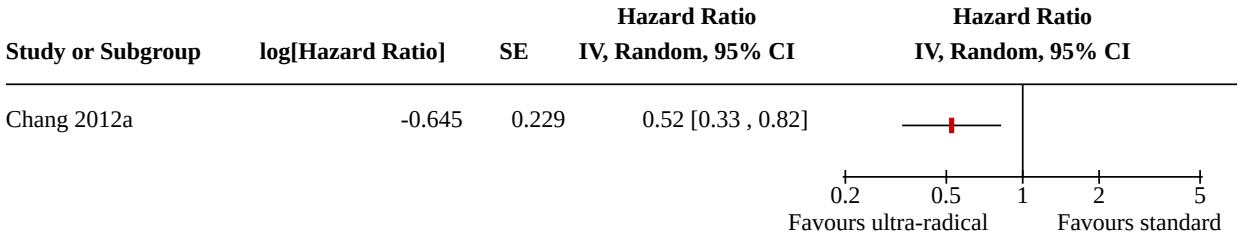
Analysis 1.2. Comparison 1: Ultra-radical versus standard surgery (upfront surgery), Outcome 2: Survival: women with carcinomatosis (upfront surgery)



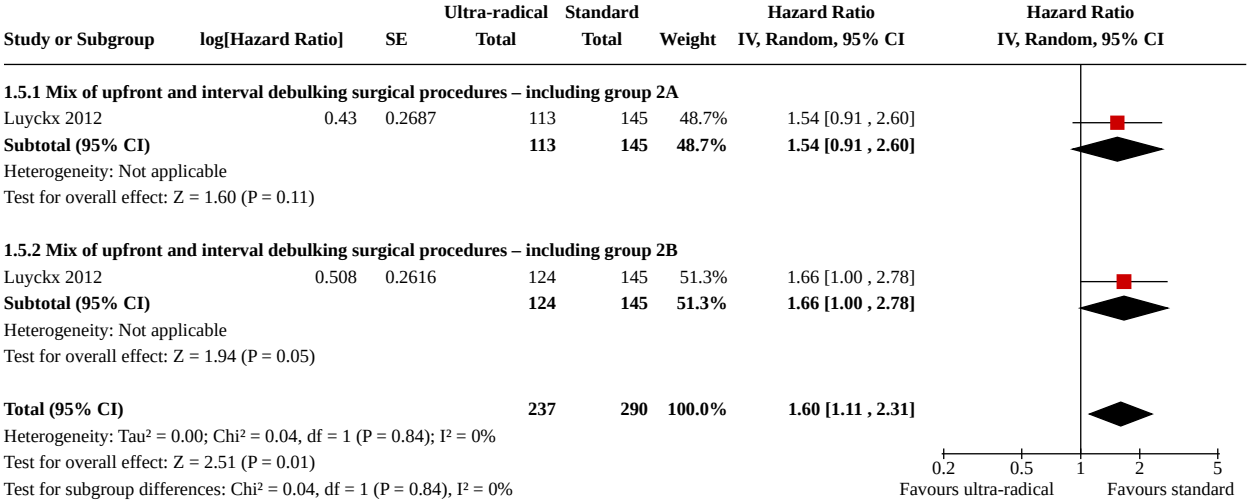
Analysis 1.3. Comparison 1: Ultra-radical versus standard surgery (upfront surgery), Outcome 3: Progression-free survival



Analysis 1.4. Comparison 1: Ultra-radical versus standard surgery (upfront surgery), Outcome 4: Progression-free survival: women with carcinomatosis (upfront surgery)



Analysis 1.5. Comparison 1: Ultra-radical versus standard surgery (upfront surgery), Outcome 5: Disease-free survival



ADDITIONAL TABLES

Table 1. Summary of Risk Of Bias In Non-randomised Studies of Interventions (ROBINS-I)

Author	Confounding	Selection bias	Classification of interventions	Deviation	Missing data	Measuring outcomes	Reporting bias
Aletti 2006a	Critical	Low	Low	Unclear	Moderate/high	Critical	Critical
Chang 2012a	Critical	Low	Low	Unclear	Moderate/high	Critical	Unclear
Luyckx 2012	Critical	Low	Low	Unclear	Moderate/high	Critical	Unclear

Risk of bias in included non-randomised studies was assessed using the ROBINS-I tool as outlined in Appendix 4(Sterne 2016).

APPENDICES

Appendix 1. CENTRAL search strategy

#1 MeSH descriptor Ovarian Neoplasms explode all trees
#2 ovar* near/5 (cancer* or neoplas* or carcinom* or malignan* or tumor* or tumour*)
#3 (#1 OR #2)
#4 MeSH descriptor Surgical Procedures, Operative explode all trees
#5 surg*
#6 Any MeSH descriptor with qualifier: SU
#7 (#4 OR #5 OR #6)
#8 debulk*
#9 cytoreduc*
#10 ultradical or ultra-radical or ultra radical
#11 MeSH descriptor Omentum explode all trees
#12 omentum
#13 bowel
#14 abdom*
#15 MeSH descriptor Spleen explode all trees
#16 spleen
#17 MeSH descriptor Liver explode all trees
#18 liver
#19 MeSH descriptor Diaphragm explode all trees
#20 diaphragm*
#21 MeSH descriptor Lymph Nodes explode all trees
#22 lymph next node*
#23 MeSH descriptor Peritoneum explode all trees
#24 peritone*
#25 MeSH descriptor Urinary Tract explode all trees
#26 urinary next tract
#27 (#8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR 23 OR #24 OR #25 OR #26)
#28 (#7 AND #27)
#29 MeSH descriptor Splenectomy explode all trees
#30 splenectomy
#31 MeSH descriptor Hysterectomy explode all trees
#32 abdom* near/5 hysterectomy
#33 abdominohysterectomy
#34 MeSH descriptor Lymph Node Excision explode all trees
#35 lymph next node next excision
#36 bilateral next salpingo next oophorectomy
#37 omentectomy
#38 MeSH descriptor Surgical Stomas explode all trees
#39 stoma
#40 (#29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39)
#41 (#28 OR #40)
#42 (#3 AND #41)

Appendix 2. MEDLINE search strategy

1. exp Ovarian Neoplasms/
2. (ovar* adj5 (cancer* or neoplas* or carcinom* or malignan* or tumor* or tumour*)).mp.
3. 1 or 2
4. exp Surgical Procedures, Operative/
5. surg*.mp.
6. surgery.fs.
7. 4 or 5 or 6
8. debulk*.mp.
9. cytoreduc*.mp.
10. (ultraradical or ultra-radical or ultra radical).mp.
11. exp Omentum/

12.omentum.mp.
13.bowel.mp.
14.abdom*.mp.
15.exp Spleen/
16.spleen.mp.
17.exp Liver/
18.liver.mp.
19.exp Diaphragm/
20.diaphragm*.mp.
21.exp Lymph Nodes/
22.(lymph adj node*).mp.
23.exp Peritoneum/
24.peritone*.mp.
25.exp Urinary Tract/
26.(urinary adj tract).mp.
27.8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26
28.7 and 27
29.exp Splenectomy/
30.splenectomy.mp.
31.exp Hysterectomy/
32.(abdom* adj5 hysterectomy).mp.
33.abdominohysterectomy.mp.
34.exp Lymph Node Excision/
35.(lymph adj node adj excision).mp.
36.(bilateral adj salpingo adj oophorectomy).mp.
37.omentectomy.mp.
38.exp Surgical Stomas/
39.stoma.mp.
40.29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39
41.28 or 40
42.3 and 41
43."randomized controlled trial".pt.
44."controlled clinical trial".pt.
45.randomized.ab.
46.randomly.ab.
47.trial.ab.
48.groups.ab.
49.exp Cohort Studies/
50.cohort*.mp.
51.(case adj series).mp.
52.43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51
53.42 and 52
54.Animals/
55.Humans/
56.54 not (54 and 55)
57.53 not 56

key: mp = title, original title, abstract, name of substance word, subject heading word

Appendix 3. Embase search strategy

1. exp Ovary Tumor/
2. (ovar* adj5 (cancer* or neoplas* or carcinom* or malignan* or tumor* or tumour*)).mp.
3. 1 or 2

4. exp Surgery/
5. surg*.mp.
6. su.fs.
7. 4 or 5 or 6
8. debulk*.mp.
9. cytoreduc*.mp.
10. (ultraradical or ultra-radical or ultra radical).mp.
11. exp Omentum/
12. omentum.mp.
13. bowel.mp.
14. abdom*.mp.
15. exp Spleen/
16. spleen.mp.
17. exp Liver/
18. liver.mp.
19. exp Diaphragm/
20. diaphragm*.mp.
21. exp Lymph Node/
22. (lymph adj node).mp.
23. exp Peritoneum/
24. peritone*.mp.
25. exp Urinary Tract/
26. (urinary adj tract).mp.
27. 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26
28. 7 and 27
29. exp Splenectomy/
30. splenectomy.mp.
31. exp Hysterectomy/
32. (abdom* adj5 hysterectomy).mp.
33. abdominohysterectomy.mp.
34. exp Lymphadenectomy/
35. (lymph adj node adj excision).mp.
36. (bilateral adj salpingo adj oophorectomy).mp.
37. omentectomy.mp.
38. exp Stoma/
39. stoma.mp.
40. 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39
41. 28 or 40
42. 3 and 41
43. exp Controlled Clinical Trial/
44. randomized.ab.
45. randomly.ab.
46. trial.ab.
47. groups.ab.
48. exp Cohort Analysis/
49. cohort*.mp.
50. (case adj series).mp.
51. 50 or 49 or 46 or 45 or 43 or 44 or 48 or 47
52. 42 and 51
53. exp Animal/
54. Human/
55. 53 not (53 and 54)

56.52 not 55

key: mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name

ab=abstract

fs=floating subheading

Appendix 4. ROBIN-1 domains

Risk of bias in included non-randomised studies was assessed using the ROBINS-I tool (Sterne 2016).

1. **Bias due to confounding**
   a. Baseline confounding – when one or more preintervention prognostic factors predict the intervention received at start of follow-up.
   b. Time-varying confounding – when the intervention received can change over time.
   c. Residual confounding – when a confounding domain is measured with error.
   d. Unmeasured confounding – when confounding domain has not been measured or controlled in the analysis.
2. **Selection bias**
   a. Bias in selection of participants into the study.
3. **Classification of interventions**
   a. Differential misclassification – intervention status is related to subsequent outcome or to the risk of the outcome.
   b. Non-differential misclassification – unrelated to outcome.
4. **Deviation from intended interventions**
   a. Considerations for co-interventions.
   b. Considerations for fidelity of implementation of intended interventions.
   c. Considerations for adherence to intervention.
5. **Bias due to missing data**
   a. Differential missingness.
   b. Whether proportions of individuals in whom adverse effects may be prevalent have been excluded.
6. **Measurement of outcomes**
   a. Differential measurement error – measurement error related intervention status.
   b. Non-differential measurement error – unrelated to the intervention received.
7. **Reporting bias**
   a. Selective outcome reporting.
   b. Selective analysis reporting.
   c. Selection of a subgroup from a larger cohort.

WHAT'S NEW

Date	Event	Description
18 May 2022	New search has been performed	Updated to include two new studies
6 April 2022	New citation required but conclusions have not changed	Updated search on 10 November 2021

HISTORY

Protocol first published: Issue 2, 2009
Review first published: Issue 4, 2011

Date	Event	Description
1 August 2016	New search has been performed	Search updated, no new studies included

Date	Event	Description
17 June 2015	New search has been performed	Search updated; two new studies included
11 February 2015	Amended	Contact details updated.
26 February 2014	Amended	Contact details updated.
28 July 2011	Amended	Author contact details updated

CONTRIBUTIONS OF AUTHORS

KG and RN drafted the clinical and discussion sections of the review.

AB, SH and PK data extracted items for inclusion in the review.

AB and SH drafted the methodological, results and discussion sections of the review.

AB and SH performed the GRADE judgements with other co-authors acting as arbiters.

SH and AB are joint first authors on the review.

RN initiated the research concept and was the lead senior clinical author.

All authors agreed the final version.

DECLARATIONS OF INTEREST

SH: none known.

AB: none known.

PK: none known.

KG: performs surgery for advanced ovarian cancer surgery, but has no conflicts of interest to declare.

RN: performs surgery for advanced ovarian cancer surgery, but has no conflicts of interest to declare.

SOURCES OF SUPPORT

Internal sources

- No sources of support provided

External sources

- NIHR, UK

  NHS Cochrane Collaboration Programme Grant Scheme CPG-506

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We added the following study constraint in the Types of studies section, as it was apparent that selection bias would have been problematic.

We added disease-free survival as a secondary outcome.

"In order to minimise selection bias, we decided to include only studies that used statistical adjustment for baseline case mix (e.g. age, performance status, grade, etc.) using multivariate analyses."

We removed discussion of unadjusted results from the data synthesis, subgroup analysis, and investigation of heterogeneity and sensitivity analysis sections as we do not plan to use unadjusted results in future updates due to the risk of selection bias.

Three studies met the inclusion criteria for the review and did not report dichotomous or continuous outcomes. Should more studies be identified for updates of the review, we will use the following methods.

Data extraction and management

Data on outcomes will be extracted as below.

- For dichotomous outcomes (e.g. adverse events or deaths, if it was not possible to use a hazard ratio), we will extract the number of patients in each treatment arm who experienced the outcome of interest and the number of participants assessed at endpoint, in order to estimate a risk ratio (RR).
- For continuous outcomes (e.g. QoL measures), we will extract the final value and standard deviation of the outcome of interest and the number of participants assessed at endpoint in each treatment arm at the end of follow-up, in order to estimate the mean difference between treatment arms and its standard error.

Measures of treatment effect

We will use the following measures of the effect of treatment.

- For dichotomous outcomes, we will use the risk ratio.
- For continuous outcomes, we will use the mean difference between treatment arms.

Data synthesis

If sufficient clinically similar studies are available, we will pool their results in a meta-analysis and use adjusted summary statistics.

- For any dichotomous outcomes, we will calculate the risk ratio for each study and then pool them.
- For continuous outcomes, we will pool the mean differences between the treatment arms at the end of follow-up if all trials measured the outcome on the same scale, otherwise we will pool standardised mean differences.

We will assess the risk of bias in included RCTs using the Cochrane RoB tool (Higgins 2019). This includes assessment of:

- sequence generation;
- allocation concealment;
- blinding (where assessment of blinding was restricted to blinding of outcome assessors, since it is generally not possible to blind participants and treatment providers to surgical interventions);
- incomplete outcome data; we coded a satisfactory level of loss to follow-up for each outcome as:
- yes, if less than 20% of participants were lost to follow-up and reasons for loss to follow-up were similar in both treatment arms;
- no, if more than 20% of participants were lost to follow-up or reasons for loss to follow-up differed between treatment groups;
- unclear if loss to follow-up was not reported;
- selective reporting of outcomes;
- other possible sources of bias.

However, we only identified three non-randomised studies, so it was more appropriate to use the ROBIN-I risk of bias tool (Sterne 2016), so this superseded the default tool used to assess risk of bias in trials (Higgins 2019).

Sensitivity analysis

We performed post hoc sensitivity analyses including only women with more extensive disease (with carcinomatosis) as there were a substantial proportion of women with this.

INDEX TERMS

Medical Subject Headings (MeSH)

Carcinoma, Ovarian Epithelial; Neoplasm Invasiveness [pathology]; Neoplasm Staging; *Neoplasms, Glandular and Epithelial [pathology] [surgery]; *Ovarian Neoplasms [pathology] [surgery]

MeSH check words

Adult; Female; Humans

Appendix 5: Publication 3: Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery

Cochrane Library

Cochrane Database of Systematic Reviews

Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery (Review)

Bryant A, Hiu S, Kunonga PT, Gajjar K, Craig D, Vale L, Winter-Roach BA, Elattar A, Naik R

Bryant A, Hiu S, Kunonga PT, Gajjar K, Craig D, Vale L, Winter-Roach BA, Elattar A, Naik R.
Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery.
Cochrane Database of Systematic Reviews 2022, Issue 9. Art. No.: CD015048.
DOI: 10.1002/14651858.CD015048.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Prognosis Review]

Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery

Andrew Bryant[1], Shaun Hiu[1], Patience T Kunonga[1], Ketankumar Gajjar[2], Dawn Craig[1], Luke Vale[1], Brett A Winter-Roach[3], Ahmed Elattar[4], Raj Naik[5]

[1]Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK. [2]Department of Gynaecological Oncology, 1st Floor Maternity Unit, City Hospital Campus, Nottingham, UK. [3]The Department of Surgery, Christie Hospital NHS Foundation Trust, Manchester, UK. [4]City Hospital & Birmingham Treatment Centre, Birmingham, UK. [5]Gynaecological Oncology, Northern Gynaecological Oncology Centre, Gateshead, UK

Contact: Andrew Bryant, andy.bryant@ncl.ac.uk.

Editorial group: Cochrane Gynaecological, Neuro-oncology and Orphan Cancer Group.
Publication status and date: New, published in Issue 9, 2022.

Citation: Bryant A, Hiu S, Kunonga PT, Gajjar K, Craig D, Vale L, Winter-Roach BA, Elattar A, Naik R. Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery. *Cochrane Database of Systematic Reviews* 2022, Issue 9. Art. No.: CD015048. DOI: 10.1002/14651858.CD015048.pub2.

ABSTRACT

Background

Ovarian cancer is the seventh most common cancer among women and a leading cause of death from gynaecological malignancies. Epithelial ovarian cancer is the most common type, accounting for around 90% of all ovarian cancers. This specific type of ovarian cancer starts in the surface layer covering the ovary or lining of the fallopian tube. Surgery is performed either before chemotherapy (upfront or primary debulking surgery (PDS)) or in the middle of a course of treatment with chemotherapy (neoadjuvant chemotherapy (NACT) and interval debulking surgery (IDS)), with the aim of removing all visible tumour and achieving no macroscopic residual disease (NMRD). The aim of this review is to investigate the prognostic impact of size of residual disease nodules (RD) in women who received upfront or interval cytoreductive surgery for advanced (stage III and IV) epithelial ovarian cancer (EOC).

Objectives

To assess the prognostic impact of residual disease after primary surgery on survival outcomes for advanced (stage III and IV) epithelial ovarian cancer. In separate analyses, primary surgery included both upfront primary debulking surgery (PDS) followed by adjuvant chemotherapy and neoadjuvant chemotherapy followed by interval debulking surgery (IDS). Each residual disease threshold is considered as a separate prognostic factor.

Search methods

We searched CENTRAL (2021, Issue 8), MEDLINE via Ovid (to 30 August 2021) and Embase via Ovid (to 30 August 2021).

Selection criteria

We included survival data from studies of at least 100 women with advanced EOC after primary surgery. Residual disease was assessed as a prognostic factor in multivariate prognostic models. We excluded studies that reported fewer than 100 women, women with concurrent malignancies or studies that only reported unadjusted results. Women were included into two distinct groups: those who received PDS followed by platinum-based chemotherapy and those who received IDS, analysed separately. We included studies that reported all RD thresholds after surgery, but the main thresholds of interest were microscopic RD (labelled NMRD), RD 0.1 cm to 1 cm (small-volume residual disease (SVRD)) and RD > 1 cm (large-volume residual disease (LVRD)).

Data collection and analysis

Two review authors independently abstracted data and assessed risk of bias. Where possible, we synthesised the data in meta-analysis. To assess the adequacy of adjustment factors used in multivariate Cox models, we used the 'adjustment for other prognostic factors' and 'statistical analysis and reporting' domains of the quality in prognosis studies (QUIPS) tool. We also made judgements about the certainty of the evidence for each outcome in the main comparisons, using GRADE.

We examined differences between FIGO stages III and IV for different thresholds of RD after primary surgery. We considered factors such as age, grade, length of follow-up, type and experience of surgeon, and type of surgery in the interpretation of any heterogeneity.

We also performed sensitivity analyses that distinguished between studies that included NMRD in RD categories of < 1 cm and those that did not. This was applicable to comparisons involving RD < 1 cm with the exception of RD < 1 cm versus NMRD. We evaluated women undergoing PDS and IDS in separate analyses.

Main results

We found 46 studies reporting multivariate prognostic analyses, including RD as a prognostic factor, which met our inclusion criteria: 22,376 women who underwent PDS and 3697 who underwent IDS, all with varying levels of RD.

While we identified a range of different RD thresholds, we mainly report on comparisons that are the focus of a key area of clinical uncertainty (involving NMRD, SVRD and LVRD). The comparison involving any visible disease (RD > 0 cm) and NMRD was also important.

SVRD versus NMRD in a PDS setting

In PDS studies, most showed an increased risk of death in all RD groups when those with macroscopic RD (MRD) were compared to NMRD. Women who had SVRD after PDS had more than twice the risk of death compared to women with NMRD (hazard ratio (HR) 2.03, 95% confidence interval (CI) 1.80 to 2.29; I^2 = 50%; 17 studies; 9404 participants; moderate-certainty). The analysis of progression-free survival found that women who had SVRD after PDS had nearly twice the risk of death compared to women with NMRD (HR 1.88, 95% CI 1.63 to 2.16; I^2 = 63%; 10 studies; 6596 participants; moderate-certainty).

LVRD versus SVRD in a PDS setting

When we compared LVRD versus SVRD following surgery, the estimates were attenuated compared to NMRD comparisons. All analyses showed an overall survival benefit in women who had RD < 1 cm after surgery (HR 1.22, 95% CI 1.13 to 1.32; I^2 = 0%; 5 studies; 6000 participants; moderate-certainty). The results were robust to analyses of progression-free survival.

SVRD and LVRD versus NMRD in an IDS setting

The one study that defined the categories as NMRD, SVRD and LVRD showed that women who had SVRD and LVRD after IDS had more than twice the risk of death compared to women who had NMRD (HR 2.09, 95% CI 1.20 to 3.66; 310 participants; I^2 = 56%, and HR 2.23, 95% CI 1.49 to 3.34; 343 participants; I^2 = 35%; very low-certainty, for SVRD versus NMRD and LVRD versus NMRD, respectively).

LVRD versus SVRD + NMRD in an IDS setting

Meta-analysis found that women who had LVRD had a greater risk of death and disease progression compared to women who had either SVRD or NMRD (HR 1.60, 95% CI 1.21 to 2.11; 6 studies; 1572 participants; I^2 = 58% for overall survival and HR 1.76, 95% CI 1.23 to 2.52; 1145 participants; I^2 = 60% for progression-free survival; very low-certainty). However, this result is biased as in all but one study it was not possible to distinguish NMRD within the < 1 cm thresholds. Only one study separated NMRD from SVRD; all others included NMRD in the SVRD group, which may create bias when comparing with LVRD, making interpretation challenging.

MRD versus NMRD in an IDS setting

Women who had any amount of MRD after IDS had more than twice the risk of death compared to women with NMRD (HR 2.11, 95% CI 1.35 to 3.29, I^2 = 81%; 906 participants; very low-certainty).

Authors' conclusions

In a PDS setting, there is moderate-certainty evidence that the amount of RD after primary surgery is a prognostic factor for overall and progression-free survival in women with advanced ovarian cancer. We separated our analysis into three distinct categories for the survival outcome including NMRD, SVRD and LVRD.

After IDS, there may be only two categories required, although this is based on very low-certainty evidence, as all but one study included NMRD in the SVRD category. The one study that separated NMRD from SVRD showed no improved survival outcome in the SVRD category, compared to LVRD. Further low-certainty evidence also supported restricting to two categories, where women who had any amount of MRD after IDS had a significantly greater risk of death compared to women with NMRD.

Therefore, the evidence presented in this review cannot conclude that using three categories applies in an IDS setting (very low-certainty evidence), as was supported for PDS (which has convincing moderate-certainty evidence).

PLAIN LANGUAGE SUMMARY

The impact of remaining (residual) disease after surgery on the survival prognosis for women with advanced epithelial ovarian cancer

Review question

We aimed to assess the effect on survival (the 'prognostic impact') of the amount of disease remaining after surgery (residual disease) during the initial treatment stage for women with advanced ovarian cancer. We looked at both surgery before chemotherapy ('primary debulking surgery') followed by adjuvant (additional) chemotherapy and chemotherapy first ('neoadjuvant chemotherapy') followed by surgery ('interval debulking surgery'). This review should help to determine the prognostic impact of residual disease after surgery on survival and work out acceptable definitions of residual disease thresholds.

Background

Ovarian cancer is the seventh most common cancer among women and a leading cause of death in women with gynaecological cancers. Ovarian cancers can develop from different cell types within the ovary/fallopian tubes. Most ovarian cancers are 'epithelial', arising from either the surface layer of the ovary or the lining of the fallopian tube. Newly diagnosed ovarian cancer is treated with a combination of surgery and chemotherapy, with surgery performed either before (called upfront or primary debulking surgery) or around the mid-point of chemotherapy (called interval debulking surgery). Ovarian cancer has normally spread throughout the abdominal cavity by the time of diagnosis, so, unlike many other cancers, surgery is still performed, even though it may not remove the cancer in its entirety. The aim of surgery is to remove as much of the visible (macroscopic) cancer tissue as possible, which is called debulking or cytoreductive surgery. Studies have shown that the amount of the visible cancer that can be removed is likely to be an important prognostic factor for survival of women with advanced epithelial ovarian cancer. The aim of this review was to investigate how well the amount of remaining (residual) disease after surgery for newly diagnosed ovarian cancer predicts how long women will survive following a diagnosis of epithelial ovarian cancer (prognosis).

Review methods

We searched electronic databases up to the end of August 2021 and we also searched for unpublished studies. We included studies that reported residual disease as a prognostic factor, which also examined other prognostic factors at the same time.

Key results

We found 46 studies (including 22,376 women in 31 primary debulking surgery studies and 3697 women in 15 interval debulking surgery studies). Each study included more than 100 women, used statistical adjustment for important prognostic factors (multivariate analysis) and met our inclusion criteria. Our analyses showed the prognostic importance of surgery leaving no visible tumour deposits ('no macroscopic residual disease') both when women had upfront debulking surgery or interval debulking surgery. Both overall survival and progression-free survival (survival without disease worsening, which was reported for upfront debulking surgery) were prolonged if this was achieved.

Primary debulking surgery for newly diagnosed ovarian cancer

Complete surgical removal of all visible tumour after upfront or primary debulking surgery improved survival, and this was also the case for those with a small amount of residual disease (0.1 cm to 1 cm). There was evidence to suggest that three categories of residual disease should be used (no macroscopic residual disease, small-volume and large-volume residual disease (more than 1 cm).

Interval debulking surgery for newly diagnosed ovarian cancer

When chemotherapy was given before surgery (interval debulking surgery), there was an association with improved survival if the remaining tumour was reduced to 'no macroscopic residual disease' (removal of all visible tumour). Women with small-volume residual disease had no survival advantage compared to those with large-volume residual disease, with both groups having a poorer prognosis compared to those with no visible tumour deposits; however, this evidence was of very low certainty. Any visible residual disease after interval debulking surgery was associated with poorer survival compared to women with none.

Most interval debulking surgery studies included no visible tumour deposits in the small-volume residual disease category, which limits our interpretation of these findings.

Certainty of the evidence

We judged our certainty of the evidence as 'moderate' for overall survival and progression-free survival in the analyses involving primary debulking surgery studies. For the interval debulking surgery studies, the certainty of evidence was very low for overall survival in all

comparisons and those that involved progression-free survival. This was largely due to all but one study including 'no macroscopic residual disease' in the small-volume residual disease category.

Main conclusions

The evidence in the review suggests that following primary debulking surgery three categories for the amount of residual disease should be used: no macroscopic residual disease, small-volume and large-volume residual disease. The evidence is more limited for interval debulking surgery and further studies are needed, but there may not be a survival difference between those with small- and large-volume residual disease. Until there is evidence for a survival benefit for those with small-volume compared to large-volume residual disease, it may only be important to use two residual disease categories when classifying surgical outcomes: 'no macroscopic residual disease' and 'macroscopic residual disease' (remaining visible disease of more than 0 cm). However, this is based on very low-certainty evidence and more information may change this finding.

SUMMARY OF FINDINGS

Summary of findings 1. Small-volume residual disease (SVRD) < 1 cm versus NMRD in PDS studies

Small-volume residual disease (SVRD) (< 1 cm) compared with NMRD after upfront primary debulking surgery (PDS) in women with advanced epithelial ovarian cancer (EOC)

Population: women with advanced EOC after PDS

Settings: all settings in adult women aged 18 years or older worldwide

Prognostic factor: SVRD compared with NMRD

Outcomes	Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
Overall survival: Median length of follow-up[1]: Range: 28 to 77.7 months	**Adjusted HR 2.03** (1.80 to 2.29)	9404 participants (17 studies)	⊕⊕⊕⊖ **moderate**[2]	We could not present illustrative absolute effects because a representative control group risk could not be ascertained from the studies. The HR estimates were adjusted for in multivariable analyses and this cannot be done in absolute terms so we did not attempt it, as the numbers were likely to mislead with any bias potentially favouring the NMRD threshold.
Progression-free survival: Median length of follow-up[1]: Range: 28 to 77.7 months	**Adjusted HR 1.88** (1.63 to 2.16)	6596 participants (10 studies)	⊕⊕⊕⊖ **moderate**[2]	There were no concerns with inconsistency and imprecision across studies due to restrictive inclusion criteria in a generally representative cohort of women with advanced EOC. Data were considerable in size in PDS studies with > 9000 and > 6500 women in the analyses of OS and PFS, respectively. The percentage of the variability in effect estimates that was due to heterogeneity rather than sampling error (chance) may appear to represent moderate heterogeneity (as measured by the I^2 statistic), but we had no major concerns as the direction of effect was consistent throughout. There did not appear to be any evidence of small study biases, such as publication bias, or any irregularities with the data by visual inspection of funnel plots. While publication bias cannot be dismissed, it would take a lot of large statistically insignificant studies to overhaul the current results. Furthermore, studies showing harmful survival in women with NMRD compared to other thresholds of RD is implausible.

CI: confidence interval; **HR:** hazard ratio; **EOC:** epithelial ovarian cancer; **NMRD:** no macroscopic residual disease; **OS:** overall survival; **PDS:** upfront primary debulking surgery; **PFS:** progression-free survival; **SVRD:** small-volume residual disease

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.

Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1]Range in Klar 2016 was 0 to 144 months.
[2]Downgraded by one level because was assessed the statistical analysis and reporting domain in the QUIPS tool as being at high or unclear risk of bias in all included studies. Either no conceptual framework was reported, where the variable selection criteria in multivariate model was unclear, or quite often the authors reported that significant variables from the univariate analysis were included in the multivariable model, but with no further details. This was the most serious bias from the QUIPS domains that could influence the effect estimates.

Summary of findings 2. Large-volume residual disease (LVRD) > 1 cm versus no macroscopic residual disease (NMRD) in PDS studies

LVRD (> 1 cm) compared with NMRD after upfront primary debulking surgery (PDS) in women with advanced epithelial ovarian cancer (EOC)

Population: women with advanced ovarian cancer after PDS

Settings: all settings in adult women aged 18 years or older worldwide

Prognostic factor: LVRD > 1 cm compared with NMRD

Outcomes	Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
Overall survival: Median length of follow-up: Range: 28 to 77.7 months	**Adjusted HR** **2.50** (2.13 to 2.94)	7988 participants (14 studies)	⊕⊕⊕⊖ **moderate**[1]	We could not present illustrative absolute effects because a representative control group risk could not be ascertained from the studies. The HR estimates were adjusted for in multivariable analyses and this cannot be done in absolute terms so we did not attempt it, as the numbers were likely to mislead with any bias potentially favouring the NMRD threshold.
Progression-free survival: Median length of follow-up: Range: 28 to 77.7 months	**Adjusted HR** **2.10** (1.84 to 2.40)	2629 participants (6 studies)	⊕⊕⊕⊖ **moderate**[1]	There were no concerns with inconsistency and imprecision across studies due to restrictive inclusion criteria in a generally representative cohort of women with advanced EOC. Data were considerable in size in PDS studies with nearly n = 8000 in the analysis of OS and to lesser extent > 2500 for PFS. The percentage of the variability in effect estimates that was due to heterogeneity rather than sampling error (chance) may appear to represent moderate heterogeneity (as measured by the I^2 statistic), but we had no major concerns as the direction of effect was consistent throughout. There did not appear to be any evidence of small study biases, such as publication bias, or any ir-

	regularities with the data by visual inspection of funnel plots. While publication bias cannot be dismissed, it would take a lot of large statistically insignificant studies to overhaul the current results. Furthermore, studies showing harmful survival in women with NMRD compared to other thresholds of RD is implausible.

CI: confidence interval; **HR:** hazard ratio; **EOC:** epithelial ovarian cancer; **LVRD**: large-volume residual disease; **NMRD:** no macroscopic residual disease; **OS:** overall survival; **PDS:** upfront primary debulking surgery; **PFS:** progression-free survival

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1]Downgraded by one level because we assessed the statistical analysis and reporting domain in the QUIPS tool as being at high or unclear risk of bias in all included studies. Either no conceptual framework was reported, where the variable selection criteria in multivariate model was unclear, or quite often the authors reported that significant variables from the univariate analysis were included in the multivariable model, but with no further details. This was the most serious bias from the QUIPS domains that could influence the effect estimates.

Summary of findings 3. Large-volume residual disease (LVRD) > 1 cm versus small-volume residual disease (SVRD) < 1 cm in PDS studies

LVRD (> 1 cm) compared with SVRD (< 1 cm) after upfront primary debulking surgery (PDS) in women with advanced epithelial ovarian cancer (EOC)

Population: women with advanced EOC after PDS

Settings: all settings in adult women aged 18 years or older worldwide

Prognostic factor: LVRD > 1 cm compared with SVRD < 1 cm

Outcomes	Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
Overall survival: Median length of follow-up[1]: Range: 28 to 34.1 months	**Adjusted HR 1.22** (1.13 to 1.32)	6000 participants (5 studies)	⊕⊕⊕⊖ **moderate**[2]	We could not present illustrative absolute effects because a representative control group risk could not be ascertained from the studies. The HR estimates were adjusted for in multivariable analyses and this cannot be done in absolute terms, so we did not attempt it as the numbers were likely to mislead with any bias potentially favouring the NMRD threshold.
Progression-free survival: Median length of follow-up[1]: 28 months	**Adjusted HR** 1.30 (1.08 to 1.56)	3402 participants (2 studies)	⊕⊕⊕⊖ **moderate**[2]	There were no concerns with inconsistency and imprecision across studies (the smallest study comparison (n = 100) was imprecise but there were only n = 23 women with sub-optimal RD) due to restrictive inclusion criteria in a generally representative cohort of women with advanced EOC. Data were considerable in size in PDS studies with n > 6000 in the analysis of OS and to lesser extent > 3000 for PFS.

	The percentage of the variability in effect estimates that was due to heterogeneity rather than sampling error (chance) may not be important (as measured by the I^2 statistic) in meta-analyses including PDS studies.

CI: confidence interval; **HR:** hazard ratio; **EOC:** epithelial ovarian cancer; **LVRD**: large-volume residual disease; **NMRD:** no macroscopic residual disease; **OS:** overall survival; **PDS:** upfront primary debulking surgery; **PFS:** progression-free survival; **SVRD:** small-volume residual disease

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1]Range in Klar 2016 was 0 to 144 months.
[2]Downgraded by one level because we assessed the statistical analysis and reporting domain in the QUIPS tool as being at high or unclear risk of bias in all included studies. Either no conceptual framework was reported, where the variable selection criteria in multivariate model was unclear, or quite often the authors reported that significant variables from the univariate analysis were included in the multivariable model, but with no further details. This was the most serious bias from the QUIPS domains that could influence the effect estimates.

Summary of findings 4. Small-volume residual disease (SVRD) (< 1 cm) versus NMRD in IDS studies

SVRD (< 1 cm) compared with NMRD after primary interval debulking surgery (IDS) in women with advanced epithelial ovarian cancer (EOC)

Population: women with advanced EOC after primary IDS

Settings: all settings in adult women aged 18 years or older worldwide

Prognostic factor: SVRD < 1 cm compared with NMRD

Outcomes	Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
Overall survival: Median length of follow-up Not reported	**Adjusted HR 2.09** (1.20 to 3.60)	310 participants (1 study reporting on 2 groups)	⊕⊖⊖⊖ **very low[123]**	We could not present illustrative absolute effects because a representative control group risk could not be ascertained from the studies. The HR estimates were adjusted for in multivariable analyses and this cannot be done in absolute terms so we did not attempt it, as the numbers were likely to mislead with any bias potentially favouring the NMRD threshold.
Progression-free survival: Median length of follow-up: 47 months	P = 0.001	322 participants (1 study)	⊕⊖⊖⊖ **very low[123]**	The authors of Petrillo 2014 found that the risk of disease progression for women with RD < 1 cm after IDS was significantly higher than those with complete cytoreduction, but the magnitude of effect was not reported.

	Range: 3 to 181 months			

CI: confidence interval; **HR:** hazard ratio; **EOC:** epithelial ovarian cancer;**IDS:** interval debulking surgery; **NMRD:** no macroscopic residual disease; **OS:** overall survival; **PDS:** upfront primary debulking surgery; **PFS:** progression-free survival; **SVRD:** small-volume residual disease

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1]Downgraded by one level because we assessed the statistical analysis and reporting domain in the QUIPS tool as being at high or unclear risk of bias in all included studies. Either no conceptual framework was reported, where the variable selection criteria in multivariate model was unclear, or quite often the authors reported that significant variables from the univariate analysis were included in the multivariable model, but with no further details. This was the most serious bias from the QUIPS domains that could influence the effect estimates.
[2]Downgraded by one level for sparse data.
[3]Downgraded by one level for lack of generalisability and validity of results as reported in single analysis or very few included studies.

Summary of findings 5. Large-volume residual disease (LVRD) > 1 cm versus no macroscopic residual disease (NMRD) in IDS studies

Large-volume residual disease (LVRD) (> 1 cm) compared with NMRD after primary interval debulking surgery (IDS) in women with advanced epithelial ovarian cancer (EOC)

Population: women with advanced EOC after primary IDS

Settings: all settings in adult women aged 18 years or older worldwide

Prognostic factor: LVRD > 1 cm compared with NMRD

Outcomes	Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
Overall survival: Median length of follow-up: Not reported	**Adjusted HR 2.23** (1.49 to 3.34)	343 participants (1 study reporting on 2 groups)	⊕⊖⊖⊖ **very low**[123]	We could not present illustrative absolute effects because a representative control group risk could not be ascertained from the studies. The HR estimates were adjusted for in multivariable analyses and this cannot be done in absolute terms, so we did not attempt it as the numbers were likely to mislead with any bias potentially favouring the NMRD threshold.
Progression-free survival	Not reported			

CI: confidence interval; **HR:** hazard ratio; **EOC:** epithelial ovarian cancer;**IDS:** interval debulking surgery; **LVRD:** large-volume residual disease; **NMRD:** no macroscopic residual disease; **OS:** overall survival; **PDS:** upfront primary debulking surgery; **PFS:** progression-free survival

GRADE Working Group grades of evidence

High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1]Downgraded by one level because we assessed the statistical analysis and reporting domain in the QUIPS tool as being at high or unclear risk of bias in all included studies. Either no conceptual framework was reported, where the variable selection criteria in multivariate model was unclear, or quite often the authors reported that significant variables from the univariate analysis were included in the multivariable model, but with no further details. This was the most serious bias from the QUIPS domains that could influence the effect estimates.
[2]Downgraded by one level for sparse data.
[3]Downgraded by one level for lack of generalisability and validity of results as reported in single analysis or very few included studies.

Summary of findings 6. Large-volume residual disease (LVRD) > 1 cm versus small-volume residual disease (SVRD) < 1 cm in IDS studies

Large-volume residual disease (LVRD) > 1 cm compared with small-volume residual disease (SVRD) < 1 cm after primary interval debulking surgery (IDS) in women with advanced epithelial ovarian cancer (EOC)

Population: women with advanced EOC after primary IDS

Settings: all settings in adult women aged 18 years or older worldwide

Prognostic factor: LVRD > 1 cm compared with SVRD < 1 cm

Outcomes	Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
Overall survival: Median length of follow-up: Range: 34.3 to 43.5 months	**Adjusted HR 1.60** (1.21 to 2.11)	1572 participants (6 studies)	⊕⊕⊕⊖ **verylow**[123]	We could not present illustrative absolute effects because a representative control group risk could not be ascertained from the studies. The HR estimates were adjusted for in multivariable analyses and this cannot be done in absolute terms, so we did not attempt it as the numbers were likely to mislead with any bias potentially favouring the NMRD threshold. The percentage of the variability in effect estimates that was due to heterogeneity rather than sampling error (chance) may represent substantial heterogeneity (as measured by the I^2 statistic) in meta-analyses.
Progression-free survival: Median length of follow-up Range: 38 to 43.5 months	**Adjusted HR 1.76** (1.23 to 2.52)	1145 participants (4 studies)	⊕⊕⊕⊖ **verylow**[123]	

CI: confidence interval; **HR:** hazard ratio; **EOC:** epithelial ovarian cancer;**IDS:** interval debulking surgery; **LVRD:** large-volume residual disease; **NMRD:** no macroscopic residual disease; **OS:** overall survival; **PDS:** upfront primary debulking surgery; **PFS:** progression-free survival; **SVRD:** small-volume residual disease

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

Very low quality: We are very uncertain about the estimate.

[1]Downgraded by one level because we assessed the statistical analysis and reporting domain in the QUIPS tool as being at high or unclear risk of bias in all included studies. Either no conceptual framework was reported, where the variable selection criteria in multivariate model was unclear, or quite often the authors reported that significant variables from the univariate analysis were included in the multivariable model, but with no further details. This was the most serious bias from the QUIPS domains that could influence the effect estimates.
[2]Downgraded by one level for heterogeneity across studies.
[3]Only one study reported a comparison of SVRD < 1 cm versus LVRD > 1 cm in the strict sense that SVRD < 1 cm was mutually exclusive of NMRD (Phillips 2018).

Summary of findings 7. Residual disease (RD) > 0 cm versus NMRD in IDS studies

Any remaining residual disease (RD) (> 0 cm) compared with NMRD after primary interval debulking surgery (IDS) in women with advanced epithelial ovarian cancer (EOC)

Population: women with advanced EOC after primary IDS

Settings: all settings in adult women aged 18 years or older worldwide

Prognostic factor: RD > 0 cm compared with NMRD

Outcomes	Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
Overall survival: Median length of follow-up: range: 37 to 39 (reported in 2 studies)	**Adjusted HR 2.11** (1.35 to 3.29)	906 participants (4 studies)	⊕⊖⊖⊖ **very low[123]**	We could not present illustrative absolute effects because a representative control group risk could not be ascertained from the studies. The HR estimates were adjusted for in multivariable analyses and this cannot be done in absolute terms, so we did not attempt it as the numbers were likely to mislead with any bias potentially favouring the NMRD threshold. The percentage of the variability in effect estimates that was due to heterogeneity rather than chance may represent considerable heterogeneity (I^2 = 81%). The authors of Lecuru 2019 additionally found that the risk of death for women with any remaining RD (> 0 cm) after IDS was significantly higher than those with NMRD (n = 163, P < 0.01), but the magnitude of effect was not reported.
Progression-free survival: Median length of follow-up: not reported	**Adjusted HR 1.36** (1.05 to 1.76)	471 participants (1 study)	⊕⊖⊖⊖ **very low[123]**	The authors of Lecuru 2019 additionally found that the risk of disease progression for women with RD > 0 cm after IDS was significantly higher than those with NMRD (n = 163, P < 0.01), but the magnitude of effect was not reported.

CI: confidence interval; **HR:** hazard ratio; **EOC:** epithelial ovarian cancer;**IDS:** interval debulking surgery; **NMRD:** no macroscopic residual disease; **OS:** overall survival; **PDS:** upfront primary debulking surgery; **PFS:** progression-free survival

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.

Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1]Downgraded by one level because we assessed the statistical analysis and reporting domain in the QUIPS tool as being at high or unclear risk of bias in all included studies. Either no conceptual framework was reported, where the variable selection criteria in multivariate model was unclear, or quite often the authors reported that significant variables from the univariate analysis were included in the multivariable model, but with no further details. This was the most serious bias from the QUIPS domains that could influence the effect estimates.
[2]Downgraded by one level for heterogeneity across studies.
[3]Downgraded by one level for lack of generalisability and validity of results as reported in single analysis or very few included studies.

BACKGROUND

Description of the health condition and context

Ovarian cancer is the seventh most common cancer among women and a leading cause of death in women with gynaecological malignancies (GLOBOCAN 2018). Globally, there are approaching 300,000 new cases per year, with approximately 6.6 new cases per 100,000 women per year. A woman's cumulative risk of developing ovarian cancer by the age of 75 years is 0.72%: 0.52% in low-income countries and 0.92% in high-income countries (GLOBOCAN 2018). Ovarian cancer is rare in women under 40 years of age and most cancers in this age group are germ cell tumours. Above age 40, more than 90% are epithelial tumours and the risk increases with age (Kurman 2014; Webb 2017). Epithelial ovarian cancer is the most common type, accounting for around 90% of all ovarian cancers. This specific type of ovarian cancer starts in the surface layer covering the ovary or lining of the fallopian tube.

Ovarian cancer is best regarded as a peritoneal malignancy. The current understanding on the pathogenesis of epithelial ovarian cancer (EOC) recognises two pathways and two clinical groupings, classified as Type 1 and Type 2. Type 1 tumours comprise low-grade serous, low-grade endometrioid, clear-cell and mucinous carcinomas, and Brenner tumours. Type 2 tumours comprise the high-grade serous and endometrioid carcinomas, mixed mullerian tumours and undifferentiated carcinomas. Type 2 tumours are more common and are thought to have their origin within the fallopian tube (Perets 2016). They are associated with the BRCA (breast cancer gene) germline and somatic mutations, and histopathologically identified with aberrant p53 expression and other characteristic immunohistochemical features (Kurman 2010; Kurman 2011).

The extent of dissemination of the disease is described using the International Federation of Gynecology and Obstetrics (FIGO) staging system; stage I disease is confined to the ovaries; stage II disease is confined to the true pelvis, stage III disease is an abdominal disease where there is spread to the lining (peritoneum) of the abdominal cavity outside the pelvis or regional lymph node spread; whilst stage IV disease is outside the abdomen or parenchymatous metastases, e.g. disease with spread to distant organs such as the chest or liver (Berek 2018). Thirty per cent of women with ovarian cancer present with early-stage disease, whilst 70% have advanced stage at presentation (Torre 2018). In Europe, just over a third of women with ovarian cancer are alive five years after diagnosis (EUROCARE 2015), largely because most women with ovarian cancer are diagnosed when the cancer is already at an advanced stage (Jemal 2017). This is, in part, due to the biology of the disease and immediate acces to the abdominal cavity and non-specific symptoms, which include progressive feelings of: abdominal distension, bloating, indigestion, urinary frequency, urgency, early satiety, weight loss, reduced appetite, abdominal and pelvic pain and, less commonly, vaginal bleeding (Shafi 2018).

Description of the surgical interventions and residual disease as a prognostic factor

Surgery and chemotherapy are the mainstay of treatment for the 70% of women who present with advanced disease (FIGO stage III/IV) when surgery alone cannot be curative (Fader 2007; Torre 2018).

Appropriate initial investigations usually include ultrasonography, tumour markers and a CT scan, if malignancy is suggested by tumour markers and ultrasound. If required, an ultrasound-guided biopsy of metastatic spread is carried out to obtain histological diagnosis (Shafi 2018).

Traditionally, upfront debulking surgery (PDS) is performed to remove as much visible disease as possible, as the amount of residual tumour is one of the most important prognostic factors for survival of epithelial ovarian cancer (Bristow 2002; Chang 2013; du Bois 2009; Griffiths 1975; Hoskins 1994; Wimberger 2010). Platinum-based chemotherapy is the standard of care, in combination with debulking surgery (Colombo 2019; National Comprehensive Cancer Network 2020).

Chemotherapy followed by interval debulking surgery (IDS) is an alternative primary treatment option for women diagnosed with advanced ovarian cancer. A Cochrane Review, which comprised five randomised controlled trials (RCTs), comprehensively reviewed the evidence in this area (Coleridge 2021). The review assessed survival, quality of life and morbidity outcomes in trials that compared upfront primary and interval debulking surgery. The five trials included two large, well-documented RCTs (CHORUS (Kehoe 2015) and EORTC 55971 (Vergote 2010)), which reported no significant difference in survival between IDS compared with PDS. It was suggested that IDS may have better overall survival in stage IV disease. One included study suggested that women with FIGO stage IIIC disease with extrapelvic metastases smaller than 5 cm may have better progression-free survival after upfront debulking (Vergote 2018). The selection of women with advanced ovarian cancer for PDS or IDS remains controversial (Vergote 2013). An investigation of maximum effort cytoreductive surgery during the initial treatment of epithelial ovarian cancer comparing PDS versus IDS is being investigated in the TRUST trial (Trial of Radical Upfront Surgical Therapy in advanced ovarian cancer (ENGOT ov33/AGO-OVAR OP7)), and results are expected in 2024 (Reuss 2019).

The terms cytoreductive and debulking surgery are often used interchangeably to indicate surgical efforts aimed at removing the bulk of the tumour. No macroscopic residual disease (NMRD) (also known as 'complete' macroscopic resection or R0) is achieved when there is no visible tumour left at the end of surgery. Previously, the term 'optimal cytoreduction' had been variably defined as referring to a maximal diameter of residual tumour left behind after surgery measuring 0 to 2 cm, and in 1994 the Gynaecologic Oncology Group (GOG) defined optimal cytoreduction as having residual disease < 1 cm (Hoskins 1994). However, in 2010 the Gynaecological Cancer Inter-Group defined 'optimal' as having no visible residual tumour nodules, i.e. NMRD ('complete' is a misnomer as microscopic disease remains in the majority of patients) (Stuart 2011), which has been shown to result in better survival than small-volume residual disease (SVRD) to < 1 cm (also referred to as near-optimal) and large-volume residual disease (LVRD) which is > 1 cm (also referred to as suboptimal) and to be a better predictor of survival (Bookman 2009; Chang 2013; du Bois 2009; Sørensen 2019; Wimberger 2010). While there is now less controversy about the prognostic importance of maximum cytoreduction, there remains divided opinion about the effects of any remaining residual disease after PDS or IDS, and about what attempts should be made for maximal efforts at debulking. All women would potentially do better if there was NMRD after surgery, and obviously no surgeon sets out for suboptimal cytoreduction from the onset. However,

different philosophies are evident within the surgical community and there are also other important considerations, such as surgical skills and training, surgical and critical care resources, the woman's fitness for more radical treatment, morbidity, mortality and quality of life. The questions about PDS in ovarian cancer that appear to have become more important and relevant over the last 10 years of practice as other evidence has emerged relate to the timing of maximal surgical effort (still within initial treatment phase), and to consideration of whether there are some histological subtypes that may have better outcomes with PDS. In this review we only consider the epithelial subtype of ovarian cancer, since it comprises 90% of histological subtypes.

Surgery to achieve NMRD appears to be associated with the best chance of prolonged survival (Bookman 2009). An attempt to achieve NMRD is the recommended standard for cytoreductive surgery for advanced ovarian cancer, as advised by the British Gynaecological Cancer Society (BGCS) (BGCS 2017), European Society of Medical Oncology (ESMO) and European Society of Gynaecological Oncology (ESGO) (Colombo 2019), and the National Comprehensive Cancer Network (NCCN) (National Comprehensive Cancer Network 2020).

A Cochrane Review assessed the role of a further attempt at cytoreductive effort after LVRD remained after primary surgery (Tangjitgamol 2016). The results from three studies in the review found that a further attempt at cytoreductive surgery after chemotherapy in first-line treatment was only of benefit to those who had not had their initial surgery performed by a gynaecological oncologist (Redman 1994; Rose 2004; Van der Burg 1995).

Over the last few decades, efforts have been made to increase NMRD resection rates. It has been shown that surgery performed by gynaecologists with training in gynaecological oncology, by high-volume surgeons and high-volume centres, is associated with increased likelihood of NMRD (Bristow 2009; Greggi 2016; Woo 2012).

There is a widespread belief that tumour biology has a significant role to play in ovarian cancer outcomes. The relationship between surgical outcome and tumour biology is complex and remains unclear. The biological rationale behind the benefit of surgical cytoreduction is that removal of certain ovarian cancer tumour cells will create a supportive microenvironment to enhance chemotherapy effect (Covens 2000; Napoletano 2010). Whether it is the intrinsic biological behaviour of the tumour or the surgeon's ability to cytoreduce that determines optimal cytoreduction is not well studied. However, among the relevant prognostic factors, the extent of surgery and consequent residual disease are the most important prognostic factors. The extent of surgical effort (standard versus extensive surgery) to achieve NMRD and its impact on survival is not fully understood, as determined by a previous Cochrane Review (Hui 2022).

Within the advanced ovarian cancer group, women with stage IV ovarian cancer represent a heterogeneous group with extraperitoneal metastases. While it has been shown in a previously published guideline that NMRD resection is associated with the best chance of prolonged survival (Vergote 2016), the data are not as convincing for stage IV ovarian cancer. The presence of microscopic disease in the extraperitoneal locations has not been assessed and can potentially be even more frequent. While some stage IV diseases could be amenable to resection to NMRD (isolated splenic parenchymal lesion or resectable liver metastasis), others could be difficult (extensive mediastinal, axillary, or supraclavicular nodes or multiple, unresectable hepatic metastases). Therefore, it is worth investigating the impact of residual disease in stage IV cancers, and in particular in relation to extra-peritoneal residual disease (thoracic, mediastinum, groin, axilla, neck). The EORTC55971 trial confirmed that neoadjuvant chemotherapy results in superior survival compared with primary debulking surgery in the management of women with stage IV disease (Vergote 2010). However, there is a need for further investigation into the impact of residual disease on survival between the PDS and IDS subgroups.

This review sets out to determine the prognostic impact of residual disease on survival rates in women with advanced epithelial ovarian cancer. There are no universally established patient selection criteria, but certain baseline characteristics are important when investigating the impact of residual disease on prognosis. These include age, nutritional status, FIGO stage, comorbidities, ASA score (American Society of Anaesthesiologists' (ASA) classification of Physical Health), ECOG (Eastern Cooperative Oncology Group) performance status (score of symptom and functional status with respect to ambulatory status and need for care), BRCA status, presence of ascites on preoperative imaging and histological grade (du Bois 2009). To date, there are no specific predictive models for surgical success that are clinically useful, and the majority of previous studies have limitations in design that make their interpretation difficult (Borley 2012).

If the surgical outcome and prognosis are to be determined by tumour biology alone, the residual disease after surgery may have little influence on overall survival. However, tumour biology and the extent of disease may influence the likelihood of achieving NMRD after surgery (Colombo 2019). The extent of residual disease and prognosis could be influenced by the extent of disease measured intraoperatively by the peritoneal cancer index (PCI) score, surgical complexity score (SCS) (Elzarkaa 2018), type and extent of surgery (Aletti 2007), characteristics of the surgical team (gynaecological oncologist in a specialist centre with a high volume of cases) (Bristow 2009) and presence of ascites during surgery (du Bois 2009).

Why it is important to do this review

A greater understanding of the biology of ovarian cancer variants, especially with respect to BRCA gene mutations, has led to more sophisticated treatment regimens. These include the emergence of tailored adjuvant and maintenance chemotherapeutic options for women with BRCA somatic and germline mutations, and greater options for the chemotherapeutic approach to recurrent disease (Colombo 2019).

While the place of surgery in the context of treatment of ovarian cancer is well established, the distinctive biological phenotypes (e.g. type and grade of disease, extent of disease) should be anticipated to lead to some heterogeneity in the level of benefit derived from maximal surgical effort. There may be a greater willingness to rely on PDS for women with known subtypes of disease, such as low-grade serous cancer, that are known to be less chemo-responsive (Grabowski 2016). PDS for highly chemo-responsive disease has also been questioned by a growing acceptance of the non-inferiority of interval debulking surgery

(Coleridge 2021). The current position in many settings, in the UK and elsewhere, is to reserve PDS in advanced disease for those women who have a good performance status, and in whom it is anticipated that NMRD or SVRD can be achieved. Performance status is relevant in consideration of PDS. Though true advocates of PDS remain, many clinicians recognise that women presenting with poor performance status are likely to be too frail to undergo a PDS without significant comorbidity. In such a situation, clinical optimisation and initiation of treatment with chemotherapy is preferable with a possible benefit of reduced morbidity by reduction in disease burden with chemotherapy (Kumar 2017).

There is consensus that the surgery performed during the initial treatment of ovarian cancer, whether PDS or IDS, should aim to leave NMRD, if possible. The need for clarity on the location (cancer centre or unit) and timing from diagnosis of first look surgery (intensive staging and cytoreductive surgery) for advanced ovarian cancer has never been more relevant. Women, clinicians and commissioners of specialist cancer services need to know what the overall benefit of cytoreductive surgery for ovarian cancer is, and to determine if there are subgroups of women for whom this intervention is of greater value. Given the diversity recognised within the overall group of women with advanced-stage ovarian cancer, it is anticipated that an ethos of individualised surgical planning, whilst recognising overarching principles, would be appropriate. One recent cohort study compared operative approaches/philosophies, where an ultra-radical approach to surgery was introduced at a population level (Falconer 2020). In this population-based cohort study, all women with suspected EOC in a region of Stockholm in two national cancer registries were selected in two three-year cohorts, based on year of diagnosis (before (cohort 1) or after (cohort 2) change in surgical treatment algorithm) and followed for at least three years. The study reported five-year overall survival in non-surgically and surgically treated women. A similar study into system reorganisation that uses either a controlled before-and-after component or interrupted time series design would be able to look at the impact of any centralisation of more radical surgery on survival.

Although the size of residual tumour mass after surgery has been shown to be an important prognostic factor for advanced ovarian cancer, there is limited evidence to support the conclusion that the surgical procedure is directly responsible for the superior outcome associated with less residual disease (Girling 1996; Hunter 1992).

Whether optimal cytoreduction is more feasible in women with biologically less aggressive tumours is a subject of continued debate. Tumour biology is not thought to be the only factor affecting prognosis (Sørensen 2019), and its impact seems to be partially overruled by the extent of residual disease, i.e. whether NMRD or SVRD was achieved (du Bois 2009). It has also been suggested that further evaluation of biological factors may help select women who are most likely to benefit from PDS (du Bois 2009; Markar 2016). It has been suggested that women whose cancer is cytoreduced to NMRD and SVRD at PDS may have super-imposable progression-free survival, meaning that women with high tumour load, completely resected at the time of surgery, may have micro/macroscopic unrecognised residual disease (Fagotti 2020). In this review, we will analyse PDS and IDS separately, as PDS achieving cytoreduction to < 1 cm may be equivalent to IDS achieving cytoreduction to NMRD.

The aim of this review is to investigate the effects of residual disease in women who received PDS or IDS for advanced epithelial ovarian cancer. This review should help to determine the prognostic impact of residual disease after surgery on survival.

OBJECTIVES

To assess the prognostic impact of residual disease after primary surgery on survival outcomes or advanced (stage III and IV) epithelial ovarian cancer. In separate analyses, primary surgery included both upfront primary debulking surgery (PDS) followed by adjuvant chemotherapy and neoadjuvant chemotherapy followed by interval debulking surgery (IDS). Each residual disease threshold is considered as a separate prognostic factor.

Investigation of sources of heterogeneity

We examined differences between FIGO stages III and IV in different thresholds of residual disease after primary surgery. We considered factors such as age, grade, length of follow-up, type and experience of surgeon, and type of surgery in the interpretation of any heterogeneity.

We also performed sensitivity analyses that distinguished between studies that included NMRD in residual disease (RD) categories of < 1 cm and those that did not. This was applicable to comparisons involving RD < 1 cm with the exception of RD < 1 cm versus NMRD.

We evaluated women undergoing PDS and IDS in separate analyses.

METHODS

Criteria for considering studies for this review

Types of studies

We included data from RCTs, prospective and retrospective cohort studies, and unselected case series of 100 or more women that included a concurrent comparison of different RD thresholds after primary surgical intervention. Any data collected from RCTs were retrospective and taken from trials that randomised groups of women to various chemotherapy protocols after primary or interval debulking surgery. We categorised the surgical outcome as macroscopic, optimal and suboptimal debulking, based on the maximum size of postoperative residual disease.

In order to minimise bias, we only included studies of multivariate Cox regression models that used sensible adjustment factors associated with survival in women with advanced EOC (e.g. age, stage, grade, extent of disease at diagnosis). We excluded studies that only reported unadjusted results. To assess the adequacy of adjustment factors used in multivariate Cox models, we used the 'adjustment for other prognostic factors' and 'statistical analysis and reporting' domains of the quality in prognosis studies (QUIPS) tool (Riley 2019). Therefore, in theory, only one other factor needed to be adjusted for the study to meet the criteria for inclusion in the review, but we judged such studies as being at high risk of bias in these domains.

We excluded case-control studies, studies that did not have concurrent comparison groups and case series of fewer than 100 women. This was to attempt to optimise the quality of the review, as poor study designs would have introduced additional forms of bias. The inclusion of adequately sized studies, although pragmatic, may

also provide more reliable estimates due to restricting results to those reporting multiple adjustments in statistical models.

Types of participants

We included adult women (over 18 years of age) with surgically staged advanced epithelial ovarian cancer (FIGO stages III and IV) who had confirmed histological diagnoses. We excluded women with other concurrent malignancies.

Women were included into two distinct groups: those who received primary debulking surgery (PDS) followed by platinum-based chemotherapy and those who received interval debulking surgery (IDS), which involves receiving the surgery sandwiched between a schedule of chemotherapy. We analysed these distinct groups separately.

Details of prognostic factor

The surgical intervention for which we assessed the resulting prognostic factor was primary debulking surgery (upfront and interval debulking).

We included studies that reported all RD thresholds after surgery but we defined optimal RD as surgery leading to residual tumours with a maximum diameter of any threshold up to 1 cm. The main RD thresholds of interest were microscopic RD (labelled as no macroscopic residual disease (NMRD)); RD < 1 cm and exclusive of 0 cm, categorised as small-volume residual disease (SVRD); and RD > 1 cm, categorised as large-volume residual disease (LVRD). However, we included studies reporting any size of RD but restricted to the most pertinent comparisons in key summary sections. We noted details of any women who had primary surgery that resulted in RD that did not meet the criteria specified in the study as 'optimal', namely not categorised as NMRD or SVRD cytoreduction.

We applied the above RD thresholds to both PDS (primary debulking surgery followed by platinum-based chemotherapy) and IDS (platinum-based chemotherapy followed by interval debulking surgery) settings.

- No macroscopic residual disease (NMRD) after PDS (RD = 0 cm).
- Small-volume residual disease (SVRD) after primary cytoreduction (RD 0.1 cm to 1 cm).
- Large-volume residual disease (LVRD) after cytoreduction (RD > 1 cm).

Types of outcome measures

- Overall survival: survival until death from any cause. We assessed survival from the time at which women were enrolled in the study.
- Progression-free survival.

We extracted survival estimates as time-to-event data from an adjusted multivariate Cox model (as outlined above in 'Types of studies'). This is the most appropriate way to analyse these outcomes as it accounts for any loss to follow-up and will correctly allow for censoring.

Search methods for identification of studies

We sought papers in all languages and translated them when necessary.

We searched the following electronic databases on 30 August 2021:

- the Cochrane Central Register of Controlled Trials (CENTRAL; 2021, Issue 8), in the Cochrane Library;
- MEDLINE via Ovid (1950 to 30 August 2021);
- Embase via Ovid (1950 to 2021 week 34).

The MEDLINE, EMBASE and CENTRAL search strategies were based on terms related to the review topic and are presented in Appendix 1, Appendix 2 and Appendix 3, respectively. We searched the databases from 1950 up to end of August 2021.

We identified all relevant articles found on PubMed and used the 'related articles' feature to carry out a further search for newly published articles.

Searching other resources

Unpublished and grey literature

We searched metaRegister, Physicians Data Query, www.controlled-trials.com/rct, www.clinicaltrials.gov and www.cancer.gov/clinicaltrials for ongoing trials.

Handsearching

We checked the citation lists of relevant publications, abstracts of scientific meetings and included studies through handsearching, and we contacted experts in the field to identify further reports of studies. We handsearched reports of conferences from the following sources.

- *Gynecologic Oncology* (Annual Meeting of the American Society of Gynecologic Oncologists).
- *International Journal of Gynecological Cancer* (Annual Meeting of the International Gynecologic Cancer Society).
- *British Journal of Cancer*.
- British Cancer Research Meeting.
- Annual Meeting of European Society of Medical Oncology (ESMO).
- Annual Meeting of the American Society of Clinical Oncology (ASCO).

Correspondence

We contacted authors of relevant trials to ask if they knew of further data, which may or may not have been published.

Data collection and analysis

Selection of studies

We downloaded all titles and abstracts retrieved by electronic searching to the reference management database Endnote. After removing duplicates, three review authors (AB, PK, SH) examined the remaining references independently. We excluded those studies that clearly did not meet the inclusion criteria and obtained copies of the full text of potentially relevant references. Three review authors (AB, PK, SH) assessed the eligibility of retrieved papers independently. We resolved disagreements by discussion between the three review authors or, when necessary, by appeal to a fourth review author (RN, KG). We documented the reasons for exclusion.

Data extraction and management

For included studies, we extracted items relevant to prognostic factor studies, derived from the checklist for critical appraisal and data extraction for systematic reviews of prediction modelling studies (CHARMS) (Moons 2014). This included data on the following:

- Author, year of publication and journal citation (including language).
- Country.
- Setting.
- Inclusion and exclusion criteria.
- Study design, methodology.
- Study population:
  - total number enrolled in each group;
  - participant characteristics;
  - age;
  - comorbidities.
- Ovarian cancer details at diagnosis:
  - FIGO stage (III or IV);
  - histological cell type;
  - preoperative tumour volume;
  - ascites (large or small volume);
  - tumour grade;
  - extent of disease.
- Surgical intervention details:
  - details of primary optimal cytoreductive surgery;
  - upfront and interval debulking settings.
- Details of platinum-based chemotherapy:
  - dose;
  - number of chemotherapy cycles before and after surgery;
  - type of surgeon (gynaecological oncologist, gynaecologist, general surgeon);
  - experience of surgeon;
  - type of surgery (ultra-radical or standard).
- Details of prognostic factor:
  - details of residual disease;
  - definition of residual disease thresholds in study;
  - covariates included in multivariate Cox models for survival that include residual disease.
- Risk of bias in study (see 'Assessment of risk of bias in included studies').
- Duration of follow-up.
- Outcomes (see 'Types of outcome measures').

For time-to-event data (survival and progression-free survival), we extracted the log of the hazard ratio (log(HR)) and its standard error from study reports. If the study did not report these, we did not attempt to estimate the log(HR) and its standard error using the methods of Parmar 1998, as we only included adjusted analyses.

We noted the time points at which outcomes were collected and reported.

Three review authors (AB, PK, SH) independently extracted data using a data collection form specially designed for the review. We resolved differences between review authors by discussion or by appeal to a fourth review author (KG), when necessary.

Assessment of risk of bias in included studies

Three review authors independently extracted data and assessed risk of bias. We extracted the data using the CHARMS-PF (checklist for critical appraisal and data extraction for systematic reviews - prognostic factor studies; Riley 2019). We assessed the risk of bias for each outcome (overall survival and progression-free survival) in each study. We assessed risk of bias (and appraised quality) in the prognostic assessment of residual disease in the included studies using the quality in prognosis studies (QUIPS) tool (Appendix 4). QUIPS is a tool designed to assess risk of bias in prognostic factor studies (Riley 2019). It assesses bias across the following six domains using intermediate signalling questions to aid the decision-making process.

1. Participant selection
2. Study attrition
3. Prognostic factor measurement
4. Outcome measurement
5. Adjustment for other prognostic factors
6. Statistical analysis and reporting

In addition, we considered the applicability of the study for four of the domains, as reported in other tools (Whiting 2011; Wolff 2019). We judged risk of bias and concerns regarding applicability using the tools shown in Appendix 4. The questions regarding applicability included the following.

- Domain 1: participant selection. Are there concerns that the included women do not match the review question?
- Domain 3: prognostic factor measurement. Are there concerns that residual disease, the way that it is measured, or the way that it is interpreted, differ from the review question?
- Domain 4: outcome measurement. Are there concerns that the outcome does not match the review question or that follow-up was not of sufficient duration?
- Domain 5: adjustment for other prognostic factors. Did the prognostic factors adjusted for match the review question?

Three review authors (AB, PK, SH) applied the risk of bias tool independently and resolved differences by discussion or by appeal to a fourth review author (KG). We presented the results in a risk of bias summary table. We interpreted the results of meta-analyses in light of the findings with respect to risk of bias.

Measures of effect

For time-to-event data (overall and progression-free survival), we used the adjusted hazard ratio (HR). We did not use unadjusted results, as outlined above in 'Types of studies'.

Dealing with missing data

We did not impute missing outcome data for any of the outcomes.

Assessment of heterogeneity

We assessed heterogeneity between studies by visual inspection of forest plots, by estimation of the percentage of heterogeneity between trials that cannot be ascribed to sampling variation (Higgins 2003), by a formal statistical test of the significance

of the heterogeneity (Deeks 2001), and, where possible, by subgroup analyses (see 'Subgroup analysis and investigation of heterogeneity'). If there was evidence of substantial heterogeneity, we investigated and reported the possible reasons for this.

Assessment of reporting biases

We examined the symmetry of funnel plots corresponding to meta-analyses of overall survival to assess the potential for small study effects in analyses containing 10 or more studies. We tested for asymmetry where evidence of asymmetry may have been an indicator of publication bias (Debray 2018; Sterne 2011).

Data synthesis

If sufficient clinically similar studies were available, we pooled their adjusted results in meta-analyses. We reported results by FIGO stage (see 'Subgroup analysis and investigation of heterogeneity').

- For time-to-event data, we pooled hazard ratios (HRs) using the generic inverse variance facility of Review Manager 2020.
- We used random-effects models with inverse variance weighting for all meta-analyses (DerSimonian 1986).
- We reported analyses separately for women who received upfront and interval debulking surgery.

Subgroup analysis and investigation of heterogeneity

We considered factors such as age, grade, length of follow-up, type and experience of surgeon, and type of surgery in the interpretation of any heterogeneity.

We performed subgroup analyses grouping studies by women with FIGO stage III versus stage IV disease.

We analysed women undergoing PDS and IDS in separate analyses (see above).

Sensitivity analysis

We had planned to perform sensitivity analysis that restricted the analyses to studies we judged to be at an overall low risk of bias. However, the overall profiles of the included studies were largely very similar.

We performed a sensitivity analysis that distinguished between studies that included NMRD in residual disease categories of < 1 cm and those that did not. This was applicable to some comparisons involving RD < 1 cm, with the exception of SVRD versus NMRD. In this area, RD <1 cm should be exclusive of NMRD and is often described as RD = 0.1 cm to 1 cm in the literature, for clarity.

We also conducted a number of post hoc sensitivity analyses. This included excluding one study (Klar 2016), which included a proportion of women with early and unknown stage disease.

Summary of findings and assessment of the certainty of the evidence

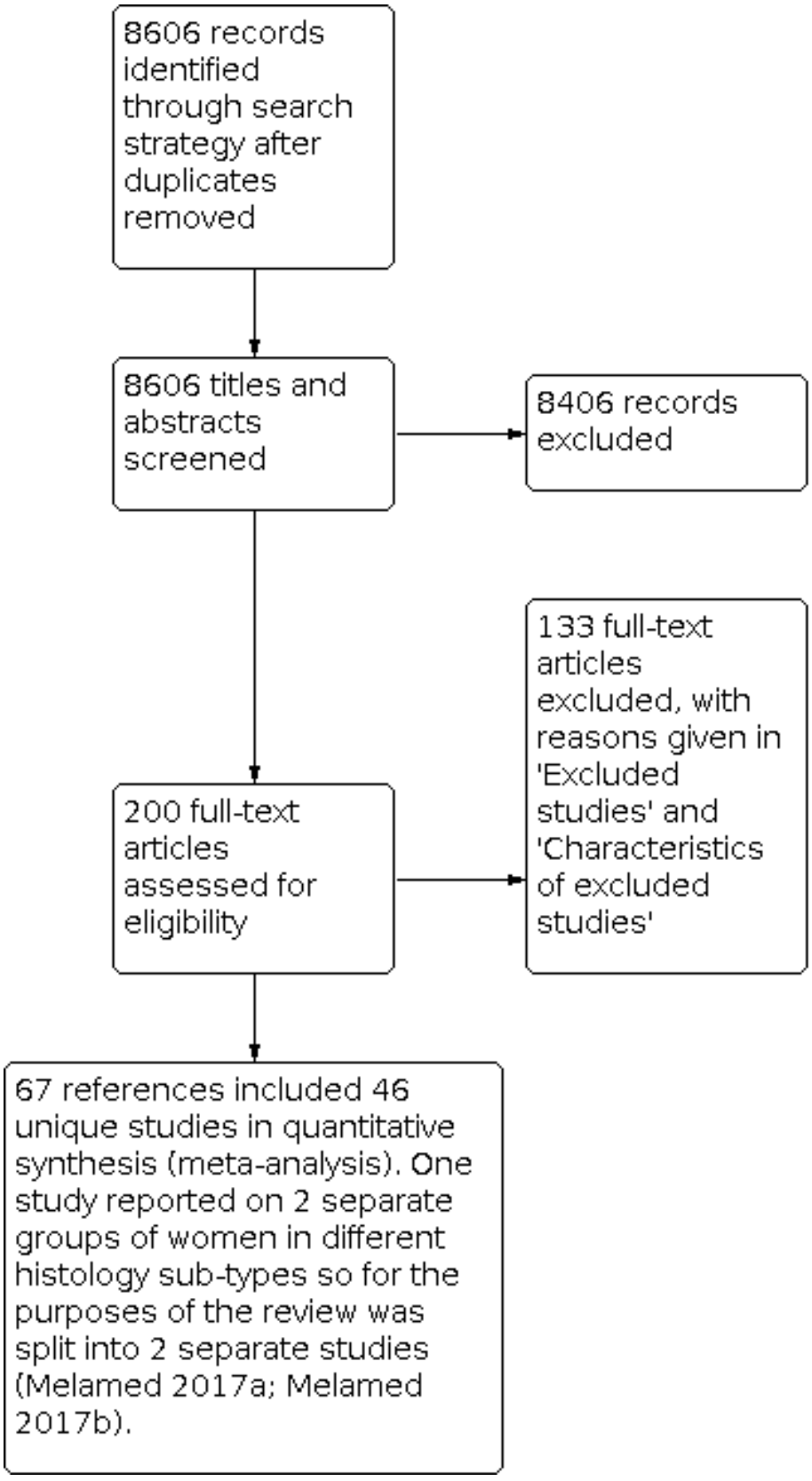
Guidance on the use of GRADE for prognostic factor studies has not yet been published (Foroutan 2020; GRADE Working Group), but we attempted to appraise the quality and certainty of the evidence where possible. We constructed summary of findings tables to present the results of outcomes in the review for the main comparisons involving prognostic factor thresholds of NMRD, SVRD (0.1 cm to 1cm) and LVRD. We used the GRADE system to rank the certainty of the evidence (Foroutan 2020; GRADE Working Group). Two review authors (AB, SH) independently graded the evidence and resolved differences by discussion or by involving a third review author (PK). We based our judgements on the strength of the body of evidence based on the domains presented in Appendix 5. Where the evidence was based on single studies, or where there was no evidence on a specific outcome for comparisons, we included the outcome in the summary of findings table and graded or explained in a narrative account accordingly. We gave the rationale for each judgement in the table footnotes. We interpreted the results of the review in light of this graded evidence. Summary of findings tables are given for PDS studies in Summary of findings 1, Summary of findings 2 and Summary of findings 3 and in IDS studies in Summary of findings 4, Summary of findings 5 and Summary of findings 6. The comparison involving any remaining macroscopic disease (RD > 0 cm) and NMRD in an IDS setting was also an important comparison so we additionally gave this a certainty of evidence judgement (Summary of findings 7).

RESULTS

Results of the search

The search strategy identified 8606 unique references (Figure 1). The title and abstract screening of these references identified 200 studies as potentially eligible for the review. The full-text screening of the 200 references identified 13 references, reporting on two RCTs (Kehoe 2015; Vergote 2010), but these trials did not meet the inclusion criteria as they did not report results across residual disease thresholds; instead they gave comparisons of residual disease by type of surgery. These trials were reported in a recent Cochrane Review (Coleridge 2021), which assessed chemotherapy versus surgery for initial treatment in advanced ovarian epithelial cancer along with another three trials (Chekman 2015; Fagotti 2020; Onda 2020), which did not report any of their outcomes for extent of disease by type of initial surgery.

Figure 1. Study flow diagram.



We excluded 133 references reporting on 115 studies that investigated the effects of residual disease after primary surgery for the reasons described in the table Characteristics of excluded studies. The remaining 67 references, reporting on 46 unique studies, met our inclusion criteria and are described in the table Characteristics of included studies. Fifty-two of these, reporting on 30 unique studies, reported on residual disease for PDS. One included publication, Klar 2016, reported results based on four individual RCTs but each one alone did not meet the inclusion criteria due to different scope so we included the combined analysis reported in Klar 2016. One study reported on two separate groups of women in different histology sub-types so for the purposes of the review we split it into two separate studies (Melamed 2017a; Melamed 2017b), therefore we refer to 31 included studies throughout. The other 15 studies reported on residual disease for IDS.

Searches of the grey literature did not identify any additional relevant trials.

There were three RCTs evaluating the effectiveness of surgery in advanced-stage epithelial ovarian cancer (Redman 1986; Rose 2004; Van Der Burg 1996). However, we excluded all three of these trials as they were designed to evaluate the benefits of surgery after an induction period with chemotherapy treatment, where the surgery was performed as a secondary procedure after initial (primary) surgery and they have been evaluated in a separate Cochrane Review (Tangjitgamol 2016).

Characteristics of included studies

See Characteristics of included studies table.

Residual disease after upfront primary debulking surgery (PDS)

The 31 included studies assessed a total of 22,376 women (Akahira 2001; Aletti 2006; Ataseven 2016; Bristow 2011; Chan 2003; Chang 2012a; Chang 2012b; Chi 2001; Chi 2006; Cuylan 2018; Eisenkop 2003; Feng 2016; Hofstetter 2013; Kahl 2017; Klar 2016; Langstraat 2011; Luger 2020; McGuire 1995; Melamed 2017a; Melamed 2017b; Paik 2018; Peiretti 2010; Peiretti 2012; Polterauer 2012; Shim 2016; Tewari 2016; Tseng 2018; Van Geene 1996; Wimberger 2010; Winter 2007; Winter 2008). Three studies included a small proportion of women with early-stage (predominantly stage II) or unknown disease. Although not stringently part of our initial inclusion criteria, we included a study if the proportion with unknown or early-stage disease in the entire cohort was small. The proportion of women with early or unknown stage of disease in Feng 2016 (9.3%), Polterauer 2012 (6.6%) and Klar 2016 (12.5%) was not going to affect the applicability of the results. The analyses in Klar 2016 included 1182 women with stage IIB to IIIB disease and 3684 had stage IIIC to IV disease. The study contributed heavily to the analyses, but the results were robust to its exclusion in a sensitivity analysis. The four individual RCTs used in the analyses could not be included separately because residual disease (RD) was not reported.

Four studies reported exclusively on women with stage IV epithelial ovarian cancer (EOC) and included 225, 326, 573 and 360 stage IV women respectively (Akahira 2001; Ataseven 2016; Wimberger 2010; Winter 2008).

Five studies reported exclusively on women with stage IIIC EOC (Aletti 2006; Bristow 2011; Chang 2012b; Chi 2006; Eisenkop 2003); whereas Cuylan 2018 and Winter 2007 reported women with stage IIIA to C disease; whilst 16 studies reported on both stage III and IV EOC (Chan 2003; Chang 2012a; Chi 2001; Hofstetter 2013; Langstraat 2011; McGuire 1995; Melamed 2017a; Melamed 2017b; Paik 2018; Peiretti 2010; Peiretti 2012; Polterauer 2012; Shim 2016; Tewari 2016; Tseng 2018; Van Geene 1996).

The number of women included in all studies varied from 104 in the Chan 2003 study to 5055 women in the Klar 2016 analysis. The larger studies tended to combine results from primary studies but generally it was not possible to report the results of these separately due to the scope of the original publications that had a different focus.

For a summary of the total number of women included in each study, as well as stage and residual disease details see Table 1.

Design

All analyses examining residual disease thresholds following surgery were retrospective in nature.

Four studies were primarily prospective cohort studies (Eisenkop 2003; Hofstetter 2013; Polterauer 2012; Van Geene 1996).

The Winter 2007, Winter 2008 and Klar 2016 studies were retrospective analyses of six, four and four randomised controlled trials of various chemotherapy protocols, respectively. The Winter 2007 study reported on women with stage III EOC, Winter 2008 reported on women with stage IV EOC and Klar 2016 a mix of stages included a small proportion of early and unknown. Winter 2007 included women from GOG protocols 111, 114, 132, 152, 158 and 172 (Armstrong 2006; Markman 2001; McGuire 1996; Muggia 2000; Ozols 2003; Rose 2004), Winter 2008 included women from GOG protocols 111, 132, 152 and 162 (McGuire 1996; Muggia 2000; Rose 2004; Spriggs 2007) and Klar 2016 reported a combined analysis of four individual RCTs (OVAR 3, 5, 7 and 9). Likewise, the McGuire 1995 study was a retrospective analysis of a randomised controlled trial of two different chemotherapy protocols.

All remaining studies were analyses of retrospective data from hospital databases, medical records and cancer registries.

Participant characteristics

Fourteen studies were conducted in the USA (Aletti 2006; Bristow 2011; Chan 2003; Chi 2001; Chi 2006; Eisenkop 2003; Langstraat 2011; Melamed 2017a; Melamed 2017b; McGuire 1995; Tewari 2016; Tseng 2018; Winter 2007; Winter 2008), whilst four were set in South Korea (Chang 2012a; Chang 2012b; Paik 2018; Shim 2016), nine set predominantly in Europe including Germany, Belgium, France, Spain, Italy, Austria and the UK (Ataseven 2016; Hofstetter 2013; Kahl 2017; Klar 2016; Luger 2020; Peiretti 2010; Polterauer 2012; Van Geene 1996; Wimberger 2010); the study Cuylan 2018 was set in Turkey, Feng 2016 in China and the Akahira 2001 study was conducted in 24 centres in Japan. One of the studies included populations from multiple locations: Peiretti 2012 (Italy and the USA).

The mean or median age reported for women with advanced EOC varied between 50.9 years (Tewari 2016) to 73.5 (Langstraat 2011) years with the range between 16 to 91 years.

Details of PDS reported in studies

RD thresholds ranged from NMRD up to > 5 cm across the included studies. The most common comparisons were of RD thresholds NMRD, SVRD (described in most studies as being < 1 cm, but exclusive of NMRD) and LVRD. We did identify studies where optimal RD was defined up to < 2 cm, but more recent studies and guidelines (BGCS 2017; du Bois 2009) state that surgery should not be considered optimal beyond 1 cm (however, we assessed RD as a prognostic factor and we included studies that included all RD thresholds, but only reported the most pertinent comparisons in the key sections of the review).

Women in all the studies described above underwent PDS followed by platinum-based adjuvant chemotherapy. All women were confirmed histologically to have invasive epithelial ovarian cancer.

The speciality of the surgeon who performed PDS (for example, general surgeon, gynaecologic surgeon or specialist gynaecologic oncology surgeon) was not reported in 20 of the included studies (Akahira 2001; Aletti 2006; Chang 2012a; Feng 2016; Hofstetter 2013; Klar 2016; Langstraat 2011; McGuire 1995; Melamed 2017a; Melamed 2017b; Paik 2018; Peiretti 2010; Polterauer 2012; Shim 2016; Tewari 2016; Tseng 2018; Van Geene 1996; Wimberger 2010; Winter 2007; Winter 2008); whereas specialist gynaecologic oncology surgeons undertook PDS in 11 studies (Ataseven 2016; Bristow 2011; Chan 2003; Chang 2012b; Chi 2001; Chi 2006; Cuylan 2018; Eisenkop 2003; Kahl 2017; Luger 2020; Peiretti 2012).

The mean duration of PDS was reported to be 210 minutes (range: 40 to 480 minutes) in Aletti 2006. Similarly the median duration of PDS was reported to be 194 minutes (range: 60 to 750 minutes) and 180 minutes (range: 55 to 480 minutes) in the Chi 2006 and Eisenkop 2003 studies respectively. All three studies reported on women with stage IIIC disease. On the other hand, the Akahira 2001 study reported on women with stage IV disease and the median duration of PDS was found to be 240 minutes (range 40 to 780 minutes). Two studies reported on the mean duration of PDS on women with stage III and IV disease: 270 minutes (range: 70 to 480 minutes) in Peiretti 2010 and 280 minutes (range: 36 to 893 minutes) in Tseng 2018.

The duration of PDS was not reported in the remaining 25 studies (Ataseven 2016; Bristow 2011; Chan 2003; Chang 2012a; Chang 2012b; Chi 2001; Cuylan 2018; Feng 2016; Hofstetter 2013; Kahl 2017; Klar 2016; Langstraat 2011; Luger 2020; McGuire 1995; Melamed 2017a; Melamed 2017b; Paik 2018; Peiretti 2012; Polterauer 2012; Shim 2016; Tewari 2016; Van Geene 1996; Wimberger 2010; Winter 2007; Winter 2008).

The median estimated operative blood loss was 500 mL (range 20 mL to 7500 mL); 850 mL (range 30 mL to 5000 mL) and 1085 mL (range 40 mL to 11,000 mL) in the Chi 2006, Eisenkop 2003 and Akahira 2001 studies, respectively. In the latter study, blood transfusion was given to 112 women (50%) intra- and postoperatively. Peiretti 2010 and Peiretti 2012 reported the estimated blood loss using different measures as 700 mL (range 50 mL to 6000 mL) and 1000 mL (range 200 mL to 8500 mL), respectively. Intraoperative blood transfusion was given to 112 (43.2%) and 152 (64%) women in Peiretti 2010 and Peiretti 2012 respectively, while postoperative blood transfusion was given to 140 (50.1%) women in Peiretti 2010 and 150 (63%) women in Peiretti 2012. The Hofstetter 2013 study did not report on the estimated blood loss, however they reported that nine of 185 women (4.86%) required blood transfusion.

Only five studies reported on the length of hospital stay (LHS). In the studies by Chi 2006, Eisenkop 2003 and Peiretti 2012 the median LHS was 10 days, with a range of 0 to 59, 0 to 93 and 4 to 24 days, respectively. The median LHS was 9 days and 8 days (range: 1 to 22 days) in Peiretti 2010 and Tseng 2018, respectively.

Postoperative mortality within 30 days of PDS ranged from 0.4% to 4.3% in eight studies reporting this outcome (Ataseven 2016; Aletti 2006; Bristow 2011; Chi 2001; Chi 2006; Eisenkop 2003; Langstraat 2011; Tseng 2018). One study reported a postoperative mortality rate of 45% but this was during a median follow-up period of 49.6 months (interquartile range (IQR) 32.9 to 66.3) (Luger 2020).

Postoperative mortality and morbidity were not reported in 19 studies (Akahira 2001; Chan 2003; Chang 2012a; Chang 2012b; Feng 2016; Hofstetter 2013; Klar 2016; Melamed 2017a; Melamed 2017b; McGuire 1995; Paik 2018; Peiretti 2010; Peiretti 2012; Polterauer 2012; Shim 2016; Van Geene 1996; Wimberger 2010; Winter 2007; Winter 2008).

Two studies used a postoperative residual disease cutoff of < 2 cm to define an optimal level of remaining RD after surgery (Akahira 2001; Van Geene 1996). Eighteen studies considered that an optimal outcome was achieved only if NMRD was left behind at the conclusion of PDS (Ataseven 2016; Chang 2012a; Chang 2012b; Cuylan 2018; Eisenkop 2003; Feng 2016; Hofstetter 2013; Kahl 2017; Langstraat 2011; Luger 2020; Melamed 2017a; Melamed 2017b; Paik 2018; Peiretti 2010; Peiretti 2012; Tewari 2016; Tseng 2018; Wimberger 2010). Four studies used a postoperative RD cutoff of < 1 cm to define the optimal level of remaining RD (Aletti 2006; Bristow 2011; Chan 2003; Klar 2016). The remaining seven studies did not define what is considered optimal in the study methodology but analysed the outcome by a range of postoperative RD (Chi 2001; Chi 2006; McGuire 1995; Polterauer 2012; Shim 2016; Winter 2007; Winter 2008).

Four studies did not make direct comparisons against NMRD. These studies included NMRD in the RD < 1 cm (Chi 2001; Chan 2003) and RD < 2 cm categories (Akahira 2001; McGuire 1995). None of the studies reported the proportion of participants with NMRD. While Winter 2008 did give a breakdown of various RD categories, the authors additionally reported a comparison involving RD > 1 cm versus < 1 cm with the latter including NMRD (n = 29/107).

The rate of NMRD after surgery was reported in 20 studies (Aletti 2006; Ataseven 2016; Chang 2012b; Chi 2006; Cuylan 2018; Eisenkop 2003; Kahl 2017; Langstraat 2011; Luger 2020; Melamed 2017a; Melamed 2017b; Paik 2018; Peiretti 2010; Peiretti 2012; Polterauer 2012; Tewari 2016; Tseng 2018; Winter 2007; Winter 2008). It was achieved in 4906 out of 15,246 women (32.2%) with the lowest macroscopic disease rate reported by Tewari 2016 (4.9%) and the highest (86%) reported by Eisenkop 2003.

Postoperative RD < 1 cm (SVRD) was achieved in 8201 out of 19,185 women (42.75%) as calculated from 19 studies (Aletti 2006; Ataseven 2016; Bristow 2011; Chan 2003; Chang 2012a; Chi 2001; Chi 2006; Cuylan 2018; Eisenkop 2003; Klar 2016; Langstraat 2011;

Melamed 2017a; Melamed 2017b; Paik 2018; Polterauer 2012; Tewari 2016; Wimberger 2010; Winter 2007; Winter 2008). The lowest rate for RD < 1 cm was 25.3% (71/281) in the Chi 2001 study and the highest was 96% (392/408) in the Eisenkop 2003 study.

In 26 studies all women received postoperative platinum-based chemotherapy (Aletti 2006; Ataseven 2016; Bristow 2011; Chan 2003; Chang 2012a; Chang 2012b; Cuylan 2018; Eisenkop 2003; Feng 2016; Hofstetter 2013; Kahl 2017; Klar 2016; Langstraat 2011; Luger 2020; McGuire 1995; Melamed 2017a; Melamed 2017b; Paik 2018; Peiretti 2010; Peiretti 2012; Polterauer 2012; Tewari 2016; Van Geene 1996; Wimberger 2010; Winter 2007; Winter 2008). In four studies the majority of women (95.1%, 96%, 97%, 98.4%, 99% respectively) received postoperative platinum-based chemotherapy (Akahira 2001; Chi 2001; Chi 2006; Tseng 2018). The main reason for not receiving postoperative chemotherapy was postoperative death within 30 days of surgery and absent records (Chi 2001). Other reasons for not receiving postoperative chemotherapy or receiving non-platinum-based chemotherapy were poorly reported. The study by Shim 2016 did not report the number of women who received postoperative chemotherapy.

Fourteen studies reported the survival outcome for NMRD (Aletti 2006; Ataseven 2016; Bristow 2011; Chi 2006; Cuylan 2018; Eisenkop 2003; Feng 2016; Hofstetter 2013; Kahl 2017; Langstraat 2011; Paik 2018; Tewari 2016; Winter 2007; Winter 2008).

Outcomes

The median duration of follow-up varied from 28 months (Winter 2008) to 77.7 months (Tseng 2018), with a range between 1 and 199 months (Chi 2006). The duration of follow-up was not reported in seven studies (Chang 2012b; McGuire 1995; Peiretti 2012; Shim 2016; Tewari 2016; Van Geene 1996; Wimberger 2010).

Only two studies did not report overall survival (Peiretti 2010; Shim 2016), while 16 studies reported progression-free survival and used appropriate statistical techniques (hazard ratios to correctly allow for censoring) (Chang 2012a; Chang 2012b; Cuylan 2018; Feng 2016; Klar 2016; Luger 2020; McGuire 1995; Paik 2018; Peiretti 2010; Polterauer 2012; Shim 2016; Tewari 2016; Tseng 2018; Wimberger 2010; Winter 2007; Winter 2008). Prognostic factors were adjusted for in the analysis of survival outcomes in each study using Cox regression. Between them, the 30 studies (31 with Melamed split (Melamed 2017a; Melamed 2017b)) included 29 different prognostic factors in the analysis. The number of prognostic factors included in the analysis ranged from two in Eisenkop 2003 to 10 in Tewari 2016. The prognostic factors most frequently included in the analyses are (in order of frequency) residual disease (26 studies), age (23 studies), stage (21 studies), performance status (nine studies), histology (nine studies) and tumour grade (six studies). A list of the different prognostic factors is shown in Appendix 6.

For the distribution of these factors at baseline for each study and by residual disease, see the table Characteristics of included studies.

Residual disease after interval debulking surgery (IDS)

The 15 included studies assessed a total of 3697 women (Bixel 2020; Cioffi 2018; Davidson 2019; Iwase 2015; Kaban 2017; Lecointre 2020; Lecuru 2019; Liu 2020; Lorusso 2016; Petrillo 2014; Phillips 2018; Shibutani 2020; Stoeckle 2014; Zhang 2018; Zhu 2016). One study, whilst it reported descriptive statistics for 102 women, only had 85 women who underwent interval debulking surgery (IDS) (Cioffi 2018). Although this was not strictly part of our inclusion criteria (i.e. n ≥ 100), we noted this study as a caveat. Additionally, adjusted hazard ratios (HRs) and 95% confidence intervals (CIs) were not reported in Petrillo 2014 and Lecuru 2019 in their multivariate Cox models; however, P values were reported in both. Two of the included studies were abstracts only (Lecuru 2019; Lorusso 2016).

All studies included women with advanced EOC who underwent IDS (neoadjuvant chemotherapy (NACT) given prior to surgery). Twelve of the studies provided descriptive statistics of FIGO stage - all of which included samples of women with FIGO stages III and IV (Bixel 2020; Cioffi 2018; Davidson 2019; Iwase 2015; Lecointre 2020; Liu 2020; Petrillo 2014; Phillips 2018; Shibutani 2020; Stoeckle 2014; Zhang 2018; Zhu 2016). For the three remaining studies, only Kaban 2017 and Lecuru 2019 reported in their methods that women with stage IIIC and IV ovarian cancer were included; we could not determine FIGO staging for Lorusso 2016.

Study sample size varied from 102 (Cioffi 2018) to 672 (Zhu 2016).

For a summary of the total number of women included in each study, as well as stage and residual disease details see Table 2.

Design

All analyses examining RD thresholds were retrospective in nature with data collected from past medical records and databases. The exceptions were Lecuru 2019, which was a secondary analysis of the CHIVA double-blind randomised phase II GINECO study that sought to examine the effects of nintedanib in combination with NACT (Ferron 2019); Davidson 2019, whose sample comprised data collected retrospectively from medical records as well as prospective participants(the purpose of the prospective data collection being to explore the role of minimally invasive surgery following NACT); and Lecointre 2020, whose sample was from a multicentre cohort study of women with histologically confirmed advanced epithelial ovarian cancer who all consented to participation.

Participant characteristics

Three of the studies were conducted in Italy (Cioffi 2018; Lorusso 2016; Petrillo 2014), three in France (Lecointre 2020; Lecuru 2019; Stoeckle 2014), three in China (Liu 2020; Zhang 2018; Zhu 2016), two in the USA (Bixel 2020; Davidson 2019), two in Japan (Iwase 2015; Shibutani 2020), and one study was conducted in Turkey (Kaban 2017), and the UK (Phillips 2018) each. Five of the studies were conducted across multiple centres: Lecointre 2020 collected data from nine French referral centres, Davidson 2019 from three US institutions, Bixel 2020 from two US institutions, Lorusso 2016 from five Italian centres, and Zhu 2016 from two Chinese institutions.

The median age reported for women with advanced EOC varied between 55 years (Zhu 2016) and 64 years (Stoeckle 2014) with the range between 28 and 88 years.

Details of interval debulking surgery reported in studies

RD thresholds ranged from NMRD up to > 2 cm across the included studies. The most common comparisons were of RD thresholds NMRD, ≤ 1 cm (although the majority included NMRD in this threshold, rather than 0.1 cm to 1 cm, which we defined as SVRD), and > 1 cm (LVRD). Optimal RD was commonly defined as less than 1 cm (RD < 1) or less than or equal to 1 cm (≤ 1 cm), consistent

with recent studies and guidelines (BGCS 2017; du Bois 2009), which state that surgery should not be considered optimal beyond 1 cm. Four studies did not provide an explicit definition of optimal RD (Lecointre 2020; Lecuru 2019; Lorusso 2016; Petrillo 2014). This was due to the nature of the information for the middle two cases (i.e. abstracts). For Petrillo 2014, although no definition of optimal RD was given, thresholds of NMRD, RD ≤ 1 cm and RD > 1 cm were provided. For Lecointre 2020, thresholds of NMRD, RD ≤ 0.25 cm, and RD 0.25 cm to 2.5cm were used. Davidson 2019 utilised two definitions of optimal RD (NMRD and RD ≤ 1 cm) in their study although only the latter was used in their multivariate Cox model.

Six studies compared SVRD versus LVRD (Cioffi 2018; Davidson 2019; Kaban 2017; Phillips 2018; Zhang 2018; Zhu 2016). Six of the studies did not make direct comparisons against NMRD and included NMRD in their SVRD category (Cioffi 2018; Davidson 2019; Kaban 2017; Shibutani 2020; Zhang 2018; Zhu 2016). Consequently, comparisons of SVRD (0.1 cm to 1 cm) and LVRD (> 1 cm) suffered from serious bias as a result of the inclusion of NMRD in the near-optimal category. Of these six studies, only three reported the number of participants with NMRD within the SVRD category: Cioffi 2018 (n = 37/57 participants with SVRD), Davidson 2019 (n = 165/228) and Zhang 2018 (n = 59/156). Only one study appropriately treated NMRD as a distinct category from SVRD (Phillips 2018).

Women in all the studies were treated by platinum-based neoadjuvant chemotherapy followed by IDS. One possible exception may be Lorusso 2016, but it was assumed that the NACT was platinum-based. All women were confirmed histologically to have invasive EOC.

The median number of NACT cycles varied from three (Zhang 2018) to six (Iwase 2015), with a range of 1 to 13. The large range is partially contributed by Stoeckle 2014, which was conducted in women receiving delayed IDS (after six or more cycles). Two studies did not provide descriptive statistics of NACT cycles (Lecuru 2019; Zhu 2016), but reported in their methodology that women received three or between three to four cycles. Information on the NACT regimen was provided in all but one study (Lorusso 2016). Carboplatin plus paclitaxel was most commonly reported and varied between 37.2 % (Zhu 2016), 96.6% (Stoeckle 2014), and 100%; although no details were reported for Kaban 2017, Lecuru 2019 (with reference to Ferron 2019), and Zhang 2018, they reported all women received carboplatin plus paclitaxel in their methods. Route of administration was reported in Bixel 2020 in which NACT was administrated intraperitoneally in 28% and intravenously in 72%, and Zhang 2018 in which NACT was administrated intraperitoneally in 45% and intravenously in 55%. Response to NACT according to RECIST criteria was reported in three studies in which complete/partial response was observed in 66.6% (Cioffi 2018), 66.1% (Zhu 2016), and in all participants in Lecointre 2020 (however, this was based on n = 380/501 with data on NACT response).

Information on the specialty of the surgeon performing the IDS was only reported in Stoeckle 2014 where all 118 surgeries were conducted by two surgeons with experience in ovarian cancer surgery and Shibutani 2020 where gynaecologic oncologists were involved in all surgeries. Duration of IDS was only reported in two studies and varied from a median of 194 minutes (Davidson 2019) to 419 minutes (Iwase 2015), with a range of 45 to 611 minutes. Length of hospital stay (LHS) was only reported in Stoeckle 2014 with a median of 10 days (range: 2 to 44) and Lecointre 2020 (median of 10 days (range: 6 to 13) in the group with ≤ 4 NACT cycles and median of 11 days (range: 7 to 14) in the group with > 4 NACT cycles). Postoperative morbidity/complications and mortality (defined as death within 30 days of IDS) was only reported in two studies (Davidson 2019; Stoeckle 2014). Postoperative mortality varied from 0% to 1.7%, whilst postoperative morbidity/complications varied from 18% to 22% in these studies. Complications after discharge and within 30 days of surgery were reported only in Davidson 2019. Approximately 11% experienced post-discharge complications of whom 6.4% were re-admitted. Operative blood loss was reported in Iwase 2015, with a median blood loss of 1291 mL (range: 220 mL to 5640 mL) and Lecointre 2020, where 57% of patients required blood transfusion (based on n = 77/501 with available data). Lecointre 2020 reported intraoperative complications in 15% of patients (based on n = 387/501 with available data). Lecointre 2020 also reported postoperative complications in 22% of participants (based on n = 421/501 patients with available data) but this was across an undefined time frame.

Information on postoperative chemotherapy following IDS was reported in 11 studies, albeit with varying levels of detail (Bixel 2020; Cioffi 2018; Iwase 2015; Kaban 2017; Lecuru 2019; Liu 2020; Petrillo 2014; Phillips 2018; Shibutani 2020; Stoeckle 2014; Zhang 2018). Clear reporting of platinum-based adjuvant chemotherapy was observed in five studies (Bixel 2020; Iwase 2015; Lecuru 2019; Petrillo 2014 (with reference to Ferron 2019); Zhang 2018), whilst it was implied (Kaban 2017; Liu 2020; Phillips 2018; Shibutani 2020; Stoeckle 2014) or unstated (Cioffi 2018) in the remaining six studies. Six of the studies did not provide descriptive statistics for adjuvant chemotherapy cycles or regimen and only reported in their methods that participants received chemotherapy following IDS (Cioffi 2018; Kaban 2017; Lecuru 2019 (with reference to Ferron 2019); Petrillo 2014; Shibutani 2020; Stoeckle 2014). However, with the exception of Cioffi 2018, they did report in their methods that their participants received two (Petrillo 2014; Liu 2020; Stoeckle 2014), two to three (Lecuru 2019 (with reference to Ferron 2019)), or two to six (Kaban 2017) cycles of adjuvant chemotherapy. Shibutani 2020 did not report the number of adjuvant cycles but did report the total (NACT + adjuvant chemotherapy) cycles. Six studies reported descriptive statistics (Bixel 2020; Iwase 2015; Liu 2020; Phillips 2018; Shibutani 2020; Zhang 2018). The median number of cycles ranged from three (Iwase 2015; Phillips 2018) to five (Zhang 2018), and ranged from one to eight in these three studies.

Optimal RD was most commonly defined as RD < 1 cm (Cioffi 2018; Iwase 2015; Phillips 2018; Shibutani 2020; Stoeckle 2014; Zhang 2018) or RD ≤ 1 cm (Bixel 2020; Davidson 2019; Kaban 2017; Liu 2020; Zhu 2016). Four studies did not provide a definition of optimal RD in their methodology but included RD thresholds in their multivariable Cox models (Lecointre 2020; Lecuru 2019; Lorusso 2016; Petrillo 2014). Davidson 2019 utilised two definitions of optimal RD (NMRD and SVRD) in their study, although only the latter was used in their multivariate Cox model. NMRD was reported in 12 studies (Bixel 2020; Cioffi 2018; Davidson 2019; Iwase 2015; Lecointre 2020; Lecuru 2019; Liu 2020; Lorusso 2016; Petrillo 2014; Phillips 2018; Stoeckle 2014; Zhang 2018), however descriptive statistics for the rate of NMRD were only reported in 10 studies (Bixel 2020; Cioffi 2018; Davidson 2019; Iwase 2015; Lecointre 2020; Liu 2020; Petrillo 2014; Phillips 2018; Stoeckle 2014; Zhang 2018). Lecointre 2020 reported missing data for RD in n = 30/501 women and did not report any imputation method. Rate of

NMRD varied from the lowest of 29.5% (Zhang 2018) to the highest of 79% (Iwase 2015). Across the 10 studies that reported descriptive statistics, NMRD was achieved in 1451 out of 2237 women (64.9%).

Across the six studies that provided descriptive statistics for RD < 1 cm (Cioffi 2018; Iwase 2015; Phillips 2018; Shibutani 2020; Stoeckle 2014; Zhang 2018), RD < 1 cm was achieved in 897 out of 1096 women (81.8%). Rates per study varied from 71% (Cioffi 2018) to 94% (Stoeckle 2014).

Across the four studies that provided descriptive statistics for RD ≤ 1 cm (Davidson 2019; Kaban 2017; Petrillo 2014; Zhu 2016), RD ≤ 1 cm was achieved in 1151 out of 1466 women (78.5%). Rates per study varied from 72% (Zhu 2016) to 84% (Davidson 2019; Petrillo 2014).

Nine studies reported the survival outcome in models comparing RD threshold(s) against NMRD (Bixel 2020; Iwase 2015; Lecointre 2020; Lecuru 2019; Liu 2020; Lorusso 2016; Petrillo 2014; Phillips 2018; Stoeckle 2014).

Outcomes

The median duration of follow-up was reported in nine studies (Bixel 2020; Iwase 2015; Kaban 2017; Lecuru 2019; Petrillo 2014; Shibutani 2020; Stoeckle 2014; Zhang 2018; Zhu 2016), and varied from a median of 29.5 months (Bixel 2020) to 47 months (Petrillo 2014), with a range between 1 and 181 months. The duration of follow-up was not reported in four studies (Cioffi 2018; Davidson 2019; Lecointre 2020; Liu 2020; Lorusso 2016; Phillips 2018).

Only one study did not report overall survival (Davidson 2019). Three studies did not provide adjusted HRs and 95% confidence intervals from their multivariate survival models predicting overall survival (Bixel 2020; Lecuru 2019; Petrillo 2014). One study only brought RD forward into the "multivariate" model for overall survival after univariate analysis, however the criteria for selection was not mentioned in the methods (Liu 2020). Eight studies reported progression-free survival and used appropriate statistical techniques (hazard ratios to correctly allow for censoring) (Cioffi 2018; Lecointre 2020; Lecuru 2019; Liu 2020; Petrillo 2014; Zhang 2018; Zhu 2016). One study reported using multivariate logistic regression to predict progression-free survival in their methods but reported hazard ratios in their results, so it may be inferred that multivariate Cox regression had actually been used (Bixel 2020). Disease-specific overall survival (DSS) was reported in Davidson 2019. Disease-free survival (DFS) was reported in Liu 2020. Prognostic factors were adjusted for in the analysis of survival outcomes in each study using Cox regression. Between them, the 15 studies included 29 different prognostic factors in the analysis. The precise prognostic factors used in Lorusso 2016 could not be determined beyond the complete cytoreduction, ECOG performance status and number of NACT cycles. The number of prognostic factors included in the analysis ranged from one in Petrillo 2014 to nine in Cioffi 2018. The prognostic factors most frequently included in the analyses are (in order of frequency): residual disease (15 studies), number of NACT cycles (eight studies), age (seven studies), FIGO stage (seven studies), performance status (six studies), ascites (four studies), response to NACT (four studies), NACT regimen (three studies), CA-125 (two studies) and lymphadenectomy (two studies). A list of the different prognostic factors is shown in Appendix 7.

One study, which included 501 women, had missing RD data for 30 (6%) (Lecointre 2020). Furthermore, other variables in the multivariate Cox model for overall survival had larger rates of missing data such as the Charlson Index (missing data for n = 203, 41%) and response to NACT (missing for n = 121, 24%). It is likely that the multivariate Cox model was based on a complete case analysis and therefore the estimates reported are based on ≤ 298 women, but the exact number cannot be known. For the multivariate model for progression-free survival, the estimates are based on ≤ 380 women as response to NACT was included as a covariate.

For the distribution of these factors at baseline for each study and by RD threshold see the table Characteristics of included studies.

Excluded studies

We excluded 133 references reporting on 115 studies after obtaining the full text, for the following primary reasons.

- We excluded 42 references reporting on 40 studies because they did not include at least 100 women with advanced epithelial ovarian cancer (Alphs 2006; Andersen Soegaard 2005; Benedetti-Panici 1996; Bristow 1999; Cai 2007; Ceresoli 2018; Colozza 1997; Del Campo 1994; Gao 2001; Gershenson 1989; Gershenson 1995; Grem 1991; Hainsworth 1990; Hakes 1992; Hamid 2002; Hardy 1991; Hoskins 1996; Kaern 2005; Kirmani 1994; Kristensen 1995; Loizzi 2016; Lorusso 1998; Malik 1998; Marchetti 1993; Ngan 1989; Palmer 1992; Risum 2012; Redman 1986; Rutten 2014; Shapiro 1998; Son 2017; Strauss 1996; Sutton 1989; Tay 1996; Taylor 1994; Vallejos 1997; Willemse 1992; Wils 1990; Zang 1999; Zhang 2015).
- Twenty-two studies either did not report multivariate analyses or did not include or adequately report residual disease as a variable to enable an analysis (Alberts 1996; Altman 2012; Bertelsen 1990; Bian 2016; Brinkhuis 1996a; Clamp 2018; Greggi 2016; Heitz 2016; Kessous 2017; Keyver-Paik 2016; Lee 2018; McGuire 1996; Piver 1991; Raspagliesi 2018; Rodriguez 2013; Sessa 1991; Sioulas 2017; Stewart 2016; Suidan 2015; Vidal 2016; Wallace 2017; Wimberger 2007).
- Fourteen studies did not report survival by residual disease (Alberts 1993; Bertelsen 1993; Brinkhuis 1996b; Conte 1991; Conte 1996; Creasman 1990; Gershenson 1992; Hoskins 1992; Hoskins 1997; Itamochi 2002; Solmaz 2015; Uyar 2005; Wadler 1996; Warwick 1995).
- Non-platinum based chemotherapy was given to all women in one study (Van Driel 2017), a proportion of women in four studies (Barda 2004; Bonnefoi 1999; de Oliviera 1990; Tingulstad 2003), and chemotherapy data were absent in the Bailey 2006 study. Women received preoperative chemotherapy in two studies (Shinozuka 1999; Sun 2000).
- Four studies included women who received neoadjuvant chemotherapy and interval debulking surgery but did not report an appropriate comparison by extent of disease (Dao 2016; Todo 2003; Van Der Burg 1996; van Vliet 2015).
- Seven studies included women with early-stage disease and it was not possible to distinguish between early- and advanced-stage participants (Crawford 2005; di Re 1996; Geisler 2004; Skarlos 1996; Smits 2015; Takano 2006; Takano 2007). The Le 1997 study did not report the survival data from the stage IIIC and IV subgroup and the authors no longer had access to these data.

- Two studies reported a HR for overall survival but did not include the corresponding 95% confidence interval, standard error (SE) (lnHR) or exact P value (Baker 1994; Omura 1989).
- The study Rose 2004 reported on outcomes after secondary debulking surgery. However, the trial statistician (Dr Mark Brady) of the included study Winter 2007 alerted us to the results of GOG 152, which reported by residual disease after primary cytoreductive surgery.
- Salani 2007 was excluded because it was a case-control study.
- The Yamamoto 2007 study included 67 selected women with rare histological subtypes and the Gasimli 2016 study included a selective group of women with cytoreduction of tumour to macroscopic optimal disease (0 cm).
- The Anuradha 2016 study focused only on the time interval between surgery and chemotherapy and the Michaan 2018 study focused on chemotherapy response score as an outcome, which is a histopathological scoring system based on morphological features of cancer tissue removed at IDS, but the same as optimal cytoreduction.
- Six references reporting on three RCTs comparing upfront versus delayed surgery did not report outcomes for extent of residual disease by type of initial primary surgery (Chekman 2015; Fagotti 2020; Onda 2020).
- Sixteen references reporting on three studies compared the threshold of residual disease based on type of intervention delivered (Kehoe 2015; Vergote 2010; Vergote 2018).
- Four studies were excluded because there was inadequate reporting and/or the full text was not available (Cummins 2019; Elgamal 2019; Stewart 2015; Trhlík 2013).
- One study did not distinguish between upfront and interval debulking primary surgery (Ruscito 2016).

For further details of all the excluded studies see the Characteristics of excluded studies table.

Risk of bias and quality appraisal in included studies

We assessed the risk of bias at outcome level for overall survival and progression-free survival for each study using the QUIPS tool (Riley 2019). Most studies reported overall survival (only two of all PDS studies (Peiretti 2010; Shim 2016), and just one study of all IDS studies (Davidson 2019) did not report overall survival). The detailed assessments are depicted in the 'Risk of bias (QUIPS)' section in the Characteristics of included studies.

We judged most studies included in the review as being at an overall 'moderate' risk of bias as they satisfied some but not all of the domains using the QUIPS tool. (See Table 3; Table 4; Table 5; Table 6 for risk of bias assessment using the QUIPS tool for overall survival and progression-free survival in the PDS and IDS studies).

Study participation

Most studies provided adequate details of study participation, which included details of eligible women, descriptions of the population and of the baseline study sample and recruitment, period and place of recruitment, and a description of inclusion and exclusion criteria. We assessed four studies as 'unclear' for this domain (two PDS studies (Hofstetter 2013; Van Geene 1996) and two IDS studies (Iwase 2015; Kaban 2017)), mostly due to a lack of detailed reporting of inclusion criteria. We assessed three studies (one PDS study (Shim 2016) and two IDS studies (Lecuru 2019; Lorusso 2016)) as being at a high risk of bias because they were in abstract form only, providing insufficient information on study participation.

Applicability: Are there concerns that the included women do not match the review question?

All studies matched the review question and there were no applicability concerns. Many studies reported one particular stage of advanced disease, but we were not concerned about this as we performed subgroup analyses by stage.

Ten PDS studies appeared to include a strictly representative sample of women with advanced epithelial ovarian cancer, by including stages III and IV combined (Chan 2003; Chang 2012a; Chi 2001; Hofstetter 2013; McGuire 1995; Peiretti 2010; Peiretti 2012; Shim 2016; Tewari 2016; Van Geene 1996). The Polterauer 2012 study included a small proportion of women with stage II disease (6.6%) and Feng 2016 included 9.3% early stage (I to II) disease, however both were otherwise representative of advanced disease. Klar 2016 included a small proportion of women with early-stage (IA to IIA) disease (3.6%) and an unknown proportion with stage IIB but the main scope was advanced disease so this was likely to be relatively few. The results of the meta-analyses were robust to the exclusion of this study in sensitivity analyses, so we did not deem the decision to include Klar 2016 in the review as being associated with any bias or issues with representativeness of women.

Of the 15 IDS included studies, four included a strictly representative sample of participants with advanced ovarian cancer (Iwase 2015; Petrillo 2014; Phillips 2018; Zhu 2016).

Study attrition

It was unclear if women with incomplete follow-up were excluded before arriving at the stated sample size in each study. There was insufficient information to permit judgement in all cases as many studies did not examine RD as a prognostic factor as their primary objective.

Prognostic factor measurement

Most studies reported a valid and reliable measurement of RD and we assessed these as being at a low risk of bias for the prognostic factor measurement domain. Even though multicentre studies are advantageous in terms of recruitment options and generalisability of participants as well as other positive features, we cautiously assessed the prognostic factor measurement to be unclear in 12 studies (eight PDS studies (Akahira 2001; Chan 2003; Cuylan 2018; Kahl 2017; Klar 2016; Peiretti 2012; Polterauer 2012; Van Geene 1996) and four IDS studies (Bixel 2020; Davidson 2019; Lecointre 2020; Zhu 2016)) that had this design, but these may well have been at a low risk too.

Applicability: Are there concerns that residual disease, the way that it is measured, or the way that it is interpreted, differ from the review question?

RD is measured by the surgeons estimate in all centres and there are no guidelines on how RD should be objectively measured. Therefore, there will be some natural variability in measurement across different centres, but we did not have any concerns about applicability.

Outcome measurement

The majority of the studies reported a valid and reliable measurement of outcome for both overall survival and progression-free survival and we assessed these as being at low risk of bias for the outcome measurement domain.

Overall survival

Two studies reported an inappropriate definition of overall survival (one PDS study (Aletti 2006) and a IDS study (Davidson 2019)) by reporting disease-specific survival, rather than all-cause overall survival. Consequently, we assessed these two studies to be at a high risk of bias. Outcome measurement of overall survival was unclear in one PDS study (Van Geene 1996) (Table 3; Table 4).

Progression-free survival

All studies that reported progression-free survival will have done so based on imaging and tumour markers. However, this is a somewhat subjective outcome and in unblinded studies could be deemed as being at a greater risk of bias. Therefore we judged the outcome measurement domain to be at unclear risk of bias as the measurement of this outcome may or may not have been reliable in certain RD thresholds (Table 5; Table 6).

Applicability: Are there concerns that outcome does not match the review question or that follow-up was not of sufficient duration?

We had no applicability concerns for outcome measurement for overall survival and progression-free survival.

Adjustment for other prognostic factors

For this domain, we assessed the appropriateness of confounders and whether important ones that a study should have at least been adjusted for such as age were included in their prognostic models. In cases where other prognostic factors in models were inadequate, we rated the studies as having a high risk of bias.

Overall survival

The studies at high risk of bias included seven PDS studies (Akahira 2001; Bristow 2011; Eisenkop 2003; Melamed 2017a; Melamed 2017b; Peiretti 2012; Shim 2016) and nine IDS studies (Bixel 2020; Davidson 2019; Lecointre 2020; Lecuru 2019; Liu 2020; Lorusso 2016; Petrillo 2014; Phillips 2018; Zhu 2016). These studies did not adequately adjust for a sufficient number of other prognostic factors in multivariate models or ones included were not pertinent. Adequate adjustment for other prognostic factors was unclear in 12 PDS studies (Chang 2012b; Feng 2016; Hofstetter 2013; Kahl 2017; Klar 2016; Langstraat 2011; McGuire 1995; Paik 2018; Van Geene 1996; Wimberger 2010; Winter 2007; Winter 2008) and in three IDS studies (Kaban 2017; Stoeckle 2014; Zhang 2018) (Table 3; Table 4).

Progression-free survival

The studies at high risk of bias included two PDS studies (Peiretti 2010; Shim 2016) and six IDS studies (Bixel 2020; Lecointre 2020; Lecuru 2019; Liu 2020; Petrillo 2014; Zhu 2016). These studies did not adequately adjust for a sufficient number of other prognostic factors in multivariate models or ones included were not pertinent. Adequate adjustment for other prognostic factors was unclear in eight PDS studies (Chang 2012b; Feng 2016; Klar 2016; McGuire 1995; Paik 2018; Wimberger 2010; Winter 2007; Winter 2008) and in one IDS study (Zhang 2018) (Table 5; Table 6).

Applicability: Did the prognostic factors adjusted for match the review question?

There was no reason to doubt the applicability of prognostic factors that were adjusted for in the multivariable models. Some studies may have used a wider range and more pertinent prognostic factors in their models for both overall survival and progression-free survival, but all studies satisfied our inclusion criteria for appropriateness of prognostic factors in their prognostic models and we had no applicability concerns.

Adjusted hazard ratios for survival using multivariable Cox models were used in each study. Any imbalances at baseline between RD thresholds should therefore be accounted for and all adjustments in the included studies met the inclusion criteria for the review.

We had applicability concerns in one IDS study (Petrillo 2014), as the multivariable analyses for overall survival and progression-free survival only adjusted for pathological response to NACT, so there may still be differences between RD thresholds that have not been controlled for.

Statistical analysis and reporting

We assessed the statistical analysis and reporting domain as being at high or unclear risk of bias in all included studies for both overall survival and progression-free survival outcomes. Either no conceptual framework was reported, where the variable selection criteria in the multivariate model was unclear or quite often the authors reported that significant variables from the univariate analysis were included in the multivariable model, but with no further details. It is also questionable whether this is adequate.

Mainly applicable to an IDS setting, it was not possible to distinguish NMRD within the SVRD thresholds in all but one study reporting a comparison of NMRD and SVRD. Only one study separated NMRD from SVRD (RD = 0.1 cm to 1 cm) and all other studies included NMRD in the SVRD group, resulting in serious risk of bias. Inclusion of NMRD in the SVRD category creates a high risk of bias when comparing suboptimal RD.

Findings

Meta-analyses of survival are based on hazard ratios (HRs) that were adjusted for prognostic variables (see Appendix 6 (PDS) and Appendix 7 (IDS) for details).

The percentage of the variability in effect estimates that was due to heterogeneity rather than sampling error (chance) may appear to represent substantial or considerable heterogeneity (as measured by the I^2 statistic) in some of the analyses below, but we had no major concerns as the direction of effect was consistent throughout.

We have reported the most pertinent comparisons involving SVRD (0.1 cm to 1 cm) versus NMRD, LVRD (> 1 cm) versus NMRD, and LVRD versus SVRD for overall survival and progression-free survival; these all provided moderate-certainty evidence. These are the most pertinent comparisons as they are included in clinical guidelines (NICE 2013), and are the focus of a key area of clinical uncertainty. Other RD comparisons were prespecified and have been provided.

The certainty of the evidence assessed using the GRADE approach (GRADE Working Group) was moderate for all comparisons involving overall survival and progression-free survival in a PDS setting and very low in an IDS setting. We restricted to comparisons of the three main reported RD thresholds (NMRD, SVRD and LVRD), since there is no firm guidance for grading the evidence in reviews of prognostic factor analyses (Riley 2019). Therefore, we did not grade beyond these key RD thresholds (Summary of findings 1; Summary of findings 2; Summary of findings 3; Summary of findings 4; Summary of findings 5; Summary of findings 6). The comparison involving any remaining macroscopic disease (RD > 0 cm) and NMRD in an IDS setting was also an important comparison, so this was included in the summary of findings and GRADE assessment (Summary of findings 7).

Residual disease after upfront primary debulking (cytoreductive) surgery (PDS)

Where possible the meta-analyses subgrouped studies by FIGO stage (stage III, IIIC, IV and all advanced stages, if studies included all advanced cases together). We conducted subgroup analyses to explore the underlying clinical heterogeneity between the studies. There was no evidence of subgroup differences in any of the subgroup analyses. The results of these subgroup analyses were robust to the findings of the overall pooled estimate for all comparisons, so the results of each subgroup are not discussed in this section (see Analysis 1.1 to Analysis 11.2).

The SVRD threshold included NMRD in some studies in comparison with LVRD, but only in a small number of studies. In PDS studies, RD < 1 cm means RD 0.1 cm to 1 cm (SVRD), unless otherwise stated. Due to only being an issue in a small number of studies, it was deemed to have a negligible impact on the results and did not affect the risk of bias profiles, the certainty of the evidence or distort the results. We performed sensitivity analyses when necessary.

We performed sensitivity analyses in comparisons that included meta-analysis of more than 10 studies. The use of a fixed-effect model aided the construction of the pseudo 95% confidence interval lines on the funnel plot (e.g. expected distribution of studies in the absence of heterogeneity and biases (such as publication bias, data irregularities)), as well as allowing us to see how robust the random-effects model results were in comparison. To further test the robustness of the findings, we additionally conducted a sensitivity analysis excluding studies with the largest weight in the meta-analyses comparing main RD thresholds, where appropriate.

We were cautious about any over-interpretation of funnel plots as they are typically underpowered. Given the nature of model selection procedures, we did not dismiss the possibility of publication bias. However, it is unclear as to the direction of any bias as, for instance, many highly significant studies only reporting unadjusted analyses found strong evidence that NMRD was associated with prolonged survival compared to other thresholds including SVRD (RD < 1 cm exclusive of 0 cm).

Overall survival (risk of death from all causes)

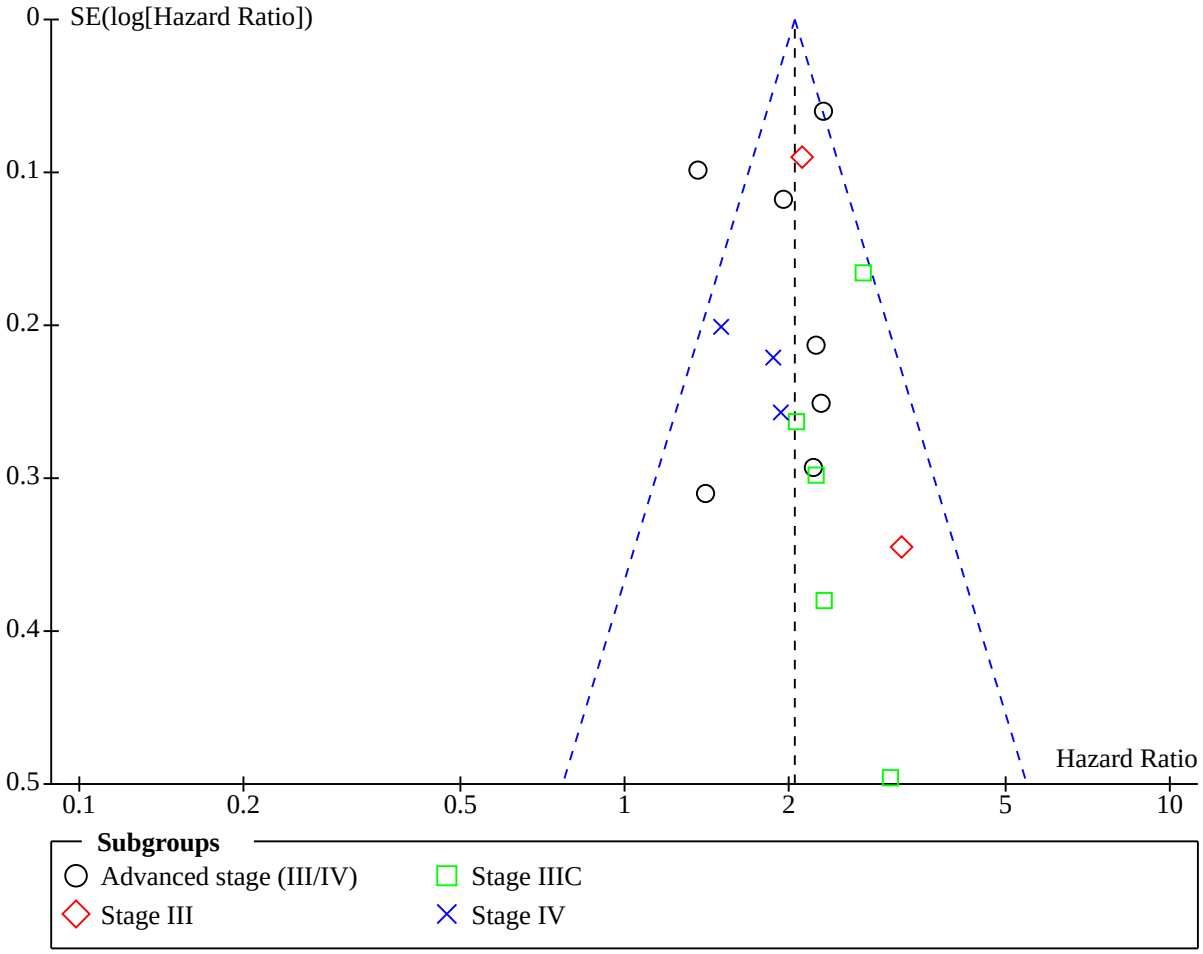
Small-volume residual disease (SVRD) versus no macroscopic residual disease (NMRD)

Meta-analysis of 17 studies, assessing 9404 participants, found that women with SVRD after PDS had more than twice the risk of death compared to women with NMRD (hazard ratio (HR) 2.03, 95% confidence interval (CI) 1.80 to 2.29). The percentage of the variability in effect estimates that was due to heterogeneity rather than chance may represent moderate heterogeneity (I^2 = 50%) (Analysis 1.1) (Summary of findings 1) (Aletti 2006; Ataseven 2016; Bristow 2011; Chang 2012a; Chang 2012b; Chi 2006; Cuylan 2018; Eisenkop 2003; Kahl 2017; Klar 2016; Langstraat 2011; Paik 2018; Tewari 2016; Tseng 2018; Wimberger 2010; Winter 2007; Winter 2008).

The results were robust to a sensitivity analysis that used a fixed-effect model and one that excluded the Klar 2016 study, which included a slight proportion of women with early or unknown stage (12.5%) disease. It also contributed the largest weight in the meta-analysis (see Analysis 1.2; Analysis 1.3).

There did not appear to be any evidence of small study biases, such as publication bias, or any irregularities with the data by visual inspection of a funnel plot (Figure 2).

Figure 2. Funnel plot of comparison: 1 SVRD (< 1 cm) versus NMRD, outcome: 1.2 Overall survival



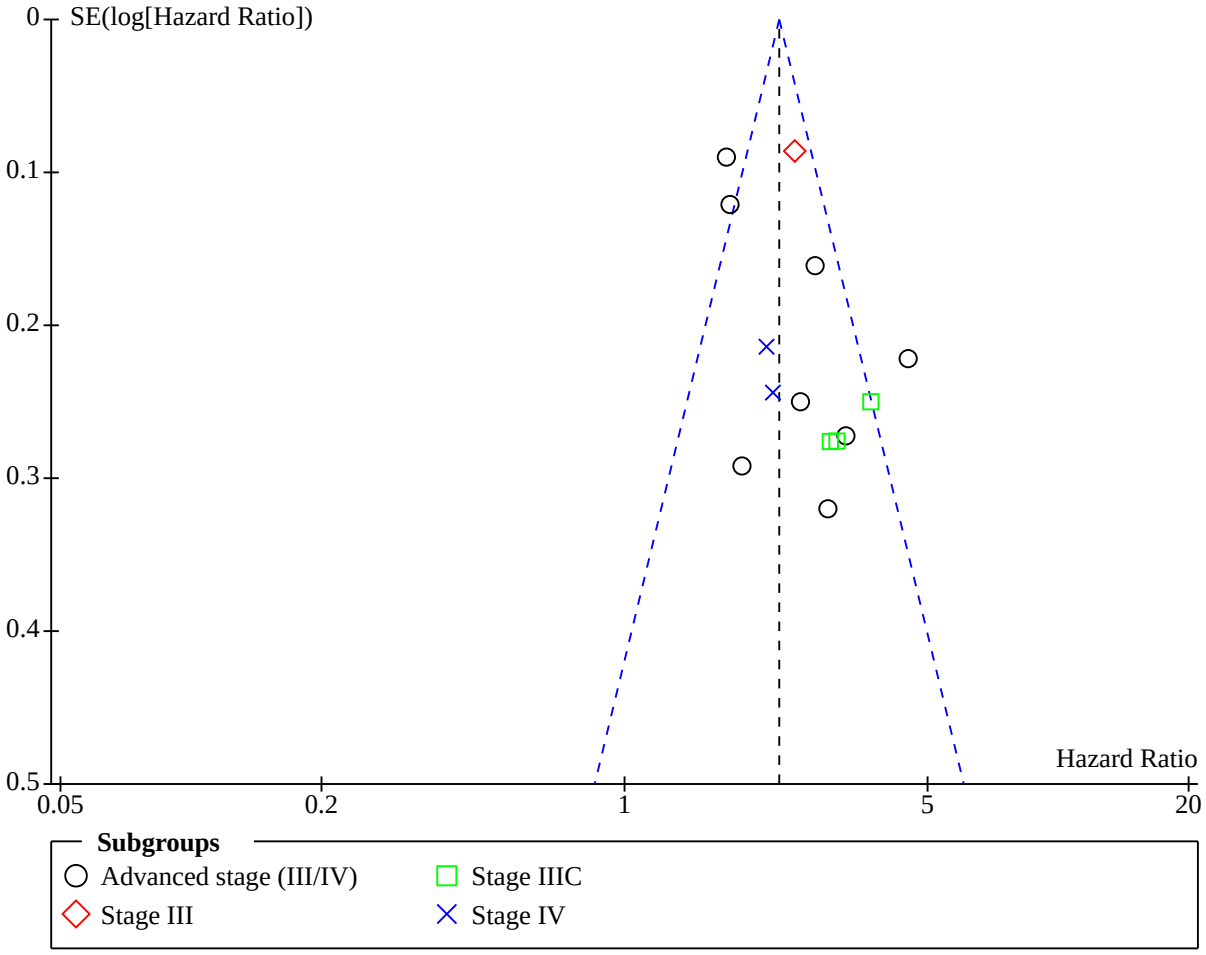
Large-volume residual disease (LVRD) (> 1 cm) versus NMRD

Meta-analysis of 14 studies, assessing 7988 participants, found that women with LVRD after PDS were associated with two and a half times the risk of death compared to women with NMRD (HR 2.50, 95% CI 2.13 to 2.94). The percentage of the variability in effect estimates that was due to heterogeneity rather than chance may represent substantial heterogeneity (I^2 = 63%) (Analysis 2.1) (Summary of findings 2) (Ataseven 2016; Chang 2012a; Chang 2012b; Chi 2006; Eisenkop 2003; Kahl 2017; Langstraat 2011; Melamed 2017a; Melamed 2017b; Paik 2018; Tewari 2016; Tseng 2018; Wimberger 2010; Winter 2007).

The results were robust to a sensitivity analysis that used a fixed-effect model and one that excluded the two studies with the largest weights in the meta-analysis (Melamed 2017b; Winter 2007) (see Analysis 2.2; Analysis 2.3).

There did not appear to be any evidence of small study biases, such as publication bias, or any irregularities with the data by visual inspection of a funnel plot (Figure 3).

Figure 3. Funnel plot of comparison: 4 LVRD (> 1 cm) versus NMRD, outcome: 2.2 Overall survival



LVRD versus SVRD

Meta-analysis of five studies, assessing 6000 participants, found that women with LVRD after PDS was associated with a greater risk of death compared to women with SVRD < 1 cm (HR 1.22, 95% CI 1.13 to 1.32; 6000 participants). The percentage of the variability in effect estimates that was due to heterogeneity rather than chance is not important (I^2 = 0%) (Analysis 3.1) (Summary of findings 3) (Chan 2003; Klar 2016; Melamed 2017a; Melamed 2017b; Winter 2008). The results were robust to a sensitivity analysis that excluded the Klar 2016 study with the largest weight in the meta-analysis (and a relatively small proportion of women with early or unknown stage (12.5%) disease) (see Analysis 3.2).

The results were also robust when only including the three studies that contributed majority of the weight in the meta-analysis and did not include NMRD in the SVRD category (HR 1.20, 95% CI 1.10 to 1.30; 5594 participants; I^2 = 0%) (Klar 2016; Melamed 2017a; Melamed 2017b)(see Analysis 3.3).

Similarly, meta-analysis of two studies that included NMRD in the SVRD category arrived at the same conclusion (HR 1.37, 95% CI 1.09 to 1.72; 435 participants; I^2 = 0%) (Chan 2003; Winter 2008).

Only Winter 2008 reported the proportion of women with NMRD (n = 29/107 of participants in the SVRD category)(see Analysis 3.4).

Residual disease (RD) > 0 cm versus NMRD

Meta-analysis of four studies, assessing 1220 participants, found that women who had RD greater than 0 cm after PDS were associated with a two-fold increase in the risk of death compared to women with NMRD (HR 1.96, 95% CI 1.44 to 2.67). The percentage of the variability in effect estimates that was due to heterogeneity rather than chance may represent moderate heterogeneity (I^2 = 49%) (Analysis 4.1) (Feng 2016; Hofstetter 2013; Luger 2020; Polterauer 2012). The authors of Peiretti 2012 additionally found that the risk of death for women with any remaining RD after PDS was higher than for those with NMRD (238 participants; P = 0.003), but the magnitude of effect was not reported.

RD 1 cm to 2 cm versus NMRD

The Aletti 2006 study, which included only women with stage IIIC disease, found that women who had RD between 1 cm and 2 cm after PDS were associated with a nearly four-fold increase in the risk of death compared to women with NMRD (HR 3.95, 95% CI 1.33 to 11.78; 68 participants) (Analysis 5.1).

RD > 2 cm versus NMRD

The Aletti 2006 study, which included only women with stage IIIC disease, found that women with LVRD > 2 cm after PDS were associated with more than eight times the risk of death compared to women with NMRD (HR 8.24, 95% CI 2.68 to 25.33; 87 participants) (Analysis 6.1).

RD 1 cm to 5 cm versus NMRD

The Winter 2008 study, which included only women with stage IV disease, found that women who had LVRD between 1 cm and 5 cm after PDS were associated with a greater risk of death compared to women with NMRD (HR 1.83, 95% CI 1.14 to 2.94; 193 participants) (Analysis 7.1).

RD > 5 cm versus NMRD

The Winter 2008 study, which included only women with stage IV disease, found that women who had LVRD > 5 cm after PDS were associated with more than two and a half times the risk of death compared to women with NMRD (HR 2.72, 95% CI 1.65 to 4.47; 118 participants) (Analysis 8.1).

RD 1 cm to 2 cm versus SVRD

The Chi 2001 study found that women who had LVRD between 1 cm and 2 cm after PDS were associated with a greater risk of death compared to women with SVRD (HR 1.70, 95% CI 1.10 to 2.60; 144 participants) (Analysis 9.1). The SVRD category in the Chi 2001 study included NMRD.

RD > 2 cm versus SVRD

The Chi 2001 study found that women with LVRD > 2 cm after PDS were associated with twice the risk of death compared to women with SVRD (HR 2.00, 95% CI 1.34 to 2.99; 208 participants) (Analysis 10.1). The SVRD category in the Chi 2001 study included NMRD.

RD > 2 cm versus RD < 2 cm

Meta-analysis of two studies, which included only women with stage IV disease and assessed 478 participants, found no statistically significant difference in the risk of death between women with LVRD > 2 cm after PDS and those with RD < 2 cm (HR 1.63, 95% CI 0.83 to 3.23). The percentage of the variability in effect estimates that was due to heterogeneity rather than chance alone may represent considerable heterogeneity (I^2 = 89%) (Akahira 2001; Winter 2008). The two studies were inconsistent: the Akahira 2001 study reported a large survival difference in favour of RD < 2 cm, whereas Winter 2008 found no difference in overall survival (Analysis 11.1). The < 2 cm category included NMRD in both studies, so this category had a mix of NMRD and SVRD < 1 cm as well as LVRD between 1 cm to 2 cm.

The authors of Van Geene 1996 reported the same comparison, but found evidence that more RD is associated with increased risk of death (HR 1.83, 95% CI not reported; 219 participants; P < 0.0001). Similarly, in two publications by McGuire 1995 in the same cohort of women, survival was significantly worse in women with LVRD > 2 cm compared to less remaining RD (1 cm to 2 cm as no women had SVRD) after PDS (n = 294 women with stage III disease, P < 0.01). The authors note that there was little notable difference in the risk of death between any volume of RD in comparisons of LVRD > 2 cm up to > 10 cm. In a further analysis including all advanced stages of disease (n = 458), women with stage III disease and LVRD > 2 cm had a lower risk of death than either those with stage III disease and LVRD > 2 cm, or those with stage IV disease (P = 0.012).

Progression-free survival (risk of disease progression)

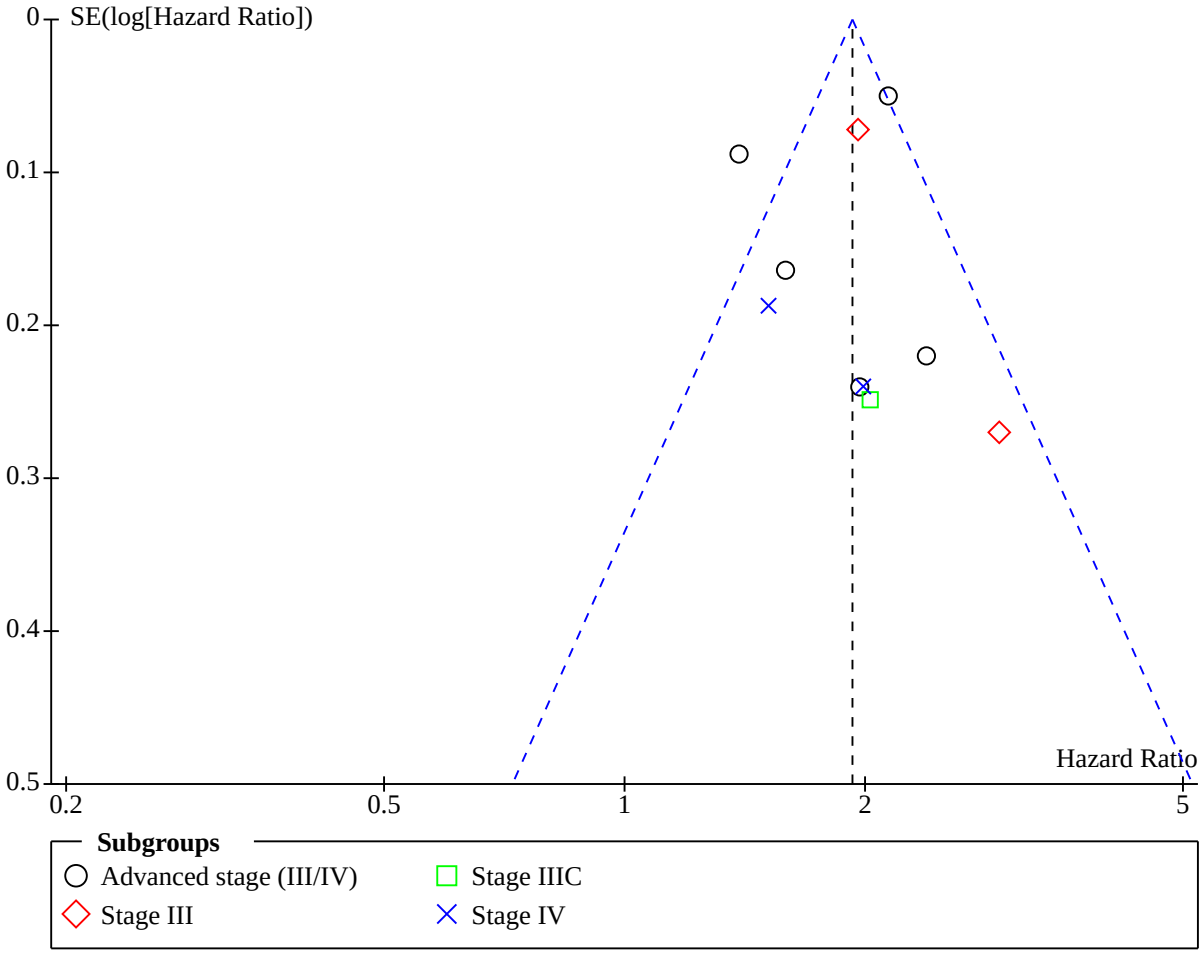
SVRD versus NMRD

Meta-analysis of 10 studies, assessing 6596 participants, found that women with SVRD after PDS were associated with nearly twice the risk of disease progression compared to women with NMRD (HR 1.88, 95% CI 1.63 to 2.16). The percentage of the variability in effect estimates that was due to heterogeneity rather than chance alone may represent substantial heterogeneity (I^2 = 63%) (Analysis 1.4) (Summary of findings 1) (Chang 2012a; Chang 2012b; Cuylan 2018; Klar 2016; Paik 2018; Shim 2016; Tseng 2018; Wimberger 2010; Winter 2007; Winter 2008).

The results were robust to a sensitivity analysis that used a fixed-effect model and one that excluded the Klar 2016 study with the largest weight in the meta-analysis (and a slight proportion of women with early or unknown stage (12.5%) disease) (see Analysis 1.5; Analysis 1.6).

There did not appear to be any evidence of small study biases, such as publication bias, or any irregularities with the data by visual inspection of a funnel plot (Figure 4).

Figure 4. Funnel plot of comparison: 1 SVRD (< 1 cm) versus NMRD, outcome: 1.5 Progression-free survival



LVRD (> 1 cm) versus NMRD

Meta-analysis of six studies, assessing 2629 participants, found that women with LVRD after PDS had more than twice the risk of disease progression compared to women with NMRD (HR 2.10, 95% CI 1.84 to 2.40). The percentage of the variability in effect estimates that was due to heterogeneity rather than chance may not be important (I^2 = 24%) (Analysis 2.4) (Summary of findings 2) (Chang 2012a; Chang 2012b; Paik 2018; Tseng 2018; Wimberger 2010; Winter 2007).

LVRD versus SVRD

Meta-analysis of two studies, assessing 3402 participants, found that women with LVRD > 1 cm after PDS had a greater risk of disease progression compared to women with SVRD (HR 1.30, 95% CI 1.08 to 1.56). The percentage of the variability in effect estimates that was due to heterogeneity rather than chance alone may represent moderate heterogeneity (I^2 = 53%) (Analysis 3.5) (Summary of findings 3) (Klar 2016; Winter 2008). Winter 2008 included NMRD in the SVRD category, but this only represented a small proportion in the analysis (n = 29/107 of participants in the SVRD category).

RD > 0 cm versus NMRD

Meta-analysis of three studies, assessing 1029 participants, found that women who had RD greater than 0 cm after PDS had more than one and a half times the risk of death compared to women with NMRD (HR 1.60, 95% CI 1.36 to 1.89). The percentage of the variability in effect estimates that was due to heterogeneity rather than chance is not important (I^2 = 0%) (Analysis 4.2) (Feng 2016; Luger 2020; Polterauer 2012). The authors of Peiretti 2010 additionally found that the risk of disease progression for women with any remaining RD was higher than those with NMRD (n = 259, P = 0.032), but the magnitude of effect was not reported.

LVRD 1 cm to 5 cm versus NMRD

The Winter 2008 study, which included only women with stage IV disease, found that women who had LVRD between 1 cm and 5 cm after PDS had more than twice the risk of disease progression compared to women with NMRD (HR 2.15, 95% CI 1.38 to 3.34; 193 participants) (Analysis 7.2).

LVRD > 5 cm versus NMRD

The Winter 2008 study, which included only women with stage IV disease, found that women who had LVRD between 1 cm and 5 cm after PDS had nearly three times the risk of disease progression

compared to women with NMRD (HR 2.96, 95% CI 1.86 to 4.71; 118 participants) (Analysis 8.2).

RD > 2 cm versus RD < 2 cm

The Winter 2008 study, which included only women with stage IV disease, found that women with LVRD > 2 cm after PDS had a slightly greater risk of disease progression compared to those with RD < 2 cm (HR 1.27, 95% CI 1.01 to 1.61; 253 participants) (Analysis 11.2).

Winter 2008 included NMRD in the < 2 cm category, but this only represented a small proportion in the analysis (n = 29/157 of participants in the RD < 2 cm category).

Residual disease after interval debulking surgery (IDS)

All meta-analyses included studies with participants with stage III and IV disease, other than in three studies where a specific breakdown was not reported (Kaban 2017; Lecuru 2019; Lorusso 2016). Therefore, we could not conduct subgroup analyses by stage to explore any underlying clinical heterogeneity between the studies as planned. However, we did perform subgroup analyses including cycle duration where possible (see Analysis 12.1 to Analysis 14.4). There was no evidence of any subgroup differences and all analyses were robust to the findings of the overall pooled estimates for all comparisons, with the exception of overall survival in the comparison of any remaining macroscopic disease versus NMRD (test for subgroup differences P = 0.01) (Analysis 15.1). However, the general direction of effect estimates was consistent and the findings were robust.

Davidson 2019 reported disease-specific survival (DSS) rather than overall survival, but this study did not appear to introduce statistical heterogeneity from visual inspection of the forest plot and the conclusions were robust to its exclusion in Analysis 14.2.

All comparisons involving SVRD included NMRD when compared to LVRD > 1 cm unless otherwise stated. The Phillips 2018 study was the only exception to this and reported an adequate comparison of LVRD > 1 cm versus SVRD using recognised RD threshold definitions, i.e. > 0 cm but < 1 cm residual disease as distinct from NMRD.

The comparison involving any remaining macroscopic disease (RD > 0 cm) and NMRD in an IDS setting was also an important comparison so we additionally gave this a certainty of the evidence judgement (Summary of findings 7).

Overall survival (risk of death from all causes)

SVRD versus NMRD

Meta-analysis of two groups of women from the same study undergoing different chemotherapy schedules found that women with SVRD after IDS had more than twice the risk of death compared to women with NMRD (HR 2.09, 95% CI 1.20 to 3.66; 310 participants) (Phillips 2018). The percentage of the variability in effect estimates that was due to heterogeneity rather than chance may represent moderate heterogeneity (I^2 = 56%) (Analysis 12.1). The magnitude of this effect was greater in this study in women who received > 4 cycles of neoadjuvant chemotherapy prior to their IDS (HR 2.78, 95% CI 1.66 to 4.65), but there was no significant difference or certainly a suggestion that there may be less of a difference between women with NMRD and those with SVRD when receiving ≤ 4 cycles of chemotherapy prior to IDS (HR 1.57, 95% CI 0.93 to 2.66) (Summary of findings 4).

The authors of Petrillo 2014 additionally found that the risk of death for women with SVRD after neoadjuvant chemotherapy (the majority received three or four cycles) before IDS was significantly higher than those with NMRD (n = 322, P = 0.001), but the magnitude of effect was not reported.

LVRD (> 1 cm) versus NMRD

Meta-analysis of two groups of women with different chemotherapy schedules, as outlined above, assessing 343 participants, found that women with LVRD > 1 cm after IDS had more than twice the risk of death compared to women with NMRD (HR 2.23, 95% CI 1.49 to 3.34). The percentage of the variability in effect estimates that was due to heterogeneity rather than chance may represent moderate heterogeneity (I^2 = 35%) (Analysis 13.1) (Phillips 2018). The magnitude of this effect was more pronounced in this study in women who received > 4 cycles of neoadjuvant chemotherapy prior to IDS (HR 2.67, 95% CI 1.76 to 4.06) (Summary of findings 5).

RD > 1 cm versus RD < 1 cm

Only Phillips 2018 compared SVRD versus LVRD > 1 cm in the strict sense that SVRD is mutually exclusive of NMRD. This was an important comparison and meta-analysis of the two groups in the study (three to six chemotherapy cycles) showed little difference in the risk of death between the SVRD and LVRD thresholds (HR 1.02, 95% CI 0.68 to 1.55; 343 participants). The percentage of the variability in effect estimates that was due to heterogeneity rather than chance was not important (I^2 = 0%) (Analysis 14.1).

The other five studies included NMRD in the SVRD category (referring to it as 'optimal') in their multivariate analyses. Nearly half of the women (261/550 (47%)) in the SVRD thresholds included NMRD in three studies (Cioffi 2018; Davidson 2019; Zhang 2018). This was not reported in the other two studies (Kaban 2017; Zhu 2016).

A sensitivity analysis that meta-analysed all six studies, assessing 1572 participants, found that women with LVRD > 1 cm after IDS had a statistically significant greater risk of death compared to women with SVRD or NMRD (HR 1.60, 95% CI 1.21 to 2.11). The percentage of the variability in effect estimates that was due to heterogeneity rather than chance may represent substantial heterogeneity (I^2 = 58%) (Analysis 14.2) (Summary of findings 6) (Cioffi 2018; Davidson 2019; Kaban 2017; Phillips 2018; Zhang 2018; Zhu 2016).

Sensitivity analysis, excluding Phillips 2018, led to an increase in effect estimates in a meta-analysis involving the five remaining studies (HR 1.84, 95% CI 1.34 to 2.52; 1429 participants; I^2 = 61%) (Analysis 14.3).

RD > 0 cm versus NMRD

Meta-analysis of four studies, assessing 906 women, found that any macroscopic RD after IDS was associated with more than twice the risk of death compared with NMRD (HR 2.11, 95% CI 1.35 to 3.29). The percentage of the variability in effect estimates that was due to heterogeneity rather than chance may represent considerable heterogeneity (I^2 = 81%) (Analysis 15.1) (Iwase 2015; Lecointre 2020; Lorusso 2016; Stoeckle 2014). For subgroup analysis by duration of NACT, we found evidence of a subgroup difference (P = 0.01, median of six cycles in two studies: N = 242, median four cycles in one study: N = 193, all range of cycles in one study: N = 471). However, the

direction of effect was consistent in all studies, showing a survival benefit in the NMRD group (Analysis 15.1).

The authors of Lecuru 2019 additionally found that the risk of death for women with any remaining RD (> 0 cm) after IDS was significantly higher than those with NMRD (n = 163, P < 0.01), but the magnitude of effect was not reported (Summary of findings 7).

Progression-free survival (risk of disease progression)

SVRD (< 1 cm) versus NMRD

Meta-analysis of two studies, assessing 248 women, found no difference in disease progression in women with SVRD after IDS and those with NMRD (HR 3.03, 95% CI 0.81 to 11.38). The percentage of the variability in effect estimates that was due to heterogeneity rather than chance alone may represent considerable heterogeneity (I^2 = 94%) (Analysis 12.2) (Bixel 2020; Liu 2020).

The authors of Petrillo 2014 found that the risk of disease progression for women with SVRD after IDS was higher than those with NMRD (n = 322, P = 0.001), but the magnitude of effect was not reported.

LVRD (> 1 cm) versus SVRD

Meta-analysis of four studies found that achieving LVRD > 1 cm after IDS was associated with a greater risk of disease progression compared to women in whom SVRD was achieved after surgery (HR 1.76, 95% CI 1.23 to 2.52; 1145 participants). The percentage of the variability in effect estimates that was due to heterogeneity rather than chance alone may represent substantial heterogeneity (I^2 = 60%) (Analysis 14.4) (Cioffi 2018; Shibutani 2020; Zhang 2018; Zhu 2016). These four studies included NMRD in the SVRD category (referring to it as 'optimal') in their multivariate analyses.

RD > 0 cm versus NMRD

The Lecointre 2020 study, assessing 471 women, found that RD > 0 cm after IDS was associated with an increased risk of disease progression compared those in whom NMRD was achieved (HR 1.36, 95% CI 1.05 to 1.76) (Analysis 15.2).

The authors of Lecuru 2019 found that the risk of disease progression for women with RD > 0 cm after IDS was higher than those with NMRD (n = 163, P < 0.01), but the magnitude of effect was not reported (Summary of findings 7).

DISCUSSION

Summary of main results

We found 46 studies reporting multivariate prognostic analyses that included residual disease (RD) as a prognostic factor, which met our inclusion criteria. These studies assessed survival after upfront primary debulking surgery (PDS) followed by adjuvant chemotherapy and neoadjuvant chemotherapy with interval debulking surgery (IDS) in advanced epithelial ovarian cancer. The review included 22,376 women who underwent PDS and 3697 women who underwent IDS, all with varying levels of RD. The main results of our review are summarised in the summary of findings tables (Summary of findings 1; Summary of findings 2; Summary of findings 3; Summary of findings 4; Summary of findings 5; Summary of findings 6; Summary of findings 7).

In PDS studies, meta- and single-study analyses demonstrate the prognostic importance of achieving no macroscopic residual disease (NMRD) after PDS for both overall and progression-free survival. Most studies showed an association with an increased risk of death in all groups with visible disease after surgery when compared to NMRD. The most pertinent comparison found that women who were debulked to leave small-volume residual disease (SVRD) after PDS had more than twice the risk of death compared to women with NMRD (meta-analysis of 17 studies: hazard ratio (HR) 2.03, 95% confidence interval (CI) 1.80 to 2.29; I^2 = 50%; 9404 participants; moderate-certainty evidence). Progression-free survival was not reported in all of the studies, but was sufficiently documented to allow conclusions to be drawn. The main comparison found that women who were debulked to SVRD after PDS had nearly twice the risk of disease progression compared to women with NMRD (meta-analysis of 10 studies: HR 1.88, 95% CI 1.63 to 2.16; I^2 = 63%; 6596 participants; moderate-certainty evidence). The fact that all of the studies included at least 100 women and used multivariate adjustment for important prognostic factors increased the level of certainty in the estimates.

When we compared large-volume residual disease (LVRD) (> 1 cm) versus SVRD cytoreduction the estimates were attenuated compared to the macroscopic RD comparisons. All analyses showed a survival benefit in women who had been debulked to leave SVRD (HR 1.22, 95% CI 1.13 to 1.32, I^2 = 0%, 6000 participants; moderate-certainty evidence). The results were robust to analyses of progression-free survival.

For neoadjuvant chemotherapy with IDS, the main comparisons involved any visible RD versus NMRD and LVRD (> 1 cm) versus SVRD. Unfortunately, it was not possible to distinguish those with NMRD after surgery within the SVRD thresholds in all but one study. A study reporting two groups of women on different chemotherapy schedules found that women who were debulked to leave SVRD and LVRD (> 1 cm) after IDS had more than twice the risk of death compared to women who had NMRD (HR 2.09, 95% CI 1.20 to 3.66; 310 participants; I^2 = 56% and HR 2.23, 95% CI 1.49 to 3.34; 343 participants; I^2 = 35%; very low-certainty evidence, for SVRD versus NMRD and LVRD versus NMRD, respectively). Women who had any amount of macroscopic RD (> 0 cm) after IDS had more than twice the risk of death compared to women with NMRD (HR 2.11, 95% CI 1.35 to 3.29, I^2 = 81%; 906 participants; very low-certainty evidence). Another study also found prolonged survival when RD was cytoreduced to NMRD (P < 0.01).

Unfortunately, in IDS studies the SVRD threshold included those with NMRD in all but one study (nearly half of women in the SVRD threshold had NMRD in three studies where it was reported). Therefore the reported comparison of NMRD or SVRD versus LVRD > 1 cm was of much lesser importance in IDS studies. Meta-analysis found that women who were debulked leaving LVRD > 1 cm had a greater risk of death and disease progression compared to women who were debulked to leave SVRD or NMRD (HR 1.60, 95% CI 1.21 to 2.11; 1572 participants; I^2 = 58% for overall survival and HR 1.76, 95% CI 1.23 to 2.52; 1145 participants; I^2 = 60% for progression-free survival; moderate-certainty evidence). The SVRD category included NMRD in all but one study, which suggests that only two categories of RD after IDS are being recognised at present, where NMRD remains of paramount prognostic importance.

Overall completeness and applicability of evidence

The evidence from this review indicates that cytoreduction to NMRD after primary surgical cytoreduction is associated with prolonged survival in advanced epithelial ovarian cancer in both PDS and IDS settings. There is more strength in the evidence from studies reporting PDS, but the results suggest that the same conclusions apply in terms of the prognostic importance of NMRD in an IDS setting. More studies, including a larger number of women, will be needed to give more certainty in the effect estimates in comparisons of other RD thresholds, but there has been an emergence of studies using IDS in the last decade, so we expect this to be the case when the review is updated in the future. Interestingly, the comparison of SVRD versus LVRD (> 1 cm) is heterogeneously reported in the PDS and IDS analyses, as in the latter (IDS studies) the SVRD threshold included NMRD in all but one of the studies in the meta-analysis. Most studies included in the PDS analyses presented mutually exclusive RD thresholds, so there was less of a problem with NMRD being included in the SVRD category. The existing evidence does not currently support three categories of RD after IDS, as was recommended for PDS.

Although this review does not enable us to determine whether prolonged survival is a direct effect of the surgical intervention whereby women with NMRD do better, it appears that every effort should be made to attempt this, where possible, in both PDS and IDS settings. It may be particularly important in the latter due to issues with chemotherapy reaching allocation and further treatment options potentially being more limited thereafter. Where NMRD is considered not achievable for PDS, attempts should be made to obtain SVRD, defined as RD greater than 0 cm and less than or equal to 1 cm. There is limited evidence in this review to suggest that this may not be the case for IDS. Further data are needed, as understanding whether there is a benefit to IDS, if NMRD cannot be achieved, would be an important clinical question. However, as this is a prognostic review, we cannot answer this question from these data. Additionally, the data are of very low certainty - we are therefore very uncertain of this finding and drawing any conclusions would be unwise. Answering this question about the benefit of IDS, if NMRD cannot be achieved, would require an intervention study randomised controlled trial (RCT), rather than retrospective analysis of prognostic factors.

We found statistical heterogeneity between the studies in some analyses, but the direction of effect was consistent throughout so we had no concerns. We also did not have too many concerns about clinical heterogeneity as we applied restrictive inclusion criteria in terms of patient population, standardised measurement of RD as a prognostic factor and standard definitions of survival. Evaluation of other prognostic factors and biomarkers can often use different criteria for the interpretation of the results and different cut-off values may introduce levels of heterogeneity. Therefore, RD as a prognostic factor is unlikely to impact on the results or introduce any bias. That is, false-positive classifications seem much more unlikely than in other prognostic areas.

One of the strengths regarding the prognostic factor studies in this review was the ease of reporting in their statistical analyses. Authors mainly reported appropriate methods for their statistical analyses, with only a few studies not reporting the magnitude of effect estimates. We used hazard ratios (HRs) as the effect measure for time-to-event data in this review. We were able to provide pooled data for the majority of the included studies in the review. Of the studies that did not report appropriate statistics to extract for inclusion in the meta-analysis, we could not estimate the HR using other available data (Parmar 1998), as we restricted studies to those using multivariate analyses. We had limited success when contacting chief investigators to provide us with additional information or data from adjusted analyses.

In order to minimise bias, we only included studies of multivariate Cox regression models that used sensible adjustment factors associated with survival in women with advanced epithelial ovarian cancer (e.g. age, stage, grade, extent of disease at diagnosis). We excluded studies that only reported unadjusted results. To assess the adequacy of adjustment factors used in multivariate Cox models, we used the 'adjustment for other prognostic factors' and 'statistical analysis and reporting' domains of the quality in prognosis studies (QUIPS) tool (Riley 2019). Therefore, we prespecified in our protocol that we would only pool adjusted associations of the index prognostic factor. We felt that it was important to suggest a set of pertinent and established covariates a priori that were important to the disease under review (Riley 2019). This meant that we could better judge which models were adequate. We took these issues around the reporting in the studies into account when we assessed risk of bias and GRADE. The reported results in univariate analyses would have potentially been at a great risk of overestimating survival of RD as a prognostic factor. It is widely accepted that adjusting the predictive effect of a specific prognostic factor for the contribution of other prognostic factors strengthens the robustness of the evidence on the clinically relevant prognostic ability of that factor (Aldin 2020; Riley 2019).

Treatment-related morbidity very often degrades the quality of the time that women live, which is especially important after the completion of treatment for advanced cancer where women have poor prognosis and will want to enjoy a comfortable standard of living during their final months. It is unlikely that studies on prognosis will measure or report adverse events, so our focus was on survival as an outcome. This needs to be considered in the context of the findings from this review in that NMRD after PDS is associated with better survival - median survival for NMRD was 85.8 months (95% CI 77.5 to 94.1 months) in the Klar 2016 study, which included the largest analysis in the review. This study did include a small proportion of women with stage I and II cancer, but not to the extent of diluting the results too much. The next largest included study reporting median overall survival (71.9 months) also suggested that the potential benefits of prolonging survival may outweigh the disadvantages of any short-term morbidities associated with the surgical procedure (Winter 2007). Similarly, median survival in the NMRD group in IDS studies ranged from 50 months (Stoeckle 2014) to 51.8 (95% CI 45 to 58.5) months (Phillips 2018), the second largest analysis of IDS in this review. In terms of the overall survival rate in the NMRD group in IDS studies, Iwase 2015 reported a two-year and five-year overall survival rate of 88.8% and 43.4% respectively.

Certainty of the evidence

Our certainty of the evidence is presented in the summary of findings tables (Summary of findings 1; Summary of findings 2; Summary of findings 3; Summary of findings 4; Summary of findings 5; Summary of findings 6; Summary of findings 7).

The 46 studies that met our inclusion criteria had reasonable risk of bias profiles when assessed using QUIPS as a prognostic risk of

bias tool (Riley 2019). We included only sufficiently large studies that controlled for various co-prognostic factors using multivariate analysis in order to reduce the possibility of bias.

The studies reported adjusted hazard ratio estimates using Cox proportional hazards models. A hazard ratio is the best statistic to summarise the difference in risk between groups over the duration of a study when there is 'censoring', that is the time to death (or disease progression) is unknown for some women as they are still alive (or disease-free) at the end of the study. Most studies were at moderate risk of bias as they satisfied some but not all of the criteria used to assess risk of bias. There were no real applicability concerns in any of the domains. This was largely due to the stringent and restrictive eligibility criteria. We were also cautious when deciding whether studies were selectively reported or whether any additional source of bias may have been present and assessed these items as being unclear.

In a PDS setting, for overall survival, all studies in the meta-analyses used adjusted results from multivariable analyses including important and well-established prognostic factors in women with advanced ovarian cancer, and the analyses all indicated the independent prognostic ability of thresholds of RD to predict overall survival. For comparisons of the three main reported RD thresholds (NMRD, SVRD and LVRD), we judged the certainty of the evidence as 'moderate' for all these comparisons (Summary of findings 1; Summary of findings 2; Summary of findings 3). We downgraded by one level for risk of bias due to some risk of bias concerns. With no firm guidance for grading the evidence in reviews of prognostic factor analyses (Riley 2019), we did not grade beyond these key RD thresholds. Similarly, progression-free survival was reported using the same methodology but in fewer studies. There was still a sufficient number to show that RD thresholds have an independent prognostic ability to predict progression-free survival. We also judged this outcomes to provide moderate-certainty evidence and we downgraded by one level for some risk of bias concerns (Summary of findings 1; Summary of findings 2; Summary of findings 3). We made the same certainty of the evidence judgements in an IDS setting for overall survival and progression-free survival. Only one study reported a comparison involving NMRD as a unique group (Phillips 2018). Furthermore, this same study was the only one to report the comparison of SVRD (< 1 cm) versus LVRD (> 1 cm) in the strict sense that SVRD was mutually exclusive of NMRD. The other studies reporting this outcome included NMRD in their SVRD group. Therefore, we downgraded overall survival and progression-free survival outcomes by one level. We also downgraded for some risk of bias concerns and insufficient and sparse data in the meta-analyses. Therefore the certainty of the evidence for overall survival and progression-free survival in an IDS setting was very low (Summary of findings 4; Summary of findings 5; Summary of findings 6; Summary of findings 7). The comparison of SVRD versus LVRD (> 1 cm) included one more study than the corresponding analysis involving PDS, but there were significantly fewer women in the analysis (less than a third) and the lack of separation of NMRD from the SVRD threshold was misleading, so that was reason it was judged to provide very low-certainty evidence (Summary of findings 6). Only one study truly reported an adequate comparison of LVRD versus SVRD.

In some cases, more data would be needed to see the full impact of leaving behind more considerable disease, although the evidence suggests that if it cannot be minimised to NMRD or SVRD it may not make a significant difference in terms of prolonged survival. The results are consistent and appear to be reliable and precise in terms of the conclusions drawn. Some comparisons were sparse, with wide confidence intervals, but even the lower 95% confidence interval would have been highly significant as a point estimate in many cases.

Further research is very unlikely to change our confidence in the estimates of effect in the larger and most pertinent meta-analyses (exclusively reported in a PDS setting), but may change the estimates for some of the comparisons involving head-to-head LVRD thresholds and in analyses that included IDS. However, in the latter the evidence base is likely to be strengthened in future years as there has been an emergence of evidence in the last decade that is expanding, given four RCTs have now demonstrated similar survival outcomes of PDS versus IDS, as reviewed by Coleridge 2021. However, this evidence needs to assess whether SVRD is associated with a survival benefit over LVRD in an IDS setting as the evidence is currently very uncertain.

Strengths and weaknesses of the review

We performed a comprehensive search, including a thorough search of the grey literature, and two review authors working independently sifted and data extracted all studies. To prevent bias in this review, at least two review authors, along with willing arbiters, also independently performed all other relevant processes, such as risk of bias and GRADE assessment, and verification of all analyses. Although the methods for grading the evidence from prognosis studies are still under development, we felt that omitting it would be less transparent and potentially create bias in the review. Therefore we followed standard methodology for grading the certainty of the evidence and used specific exemplars from the Cochrane prognostic group for guidance, as well as examining other relevant prognostic factor reviews (Aldin 2020). We were not restrictive in our inclusion criteria with regards to types of studies, but limited to prognostic models that used multivariate analyses. This was to ensure that we minimised bias in getting accurate and reflective effect estimates for the prognostic performance of RD. We restricted to studies including at least 100 women in their analyses due to limiting the analyses to multivariate ones and the potential issue with adjustment for multiple prognostic factors in sparse data (Ogundimu 2016). There was more chance of drawing satisfactory conclusions in the review as the number of women in each study was adequate. We also conducted analyses using appropriate statistical methods for survival outcomes, namely hazard ratios, which correctly allow for censoring (see above).

In the analyses comparing SVRD versus LVRD (> 1 cm) for both PDS and IDS, we included studies that either treated NMRD as a distinct category of SVRD (Phillips 2018), or included NMRD within the SVRD category (Akahira 2001; Chan 2003; Chi 2001; Cioffi 2018; Davidson 2019; Kaban 2017; McGuire 1995; Winter 2008; Zhang 2018; Zhu 2016) during analyses. In keeping with the view that there is a dose-response relationship between RD thresholds and survival, the inclusion of these latter studies will have introduced an overestimation of the survival benefit of SVRD compared to LVRD (> 1 cm) and introduced serious bias. We attempted to determine the extent of this bias by identifying the number of participants with NMRD included in these latter studies, however this information was only provided in four studies – NMRD ranged from 27% (Winter 2008) to 72% (Davidson 2019). Results of analyses in PDS studies

were robust to the exclusion of studies that included NMRD within their SVRD category. Similar sensitivity analyses were not practical in the IDS case as only Phillips 2018 adequately reported the comparison involving SVRD that did not include NMRD.

A significant threat to the validity of the review is likely to be publication bias; that is, studies that did not find a positive association with the degree of surgical debulking achieved may not have been published. Although we conducted a test for funnel plot asymmetry and there did not appear to be any evidence of small study bias, such as publication bias, this type of test is not necessarily recommended for survival data due to issues of censoring (Debray 2018). Therefore, we cannot exclude potential publication bias and the presence of small study effects in our review (Riley 2019). Further investigation is beyond the scope of this review. Most included studies included in this review were retrospective and were probably not pre-registered. Studies are also not always labelled or indexed as prognosis studies, and search filters for studies on prognosis are still under development. Therefore the search had much wider scope than was necessary, but we felt it was better to be overly inclusive to reduce the chance of missing eligible studies for inclusion in the review.

Agreements and disagreements with other studies or reviews

In our review, we included studies that have assessed residual disease (RD) as a prognostic factor after primary surgery in women with advanced epithelial ovarian cancer. Overall, the findings from this review are in agreement with similar reviews and studies that have investigated the prognostic value of NMRD in both PDS and IDS settings. They also support the findings that, in general, small-volume RD is associated with better survival after surgery. The majority of these studies reported univariate analyses and that was one of the exclusion criteria in our review. These univariate analyses widely reported larger magnitudes of effect giving greater levels of statistical significance, but our analyses restricted to estimates that adjusted for sensible covariates that were likely to give less biased and more reliable estimates. Many of these studies are documented in the Characteristics of excluded studies table.

The association with improved survival outcomes associated with NMRD categorisation consolidates use of the term 'optimal cytoreduction' by the Gynaecological Cancer Inter-Group (GCIG) to mean 'NMRD', from its former definition of < 1 cm RD, which we categorised as SVRD. Although the results of our review show that cytoreduction to SVRD is still superior to LVRD (> 1 cm).

In a PDS setting, if the term macroscopic cytoreduction is to be used solely for the group where there is NMRD, the moderate-certainty evidence in this review that women who undergo PDS and achieve SVRD still do better than women who achieve LVRD should prompt the surgical community to retain this category as well as SVRD for RD < 1 cm (and consider the term 'near-optimal'), while reserving the term LVRD (and consider using 'suboptimal') to cases where the RD is > 1 cm (a three-category classification of NMRD, SVRD and LVRD, or alternatively consider the terms 'optimal', 'near-optimal' and 'suboptimal' RD). In contrast, we obtained very low-certainty evidence from a single IDS study that showed a survival benefit for NMRD compared to SVRD (Phillips 2018). All but one study included NMRD in their comparison of SVRD versus LVRD (> 1 cm) so strong inferences were not possible. Evidence from this one study that reported a valid comparison found little difference in survival outcomes in this comparison of RD thresholds (Phillips 2018). Further evidence from a meta-analysis including four studies showed that achieving NMRD was associated with superior survival outcomes to having any remaining RD (> 0 cm) (Iwase 2015; Lecointre 2020; Lorusso 2016; Stoeckle 2014). Therefore, given the available evidence, the strongest conclusion renderable is a two-category classification following IDS (NMRD versus any RD > 0 cm).

The debate regarding whether a three-category classification should hold in both PDS and IDS has also surfaced amongst the surgical community in recent publications. To our knowledge, two retrospective studies of women with advanced epithelial ovarian cancer provided evidence pertinent to this debate (Ghirardi 2020; Kobal 2018). One rationale behind these studies was to address whether women in whom PDS achieved SVRD would be conferred similar or better survival compared to those in whom NMRD was achieved following IDS. In the Kobal 2018 study, amongst women achieving NMRD, the IDS group had poorer overall survival (36.3 versus 54.7 months; P = 0.012) but similar progression-free survival (19.9 versus 20.7 months; P = 0.251) compared to the PDS group. On the other hand, achieving NMRD following IDS was associated with similar overall survival (36.3 versus 34.7 months; P = 0.073), but better progression-free survival (19.9 versus 11.2 months; P = 0.005) compared to achieving SVRD following PDS. In contrast, Ghirardi 2020 found that achieving NMRD following IDS was associated with poorer overall survival compared to achieving SVRD following PDS (41.4 versus 52.4 months; P = 0.022). Given the unadjusted estimates and retrospective nature of these studies, and that these compare prognostic factors and not treatment effects, conclusions about the relative merits of different treatments cannot be made. However, they do reflect an ongoing point of discussion, and contribute towards a burgeoning empirical basis for either a two-threshold 'all-or-nothing' classification system following IDS (NMRD versus RD > 0 cm) or the retention of the three-threshold classification. The results of our review appear to lend support for the two-threshold classification following IDS based on the conduct of the included studies, although this is more on the grounds that there is a lack of evidence of significant differences in survival between SVRD and LVRD (> 1 cm) thresholds due to lack of reporting of this comparison.

A Cochrane Review by Coleridge 2021 compared intervention RCTs directly comparing PDS versus IDS (Chekman 2015; Fagotti 2020; Kehoe 2015; Onda 2020; Vergote 2010). The included studies did not meet our inclusion criteria, as they did not report results across RD thresholds. Within this review, Kehoe 2015 and Vergote 2010 randomised 1270 participants (of which 1220 were assessed), compared PDS versus IDS and provided a breakdown of extent of disease by type of surgery (but did not give breakdown of differences within RD thresholds for each type of surgery, so did not meet our inclusion criteria). Both trials recruited participants with stage IIIC and IV epithelial ovarian cancer. Both trials reported RD thresholds that included NMRD (optimal), SVRD (RD < 1 cm) and LVRD (RD > 1 cm). The two trials found no significant difference in overall survival for the comparison of extent of RD threshold (NMRD, SVRD and LVRD) by primary surgery (upfront versus interval). The trial Vergote 2010 reported no significant difference between PDS and IDS for SVRD or NMRD (RD < 1cm including 0 cm) (HR 1.17, 95% CI 0.82 to 1.67). There were also no significant differences observed for SVRD (< 1 cm) (HR 1.22, 95% CI 0.84 to 1.77) and LVRD thresholds (HR 0.91, 95% CI 0.89 to 1.30) by type of surgery. Similarly, the authors of Kehoe 2015 reported a P value of 0.98 for the interaction

between treatment and extent of RD after debulking. It should be emphasised that these studies were RCTs designed to measure the effect of PDS versus IDS and were not designed to evaluate the intervention of differing degrees of surgical effort.

The results of the SOCQER-2 study assessing quality of life and progression-free survival found that patients with late-stage ovarian cancer had no important differences in EORTC QLQ-C30 global scores measured across six weeks, six months and 12 months post-surgery when undergoing surgery of varying complexity, despite a higher preoperative disease burden in patients undergoing more radical surgical procedures (Sundar 2022). The authors of the study found that patients who underwent low-complexity surgery had higher rates of residual disease and lower survival compared with those with a similar disease burden undergoing surgery of intermediate complexity. However, no statistical adjustment was performed in these analyses. Postoperative residual disease was associated with poorer overall survival, particularly in patients undergoing low-complexity surgery, but again no statistical adjustment was made and, as this was not an intervention study, it is not able to determine the causal effect of this relationship.

Women with FIGO stage IIIC disease with extra pelvic metastases smaller than 5 cm have been shown to have better progression-free survival after upfront debulking (Vergote 2018). An investigation of NMRD during the initial treatment of epithelial ovarian cancer comparing PDS versus IDS has been investigated in a TRUST (Trial of Radical Upfront Surgical Therapy in advanced ovarian cancer (ENGOT ov33/AGO-OVAR OP7)) trial, which is due to report in 2024 (Reuss 2019).

AUTHORS' CONCLUSIONS

Implications for practice

In a primary debulking surgery (PDS) setting, this review provides moderate-certainty evidence that residual disease (RD) after primary surgery is a strong prognostic factor for overall and progression-free survival in women with advanced ovarian cancer. The certainty of the evidence for these outcomes was very low for studies involving interval debulking surgery (IDS). We conclude that there should be three distinct categories of RD after PDS, including no macroscopic residual disease (NMRD) (labelled as optimal), < 1 cm (labelled as small-volume residual disease (SVRD) and strictly meaning 0.1 cm to 1 cm) and > 1 cm (large-volume residual disease (LVRD)).

After IDS, there may be only two categories required, although this is based on very low-certainty evidence and it would be unwise to make any firm inferences or conclusions until further studies are added to the evidence base.

It is acknowledged that there is considerable variation in achieving NMRD or SVRD between different surgeons and centres. Predicting the achievement of NMRD or SVRD prior to surgery will be dependent on this variation, resulting in difficulties in developing models of prediction, so deciding on whether to perform PDS or IDS at present is down to clinician preference.

NMRD remains a key prognosticator of survival in advanced ovarian cancer. Whether PDS or IDS is the primary treatment, the surgical goal should be to completely remove all visible disease, although SVRD should still be regarded as a favourable outcome after PDS, as shown in this systematic review, although this is not clear following IDS.

The evidence on the ability of different thresholds of RD to distinguish between a good and bad prognosis can aid decision-making for clinicians and diagnosed individuals, where the survival advantage can be considered alongside any potential morbidity or adverse event trade-offs.

Implications for research

The purpose of this systematic review was to assess RD as a prognostic factor in women who received primary surgery (PDS and IDS) for advanced epithelial ovarian cancer (stages III and IV). The results should encourage the surgical community to make trials in this area a priority. Future research should focus on investigations that determine whether increasing attempts at achieving NMRD have a direct effect in improving survival outcomes using methodologies and trial designs that reduce or eliminate confounding effects, such as the women's performance status, disease spread and tumour biology.

Greater emphasis should be made in future studies to investigate IDS to raise the certainty of the evidence profiles. In both PDS and IDS settings, quality of life parameters and adverse effects and complications of the surgery need to be adequately addressed as there are significant deficiencies in previous studies in evaluating these outcome measures. It is unlikely that studies on prognosis will measure or report adverse events, so our focus in this review was on survival as an outcome. These additional evaluations should be given high priority, as this systematic review has identified large differences in survival outcomes associated with LVRD compared to when NMRD is achieved. The results of the SOCQER-2 study suggest that quality of life may still be reasonable even after more extensive surgery, which is reassuring, although this was an observational study (Sundar 2022). An investigation of cytoreductive surgery during the initial treatment of epithelial ovarian cancer comparing PDS versus IDS has also been investigated in a TRUST (Trial of Radical Upfront Surgical Therapy in advanced ovarian cancer (ENGOT ov33/AGO-OVAR OP7)) trial and we await the results in 2024 (Reuss 2019).

To avoid continuous confounding of results, observational studies should report the following to better assess the effect of surgical treatment in advanced ovarian cancer:

- Structural selection – the specific setting in which women are referred/seek care and which sample of the population (or population) has been chosen.
- To what extent the population of women with ovarian cancer are accounted for (selection of patients macro level).
- Institutional selection – how women were selected for surgery (choice of surgeon, patient, etc.).
- The extent of surgery needed to achieve complete resection, i.e. procedures and surgical complexity scores (surgical proficiency).
- Complete resection rates.

ACKNOWLEDGEMENTS

We thank Jo Morrison, Tracey Harrison and Gail Quinn from Cochrane Gynaecological, Neuro-oncology and Orphan Cancers

(GNOC) for their advice support and contribution to the editorial process. We thank Nicole Skoetz from the Prognostic Methods Group for her valuable input. We also thank Jo Platt, Information Manager for GNOC, for designing the search strategies.

The authors and GNOC are grateful to the following peer reviewers for their time and comments; Simon Butler-Manuel, Jennifer Hare, Sonali Kaushik, Hans Nagar and Sahar Salehi.

This project was supported by the National Institute for Health Research (NIHR), via Cochrane infrastructure funding to Cochrane Gynaecological, Neuro-oncology and Orphan Cancers. The views and opinions expressed therein are those of the review authors and do not necessarily reflect those of the Systematic Reviews Programme, NIHR, NHS, or the Department of Health.

REFERENCES

References to studies included in this review

Akahira 2001 *{published data only}*

Akahira JI, Yoshikawa H, Shimizu Y, Tsunematsu R, Hirakawa T, Kuramoto H. Prognostic factors of stage IV epithelial ovarian cancer: a multicenter retrospective study. *Gynecologic Oncology* 2001;**81**(3):398-403.

Aletti 2006 *{published data only}*

Aletti GD, Dowdy SC, Gostout BS, Jones MB, Stanhope CR, Wilson TO, et al. Aggressive surgical effort and improved survival in advanced-stage ovarian cancer. *Obstetrics and Gynecology* 2006;**107**:77–85.

Aletti GD, Dowdy SC, Podratz KC, Cliby WA. Relationship among surgical complexity, short-term morbidity, and overall survival in primary surgery for advanced ovarian cancer. *American Journal of Obstetrics and Gynecology* 2007;**197**(6):1-7.

Aletti GD, Dowdy SC, Podratz KC, Cliby WA. Surgical treatment of diaphragm disease correlates with improved survival in optimally debulked advanced stage ovarian cancer. *Gynecologic Oncology* 2006;**100**(2):283-7.

Aletti GD, Podratz KC, Jones MB, Cliby WA. Role of rectosigmoidectomy and stripping of pelvic peritoneum in outcomes of patients with advanced ovarian cancer. *Journal of the American College of Surgeons* 2006;**203**(4):521-6.

Ataseven 2016 *{published data only}*

Ataseven B, Grimm C, Harter P, Heitz F, Traut A, Prader S, et al. Prognostic impact of debulking surgery and residual tumor in patients with epithelial ovarian cancer FIGO stage IV. *Gynecologic Oncology* 2016;**140**(2):215-20.

Bixel 2020 *{published data only}*

Bixel K, Vetter M, Davidson B, Berchuck A, Cohn D, Copeland L, et al. Intraperitoneal chemotherapy following neoadjuvant chemotherapy and optimal interval tumor reductive surgery for advanced ovarian cancer. *Gynecologic Oncology* 2020;**156**:530–4.

Bristow 2011 *{published data only}*

Bristow RE, Ueda S, Gerardi MA, Ajiboye OB, Ibeanu OA. Analysis of racial disparities in stage IIIC epithelial ovarian cancer care and outcomes in a tertiary gynecologic oncology referral center. *Gynecologic Oncology* 2011;**122**(2):319-23.

Chan 2003 *{published data only}*

Chan JK, Loizzi V, Lin YG, Osann K, Brewster WR, DiSaia PJ. Stages III and IV invasive epithelial ovarian carcinoma in younger versus older women: what prognostic factors are important? *Obstetrics and Gynecology* 2003;**102**(1):156-61.

Chang 2012a *{published data only}*

Chang SJ, Bristow RE, Ryu HS. Impact of complete cytoreduction leaving no gross residual disease associated with radical cytoreductive surgical procedures on survival in advanced ovarian cancer. *Annals of Surgical Oncology* 2012;**19**(13):4059-67.

Chang 2012b *{published data only}*

Chang SJ, Bristow RE, Ryu HS. Prognostic significance of systematic lymphadenectomy as part of primary debulking surgery in patients with advanced ovarian cancer. *Gynecologic Oncology* 2012;**126**(3):381-6.

Chi 2001 *{published data only}*

Chi DS, Liao JB, Leon LF, Venkatraman ES, Hensley ML, Bhaskaran D, et al. Identification of prognostic factors in advanced epithelial ovarian carcinoma. *Gynecologic Oncology* 2001;**82**(3):532-7.

Chi 2006 *{published data only}*

Chi DS, Eisenhauer EL, Lang J, Huh J, Haddad L, Abu-Rustum NR, et al. What is the optimal goal of primary cytoreductive surgery for bulky stage IIIC epithelial ovarian carcinoma (EOC)? *Gynecologic Oncology* 2006;**103**:559–64.

Eisenhauer EL, Abu-Rustum NR, Sonoda Y, Aghajanian C, Barakat RR, Chi DS. The effect of maximal surgical cytoreduction on sensitivity to platinum-taxane chemotherapy and subsequent survival in patients with advanced ovarian cancer. *Gynecologic Oncology* 2008;**108**(2):276-81.

Cioffi 2018 *{published data only}*

Cioffi R, Bergamini A, Rabaiotti E, Petrone M, Pella F, Ferrari D, et al. Neoadjuvant chemotherapy in high-risk ovarian cancer patients: role of age. *Tumori Journal* 2018;**105**(2):168-73.

Cuylan 2018 *{published data only}*

Cuylan ZF, Meydanli MM, Sari ME, Akbayir O, Celik H, Dede M, et al. Prognostic factors for maximising or optimally cytoreduced stage III non serous epithelial ovarian carcinoma treated with carboplatin/paclitaxel chemotherapy. *Journal of Obstetrics and Gynaecology Research* 2018;**44**(7):1284-93.

Davidson 2019 *{published data only}*

Davidson BA, Broadwater G, Crim A, Boccacio R, Bixel K, Backes F, et al. Surgical complexity score and role of laparoscopy in women with advanced ovarian cancer treated with neoadjuvant chemotherapy. Gynecologic Oncology 2019;**152**(3):554-9.

Eisenkop 2003 *{published data only}*

Eisenkop SM, Friedman RL, Wang HJ. Complete cytoreductive surgery is feasible and maximizes survival in patients with advanced epithelial ovarian cancer: a prospective study. *Gynecologic Oncology* 1998;**69**(2):103-8.

Eisenkop SM, Spirtos NM, Friedman RL, Lin WC, Pisani AL, Perticucci S. Relative influences of tumor volume before surgery and the cytoreductive outcome on survival for patients with advanced ovarian cancer: a prospective study. *Gynecologic Oncology* 2003;**90**(2):390-6.

Feng 2016 *{published data only}*

Feng Z, Wen H, Bi R, Yang W, Wu X. Prognostic impact of the time interval from primary surgery to intravenous chemotherapy

in high grade serous ovarian cancer. *Gynecologic Oncology* 2016;**141**:466-70.

Hofstetter 2013 *{published data only}*

Hofstetter G, Concin N, Braicu I, Chekerov R, Sehouli J, Cadron I. The time interval from surgery to start of chemotherapy significantly impacts prognosis in patients with advanced serous ovarian carcinoma - analysis of patient data in the prospective OVCAD study. *Gynecologic Oncology* 2013;**131**(1):15-20.

Iwase 2015 *{published data only}*

Iwase H, Takada T, Iitsuka C, Nomura H, Abe A, Taniguchi T, et al. Clinical significance of systematic retroperitoneal lymphadenectomy during interval debulking surgery in advanced ovarian cancer patients. *Journal of Gynecologic Oncology* 2015;**26**(4):303-10.

Kaban 2017 *{published data only}*

Kaban A, Topuz S, Saip P, Sozen H, Celebi K, Salihoglu Y. Poor prognostic factors in patients undergoing surgery after neoadjuvant chemotherapy for ovarian, tubal, or peritoneal cancer. *Journal of Obstetrics and Gynaecology* 2017;**39**(12):1163-70.

Kahl 2017 *{published data only}*

Kahl A, du Bois A, Harter P, Prader S, Schneider S, Heitz F, et al. Prognostic value of the age-adjusted Charlson comorbidity index (ACCI) on short- and long-term outcome in patients with advanced primary epithelial ovarian cancer. *Annals of Surgical Oncology* 2017;**24**:3692-9.

Klar 2016 *{published data only}*

du Bois A, Lück H, Meier W, Adams HP, Möbus V, Costa S, et al. A randomized clinical trial of cisplatin/paclitaxel versus carboplatin/paclitaxel as first-line treatment of ovarian cancer. *Journal of the National Cancer Institute* 2003;**95**(17):1320-9.

du Bois A, Weber B, Rochon J, Meier W, Goupil A, Olbricht S, , et al. Addition of epirubicin as a third drug to carboplatin/paclitaxel in first-line treatment of advanced ovarian cancer: a prospectively randomized gynecologic cancer intergroup trial by the Arbeitsgemeinschaft Gynaekologische Onkologie Ovarian Cancer Study Group and the Groupe d'Investigateurs Nationaux pour l'Etude des Cancers Ovariens. *Journal of Clinical Oncology* 2006;**24**(7):1127-35.

du Bois A HJ, Hardy-Bessard AC, Muller HH, Harter P, Kristensen G. Phase III trial of carboplatin plus paclitaxel with or without gemcitabine in first-line treatment of epithelial ovarian cancer. *Journal of Clinical Oncology* 2010;**28**(27):4162-9.

Klar M, Hasenburg A, Hasenov M, Hilpert F, Meier W, Pfisterer J, et al. Prognostic factors in young ovarian cancer patients: an analysis of four prospective phase III intergroup trials of the AGO Study Group, GINECO and NSGO. *European Journal of Cancer* 2016;**66**:114-24.

Mahner S, Eulenburg C, Staehle A, Wegscheider K, Reuss A, Pujade-Lauraine E, et al. Prognostic impact of the time interval between surgery and chemotherapy in advanced ovarian cancer: analysis of prospective randomised phase III trials. *European Journal of Cancer* 2013;**49**(1):142-9.

Pfisterer JWB, Reuss A, Kimmig R, du Bois A, Wagner U, et al. Randomized phase III trial of topotecan following carboplatin and paclitaxel in first-line treatment of advanced ovarian cancer: a gynecologic cancer intergroup trial of the AGO-OVAR and GINECO. *Journal of the National Cancer Institute* 2006;**98**(15):1036-45.

Langstraat 2011 *{published data only}*

Langstraat C, Aletti GD, Cliby WA. Morbidity, mortality and overall survival in elderly women undergoing primary surgical debulking for ovarian cancer: a delicate balance requiring individualization. *Gynecologic Oncology* 2011;**123**(2):187-91.

Lecointre 2020 *{published data only}*

Lecointre L, Velten M, Lodi M, Saadeh R, Lavoué V, Ouldamer L, et al. Impact of neoadjuvant chemotherapy cycles on survival of patients with advanced ovarian cancer: a French national multicenter study (FRANCOGYN). *European Journal of Obstetrics & Gynecology and Reproductive Biology* 2020;**245**:64–72.

Lecuru 2019 *{published data only}*

Lecuru F, Pujade-Lauraine E, Hamizi S, Caumont-Prim A, Raban N, Malaurie E, et al. 1016P - Surrogate endpoint of progression-free (PFS) and overall survival (OS) for advanced ovarian cancer (AOC) patients (pts) treated with neo-adjuvant chemotherapy (NACT): results of the CHIVA randomized phase II GINECO study. *Abstract Book of the 44th ESMO Congress (ESMO 2019) 27 September - 1 October 2019, Barcelona, Spain* 2019;**30**:v415.

Liu 2020 *{published data only}*

Liu Y, Cao L, Chen W, Wang J, Wang W, Liang Z. Feasibility of neoadjuvant and adjuvant intraperitoneal chemotherapy in patients with advanced epithelial ovarian cancer: a single-center experience. *Medicine* 2020;**99**(36):e22100.

Lorusso 2016 *{published data only}*

Lorusso D, Bogani G, Matteucci L, Ditto A, Tamberi S, Arcangeli V, et al. Number of cycles of neoadjuvant chemotherapy might influence survival of patients undergoing interval debulking surgery for non-cytoreducible ovarian cancer: results from a multi-institutional study. *International Journal of Gynecological Cancer* 2016;**26**:647.

Luger 2020 *{published data only}*

Luger AK, Steinkohl F, Aigner F, Jaschke W, Marth C, Zeimet AG, et al. Enlarged cardiophrenic lymph nodes predict disease involvement of the upper abdomen and the outcome of primary surgical debulking in advanced ovarian cancer. *Acta Obstetricia et Gynecologica Scandinavica* 2020;**99**:1092–9.

McGuire 1995 *{published data only}*

Hoskins WJ, McGuire WP, Brady MF, Homesley HD, Creasman WT, Berman M, et al. The effect of diameter of largest residual disease on survival after primary cytoreductive surgery in patients with suboptimal residual epithelial ovarian carcinoma. *American Journal of Obstetrics and Gynecology* 1994;**170**(4):974-80.

McGuire WP, Hoskins WJ, Brady MF, Homesley HD, Creasman WT, Berman ML, et al. Assessment of dose-intensive therapy in suboptimally debulked ovarian cancer: a Gynecologic Oncology Group Study. *Journal of Clinical Oncology* 1995;**13**:1589-99.

Melamed 2017a *{published data only}*

Melamed A, Manning-Geist B, Bregar AJ, Diver EJ, Goodman A, del Carmen MG, et al. Associations between residual disease and survival in epithelial ovarian cancer by histologic type. *Gynecologic Oncology* 2017;**14**(7):250-6.

Melamed 2017b *{published data only}*

Melamed A, Manning-Geist B, Bregar AJ, Diver EJ, Goodman A, del Carmen MG, et al. Associations between residual disease and survival in epithelial ovarian cancer by histologic type. *Gynecologic Oncology* 2017;**14**(7):250-6.

Paik 2018 *{published data only}*

Paik ES, Kim JH, Kim TJ, Lee JW, Kim BG, Bae DS, et al. Prognostic significance of normal-sized ovary in advanced serous epithelial ovarian cancer. *Journal of Gynecologic Oncology* 2018;**29**(1):e13.

Peiretti 2010 *{published data only}*

Peiretti M, Zanagnolo V, Aletti GD, Bocciolone L, Colombo N, Landoni F, et al. Role of maximal primary cytoreductive surgery in patients with advanced epithelial ovarian and tubal cancer: surgical and oncological outcomes. *Gynecologic Oncology* 2010;**119**(2):259-64.

Peiretti 2012 *{published data only}*

Peiretti M, Bristow RE, Zapardiel I, Gerardi M, Zanagnolo V, Biffi R, et al. Rectosigmoid resection at the time of primary cytoreduction for advanced ovarian cancer. A multi-center analysis of surgical and oncological outcomes. *Gynecologic Oncology* 2012;**126**(2):220-3.

Petrillo 2014 *{published data only}*

Petrillo M, Zannoni GF, Tortorella L, Pedone AL, Salutari V, Ercoli A, et al. Prognostic role and predictors of complete pathologic response to neoadjuvant chemotherapy in primary unresectable ovarian cancer. *American Journal of Obstetrics and Gynecology* 2014;**211**(6):632.e1-8.

Phillips 2018 *{published data only}*

Phillips A, Sundar S, Singh K, Nevin J, Elattar A, Kehoe S, et al. Complete cytoreduction after five or more cycles of neo-adjuvant chemotherapy confers a survival benefit in advanced ovarian cancer. *European Journal of Surgical Oncology* 2018;**44**(6):760-5.

Polterauer 2012 *{published data only}*

Polterauer S, Vergote I, Concin N, Braicu I, Chekerov R, Mahner S. Prognostic value of residual tumor size in patients with epithelial ovarian cancer FIGO stages IIA-IV: analysis of the OVCAD data. *International Journal of Gynecological Cancer* 2012;**22**(3):380-5.

Shibutani 2020 *{published data only}*

Shibutani T, Nagao S, Suzuki K, Kaneda M, Yamamoto K, Jimi T. `Dose-dense` paclitaxel and carboplatin vs. conventional paclitaxel and carboplatin as neoadjuvant chemotherapy for advanced epithelial ovarian, fallopian tube, or primary peritoneal cancer: a retrospective study. *International Journal of Clinical Oncology* 2020;**25**:502–7.

Shim 2016 *{published data only}*

Shim SH, Kim DY, Jung PS, Kim JH, Kim YM, Suh DS, et al. Prognostic impact of the time interval from surgery to chemotherapy in patients with advanced ovarian cancer. *Gynecologic Oncology* 2016;**141**(S1):49.

Stoeckle 2014 *{published data only}*

Stoeckle E, Bourdarias L, Guyon F, Croce S, Brouste V, Thomas L, et al. Progress in survival outcomes in patients with advanced ovarian cancer treated by neo-adjuvant platinum/taxane-based chemotherapy and late interval debulking surgery. *Annals of Surgical Oncology* 2014;**21**(2):629-36.

Tewari 2016 *{published data only}*

Tewari KS, Java JJ, Eskander RN, Monk BJ, Burger RA. Early initiation of chemotherapy following complete resection of advanced ovarian cancer associated with improved survival: NRG Oncology/Gynecologic Oncology Group study. *Annals of Oncology* 2016;**27**:114-21.

Tseng 2018 *{published data only}*

Tseng JH, Cowan RA, Zhou Q, Lasonos A, Byrne M, Polcino T, et al. Continuous improvement in primary debulking surgery for advanced ovarian cancer: do increased complete Gross resection rates independently lead to increased progression-free and overall survival? *Gynecologic Oncology* 2018;**151**(1):24-31.

Van Geene 1996 *{published data only}*

Van Geene P, Varma R, Dunn J, Chan KK, Luesley DM. The prognostic significance of intraperitoneal growth characteristics in epithelial ovarian carcinoma. *International Journal of Gynecological Cancer* 1996;**6**(3):219-24.

Wimberger 2010 *{published data only}*

Wimberger P, Wehling M, Lehmann N, Kimmig R, Schmalfeldt B, Burges A. Influence of residual tumor on outcome in ovarian cancer patients with FIGO stage IV disease: an exploratory analysis of the AGO-OVAR (Arbeitsgemeinschaft Gynaekologische Onkologie Ovarian Cancer Study Group). *Annals of Surgical Oncology* 2010;**17**(6):1642-8.

Winter 2007 *{published data only}*

Armstrong DK, Bundy BW, Wenzel L, Huang HQ, Baergen RL, Shashikant L, et al, the Gynecologic Oncology Group. Intraperitoneal cisplatin and paclitaxel in ovarian cancer. *New England Journal of Medicine* 2006;**354**(1):34-43.

Markman M, Bundy BM, Alberts DS, Fowler JM, Clark-Pearson DL, Carson LF, et al. Phase III trial of standard-dose intravenous cisplatin plus paclitaxel versus moderately high-dose carboplatin followed by intravenous paclitaxel and intraperitoneal cisplatin in small-volume stage III ovarian

carcinoma: An Intergroup Study of the Gynecologic Oncology Group, Southwestern Oncology Group, and Eastern Cooperative Oncology Group. *Journal of Clinical Oncology* 2001;**19**(4):1001-7.

McGuire WP, Hoskins WJ, Brady MF, Kucera PR, Partridge EE, Look KY, et al. Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer. *New England Journal of Medicine* 1996;**334**(1):1-6.

Muggia FM, Braly PS, Brady MF, Sutton GN, Theodore HL, Samuel LA, et al. Phase III randomized study of cisplatin versus paclitaxel versus cisplatin and paclitaxel in patients with suboptimal stage III or IV ovarian cancer: a Gynecologic Oncology Group Study. *Journal of Clinical Oncology* 2000;**18**(1):106-15.

Ozols RF, Bundy BN, Greer BE, Fowler JM, Clarke-Pearson DB, Robert AM, et al. Phase III trial of carboplatin and paclitaxel compared with cisplatin and paclitaxel in patients with optimally Resected stage III ovarian cancer: a Gynecologic Oncology Group Study. *Journal of Clinical Oncology* 2003;**21**(17):3194-200.

Rose PG, Nerenstone S, Brady MF, Clarke-Pearson D, Olt G, Rubin SC, et al. Secondary surgical cytoreduction for advanced ovarian carcinoma. *New England Journal of Medicine* 2004;**351**(24):2489-97.

Winter WE, Maxwell III GL, Tian C, Carlson JW, Ozols RF, Rose PG, et al. Prognostic factors for stage III epithelial ovarian cancer: a Gynecologic Oncology Group Study. *Journal of Clinical Oncology* 2007;**25**(24):3621-7.

Winter 2008 *{published data only}*

McGuire WP, Hoskins WJ, Brady MF, Kucera PR, Partridge EE, Look KY, et al. Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer. *New England Journal of Medicine* 1996;**334**(1):1-6.

Muggia FM, Braly PS, Brady MF, Sutton G, Niemann HL, Lentz SL, et al. Phase III randomized study of cisplatin versus paclitaxel versus cisplatin and paclitaxel in patients with suboptimal stage III or IV ovarian cancer: a Gynecologic Oncology Group Study. *Journal of Clinical Oncology* 2000;**18**(1):106-15.

Rose PG, Nerenstone S, Brady MF, Clarke-Pearson D, Olt G, Rubin SC, et al. Secondary surgical cytoreduction for advanced ovarian carcinoma. *New England Journal of Medicine* 2004;**351**(24):2489-97.

Spriggs DR, Brady MF, Vaccarello L, Clarke-Pearson DL, Burger RA, Mannel R, et al. Phase III randomized trial of intravenous cisplatin plus a 24- or 96-hour infusion of paclitaxel in epithelial ovarian cancer: a Gynecologic Oncology Group Study. *Journal of Clinical Oncology* 2007;**25**(28):4466-71.

Winter WE, Maxwell GL, Tian C, Sundborg MJ, Rose GS, Rose PG, et al. Tumor residual after surgical cytoreduction in prediction of clinical outcome in stage IV epithelial ovarian cancer: a Gynecologic Oncology Group Study. *Journal of Clinical Oncology* 2008;**26**(1):83-9.

Zhang 2018 *{published data only}*

Zhang J, Ning L, Zhang A, Xiangxiang B. Potential risk factors associated with prognosis of neoadjuvant chemotherapy followed by interval debulking surgery in stage IIIc-IV high-grade serous ovarian carcinoma patients. *Journal of Obstetrics and Gynaecology Research* 2018;**44**(9):1808-16.

Zhu 2016 *{published data only}*

Zhu J, Wang H, Cheng-Cheng L, Lu Y, Tang H. The Glasgow Prognostic Score (GPS) is a novel prognostic indicator in advanced epithelial ovarian cancer: a multicenter retrospective study. *Journal of Cancer Research and Clinical Oncology* 2016;**142**(11):2339-45.

References to studies excluded from this review

Alberts 1993 *{published data only}*

Alberts DS, Dahlberg S, Green SJ, Garcia D, Hannigan EV, O'Toole R, et al. Analysis of patient age as an independent prognostic factor for survival in a phase III study of cisplatin-cyclophosphamide versus carboplatin-cyclophosphamide in stages III (suboptimal) and IV ovarian cancer. A Southwest Oncology Group study. *Cancer* 1993;**71**(2 Suppl):618-27.

Alberts 1996 *{published data only}*

Alberts DS, Liu PY, Hannigan EEV, O'Toole RW, Stephen DY, James AF, et al. Intraperitoneal cisplatin plus intravenous cyclophosphamide versus intravenous cisplatin plus intravenous cyclophosphamide for stage III ovarian cancer. *New England Journal of Medicine* 1996;**335**(26):1950-5.

Alphs 2006 *{published data only}*

Alphs HH, Zahurak ML, Bristow RE, Diaz-Montes TP. Predictors of surgical outcome and survival among elderly women diagnosed with ovarian and primary peritoneal cancer. *Gynecologic Oncology* 2006;**103**(3):1048-53.

Altman 2012 *{published data only}*

Altman AD, Nelson G, Chu P, Nation J, Ghatage P. Optimal debulking targets in women with advanced stage ovarian cancer: a retrospective study of immediate versus interval debulking surgery. *Journal of Obstetrics and Gynaecology Canada* 2012;**34**(6):558-66.

Andersen Soegaard 2005 *{published data only}*

Andersen ES, Knudsen A, Svarrer T, Lund B, Nielsen K, Grove A, et al. The results of treatment of epithelial ovarian cancer after centralisation of primary surgery. Results from North Jutland, Denmark. *Gynecologic Oncology* 2005;**99**(3):552-6.

Anuradha 2016 *{published data only}*

Anuradha S, Donovan PJ, Webb PM, Brand AH, Goh J, Friedlander M, et al. Variations in adjuvant chemotherapy and survival in women with epithelial ovarian cancer–a population-based study. *Acta Oncologica* 2016;**55**(2):226-33.

Bailey 2006 *{published data only}*

Bailey J, Murdoch J, Anderson R, Weeks J, Foy C. Stage III and IV ovarian cancer in the South West of England: five-year outcome

analysis for cases treated in 1998. *International Journal of Gynecological Cancer* 2006;**1**:25-9.

Baker 1994 *{published data only}*

Baker TR, Piver MS, Hempling RE. Long term survival by cytoreductive surgery to less than 1 cm, induction weekly cisplatin and monthly cisplatin, doxorubicin, and cyclophosphamide therapy in advanced ovarian adenocarcinoma. *Cancer* 1994;**74**(2):656-63.

Barda 2004 *{published data only}*

Barda G, Menczer J, Chetrit A, Lubin F, Beck D, Piura B, et al. Comparison between primary peritoneal and epithelial ovarian carcinoma: a population-based study. *American Journal of Obstetrics and Gynecology* 2004;**190**(4):1039-45.

Benedetti-Panici 1996 *{published data only}*

Benedetti-Panici P, Maneschi F, Scambia G, Cutillo G, Greggi S, Mancuso S. The pelvic retroperitoneal approach in the treatment of advanced ovarian carcinoma. *Obstetrics & Gynecology* 1996;**87**(4):532-8.

Bertelsen 1990 *{published data only}*

Bertelsen K. Tumor reduction surgery and long-term survival in advanced ovarian cancer: a DACOVA study. *Gynecologic Oncology* 1990;**38**(2):203-9.

Bertelsen 1993 *{published data only}*

Bertelsen K, Jakobsen A, Strøyer J, Nielsen K, Sandberg E, Andersen JE, et al. A prospective randomized comparison of 6 and 12 cycles of cyclophosphamide, adriamycin, and cisplatin in advanced epithelial ovarian cancer: a Danish Ovarian Study Group Trial (DACOVA). *Gynecologic Oncology* 1993;**49**:30-6.

Bian 2016 *{published data only}*

Bian C, Yao K, Li L, Yi T, Zhao X. Primary debulking surgery vs. neoadjuvant chemotherapy followed by interval debulking surgery for patients with advanced ovarian cancer. *Archives of Gynecology and Obstetrics* 2016;**293**(1):163-8.

Bonnefoi 1999 *{published data only}*

Bonnefoi H, A'Hern RP, Fisher C, Macfarlane V, Barton D, Blake P, et al. Natural history of stage IV epithelial ovarian cancer [see comment]. *Journal of Clinical Oncology* 1999;**17**(3):767-75.

Brinkhuis 1996a *{published data only}*

Brinkhuis M, Baak JPA, Van Diest PJ, Lund B, Wils J. In Dutch and Danish patients with FIGO III ovarian carcinoma, geographic survival differences are associated with differences in quantitative pathologic features. *International Journal of Gynecological Cancer* 1996;**6**(2):108-14.

Brinkhuis 1996b *{published data only}*

Brinkhuis M, Lund B, Meijer GA, Baak JPA. Quantitative pathological variables as prognostic factors for overall survival in Danish patients with FIGO stage III ovarian cancer. *International Journal of Gynecological Cancer* 1996;**6**(3):168-74.

Bristow 1999 *{published data only}*

Bristow RE, Montz FJ, Lagasse LD, Leuchter RS, Karlan BY. Survival impact of surgical cytoreduction in stage IV epithelial ovarian cancer. *Gynecologic Oncology* 1999;**72**(3):278-87.

Cai 2007 *{published data only}*

Cai HB, Zhou YF, Chen HZ, Hou HY. The role of bowel surgery with cytoreduction for epithelial ovarian cancer. *Clinical Oncology* 2007;**19**(10):757-62.

Ceresoli 2018 *{published data only}*

Ceresoli M, Verrengia A, Montori G, Busci L, Coccolini F, Ansaloni L, et al. Effect of cytoreductive surgery and hyperthermic intraperitoneal chemotherapy on relapse pattern in primary epithelial ovarian cancer: a propensity score based case-control study. *Journal of Gynecologic Oncology* 2018;**29**(3):e53.

Chekman 2015 *{published data only}*

Chekman C, Layoune R, Hocine O, Raissi N, Ferhat HA, Ali Khodja H, et al. An open prospective randomized trial comparing primary complete cytoreduction surgery to debulking surgery after chemotherapy in advanced stage (FIGO's IIIC) ovarian carcinoma. In: 19th International Meeting of the European Society of Gynaecological Oncology, ESGO 2015; 2015 Oct 24-27; Nice, France. 2015:1316. [12185451]

Clamp 2018 *{published data only}*

Clamp AR, McNeish IA, Dean A, Gallardo-Rincon D, Kim JW, O'Donnell DM, et al. Response to neoadjuvant chemotherapy in ICON8: a GCIG phase III randomised trial evaluating weekly dose-dense chemotherapy integration in first-line epithelial ovarian/fallopian tube/primary peritoneal carcinoma (EOC) treatment. In: Annals of Oncology. Vol. 29. 2018:viii336.

Colozza 1997 *{published data only}*

Colozza M, Mosconi AM, Gori S, Belsanti V, Basurto C, De Angelis V, et al. Long-term results in patients with advanced epithelial ovarian carcinoma treated with a combination of cisplatin, doxorubicin, and cyclophosphamide. *American Journal of Clinical Oncology* 1997;**20**(5):522-6.

Conte 1991 *{published data only}*

Conte PF, Bruzzone M, Carnino F, Chiara S, Donadio M, Facchini V, et al. Carboplatin, doxorubicin, and cyclophosphamide versus cisplatin, doxorubicin, and cyclophosphamide: a randomized trial in stage III-IV epithelial ovarian carcinoma. *Journal of Clinical Oncology* 1991;**9**(4):658-63.

Conte 1996 *{published data only}*

Conte PF, Bruzzone M, Carnino F, Gadducci A, Algeri R, Bellini A, et al. High-dose versus low-dose cisplatin in combination with cyclophosphamide and epidoxorubicin in suboptimal ovarian cancer: a randomized study of the Gruppo Oncologico Nord-Ovest. *Journal of Clinical Oncology* 1996;**14**(2):351-6.

Crawford 2005 *{published data only}*

Crawford SC, Vasey PA, Paul J, Hay A, Davis JA, Kaye SB. Does aggressive surgery only benefit patients with less advanced ovarian cancer? Results from an international comparison

within the SCOTROC-1 trial. *Journal of Clinical Oncology* 2005;**23**(34):8802-11.

Creasman 1990 *{published data only}*

Creasman WT, Omura GA, Brady MF, Yordan E, DiSaia PJ, Beecham J. A randomized trial of cyclophosphamide, doxorubicin, and cisplatin with or without bacillus Calmette-Guerin in patients with suboptimal stage III and IV ovarian cancer: a Gynecologic Oncology Group Study. *Gynecologic Oncology* 1990;**39**(3):239-43.

Cummins 2019 *{published data only}*

Cummins C, Hannah P, Long J, Kumar S, Sundar S, Bramley G. VP100 Ultraradical ovarian cancer surgery comparative clinical effectiveness. *International Journal of Technology Assessment in Health Care* 2019;**35**(S1):97.

Dao 2016 *{published data only}*

Dao F, Schlappe BA, Tseng J, Lester J, Nick AM, Lutgendorf SK, et al. Characteristics of 10-year survivors of high-grade serous ovarian carcinoma. *Gynecologic Oncology* 2016;**141**:260-3.

Del Campo 1994 *{published data only}*

Del Campo JM, Felip E, Rubio D, Bermejo B, Colomer R, Zanon V. Long-term survival in advanced ovarian cancer after cytoreduction and chemotherapy treatment. *Gynecologic Oncology* 1994;**53**(1):27-32.

de Oliviera 1990 *{published data only}*

de Oliveira CF, Lacave AJ, Villani C, Wolff JP, di Re F, Namer M, et al. Randomized comparison of cyclophosphamide, doxorubicin and cisplatin (CAP) versus cyclophosphamide and doxorubicin (CA) for the treatment of advanced ovarian cancer (ADOVCA). A EORTC Gynecological Cancer Cooperative Group Study. *European Journal of Gynaecologic Oncology* 1990;**11**(5):323-30.

di Re 1996 *{published data only}*

di Re F, Baiocchi G, Fontanelli R, Grosso G, Cobellis L, Raspagliesi F, et al. Systematic pelvic and paraaortic lymphadenectomy for advanced ovarian cancer: prognostic significance of node metastases. *Gynecologic Oncology* 1996;**62**(3):360-5.

Elgamal 2019 *{published data only}*

Elgamal M, Saha S, Cherry M, Saharan V, Buttar R, Wiese D, et al. Prognostic implications of lymph node metastasis in advanced ovarian cancer: analysis of the National Cancer Database 2006 to 2014. *Journal of Clinical Oncology* 2019;**0**:e17042.

Fagotti 2020 *{published data only}*

Fagotti A, Ferrandina G, Vizzielli G, Fanfani F, Gallotta V, Chiantera V, et al. Phase III randomised clinical trial comparing primary surgery versus neoadjuvant chemotherapy in advanced epithelial ovarian cancer with high tumour load (SCORPION trial): final analysis of peri-operative outcome. *European Journal of Cancer* 2016;**59**:22-33. [12185453]

Fagotti A, Ferrandina MG, Vizzielli G, Pasciuto T, Fanfani F, Gallotta V, et al. Randomized trial of primary debulking surgery versus neoadjuvant chemotherapy for advanced epithelial ovarian cancer (SCORPION-NCT01461850). *International Journal of Gynecological Cancer* 2020;**30**(11):1657-64. [PMID: 10.1136/ijgc-2020- 001640]

Fagotti A, Vizzielli G, Ferrandina G, Fanfani F, Gallotta V, Chiantera V, et al. Survival analyses from a randomized trial of primary debulking surgery versus neoadjuvant chemotherapy for advanced epithelial ovarian cancer with high tumor load (SCORPION trial). *Journal of Clinical Oncology* 2018;**36**(15):5516. [12185454]

Gao 2001 *{published data only}*

Gao J, Zheng A, Chen W, Peng Z, Cao Z. A study of prognostic factors of stage IV epithelial ovarian cancer [Chinese]. *Journal of West China University of Medical Sciences* 2001;**32**(2):309-12.

Gasimli 2016 *{published data only}*

Gasimli K, Braicu EI, Nassir M, Richter R, Babayeva A, Chekerov R, et al. Lymph node involvement pattern and survival differences of FIGO IIIC and FIGO IIIA1 ovarian cancer patients after primary complete tumor debulking surgery: a 10-year retrospective analysis of the tumor bank ovarian cancer network. *Annals of Surgical Oncology* 2016;**23**(4):1279-86.

Geisler 2004 *{published data only}*

Geisler JP, Tammela JE, Manahan KJ, Geisler HE, Miller GA, Zhou Z, et al. HSP27 in patients with ovarian carcinoma: still an independent prognostic indicator at 60 months follow-up. *European Journal of Gynaecological Oncology* 2004;**25**(2):165-8.

Gershenson 1989 *{published data only}*

Gershenson DM, Wharton J, Taylor C, Stringer LJ, Edwards CA, Creighton L, et al. Treatment of advanced epithelial ovarian cancer with cisplatin and cyclophosphamide. *Gynecologic Oncology* 1989;**32**(3):336-41.

Gershenson 1992 *{published data only}*

Gershenson DM, Mitchell MF, Atkinson N, Elvio G, Kavanagh J, Burke M, et al. The effect of prolonged cisplatin-based chemotherapy on progression-free survival in patients with optimal epithelial ovarian cancer: maintenance therapy reconsidered. *Gynecologic Oncology* 1992;**47**(1):7-13.

Gershenson 1995 *{published data only}*

Gershenson DM, Morris M, Burke TW, Levenback C, Kavanagh JJ, Fromm GL. Combined cisplatin and carboplatin chemotherapy for treatment of advanced epithelial ovarian cancer. *Gynecologic Oncology* 1995;**58**(3):349-55.

Greggi 2016 *{published data only}*

Greggi S, Falcone F, Carputo R, Raspagliesi F, Scaffa C, Laurelli G, et al. Primary surgical cytoreduction in advanced ovarian cancer: an outcome analysis within the MITO (Multicentre Italian Trials in Ovarian Cancer and Gynecologic Malignancies) Group. *Gynecologic Oncology* 2016;**140**(3):425-9.

Grem 1991 *{published data only}*

Grem J, O'Dwyer P, Elson P, Simon N, Trump D, Frontiera M, et al. Cisplatin, carboplatin, and cyclophosphamide combination chemotherapy in advanced-stage ovarian carcinoma: an Eastern Cooperative Oncology Group pilot study. *Journal of Clinical Oncology* 1989;**9**(10):1793-800.

Hainsworth 1990 *{published data only}*

Hainsworth JD, Burnett LS, Jones HW 3rd, Grosh WW, Johnson DH, Greco FA. High-dose cisplatin combination chemotherapy in the treatment of advanced epithelial ovarian carcinoma. *Journal of Clinical Oncology* 1990;**8**(3):502-8.

Hakes 1992 *{published data only}*

Hakes TB, Chalas E, Hoskins WJ, Jones WB, Markman M, Rubin SC, et al. Randomized prospective trial of 5 versus 10 cycles of cyclophosphamide, doxorubicin, and cisplatin in advanced ovarian carcinoma. *Gynecologic Oncology* 1992;**45**(3):284-9.

Hamid 2002 *{published data only}*

Hamid D, Rohr S, Baldauf JJ, Ritter J, Kurtz E, Dufour P, et al. Interest of intestinal resection for treatment of advanced ovarian carcinoma [Intérêt des exéréses digestives dans le traitement des cancers écolués de l'ovaire]. *Annales de Chirurgie* 2002;**127**(1):40-7.

Hardy 1991 *{published data only}*

Hardy JR, Wiltshaw E, Blake PR, Harper P, Slevin M, Perren TJ, et al. Cisplatin and carboplatin in combination for the treatment of stage IV ovarian carcinoma. *Annals of Oncology* 1991;**2**(2):131-6.

Heitz 2016 *{published data only}*

Heitz F, Harter P, Alesina PF, Walz MK, Lorenz D, Groeben H, et al. Pattern of and reason for postoperative residual disease in patients with advanced ovarian cancer following upfront radical debulking surgery. *Gynecologic Oncology* 2016;**141**(2):264-70.

Hoskins 1992 *{published data only}*

Hoskins WJ, Bundy BN, Thigpen JT, Omura GA. The influence of cytoreductive surgery on recurrence-free interval and survival in small-volume stage III epithelial ovarian cancer: a Gynecologic Oncology Group study. *Gynecologic Oncology* 1992;**47**(2):159-66.

Hoskins 1996 *{published data only}*

Hoskins PJ, Swenerton KD, Pike JA, McMurtrie EM, Lee N. "MECCA": a developmental, dose-intensive, non-cross-resistant platinum-based chemotherapy for advanced ovarian cancer. *Gynecologic Oncology* 1996;**63**(3):345-51.

Hoskins 1997 *{published data only}*

Hoskins WJ, McGuire WP, Brady MF, Kucera PR, Partridge EE, Look KY, et al. Combination paclitaxel (Taxol(R))-cisplatin vs cyclophosphamide-cisplatin as primary therapy in patients with suboptimally debulked advanced ovarian cancer. *International Journal of Gynecological Cancer* 1997;**7**:9-13.

Itamochi 2002 *{published data only}*

Itamochi H, Kigawa J, Sugiyama T, Kikuchi Y, Suzuki M, Terakawa N. Low proliferation activity may be associated with chemoresistance in clear cell carcinoma of the ovary. *Obstetrics & Gynecology* 2002;**100**(2):281-7.

Kaern 2005 *{published data only}*

Kaern J, Aghmesheh M, Nesland JM, Danielsen HE, Sandstad B, Friedlander M, et al. Prognostic factors in ovarian carcinoma stage III patients. Can biomarkers improve the prediction of short- and long-term survivors? *International Journal of Gynecological Cancer* 2005;**15**(6):1014-22.

Kehoe 2015 *{published data only}*

Kehoe S, Hook J, Nankivell M, Jayson GC, Kitchener H, Lopes T, et al. Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer (CHORUS): an open-label, randomised, controlled, non-inferiority trial. *Lancet* 2015;**386**(9990):249-57. [12185455]

Kehoe S, Hook J, Nankivell M. Chemotherapy or upfront surgery for newly diagnosed advanced ovarian cancer: results from the MRC CHORUS trial. *Journal of Clinical Oncology* 2013;**31 Suppl**(15):Abstract 5500. [12185456]

Kehoe S, Wheeler S. CHORUS (Chemotherapy or Upfront Surgery). A randomised feasibility trial to determine the impact of timing of surgery and chemotherapy in newly diagnosed patients with advanced epithelial ovarian, primary peritoneal or fallopian tube carcinoma. www.ctu.mrc.ac.uk/plugins/StudyDisplay/protocols/CHORUS protocol Version 2.0 - 05 June 2008.pdf; and http://www.ctu.mrc.ac.uk/research_areas/study_details.aspx?s=9 (accessed 18/6/2012). [12185457]

Law K, Murray C, Kehoe S. CHORUS - a randomised study to determine the impact of timing of surgery and chemotherapy in newly diagnosed patients with advanced epithelial ovarian, primary peritoneal or fallopian tube carcinoma. In: Annual Meeting of the British Gynaecological Cancer Society; 2006 Nov 30-Dec 1; Manchester, UK. 2006:90. [12185459]

Vergote I, Coens C, Nankivell M, Kristensen GB, Parmar MK, Ehlen T, et al. Neoadjuvant chemotherapy versus debulking surgery in advanced tubo-ovarian cancers: pooled analysis of individual patient data from the EORTC 55971 and CHORUS trials. *Lancet Oncology* 2018;**19**:1680-7. [12185460]

Kessous 2017 *{published data only}*

Kessous R, Laskov I, Abitbol J, Bitharas J, Yasmeen A, Salvador S, et al. Clinical outcome of neoadjuvant chemotherapy for advanced ovarian cancer. *Gynecologic Oncology* 2017;**144**(3):474-9.

Keyver-Paik 2016 *{published data only}*

Keyver-Paik MD, Abramian A, Domröse C, Döser A, Höller T, Friedrich M, et al. Integrated care in ovarian cancer "IgV Ovar": results of a German pilot for higher quality in treatment of ovarian cancer. *Journal of Cancer Research and Clinical Oncology* 2016;**142**(2):481-7.

Kirmani 1994 *{published data only}*

Kirmani S, Braly PS, McClay EF, Saltzstein SL, Plaxe SC, Kim S et al. A comparison of intravenous versus intraperitoneal chemotherapy for the initial treatment of ovarian cancer. *Gynecologic Oncology* 1994;**54**(3):338-44.

Kristensen 1995 *{published data only}*

Kristensen GB, Baekelandt M, Vergote IB, Trope C. A phase II study of carboplatin and hexamethylmelamine as induction chemotherapy in advanced epithelial ovarian carcinoma. *European Journal of Cancer* 1995;**31**(11):1778-80.

Le 1997 *{published data only}*

Le T, Krepart GV, Lotocki R J, Heywood MS. Does debulking surgery improve survival in biologically aggressive ovarian carcinoma? *Gynecologic Oncology* 1997;**67**(2):208-14.

Lee 2018 *{published data only}*

Lee YY, Lee JW, Lu L, Xu W, Kollara A, Brown T, et al. Impact of interval from primary cytoreductive surgery to initiation of adjuvant chemotherapy in advanced epithelial ovarian cancer. *International Journal of Gynaecology and Obstetrics* 2018;**143**(3):325-32.

Loizzi 2016 *{published data only}*

Fumarulo VV, Loizzi V, Cormio G, Murgia F, Vecchio V, Minicucci V, et al. Neoadjuvant chemotherapy in advanced ovarian cancer: a single institution experience and a review of the literature. In: International Journal of Gynecological Cancer. Vol. 26. 2016:1-1193.

* Loizzi V, Leone L, Camporeale A, Resta L, Selvaggi L, Cicinelli E, et al. Neoadjuvant chemotherapy in advanced ovarian cancer: a single-institution experience and a review of the literature. *Oncology* 2016;**91**(4):211-6.

Lorusso 1998 *{published data only}*

Lorusso V, Leone B, Di Vagno G, Manzione L, Palmeri S, Vallejo C. Combined carboplatin plus ifosfamide and cisplatin in patients with advanced ovarian carcinoma. A phase I-II study. *Gynecologic Oncology* 1998;**68**(2):172-7.

Malik 1998 *{published data only}*

Malik IM, Khan ZK, Khan WA, Hussain M, Moid I, Rizvi J. Continuous infusion of ifosfamide and cisplatin as first-line therapy of patients with suboptimally debulked stage III/IV epithelial ovarian cancer. *International Journal of Gynecological Cancer* 1998;**8**(2):138-43.

Marchetti 1993 *{published data only}*

Marchetti DL, Lele SB, Priore RL, McPhee ME, Hreshchyshyn MM. Treatment of advanced ovarian carcinoma in the elderly. *Gynecologic Oncology* 1993;**49**(1):86-91.

McGuire 1996 *{published data only}*

McGuire WP, Hoskins WJ, Brady MF, Kucera PR, Partridge EE, Look KY, et al. Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer. *New England Journal of Medicine* 1996;**334**(1):1-6.

Michaan 2018 *{published data only}*

Michaan N, Chong WY, Han NY, Lim MC, Park SY. Prognostic value of pathologic chemotherapy response score in patients with ovarian cancer after neoadjuvant chemotherapy. *International Journal of Gynecologic Cancer* 2018;**28**(9):1676-82.

Ngan 1989 *{published data only}*

Ngan HY, Choo YC, Cheung M, Wong LC, Ma HK, Collins R. A randomized study of high-dose versus low-dose cis-platinum combined with cyclophosphamide in the treatment of advanced ovarian cancer. *Chemotherapy* 1989;**35**(3):221-7.

Omura 1989 *{published data only}*

Omura GA, Bundy BN, Berek JS, Curry S, Delgado G Mortel R. Randomized trial of cyclophosphamide plus cisplatin with or without doxorubicin in ovarian carcinoma: a Gynecologic Oncology Group Study. *Journal of Clinical Oncology* 1989;**7**(4):457-65.

Onda 2020 *{published data only}*

Onda T, Matsumoto K, Shibata T, Sato A, Fukuda H, Konishi I, et al. Phase III trial of upfront debulking surgery versus neoadjuvant chemotherapy for stage III/IV ovarian, tubal and peritoneal cancers: Japan Clinical Oncology Group Study JCOG0602. *Japanese Journal of Clinical Oncology* 2008;**38**(1):74-7. [12185462]

* Onda T, Satoh T, Ogawa G, Saito T, Kasamatsu T, Nakanishi T, et al, Japan Clinical Oncology Group. Comparison of survival between primary debulking surgery and neoadjuvant chemotherapy for stage III/IV ovarian, tubal and peritoneal cancers in phase III randomised trial. *European Journal of Cancer* 2020;**130**:114-25.

Onda T, Satoh T, Saito T, Kasamatsu T, Nakanishi T, Nakamura K, et al. Comparison of treatment invasiveness between upfront debulking surgery versus interval debulking surgery following neoadjuvant chemotherapy for stage III/IV ovarian, tubal, and peritoneal cancers in a phase III randomised trial: Japan Clinical Oncology Group Study JCOG0602. *European Journal of Cancer* 2016;**64**:22-31. [12185463]

Onda T, Satoh T, Saito T, Kasamatsu T, Nakanishi T, Takehara K, et al. Comparison of survival between upfront primary debulking surgery versus neoadjuvant chemotherapy for stage III/IV ovarian, tubal and peritoneal cancers in phase III randomized trial: JCOG0602. Journal of Clinical Oncology 2018;**36**(15 Suppl). [12185464]

Palmer 1992 *{published data only}*

Palmer MC, Shepert E, Schepansky A, MacLean GD. Novel, dose intensive, single-agent cisplatin in the first-line management of advanced stage ovarian cancer. *International Journal of Gynecological Cancer* 1992;**2**(6):301-6.

Piver 1991 *{published data only}*

River MS, Fanning J, Sprance HE. Five-year survival for cisplatin-based chemotherapy versus single-agent melphalan in patients with advanced ovarian cancer and optimal debulking surgery. *Journal of Surgical Oncology* 1991;**48**(1):39-44.

Raspagliesi 2018 *{published data only}*

Raspagliesi F, Bogani G, Matteucci L, Casarin J, Sabatucci I, Tamberi S, et al. Surgical efforts might mitigate difference in response to neoadjuvant chemotherapy in stage IIIC–IV unresectable ovarian cancer: a case-control multi-institutional study. *International Journal of Gynecologic Cancer* 2018;**28**(9):1706-13.

Redman 1986 *{published data only}*

Redman JR, Petroni GR, Saigo PE, Geller NL, Hakes TB. Prognostic factors in advanced ovarian carcinoma. *Journal of Clinical Oncology* 1986;**4**(4):515-23.

Risum 2012 *{published data only}*

Risum S, Loft A, Engelholm SA, Høgdall E, Berthelsen AK, Nedergaard L, et al. Positron emission tomography/computed tomography predictors of overall survival in stage IIIC/IV ovarian cancer. *International Journal of Gynecologic Cancer* 2012;**22**(7):1163-9.

Rodriguez 2013 *{published data only}*

Rodriguez N, Miller A, Richard SD, Rungruang B, Hamilton CA, Bookman MA, et al. Upper abdominal procedures in advanced stage ovarian or primary peritoneal carcinoma patients with minimal or no gross residual disease: an analysis of Gynecologic Oncology Group (GOG). *Gynecologic Oncology* 2013;**130**(3):487-92.

Rose 2004 *{published data only}*

Rose PG, Nerenstone S, Brady MF, Clarke-Pearson D, Olt G, Rubin SC, et al and the Gynecologic Oncology Group. Secondary surgical cytoreduction for advanced ovarian carcinoma. *New England Journal of Medicine* 2004;**351**(24):89-97.

Ruscito 2016 *{published data only}*

Ruscito I, Gasparri ML, Marchetti C, Crispino S, La Russa C, Petriglia G, et al. Obesity is a key prognostic factor in stage IIIC-IV ovarian cancer diagnosed prior to 65 years of age: a 10-year survival analysis. *Gynecologic Oncology* 2016;**141**:204.

Rutten 2014 *{published data only}*

* Rutten MJ, Boldingh JH, Schuit E, Trum H, Van Driel W, Mol BW, et al. Development and internal validation of a prognostic model for survival after debulking surgery for epithelial ovarian cancer. *Gynecologic Oncology* 2014;**35**(1):13-8.

Rutten MJ, Boldingh JH, Schuit E, Trum H, Van Driel W, Mol BW, et al. Prognostic model for survival of epithelial ovarian cancer patients treated with primary or interval debulking surgery. *InInternational Journal Of Gynecological Cancer* 2013;**23**:8.

Salani 2007 *{published data only}*

Salani R, Zahurak ML, Santillan A, Giuntoli RL, Bristow RE. Survival impact of multiple bowel resections in patients undergoing primary cytoreductive surgery for advanced ovarian cancer: a case-control study. *Gynecologic Oncology* 2007;**107**(3):495-9.

Sessa 1991 *{published data only}*

Sessa C, Colombo N, Bolis G, Marsoni S, Mangioni C. Randomized comparison of hexamethylmelamine, adriamycin, cyclophosphamide (HAC) vs. cisplatin, adriamycin, cyclophosphamide (PAC) in advanced ovarian cancer: long-term results. *Cancer Treatment Reviews* 1991;**18**(18 Suppl A):37-46.

Shapiro 1998 *{published data only}*

Shapiro JD, Rothenberg ML, Sarosy GA, Steinberg SM, Adamo DO, Reed E, et al. Dose intensive combination platinum and cyclophosphamide in the treatment of patients with advanced untreated epithelial ovarian cancer. *Cancer* 1998;**83**(9):1980-8.

Shinozuka 1999 *{published data only}*

Shinozuka T, Miyamoto T, Muramatsu T, Hirasawa T, Murakami M, Makino T, et al. High dose chemotherapy with autologous stem cell support in the treatment of patients with ovarian carcinoma: long term results for 105 patients. *Cancer* 1999;**85**(7):1555-64.

Sioulas 2017 *{published data only}*

Sioulas VD, Schiavone MB, Kadouri D, Zivanovic O, Roche KL, O'Cearbhaill R, et al. Optimal primary management of bulky stage IIIC ovarian, fallopian tube and peritoneal carcinoma: are the only options complete gross resection at primary debulking surgery or neoadjuvant chemotherapy? *Gynecologic Oncology* 2017;**145**(1):15-20.

Skarlos 1996 *{published data only}*

Skarlos DV, Aravantinos G, Kosmidis P, Pavlidis N, Gennatas K, Beer M, et al. Carboplatin alone compared with its combination with epirubicin and cyclophosphamide in untreated advanced epithelial ovarian cancer: a Hellenic Co-operative Oncology Group study. *European Journal of Cancer* 1996;**32**(3):421-8.

Smits 2015 *{published data only}*

Smits A, Smits E, Lopes A, Das N, Hughes G, Talaat A, et al. Body mass index, physical activity and quality of life of ovarian cancer survivors: time to get moving? *Gynecologic Oncology* 2015;**139**(1):148-54.

Solmaz 2015 *{published data only}*

Solmaz U, Mat E, Dereli ML, Turan V, Peker N, Tosun G, et al. Does neoadjuvant chemotherapy plus cytoreductive surgery improve survival rates in patients with advanced epithelial ovarian cancer compared with cytoreductive surgery alone? *Journal of Balkan Union of Oncology* 2015;**20**(3):847-54.

Son 2017 *{published data only}*

Son JH, Kong TW, Paek J, Song KH, Chang SJ, Ryu HS. Clinical characteristics and prognostic inflection points among long-term survivors of advanced epithelial ovarian cancer. *International Journal of Gynaecology and Obstetrics* 2017;**193**(3):352-7.

Stewart 2015 *{published data only}*

Stewart JM, Tone AA, Bernardini MQ, Ferguson SE, Dodge J, Laframboise S, et al. Surgical factors do not impact survival in high-grade serous ovarian carcinoma patients treated with neoadjuvant chemotherapy. *Gynecologic Oncology* 2015;**137**(Suppl 1):43.

Stewart 2016 *{published data only}*

Stewart JM, Tone AA, Jiang H, Bernardini MQ, Ferguson S, Laframboise S, et al. The optimal time for surgery in women with serous ovarian cancer. *Canadian Journal of Surgery* 2016;**59**(4):223-32.

Strauss 1996 *{published data only}*

Strauss G, Lund B, Hansen M, Hansen OP, Hansen HH. Combined high-dose platinum and etoposide in previously untreated ovarian cancer patients with residual disease. *International Journal of Gynecological Cancer* 1996;**6**(5):410-4.

Suidan 2015 *{published data only}*

Suidan RS, Zhou Q, Iasonos A, O'Cearbhaill RE, Chi DS, Roche KC, et al. Prognostic significance of the number of postoperative intraperitoneal chemotherapy cycles for patients with advanced epithelial ovarian cancer. *International Journal of Gynecologic Cancer* 2015;**25**(4):599-606.

Sun 2000 *{published data only}*

Sun T, Feng Y, ZhuY, Zheng Y. Therapeutic strategy in the management of stage II - IV epithelial ovarian carcinoma. *Chinese Medical Journal* 2000;**113**(7):625-7.

Sutton 1989 *{published data only}*

Sutton GP, Stehman FB, Einhorn LH, Roth LM, Blessing JA, Ehrlich CE. Ten-year follow-up of patients receiving cisplatin, doxorubicin, and cyclophosphamide chemotherapy for advanced epithelial ovarian carcinoma. *Journal of Clinical Oncology* 1989;**7**:223-9.

Takano 2006 *{published data only}*

Takano MJ, Kikuchi Y, Yaegashi N, Suzuki M, Tsuda H, Sagae S, et al. Adjuvant chemotherapy with irinotecan hydrochloride and cisplatin for clear cell carcinoma of the ovary. *Oncology Reports* 2006;**16**(6):1301-6.

Takano 2007 *{published data only}*

Takano M, Sugiyama T, Yaegashi N, Suzuki M, Sagae S, Udagawa, et al. Progression-free survival and overall survival of patients with clear cell carcinoma of the ovary treated with paclitaxel-carboplatin or irinotecan-cisplatin: retrospective analysis. *International Journal of Clinical Oncology* 2007;**12**(4):256-60.

Tay 1996 *{published data only}*

Tay SK, Chang TC. An evaluation of the policy of routine treatment of advanced epithelial ovarian carcinoma by debulking surgery and combined platinum-cyclophosphamide chemotherapy in Singapore. *International Journal of Gynecological Cancer* 1996;**6**(1):44-8.

Taylor 1994 *{published data only}*

Taylor AE, Wiltshaw E, Gore ME, Fryatt I, Fisher C. Long-term follow-up of the first randomized study of cisplatin versus carboplatin for advanced epithelial ovarian cancer. *Journal Of Clinical Oncology* 1994;**12**(10):2066-70.

Tingulstad 2003 *{published data only}*

Tingulstad S, Skjeldestad FE, Hagen B. The effect of centralization of primary surgery on survival in ovarian cancer patients. *Obstetrics and Gynecology* 2003;**102**(3):499-505.

Todo 2003 *{published data only}*

Todo Y, Sakuragi N, Oikawa M, Negishi H, Yamamoto R, Yoshiaki K, et al. Cytoreductive surgery combined with organ resection for advanced ovarian carcinoma. *International Journal of Clinical Oncology* 2003;**8**(2):90-6.

Trhlík 2013 *{published data only}*

Trhlík M, Soumarová R, Bartoš P, Koláček J, Horová I, Těžká M. Neoadjuvant chemotherapy for primary advanced ovarian cancer. In: Biomedical Papers. Vol. 157. 2013:S98.

Uyar 2005 *{published data only}*

Uyar D, Frasure HE, Markman M, Von Gruenigen VE. Treatment patterns by decade of life in elderly women ([greater-than or equal to]70 years of age) with ovarian cancer. *Gynecologic Oncology* 2005;**98**(3):403-8.

Vallejos 1997 *{published data only}*

Vallejos C, Solidoro A, Castellano C, Barriga C, Galdos O, Casanova R, et al. Ifosfamide plus cisplatin as primary chemotherapy of advanced ovarian cancer. *Gynecologic Oncology* 1997;**67**(2):168-71.

Van Der Burg 1996 *{published data only}*

Van Der Burg MEL, Van Lent M, Buyse M, Kobierska A, Maggioni A, Favalli G, et al. The role of intervention debulking surgery in advanced epithelial ovarian cancer: an EORTC Gynecological Cancer Cooperative Group study. *International Journal of Gynecological Cancer* 1996;**6**:30-8.

Van Driel 2017 *{published data only}*

Van Driel W, Sikorska K, Schagen van Leeuwen J, Schreuder H, Hermans R, de Hingh I, et al. A phase 3 trial of hyperthermic intraperitoneal chemotherapy (HIPEC) for ovarian cancer. *Journal of Clinical Oncology* 2017;**35**(15 Suppl):5519.

van Vliet 2015 *{published data only}*

van Vliet MML, Schreuder HWR, Pasker-de Jong PCM, Duk MJ. Centralisation of epithelial ovarian cancer surgery: results on survival from a peripheral teaching hospital. *European Journal of Obstetrics, Gynecology, and Reproductive Biology* 2015;**192**:72-8.

Vergote 2010 *{published data only}*

Greimel E, Kristensen G, Vergote I, Hoskins P, van der Burg ME, Casado Herraez A, et al. Quality of life in advanced ovarian cancer patients: a randomized phase III study comparing upfront debulking surgery versus neo-adjuvant chemotherapy. *International Journal of Gynaecological Cancer* 2011;**21**:S620. [12185466]

Greimel E, Kristensen GB, van der Burg ME, Coronado P, Rustin G, del Rio AS, et al. Quality of life of advanced ovarian cancer patients in the randomized phase III study comparing primary debulking surgery versus neo-adjuvant chemotherapy. *Gynecologic Oncology* 2013;**131**(2):437-44. [12185467]

Vergote I, Amant F, Kristensen G, Ehlen T, Reed NS, Casado A. Primary surgery or neoadjuvant chemotherapy followed by interval debulking surgery in advanced ovarian cancer. *European Journal of Cancer* 2011;**47**(Suppl 3):S88-91. [12185468]

Vergote I, Coens C, Nankivell M, Kristensen GB, Parmar MK, Ehlen T, et al. Neoadjuvant chemotherapy versus debulking surgery in advanced tubo-ovarian cancers: pooled analysis of individual patient data from the EORTC 55971 and CHORUS trials. *Lancet Oncology* 2018;**19**:1680-7. [12185469]

Vergote I, Pecorelli S, Stuart G. Intergroup Study (EORTC 55971/NCIC OV13). Randomized phase III study comparing upfront debulking surgery versus neo-adjuvant chemotherapy in patients with stage IIIc or IV epithelial ovarian carcinoma. Trial

protocol: http://www.cancer.gov/clinicaltrials/EORTC-55971 2003 (accessed 17 June 2012). [12185470]

Vergote I, Tropé CG, Amant F, Ehlen T, Reed NS, Casado A. Neoadjuvant chemotherapy is the better treatment option in some patients with stage IIIc to IV ovarian cancer. *Journal of Clinical Oncology* 2011;**29**(31):4076-8. [12185471]

Vergote I, Trope CG, Amant F, Kristensen GB, Ehlen T, Johnson N, et al. Neoadjuvant chemotherapy or primary surgery in stage IIIC or IV ovarian cancer. *New England Journal of Medicine* 2010;**363**(10):943-53. [12185472] [Incl. Supplementary Appendix and Protocol]

Verleye L, Ottevanger PB, Kristensen GB, Ehlen T, Johnson N, van der Burg ME, et al. Quality of pathology reports for advanced ovarian cancer: are we missing essential information? An audit of 479 pathology reports from the EORTC-GCG 55971/NCIC-CTG OV13 neoadjuvant trial. *European Journal of Cancer* 2011;**47**(1):57-64. [12185473]

Vergote 2018 *{published data only}*

Vergote I, Coens C, Nankivell M, Kristensen G, Parmer M, Ehlen T, et al. Meta-analysis of the randomized EORTC and chorus neoadjuvant versus primary debulking trials in advanced tubo-ovarian cancer. *International Journal of Gynecological Cancer* 2016;**26**(Suppl 3):24-5.

Vergote I, Coens C, Nankivell M, Kristensen GB, Parmar MK, Ehlen T, et al. Neoadjuvant chemotherapy versus debulking surgery in advanced tubo-ovarian cancers: pooled analysis of individual patient data from the EORTC 55971 and CHORUS trials. *Lancet Oncology* 2018;**19**(12):1680-7.

* Vergote I, Coens C, Nankivell M, Kristensen GB, Parmar MK, Ehlen T, et al. Neoadjuvant chemotherapy versus debulking surgery in advanced tubo-ovarian cancers: pooled analysis of individual patient data from the EORTC 55971 and CHORUS trials. *Lancet Oncology* 2018;**19**(12):1680-7.

Vidal 2016 *{published data only}*

Vidal F, Guerby P, Luyckx M, Haddad P, Stoeckle E, Morice P, et al. Are early relapses in advanced-stage ovarian cancer doomed to a poor prognosis? *PlOS One* 2016;**11**(1):e0147787. [DOI: 10.1371/journal.pone.0147787]

Wadler 1996 *{published data only}*

Wadler S, Yeap B, Vogl S, Carbone P. Randomized trial of initial therapy with melphalan versus cisplatin-based combination chemotherapy in patients with advanced ovarian carcinoma: initial and long term results - Eastern Cooperative Oncology Group study E2878. *Cancer* 1996;**77**(4):733-42.

Wallace 2017 *{published data only}*

Wallace S, Kumar A, McGree M, Weaver A, Mariani A, Langstraat C, et al. Efforts at maximal cytoreduction improve survival in ovarian cancer patients, even when complete gross resection is not feasible. *Gynecologic Oncology* 2017;**145**(1):21-6.

Warwick 1995 *{published data only}*

Warwick J, Kehoe S, Earl H, Luesley D, Redman C, Chan KK. Long-term follow-up of patients with advanced ovarian cancer treated in randomised clinical trials. *British Journal of Cancer* 1995;**72**(6):1513-7.

Willemse 1992 *{published data only}*

Willemse PHB, De Vries EGE, Kloppenburg M, Fontein DL, Aalders JG, Boonstra H, et al. Carboplatin with cyclophosphamide in patients with advanced ovarian cancer: an efficacy and quality-adjusted survival analysis. *International Journal of Gynecological Cancer* 1992;**2**(5):236-43.

Wils 1990 *{published data only}*

Wils JA. Long-term follow-up of patients with advanced ovarian carcinoma treated with debulking surgery and chemotherapy consisting of cisplatin, doxorubicin, and cyclophosphamide. Gynecologic Oncology Group of the Comprehensive Cancer Center. *Oncology* 1990;**47**(2):115-20.

Wimberger 2007 *{published data only}*

Wimberger P, Lehmann N, Kimmig R, Burges A, Meier W, Du Bois A. Prognostic factors for complete debulking in advanced ovarian cancer and its impact on survival. An exploratory analysis of a prospectively randomized phase III study of the Arbeitsgemeinschaft Gynaekologische Onkologie Ovarian Cancer Study Group (AGO-OVAR). *Gynecologic Oncology* 2007;**106**(1):69-74.

Yamamoto 2007 *{published data only}*

Yamamoto S, Tsuda H, Yoshikawa T, Kudoh K, Kita T, Furuya K, et al. Clear cell adenocarcinoma associated with clear cell adenofibromatous components: a subgroup of ovarian clear cell adenocarcinoma with distinct clinicopathologic characteristics. *American Journal of Surgical Pathology* 2007;**31**(7):999-1006.

Zang 1999 *{published data only}*

Zang RY, Zhang ZY, Cai SM, Li ZT, Chen J, Tang MQ, et al. Cytoreductive surgery for stage IV epithelial ovarian cancer. *Journal of Experimental & Clinical Cancer Research* 1999;**18**(4):449-54.

Zhang 2015 *{published data only}*

Zhang LY, Li PL, Wang TZ, Zhang XC. Prognostic values of 5-hmC, 5-mC and TET2 in epithelial ovarian cancer. *Archives of Gynecology and Obstetrics* 2015;**292**(4):891-7.

Additional references

Aldin 2020

Aldin A, Umlauff L, Estcourt LJ, Collins G, Moons KG, Engert A, et al. Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies. *Cochrane Database of Systematic Reviews* 2020, Issue 1. Art. No: CD012643. [DOI: 10.1002/14651858.CD012643.pub3]

Aletti 2007

Aletti GD, Dowdy SC, Podratz KC, Cliby WA. Relationship among surgical complexity, short-term morbidity, and overall survival in primary surgery for advanced ovarian cancer. *American Journal of Obstetrics and Gynecology* 2007;**187**(6):676.

Armstrong 2006

Armstrong DK, Bundy B, Wenzel L, Huang HQ, Baergen R, Lele S, et al. Intraperitoneal cisplatin and paclitaxel in ovarian cancer. *New England Journal of Medicine* 2006;**354**(1):34-43.

Berek 2018

Berek JS, Kehoe ST, Kumar L, Friedlander M. Cancer of the ovary, fallopian tube, and peritoneum. *International Journal of Gynaecology and Obstetrics* 2018;**143**:59-78.

BGCS 2017

BGCS. https://www.bgcs.org.uk/wp-content/uploads/2019/05/BGCS-Guidelines-Ovarian-Guidelines-2017.pdf 2017.

Bookman 2009

Bookman MA, Brady MF, McGuire WP, Harper PG, Alberts DS, Friedlander M, et al. Evaluation of new platinum-based treatment regimens in advanced-stage ovarian cancer: a phase III trial of the Gynecologic Cancer Intergroup. *Journal of Clinical Oncology* 2009;**27**(9):1419-25.

Borley 2012

Borley J, Wilhelm-Benartzi C, Brown R, Ghaem-Maghami S. Does tumour biology determine surgical success in the treatment of epithelial ovarian cancer? A systematic literature review. *British Journal of Cancer* 2012;**107**(7):1069-74.

Bristow 2002

Bristow RE, Tomacruz RS, Armstrong DK, Trimble EL, Montz FJ. Survival effect of maximal cytoreductive surgery for advanced ovarian carcinoma during the platinum era: a meta-analysis. *Journal of Clinical Oncology* 2002;**20**(5):1248-59.

Bristow 2009

Bristow RE, Zahurak ML, Diaz-Montes TP, Giuntoli RL, Armstrong DK. Impact of surgeon and hospital ovarian cancer surgical case volume on in hospital mortality and related short-term outcomes. *Gynecologic Oncology* 2009;**115**(3):334-8.

Chang 2013

Chang SJ, Hodeib M, Chang J, Bristow RE. Survival impact of complete cytoreduction to no gross residual disease for advanced-stage ovarian cancer: a meta-analysis. *Gynecologic Oncology* 2013;**130**(3):493–8.

Coleridge 2021

Coleridge SL, Bryant A, Kehoe S, Morrison J. Neoadjuvant chemotherapy before surgery versus surgery followed by chemotherapy for initial treatment in advanced ovarian epithelial cancer. *Cochrane Database of Systematic Reviews* 2021, Issue 7. Art. No: CD005343. [DOI: 10.1002/14651858.CD005343]

Colombo 2019

Colombo N, Sessa C, du Bois A, Ledermann J, McCluggage WG, McNeish I, et al. ESMO-ESGO consensus conference recommendations on ovarian cancer: pathology and molecular biology, early and advanced stages, borderline tumours and recurrent disease. *Annals of Oncology* 2019;**30**(5):672–705.

Covens 2000

Covens AL. A critique of surgical cytoreduction in advanced ovarian cancer. *Gynecologic Oncology* 2000;**78**:269-74.

Debray 2018

Debray TPA, Moons KGM, Riley RD. Detecting small-study effects and funnel plot asymmetry in meta-analysis of survival data: a comparison of new and existing tests. *Research Synthesis Methods* 2018;**9**(1):41-50.

Deeks 2001

Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, Altman DG, editors(s). Systematic Reviews in Health Care: Meta-Analysis in Context (2nd edition). London: BMJ Publication Group, 2001.

DerSimonian 1986

DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986;**7**:177-88.

du Bois 2009

du Bois A, Reuss A, Pujade-Lauraine E, Harter P, Ray-Coquard I, Pfisterer J. Role of surgical outcome as prognostic factor in advanced epithelial ovarian cancer: a combined exploratory analysis of 3 prospectively randomized phase 3 multicenter trials: by the Arbeitsgemeinschaft Gynaekologische Onkologie Studiengruppe Ovarialkarzinom (AGO-OVAR) and the Groupe d'Investigateurs Nationaux Pour les Etudes des Cancers de l'Ovaire (GINECO). *Cancer* 2009;**115**(6):1234–44.

Elzarkaa 2018

Elzarkaa AA, Shaalan W, Elemam D, Mansour H, Melis M, Malik E, et al. Peritoneal cancer index as a predictor of survival in advanced stage serous epithelial ovarian cancer: a prospective study. *Journal of Gynecologic Oncology* 2018;**29**(4):e47.

EUROCARE 2015

Sant M, Chirlaque Lopez MD, Agresti R, Sánchez Pérez MJ, Holleczek B, Bielska Lasota M, et al, EUROCARE-5 Working Group. Survival of women with cancers of breast and genital organs in Europe 1999-2007: results of the EUROCARE-5 study. *European Journal of Cancer* 2015;**51**(15):2191-205.

Fader 2007

Fader AN, Rose PG. Role of surgery in ovarian carcinoma. *Journal of Clinical Oncology* 2007;**25**(20):2873-83.

Falconer 2020

Falconer H, Joneborg U, Krawiec K, Palsdottir K, Bottai M, Salehi S. Ultra-radical upfront surgery does not improve survival in women with advanced epithelial ovarian cancer; a natural experiment in a complete population. *Gynecologic Oncology* 2020;**159**(1):58-65.

Ferron 2019

Ferron G, De Rauglaudre G, Chevalier A, Combe P, Joly F, Lortholary A, et al. Impact of adding nintedanib to neoadjuvant chemotherapy (NACT) for advanced epithelial ovarian cancer (EOC) patients: the CHIVA double-blind randomized phase II GINECO study. *Journal of Clinical Oncology* 2019;**37**(15 Suppl):5512.

Foroutan 2020

Foroutan F, Guyatt G, Zuk V, Vandvik PO, Alba AC, Mustafa R, et al. GRADE Guidelines 28: Use of GRADE for the assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks. *Journal of Clinical Epidemiology* 2020;**121**:62-70. [DOI: 10.1016/j.jclinepi.2019.12.023; PubMed: 31982539]

Ghirardi 2020

Ghirardi V, Moruzzi MC, Bizzarri N, Vargiu V, D'Indinosante M, Garganese G, et al. Minimal residual disease at primary debulking surgery versus complete tumor resection at interval debulking surgery in advanced epithelial ovarian cancer: a survival analysis. *Gynecologic Oncology* 2020;**157**(1):209-13.

Girling 1996

Girling JC, Soutter WP. Cytoreductive surgery in the primary management of advanced epithelial ovarian carcinoma: a topic for debate. *International Journal of Gynecological Cancer* 1996;**1**:81-4.

GLOBOCAN 2018

(GLOBOCAN 2018) Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. Available from: http://gco.iarc.fr/today/home 2018;**68**(394-424). [DOI: 10.3322/caac.21492]

Grabowski 2016

Grabowski JP, Harter P, Heitz F, Pujade-Lauraine E, Reuss A, Kristensen G, et al. Operability and chemotherapy responsiveness in advanced low-grade serous ovarian cancer. An analysis of the AGO Study Group metadatabase. *Gynecologic Oncology* 2016;**140**(3):457-62.

GRADE Working Group

GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;**328**:1490-4.

Griffiths 1975

Griffiths. Surgical resection of tumor bulk in the primary treatment of ovarian carcinoma. *Journal of the National Cancer Institute Monographs* 1975;**42**:1014.

Guyatt 2008

Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;**336**(7650):924-6.

Higgins 2003

Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**:557-60.

Hoskins 1994

Hoskins WJ, McGuire WP, Brady MF, Homesley HD, Creasman WT, Berman M, et al. The effect of diameter of largest residual disease on survival after primary cytoreductive surgery in patients with suboptimal residual epithelial ovarian carcinoma: a Gynecologic Oncology Group study. *American Journal of Obstetrics and Gynecology* 1994;**170**:974-80.

Hui 2022

Hui S, Bryant A, Kunonga P, Gajjar K, Naik R. Ultra-radical (extensive) surgery versus standard surgery for the primary cytoreduction of advanced epithelial ovarian cancer. *Cochrane Database of Systematic Reviews* 2022, Issue In Press.

Hunter 1992

Hunter RW, Alexander ND, Soutter WP. Meta-analysis of surgery in advanced ovarian carcinoma: Is maximum cytoreductive surgery an independent determinant of prognosis? *American Journal of Obstetrics and Gynecology* 1992;**166**:504-11.

Jemal 2017

Jemal A, Ward E, Johnson C, Cronin K, Ma J, Ryerson B, et al. Annual Report to the Nation on the Status of Cancer, 1975–2014, Featuring Survival Cancer Statistics. *Journal of the National Cancer Institute* 2017;**109**:https://doi.org/10.1093/jnci/djx030.

Kobal 2018

Kobal B, Noventa M, Cvjeticanin B, Barbic M, Meglic L, Herzog M, et al. Primary debulking surgery versus primary neoadjuvant chemotherapy for high grade advanced stage ovarian cancer: comparison of survivals. *Radiology and Oncology* 2018;**52**(3):307-19.

Kumar 2017

Kumar A, Langstraat CL, DeJong SR, McGree ME, Bakkum-Gamez JN, Weaver AL, et al. Functional not chronologic age: frailty index predicts outcomes in advanced ovarian cancer. *Gynecologic Oncology* 2017;**147**(1):104-9.

Kurman 2010

Kurman RJ, Shih IeM. The origin and pathogenesis of epithelial ovarian cancer: a proposed unifying theory. *American Journal of Surgical Pathology* 2010;**34**(3):433-43.

Kurman 2011

Kurman RJ, Shih IeM. Molecular pathogenesis and extra ovarian origin of epithelial ovarian cancer-shifting the paradigm. *Human Pathology* 2011;**42**(7):918-31.

Kurman 2014

Kurman RJ, Carcangiu ML, Herrington CS. WHO Classification of Tumours of Female Reproductive Organs. 4 edition. Lyon: WHO Press, 2014.

Markar 2016

Makar AP, Tropé CG, Tummers P, Denys H, Vandecasteele K. Advanced ovarian cancer: primary or interval debulking? Five

categories of patients in view of the results of randomized trials and tumor biology: Primary ebulking surgery and interval debulking surgery for advanced ovarian cancer. *Oncologist* 2016;**21**(6):745-54.

Markman 2001

Markman M, Bundy BN, Alberts DS, Fowler JM, Clark-Pearson DL, Carson LF, et al. Phase III trial of standard-dose intravenous cisplatin plus paclitaxel versus moderately high-dose carboplatin followed by intravenous paclitaxel and intraperitoneal cisplatin in small-volume stage III ovarian carcinoma: an Intergroup Study of the Gynecologic Oncology Group, Southwestern Oncology Group, and Eastern Cooperative Oncology Group. *Journal of Clinical Oncology* 2001;**19**(4):1001-7.

Moons 2014

Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLOS Medicine* 2014;**11**(10):e1001744.

Muggia 2000

Muggia FM, Braly PS, Brady MF, Sutton GN, Theodore HL, Samuel AL, et al. Phase III randomized study of cisplatin versus paclitaxel versus cisplatin and paclitaxel in patients with suboptimal stage III or IV ovarian cancer: a Gynecologic Oncology Group Study. *Journal of Clinical Oncology* 2000;**18**(1):106-15.

Napoletano 2010

Napoletano C, Bellati F, Landi R, Pauselli S, Marchetti C, Visconti V, et al. Ovarian cancer cytoreduction induces changes in T cell population subsets reducing immunosuppression. *Journal of Cellular and Molecular Medicine* 2010;**14**(12):2748-59.

National Comprehensive Cancer Network 2020

National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology, ovarian cancer including fallopian tube cancer and primary peritoneal cancer. http://www.nccn.org/professionals/physician_gls/pdf/ovarian.pdf 2020.

NICE 2013

National Institute for Health and Care Excellence. Ultra-radical (extensive) surgery for advanced ovarian cancer. NICE Interventional procedure guidance [IPG 470]. https://www.nice.org.uk/guidance/ipg470 2013.

Ogundimu 2016

Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. *Journal of Clinical Epidemiology* 2016;**76**:175-82. [DOI: 10.1016/j.jclinepi.2016.02.031]

Ozols 2003

Ozols RF, Bundy BN, Greer BE, Fowler JM, Clarke-Pearson DB, Mannel RA, et al. Phase III trial of carboplatin and paclitaxel compared with cisplatin and paclitaxel in patients with optimally resected stage III ovarian cancer: a Gynecologic Oncology Group Study. *Journal of Clinical Oncology* 2003;**21**(17):3194-200.

Parmar 1998

Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Statistics in Medicine* 1998;**17**(24):2815-34.

Perets 2016

Perets R, Drapkin R. It's totally tubular. Riding the new wave of ovarian cancer research. *Cancer Research* 2016;**76**(1):10-7.

Redman 1994

Redman CW, Warwick J, Luesley DM, Varma R, Lawton FG, Blackledge GR. Intervention debulking surgery in advanced epithelial ovarian cancer. *British Journal of Obstetrics and Gynaecology* 1994;**101**:142-6.

Reuss 2019

Reuss A, du Bois A, Harter P, Fotopoulou C, Sehouli J, Aletti G, et al. TRUST: trial of radical upfront surgical therapy in advanced ovarian cancer (ENGOT ov33/AGO-OVAR OP7). *International Journal of Gynecologic Cancer* 2019;**29**:1327-31.

Review Manager 2020 [Computer program]

The Cochrane Collaboration Review Manager 5 (RevMan 5). Version 5.4. Copenhagen: The Cochrane Collaboration, 2020.

Riley 2019

Riley RD, Moons KGM, Snell KIE, Ensor J, Hooft L, Altman DG, et al. A guide to systematic review and meta-analysis of prognostic factor studies. *BMJ* 2019;**364**:k4597.

Shafi 2018

Shafi M, Bolton H, Gajjar K, editor(s). Gynaecological Oncology for the MRCOG. Cambridge: Cambridge University Press, 2018. [DOI: 10.1017/9781316986844]

Spriggs 2007

Spriggs DR, Brady M, Vaccarello L, Clarke-Pearson DL, Burger RA, Mannel R, et al. Phase III randomized trial of intravenous cisplatin plus a 24- or 96-hour infusion of paclitaxel in epithelial ovarian cancer: a Gynecologic Oncology Group Study. *Journal of Clinical Oncology* 2007;**25**(28):4466-71.

Sterne 2011

Sterne JAC, Egger M, Moher D (editors). Chapter 10: Addressing reporting biases. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from training.cochrane.org/handbook.

Stuart 2011

Stuart GC, Kitchener H, Bacon M, duBois A, Friedlander M, Ledermann J, et al. Gynecologic Cancer InterGroup (GCIG) consensus statement on clinical trials in ovarian cancer: report from the Fourth Ovarian Cancer Consensus Conference. *International Journal of Gynecological Cancer* 2011;**21**(4):750–5.

Sundar 2022

Sundar S, Cummins C, Kumar S, Long J, Arora V, Balega J, et al. Quality of life from cytoreductive surgery in advanced ovarian cancer: Investigating the association between disease burden and surgical complexity in the international,

prospective, SOCQER-2 cohort study. *British Journal of Obstetrics and Gynaecology* 2022;**129**(7):1122-32. [DOI: 10.1111/1471-0528.17041] [PMCID: PMC9306902] [PMID: 34865316]

Sørensen 2019

Sørensen SM, Schnack TH, Høgdall C. Impact of residual disease on overall survival in women with federation of gynecology and obstetrics stage IIIB-IIIC vs stage IV epithelial ovarian cancer after primary surgery. *Acta Obstetricia et Gynecologica Scandinavica* 2019;**98**(1):34–43.

Tangjitgamol 2016

Tangjitgamol S, Manusirivithaya S, Laopaiboon M, Lumbiganon P, Bryant A. Interval debulking surgery for advanced epithelial ovarian cancer. *Cochrane Database of Systematic Reviews* 2016, Issue 1. Art. No: CD006014. [DOI: 10.1002/14651858.CD006014]

Torre 2018

Torre LA, Trabert B, DeSantis CE, Miller K, Samimi G, Runowicz C, et al. Ovarian cancer statistics. *CA: a Cancer Journal for Clinicians* 2018;**68**(4):284–96.

Van der Burg 1995

van der Burg ME, van Lent M, Buyse M, Kobierska A, Colombo N, Favalli G, et al. The effect of debulking surgery after induction chemotherapy on the prognosis in advanced epithelial ovarian cancer. Gynecological cancer co-operative group of the European organization for research and treatment of cancer. *New England Journal of Medicine* 1995;**332**(10):629-34.

Vergote 2013

Vergote I, du Bois A, Amant F, Heitz F, Leunen K, Harter P. Neoadjuvant chemotherapy in advanced ovarian cancer: on what do we agree and disagree? *Gynecologic Oncology* 2013;**128**:6-11.

Vergote 2016

Vergote I, Vlayen J, Heus P, Hoogendam JP, Damen JA, Van de Wetering FT, et al. Ovarian cancer: diagnosis, treatment and follow-up; KCE Report 28Cs; 2016. Available at: kce.fgov.be/sites/default/files/atoms/files/KCE_268Cs_Ovarian_cancer_summary.pdf 2016.

Webb 2017

Webb PM, Jordan SJ. Epidemiology of epithelial ovarian cancer. *Clinical Obstetrics & Gynaecology* 2017;**41**:3-14.

Whiting 2011

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks, JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529-36.

Wolff 2019

Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine* 2019;**170**(1):51-8.

Woo 2012

Woo YL, Kyrgiou M, Bryant A, Everett T, Dickinson HO. Centralisation of services for gynaecological cancer. *Cochrane Database of Systematic Reviews* 2012, Issue 3. Art. No: CD007945. [DOI: 10.1002/14651858.CD007945.pub2]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Akahira 2001

Study characteristics	
Methods	Multicentre retrospective analysis: 24 Japanese institutions received questionnaires regarding stage IV epithelial ovarian cancer women
Participants	225 women with stage IV ovarian cancer whose disease had been confirmed by exploration and only women with complete medical records were included. Stage IV disease was defined according to FIGO. Only women who underwent an initial attempt at surgical debulking were analysed. The median age in the study was 54 years (range: 26 to 85 years) All 225 women had FIGO stage IV disease Histological cell type: serous: 136 (60.5%), mucinous: 16 (7%), clear cell 26 (11.5%), endometrioid 27 (12%), transitional 4 (2%), undifferentiated 12 (5%), other 4 (2%) Extent of disease: pleural effusion: 89 (39.5%), liver: 34 (15%), lung: 8 (3.5%), lymph node: 44 (19.5%), other: 15 (6.5%), multiple sites: 35 (15%)

Akahira 2001 *(Continued)*

	Performance status: 0: 26 (11%), 1: 76 (34%), 2: 49 (22%), 3: 67 (30%), 4: 7 (3%)
Residual disease details	**Intervention group:** **'Optimal' cytoreduction** was defined as no gross residual tumour greater than 2 cm in diameter **Comparison group:** **LVRD** was defined as any gross residual disease remaining greater than 2 cm in diameter
Outcomes	Overall survival: HR adjusted for histology and performance status: • < 2 cm versus > 2 cm; HR 0.42 (95% CI 0.31 to 0.64), or > 2 cm vs < 2c m; HR 2.39 (95% CI 1.68 to 3.40) so that reference group is consistent throughout review Adverse events; median blood loss, blood transfusions
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): unclear risk Valid and reliable measurement of RD. Multicentre design may introduce heterogeneity in measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): high risk HR for OS was adjusted for histology, performance status and RD in multivariable Cox model 6. Statistical analysis and reporting (a-d): unclear risk In methods, authors reported that significant variables from the univariate analysis were included in the multivariable model **Outcome: progression-free survival** Not reported
Notes	There were 70 women (31.1%) in the 'optimal' group and 155 (68.9%) in the LVRD group The median follow-up time was 47.5 months (range: 13 to 112 months) The median survival for all women with stage IV ovarian cancer was 20 months, with an estimated 5-year survival rate of 19.6% Mean survival in the optimal group was 32 months and 16 months in the suboptimal group (P < 0.0001) MV analysis included the histology and performance status as covariates in the model The median duration of the debulking surgery was 240 minutes (range: 40 to 780 minutes.

Akahira 2001 *(Continued)*

	The median estimated blood loss was 1085 mL (range 40 to 11,000 mL), and 112 women (50%) received blood transfusions intra- and postoperatively

Aletti 2006

Study characteristics	
Methods	Retrospective cohort study of consecutive women identified from surgical records
Participants	Women with FIGO stage IIIC ovarian cancer, where disease status was extracted from surgical exploration notes The mean and median age at study entry was 64.4 and 64 years respectively (range: 24 to 87) All women presented with FIGO stage IIIC - 194 (100%) Tumour cell type: serous 126 (64.9%), mucinous: 4 (2.1%), endometrioid: 18 (9.3%), clear cell: 7 (3.6%), mixed: 17 (8.8%), seroanaplastic: 17 (8.8%), mullerian origin: 2 (1%) Tumour grade: 1: 1 (0.5%), 2: 13 (6.7%), 3: 180 (92.8%) ASA score: 1: 7 (3.6%), 2: 87 (44.8%), 3: 88 (45.4%), 4: 7 (3.6%), unknown: 5 (2.6%) Ascites: mean: 2076 mL, median 1000 mL, (range: 0 to 12,000 mL) Extent of disease: carcinomatosis: 144 (74.2%), diaphragm involvement: 137 (70.6%), mesentery: 138 (71.1%), cul-de-sac: 163 (84%), omentum 168: (86.6%), ascites 160: (82.5%)
Residual disease details	Residual disease was noted as follows: 1. NMRD: 46 (23.7%) 2. SVRD: 85 (43.8%) 3. Residual disease of 1 cm to 2 cm: 22 (11.3%) 4. Residual disease larger than 2cm: 41 (21.1%) Optimal cytoreduction was defined as residual disease < 1 cm All women were scheduled for treatment with first-line postoperative platinum-based chemotherapy (paclitaxel or cyclophosphamide for 6 to 8 courses, every 3 to 4 weeks)

Outcomes	• Overall survival, HR adjusted for several prognostic categories:
	  ○ SVRD vs NMRD: HR 3.89 (95% CI 2.27 to 7.11)
	  ○ 1 cm to 2 cm vs NMRD: HR 6.25 (95% CI 3.16 to 12.61)
	  ○ > 2 cm vs NMRD: HR 13.00 (95% CI 7.14 to 24.87)
	• Adverse events:
	  ○ Perioperative mortality rate, defined as the percentage of women who died within 30 days of surgery, was 1.5% (3/194; 95% CI 0.5 to 4.4%). However, there was no breakdown by treatment arm.

Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk

Aletti 2006 *(Continued)*

	Valid and reliable measurement of RD
	Outcome level assessment:
	Outcome: overall survival
	4. Outcome measurement (a-c): high risk
	Overall survival not used as outcome. Rather, disease-specific survival was used.
	5. Adjustment for other prognostic factors (a-g): low risk
	HR for disease-specific survival was adjusted for residual disease, age, American Society of Anesthesiology (ASA) score, histological grade, operative time and aggressive surgery in multivariable Cox model
	6. Statistical analysis and reporting (a-d): unclear risk
	In methods, authors reported that significant variables from the univariate analysis were included in the multivariable model
	Outcome: progression-free survival
	Not reported
Notes	Median length of follow-up: 2.7 years
	Mean length of follow-up: 3.5 years (range 0.02 to 10.5 years)
	5-year disease-specific death rate:
	Optimal group: 70/131 (53.4%)
	Suboptimal group: 56/63 (88.9%)

Ataseven 2016

Study characteristics	
Methods	Prospective cohort study
Participants	326 consecutive women with FIGO IV
	Median age in the study was 61 years (range: 19 to 88 years)
	All 326 women presented with FIGO stage IV disease
	Histological cell type: high grade serous: 287 (88.0%), others: 39 (12.0%)
	Ascites: ≤ 500 mL: 149 (45.7%), > 500 mL: 177 (54.3%)
	Performance status: ECOG 0: 248 (76.1%), ECOG > 0: 78 (23.9%)
	Localization of metastasis: • Pleural effusion/involvement: 134 (41.1%) • Abdominal wall: 133 (40.8%) • Extraregional lymph node: 63 (19.3%) • Liver: 45 (13.8%) • Spleen: 22 (6.7%) • Others: 19 (5.8%)

Ataseven 2016 *(Continued)*

	Germany
Residual disease details	Surgery was performed by accredited gynaecological oncologists Cohort 1 included 286 women who underwent primary debulking surgery Postoperative chemotherapy was administered in 92% (263/286) Cohort 2 included 40 women who underwent either no surgery or only diagnostic procedures without cytoreductive intention (NoCS - no cytoreductive surgery) In cohort 2, platinum-based chemotherapy was given to 87.5% (35/40) of women Residual disease for total cohort was noted as follows, n (%): • NMRD: 157 (48.2%) • SVRD: 88 (27.0%) • LVRD (> 10 mm): 41 (12.6%) • No cytoreduction: 40 (12.3%)
Outcomes	Overall survival: HR adjusted for age, performance status, residual tumour, tumour stage and ascites NMRD: HR 1 SVRD: HR 1.50 (95% CI 1.01 to 2.23) LVRD (> 10 mm): HR 2.20 (95% CI 1.36 to 3.55)
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): low risk HR for OS was adjusted for age, performance status, residual tumour, tumour stage and ascites in a multivariable Cox model 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria in multivariate model **Outcome: progression-free survival** Not reported
Notes	Follow-up time: up to 4 years (mean: 46 months; median: 34 months; interquartile range: 12 to 70 months)

Ataseven 2016 *(Continued)*

	In total, 28 women (8.6%) did not receive chemotherapy 30-day mortality was observed in: 12/326 (3.68%) Median OS for all women was 50.3 months In cohort 1, complete resection was achieved in 54.9% (n = 157/286; RD0), cytoreduction to 1 mm to 10 mm in 30.7% (n = 88/286; RD1-10) and bulky residual disease exceeding 10 mm in 14.3% (n = 41/286; RD > 10) Risk factors for residual disease after debulking surgery in women with EOC FIGO stage IV included: • Age (OR 1.85, 95% CI 1.13 to 3.03; P = 0.015) • Poor performance status (OR 3.46, 95% CI 1.67 to 7.18; P = 0.001) • Large volume ascites > 500 mL (OR 2.37, 95% CI 1.06 to 4.22; P = 0.035) • Presence of liver metastasis (OR 6.17, 95% CI 2.78 to 13.7; P < 0.001) Length of hospital stay not reported

Bixel 2020

Study characteristics	
Methods	Retrospective analysis of past medical data from The Ohio State University Wexner Medical Center and Duke University Health System between January 2004 and April 2017 Multicentre study USA
Participants	134 patients diagnosed with stage III to IV ovarian, fallopian tube or primary peritoneal cancer Median age (range): 64.3 (21 to 87) Median BMI (range): 28.1 (16 to 52.5) Ethnicity: 110 white (82%) FIGO III: 49 (36%) FIGO IV: 54 (40%) FIGO stage not otherwise specified but considered advanced: 31 (24%) Serous histology: 112 (83%) Tumour grade 1: 3 (2%) Tumour grade 2: 123 (92%) Tumour grade unknown: 8 (6%)
Residual disease details	Women underwent interval debulking surgery Optimal RD defined as RD ≤ 1 cm NMRD: 89 (66%) SVRD: 45 (34%)
Outcomes	Overall survival Median OS: 35.3 (95% CI 28.6 to 42.9) There was no multivariate model for overall survival despite there being progression-free survival Progression-free survival Disease recurrence: 117 (87%) Median PFS: 12.2 (95% CI 11.3 to 13.7)

Bixel 2020 *(Continued)*

	After controlling for NACT cycles, route of postoperative chemotherapy administration (intraperitoneal or intravenous), maintenance therapy (yes/no); residual disease (SVRD vs NMRD) (adjusted HR 1.564 (1.055 to 2.287)) 2 (1%) patients died during treatment: 1 patient in the IP group died from a myocardial infarction and 1 patient in the IV group died as a result of sepsis with resulting complications
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): unclear risk Adequate cut-off for residual disease used (< 1 cm). Multicentre design may introduce heterogeneity in measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of OS 5. Adjustment for other prognostic factors (a-g): high risk OS was reported in KM curve but was not used in any multivariable modelling 6. Statistical analysis and reporting (a-d): high risk There was only a multivariate model for PFS but not OS **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of PFS 5. Adjustment for other prognostic factors (a-g): high risk Model for PFS adjusted for NACT cycles, route of administration (IP or IV), maintenance therapy. However, none deemed to be critically important prognostic factors. 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; data driven based on P values of univariate associations. Unclear if multivariate Cox was used as logistic regression mentioned in methods but hazard ratios reported. There was only a multivariate model for PFS but not OS.
Notes	37 (28%) patients receiving IP and 97 (72%) patients receiving IV chemotherapy Median NACT cycles: 3 (range 1 to 6) NACT regime Platinum/taxane: 133 (99%) Platinum/other: 1 (1%) Adjuvant chemotherapy regime

Bixel 2020 *(Continued)*

	Platinum/taxane: 122 (91%) Platinum/other: 3 (2%) Non platinum: 9 (7%) Adjuvant chemotherapy cycles: Intraperitoneal group: median 4 (range 2 to 6) Intravenous group: median 3 (range 1 to 6) Maintenance therapy following completion of planned chemotherapy: 10 (7%) At the time of surgery, 32 (24%) patients underwent a bowel resection and 15 (11%) underwent extensive upper abdominal debulking procedures

Bristow 2011

Study characteristics	
Methods	Retrospective chart review at Johns Hopkins Hospital, USA Women enrolment was between January 1995 and December 2008
Participants	405 women with FIGO stage IIIC epithelial ovarian cancer based on intraoperative findings or radiographic imaging coupled with fine-needle biopsy diagnosis. All epithelial histological subtypes were included. Borderline ovarian tumours of low malignant potential were excluded. Women characteristics reported as Whites (n = 366) vs African-Americans (n = 39) Median age: 59 vs 59 years ASA class, I/II/III/IV: 5/124/232/5 vs 0/4/31/4 Histology, serous/non-serous: 314/52 vs 31/8 Tumour grade, 1/2/3: 39/33/294 vs 2/4/33 Optimal RD (defined as ≤ 1 cm)/no gross RD: 267/188 vs 18/21
Residual disease details	All women underwent attempted surgical cytoreduction either primarily Residual disease was defined as: SVRD (RD 0.1 cm to 1.0 cm) NMRD (no gross RD) Residual disease was noted as follows: Optimal (≤ 1 cm): White, n (%): 178 (44%); African-American; n (%): 18 (4.5%) NMRD: White, n (%): 188 (46.5%); African-American; n (%):21 (5%)
Outcomes	SVRD vs NMRD: HR for OS 2.74 (95% CI 1.98 to 3.71) (HR adjusted for age, race, tumour grade, histology, ASA score, surgical complexity score, serum albumin, administration of platinum-based chemotherapy and significant peri-operative morbidity) OS was calculated from the date of diagnosis using Kaplan–Meier curves and compared using the log-rank test and Cox proportional hazards model
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly.

Bristow 2011 *(Continued)*

	2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): high risk HR for OS was adjusted for race, tumour grade 3, non-serous histology, ASA score >3, surgical complexity score, serum albumin < 3.0 g/dL, platinum-based therapy, residual disease and perioperative morbidity in multivariable Cox model 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria in multivariate model **Outcome: progression-free survival** Not reported
Notes	A total of 433 ovarian cancer women were identified with stage IIIC disease. Of these, 28 women were variously classified as either Asian-Pacific Islander, Hispanic, unknown or other and were excluded from further study. Source of funding: the Queen of Hearts Foundation for Ovarian Cancer Research Declaration of interest: none declared Median follow-up: 33.0 months The 30-day mortality rate for all 405 women was 1.5% Retrospective non-randomised study. Blinding not reported (but not applicable). Adjusted HRs are derived from a prognostic model. No details on how modelling was performed, but this seems to have been done based on significance testing (and not on including putative confounders in the analysis, irrespective of statistical significance). Women and disease characteristics not reported according to debulking status. NB: study only included women with stage IIIC ovarian cancer/possible overlap with Peiretti 2012.

Chan 2003

Study characteristics	
Methods	Retrospective cohort study
Participants	All consecutive cases of advanced-stage epithelial ovarian carcinoma diagnosed in younger women (range 22 to 45 years) were identified from tumour registry databases and a comparable group of 52 women who averaged 21 years older (range 46 to 85 years) was selected as controls. One-to-one matching from the same database was performed based on the date of diagnosis and stage of disease

Chan 2003 *(Continued)*

	during the same period in the same institution. Thus, the controls were similarly distributed across 17 years. The mean age at study entry was 50.5 years with a range between 22 and 85 years (40 (SD 5.7) and 61 years (SD 8.7) for younger and older women respectively) 5 (4.8%) women had FIGO stage IIIA, 5 (4.8%) had stage IIIB, 74 (71.1%) women had stage IIIC and 20 (19.2%) had stage IV disease Tumour cell type: papillary serous 72 (63.16%), mucinous: 3 (2.63%), endometrioid: 17 (14.9%), clear cell: 1 (0.88%), small cell: 3 (2.63%), undifferentiated: 8 (7%) Tumour grade: 1: 8 (7%), 2: 24 (21.1%), 3: 72 (63.2%) Performance status: 0: 65 (57%), 1 to 2: 35 (30.7%), unknown: 4 (3.51%)
Residual disease details	Residual disease was noted as follows: 1. SVRD: 71 (62.3%) 2. LVRD (> 1 cm): 43 (37.7%) Women were divided into SVRD (defined as optimal) and 1 cm or more (defined as suboptimal) groups based on residual disease after initial surgery. Optimal debulking was achieved in 36 (69%) and 35 (67%) women in younger in older groups respectively. All women received either a platinum/paclitaxel or a platinum/cyclophosphamide regimen for primary chemotherapy and women who underwent neoadjuvant chemotherapy with interval debulking were removed from the study. Gynaecology oncologists from the academic institution surgically staged all women.
Outcomes	A multivariable analysis which included older versus younger age, stage (IV vs III), performance status (1 to 2 vs 0) and residual disease (LVRD (> 1 cm) vs SVRD) was performed to evaluate all factors that were significant in the univariate analysis Overall survival: HR adjusted for prognostic categories (see above): • LVRD (> 1 cm) vs SVRD HR 1.67 (95% CI 1.03 to 2.72)
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): unclear risk Valid and reliable measurement of RD. Multicentre design may introduce heterogeneity in measurement of RD. **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Definition of OS not provided but no reason to doubt they used standard definition 5. Adjustment for other prognostic factors (a-g): low risk

Chan 2003 *(Continued)*

	HR for OS was adjusted for residual disease, age (older versus younger), stage (IV versus III) and performance status (1 to 2 versus 0) in a multivariable Cox model 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria in multivariate model **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Definition of PFS not provided but no reason to doubt they used standard definition 5. Adjustment for other prognostic factors (a-g): high risk PFS was reported in table comparing younger vs older patients but was not used in any multivariable modelling 6. Statistical analysis and reporting (a-d): high risk There was only a multivariate model for OS but not PFS
Notes	The median follow-up after surgery was 33 months (range 6 to 142 months) 5-year survival: of younger and older women: SVRD: 59% and 21% in young and old women respectively, LVRD (> 1 cm): 28% and 22% in young and old women respectively Median survival: SVRD: 66 months and 45 in young and old women respectively, LVRD (> 1 cm): 37 and 19 months in young and old women respectively, P = 0.003 Other variables in Cox model: Older versus younger age (HR 1.82, 95% CI 1.09 to 3.05), stage IV versus stage III disease (HR 3.00, 95% CI 1.71 to 5.25), performance status 1 to 2 versus 0 (HR 1.89, 95% CI 1.13 to 3.15) Despite the higher prevalence of poorly differentiated tumours in the older group, tumour grade (3 versus 1 to 2) was not an important prognostic factor in multivariable analysis (HR 1.06, 95% CI 0.57 to 1.97)

Chang 2012a

Study characteristics	
Methods	Retrospective review of medical records
Participants	All women underwent primary cytoreductive surgery followed by platinum-based chemotherapy. Consecutive women with stage IIIC and IV primary epithelial ovarian, fallopian tube or peritoneal cancer who underwent primary cytoreductive surgery at Ajou University Hospital between 1 January 2000 and 31 December 2011. Women received neoadjuvant chemotherapy, operated in other institution, stage IIIC due to nodal involvement were excluded N = 203 Median age was 54 years (range 30 to 78) Median BMI 23.3 (range 11.7 to 35.2) ASA 1 to 2: 114 (56.2%), 3 to 4: 80 (39.4%)

Chang 2012a *(Continued)*

	Stage IIIC: 189 (93.1%), IV: 14(6.9%) Tumour grade 1: 26 (12.8%), grade 2: 72 (35.5%), grade 3: 100 (49.3%) Histological subtype: serous: 167 (82.3%), mucinous: 4 (2.0%), endometrioid: 5 (2.5%), clear cell: 9 (4.4%), mixed: 18 (8.9%) Median pre-operative CA-125: 603.8 (range 4.5 to 21,677) Ascites < 1000 mL (54.7%), > 1000 mL (45.3%) Carcinomatosis: yes (73.4), no (26.6%) Simple procedure (58.6%), radical procedure (41.4%). Cohort was divided into simple procedures and radical procedures group for statistical analysis.
Residual disease details	Residual disease were defined: • NMRD (31.0%) • SVRD 0.1 cm to 1.0 cm (37.9%) • LVRD (> 1 cm) (31.0%)
Outcomes	Median follow-up was 43 months (range of 1 to 124) **Kaplan-Meier** Median unadjusted OS LVRD > 1 cm 37 months; SVRD 0.1 cm to 1 cm 46 months; NMRD 86 months Median unadjusted PFS LVRD > 1 cm 9 months; SVRD 0.1 cm to 1 cm 15 months; NMRD 35 months **Multivariate analysis for OS:** HR (LVRD > 1 cm vs NMRD) 3.24 (95% CI 1.90 to 5.53) HR (SVRD 0.1 cm to 1 cm vs NMRD): 2.22 (95% CI 1.25 to 3.94) **Multivariate analysis for PFS:** HR (LVRD > 1 cm vs NMRD): 3.40 (95% CI 2.00 to 5.77) HR (SVRD 0.1 cm to 1 cm vs NMRD): 2.20 (95% CI 1.26 to 3.84) HRs adjusted for age, FIGO stage and type of surgery (radical vs simple) **Morbidity** Operative time (minutes): simple: 235 (range 85 to 570), radical: 307 (range 150 to 810) Estimated blood loss: simple: 500 (range 200 to 4000), radical: 800 (range 300 to 7500) Intraoperative blood transfusion: simple (17.6%), radical (25.0%) Postoperative blood transfusion: simple (26.1%), radical (39.3%) Length of stay in ICU: simple: 0.8 (0 to 6), radical: 1.5 (0 to 6) Postoperative morbidity: simple (11.8%), radical (38.1%) Postoperative death < 30 days: simple = 0, radical = 1
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk

Chang 2012a *(Continued)*

	Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Definition of OS not provided but it usually has a standard definition 5. Adjustment for other prognostic factors (a-g): low risk HR for OS was adjusted for stage (IV), surgical procedure, residual disease and age in a multivariable Cox model 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria in multivariate model **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Definition of PFS not provided but it usually has a standard definition. 5. Adjustment for other prognostic factors (a-g): low risk HR for PFS was adjusted for stage (IV), surgical procedure, residual disease and age in a multivariable Cox model 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria in multivariate model
Notes	Subgroup analysis for 139 women with peritoneal carcinomatosis, the median unadjusted OS LVRD > 1 cm 39 months, SVRD 0.1 cm to 1 cm 50 months, NMRD 86 months

Chang 2012b

Study characteristics	
Methods	Retrospective review of medical records
Participants	Consecutive women with stage IIIC primary epithelial ovarian, fallopian tube or peritoneal cancer who underwent primary cytoreductive surgery at Ajou University Hospital between 1 January 2000 and 31 December 2011 After primary surgery, all women received adjuvant chemotherapy consisting of cisplatin (75 mg/m^2) or carboplatin (area under the curve; 5 to 7) and paclitaxel (135 mg/m^2) based systemic combination chemotherapy (every 3 weeks for 6 to 9 cycles) Exclusion: primary cytoreduction at an outside institution, neoadjuvant chemotherapy, stage IIIC disease based on lymph node metastasis only or borderline malignancy N = 191 Median age was 54 years (range 30 to 78)

Chang 2012b *(Continued)*

	Median BMI 23.2 (18.1 to 35.2) ASA 1 or 2: 107 (56.6%), 3 or 4: 74 (39.2%) Median pre-op CA-125 173.1 (range 4.5 to 21,677) Histological subtypes: serous: 155 (82%), mucinous: 4 (2.1%), endometrioid: 4 (2.1%), clear cell: 9 (4.8%), mixed: 17 (9.0%) Grade 1: 26 (13.8%), grade 2: 67 (35.4%), grade3: 5 (2.6%) Ascites < 1000 mL (57.7%), > 1000 mL (42.3%) Peritoneal carcinomatosis: yes:139 (73.5%), no: 50 (26.5%) Systematic lymphadenectomy (n = 135), no lymphadenectomy (n = 54) Lymphadenectomy; pelvic only (22.2%), pelvic and para-aortic (77.8%)
Residual disease details	Residual disease were defined: • NMRD: 61 (32.3%) • SVRD (0.1 to 1.0 cm): 67 (35.4%) • LVRD (> 1.0 cm): 61 (32.3%) Overall surgical morbidity - blood transfusion, deep vein thrombosis, sepsis, intestinal obstruction, ileus, lymphocyst or wound dehiscence was significantly higher in women who had lymphadenectomy
Outcomes	**Multivariate analysis for OS:** SVRD 0.1 cm to 1 cm vs NMRD: HR 2.25 (95% CI 1.25 to 4.03) LVRD > 1 cm vs NMRD: HR 3.09 (95% CI 1.80 to 5.30) HRs adjusted for age, performance of radical surgery and performance of lymphadenectomy
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk **Valid and reliable measurement of RD** **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Definition of OS not provided but it usually has a standard definition 5. Adjustment for other prognostic factors (a-g): low risk HR for OS was adjusted for residual disease, type of surgery, performance of lymphadenectomy and age in a multivariable Cox model 6. Statistical analysis and reporting (a-d): high risk

Chang 2012b *(Continued)*

	No conceptual framework; unclear of variable selection criteria in multivariate model **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Definition of PFS not provided but it usually has a standard definition. 5. Adjustment for other prognostic factors (a-g): low risk HR for PFS was adjusted for residual disease, type of surgery, performance of lymphadenectomy and age in a multivariable Cox model 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria in multivariate model
Notes	Systematic lymphadenectomy was performed in 135 (71.4%) of whom 105 had both pelvic and para-aortic lymphadenectomy. The mean number of dissected pelvic and para-aortic nodes were 25 (range 11 to 57) and 11 (range 3 to 35), respectively. 53.4% were found to have grossly enlarged lymph nodes during surgery. Of 135 women who underwent systematic lymphadenectomy, positive lymph nodes were found in 59%. The median unadjusted OS; lymphadenectomy 66 months, no lymphadenectomy 40 months. Sub-group analysis of NMRD: median OS 86 month versus no lymphadenectomy 46 months Of 189 women, tumour recurred in 110 women (58.2%) and 90 (47.6%) died of disease. 65 women with lymphadenectomy and 45 without lymphadenectomy had disease recurrence and there is no significant difference in the site of disease recurrence.

Chi 2001

Study characteristics	
Methods	Retrospective cohort study
Participants	282 women with stage III and IV epithelial ovarian cancer. Women with ovarian tumours of low-malignant potential were excluded from this study. All women were treated between 1987 and 1994 at Memorial Sloan-Kettering Cancer Center (MSKCC) The median age at study entry was 59 years with a range between 22 and 87 years 22 (8%) women had FIGO stage IIIA/IIIB, 194 (69%) had stage IIIC and 66 (23%) had stage IV disease Tumour cell type: serous 199 (71%), endometrioid: 46 (16%), clear cell: 19 (7%), mucinous: 10 (4%), mixed: 8 (3%) Tumour grade: 1: 13 (5%), 2: 69 (24%), 3: 184 (65%) Ascites: yes: 238 (84%), no: 43 (15%), unknown: 1 (1%)
Residual disease details	Women were treated with primary surgery followed by chemotherapy **Type of surgeon** Residual disease was noted as follows: 1. SVRD: 71 (25.2%) 2. Residual disease between 1 cm and 2 cm: 73 (26%)

Chi 2001 *(Continued)*

	3. LVRD greater than 2 cm: 137 (48.7%) The following types of chemotherapy were given to women in the study: cisplatin/cyclophosphamide: 143 (51%), carboplatin/cyclophosphamide: 65 (23%), carboplatin/paclitaxel: 31 (11%), cisplatin/paclitaxel 24 (8%), carboplatin: 7 (3%), cisplatin 1 (< 1%), none or unknown 10 (4%) Gynaecology oncologists from the academic institution surgically staged all women
Outcomes	A multivariable analysis which included age, stage (IIIC and IV vs IIIA/IIIB), ascites (yes vs no) and residual disease (1 cm to 2cm and > 2 cm vs < 1 cm) was performed to evaluate important prognostic factors Overall survival: HR adjusted for prognostic categories (see above): • 1 cm to 2 cm vs SVRD: HR 1.7 (95% CI 1.1 to 2.6) • LVRD (> 2 cm) vs SVRD: HR 2.0 (95% CI 1.3 to 2.9) **Direct surgical morbidity** 8 women (2.83%) died within 1 month of surgery
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk **Valid and reliable measurement of RD** **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Survival was calculated as the number of months from initial surgery to death or the date of last follow-up 5. Adjustment for other prognostic factors (a-g): low risk HR for OS was adjusted for residual disease, age, stage (IIIC and IV versus IIIA/IIIB) and ascites (yes versus no) in a multivariable Cox model 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria in multivariate model **Outcome: progression-free survival** Not reported
Notes	Of the 295 women who were treated for FIGO stage III and IV epithelial ovarian cancer at this centre over the period of the study, 13 (5%) were lost to follow-up, and the remaining 282 form the study group for this analysis Median follow-up in the study was 32 months (range: 1 to 139 months) The chemotherapy was platinum-based and when women who had initially had single agent therapy or combinations with cyclophosphamide recurred they were often given paclitaxel

Chi 2001 *(Continued)*

Survival was calculated as the number of months from initial surgery to death or the date of last follow-up.

214 of the 282 (76%) women were dead from disease or other causes at the time of census.

Multivariate analysis:

Only women age at diagnosis (P = 0.001), presence of ascites (P = 0.001) and the size of residual disease after primary cytoreductive surgery (1 cm vs 1 cm to 2cm vs > 2 cm (P = 0.02 and 0.001, respectively)) retained prognostic significance

Kaplan-Meier curve

Women with no more than 1 cm of residual disease after primary surgery have a 5-year survival of 50% and a median survival of 55 months. There is no statistically significant difference in survival between those women with 1 cm to 2 cm of residual disease and those with greater than 2 cm residual (P = 0.40). This combined group of women have a 5-year survival of 22% with a median survival of 28 months.

Impact of residual tumour volume for FIGO stage III

A subgroup analysis of the 216 women with stage III disease was done to examine the impact of size of residual disease on survival

56 of these women had up to 1 cm of residual disease and had 5-year survival of 50% and median survival of 56 months

73 of these women had between 1 cm and 2cm of residual disease and had 5-year survival of 28% and median survival of 31 months

87 of these women had greater than 2 cm of residual disease after surgery and had 5-year survival of 21% and a median survival of 28 months

The differences in survival are statistically significant between the women with up to 1 cm of residual disease and the women in the other 2 groups (P = 0.001). There is no statistically significant difference in survival between the women who had more than 1 cm residual disease.

Chi 2006

Study characteristics	
Methods	Retrospective study
Participants	Women with stage IIIC epithelial ovarian cancer The median age at study entry was 60 years (range: 22 to 87) All women presented with FIGO stage IIIC: 465 (100%) Tumour cell type: serous 331 (72%), endometrioid: 57 (12%), clear cell: 22 (5%), mixed: 53 (11%) Tumour grade: 1: 13 (3%), 2: 90 (19%), 3: 339 (73%), unknown: 23 (5%) Ascites: median 1600 mL (range: 0 to 17,000 mL), presence of ascites (N = 429): no = 58 (14%); yes = 371 (86%)
Residual disease details	Type of surgeon: gynaecologic oncologist Options for residual disease on the standardised operative form were as follows: 1. NMRD: 67 (14.4%) 2. Gross residual disease < 0.5 cm: 70 (15.1%)

Chi 2006 *(Continued)*

	3. SVRD of 0.6 cm to 1.0 cm: 99 (21.3%) 4. LVRD of 1 cm to 2 cm: 53 (11.4%) 5. LVRD > 2.0 cm: 176 (37.8%) Optimal is defined in 2 ways as NMRD and SVRD (< 1 cm), suboptimal defined as LVRD (> 1 cm) Postoperative chemotherapy records were available in 440/465 (95%) women. Of these 440 women, 426 (97%) were treated with primary platinum-based systemic chemotherapy with the intent to treat with at least 6 cycles.
Outcomes	Three women (0.6%) died within 30 days of surgery Overall survival: HR adjusted for age and ascites using Cox model: SVRD (< 1 cm) vs NMRD HR 2.07 (95% CI 1.23 to 3.46) LVRD (> 1 cm) vs NMRD HR 3.70 (95% CI 2.27 to 6.04)
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk **Valid and reliable measurement of RD** **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): unclear risk HR for OS was adjusted for residual disease, age and ascites in a multivariable Cox model 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria in multivariate model **Outcome: progression-free survival** Not reported
Notes	Median follow-up: 38 months (range: 1 to 199 months) 17-year death rate: 'Optimal' group: 105/236 'Suboptimal' group: 188/229 Median overall survival in relation to the 5 residual disease categories was: NMRD: 106 months; gross < 0.5 cm: 66 months; 0.6 cm to 1.0 cm: 48 months; 1 cm to 2 cm: 33 months; and > 2 cm: 34 months

Cioffi 2018

Study characteristics	
Methods	Single-centre retrospective study
Participants	N = 102 participants who received a diagnosis of International Federation of Gynecology and Obstetrics (FIGO) stage IIIC or IV EOC between 2000 and 2016, received neoadjuvant chemotherapy, and presented at least one of the following: • High tumour dissemination (assessed by laparoscopic Fagotti score > 8 or Peritoneal Cancer Index > 15): 83 (81.4%) • Stage IV: 38 (37.3%) • Comorbidities (Charlson comorbidity score ≥ 1): 27 (26.5%) • Poor performance status (ASA score ≥ 3): 58 (56.9%) Participants were stratified according to their age: ≥ 70 vs < 70 Age (mean): 74.5 (≥ 70 years) and 58.3 (< 70 years) FIGO: III - 64 (62.7%); IV - 38 (37.3%) Histology: serous - 58 (56.9%); undifferentiated - 1 (1%); endometrioid - 14 (13.7%); sero-endometrioid - 21 (20.6%); clear cell - 3 (2.9%); unknown - 5 (4.9%) Ascites (≥ 500 mL): 76 (74.5%) Tumour grade: G1 - 0; G2 - 8 (7.8%); G3 - 80 (78.4%); unknown - 14 (13.7%) CA-125 at diagnosis (median): 2934.1 (≥ 70 years) and 1462 (< 70 years)
Residual disease details	All women received platinum-based regimens, according to standard first-line protocols. After receiving 3 cycles of NACT, women were evaluated by computed tomography (CT) scan or positron emission tomography (PET)–CT scan; radiologic response was assessed according to RECIST 1.1. Women showing complete response (CR) or partial response (PR) to chemotherapy, and considered respectable by a gynaecologic oncologist team, underwent IDS. Women with either stable disease (SD) or progressive disease (PD) after 3 NACT cycles were re-evaluated after 3 further chemotherapy cycles. Women showing CR, PR or SD after 6 chemotherapy cycles underwent debulking surgery. Carboplatin AUC5 and paclitaxel 175 mg/m^2 every 3 weeks: 58 (56.9%) Carboplatin AUC5, paclitaxel 175 mg/m^2 and bevacizumab (15 mg/kg) on day 1 for 6 x 3-weekly courses followed by bevacizumab single-agent maintenance for 22 cycles or until toxicity or progression: 11 (10.8%) Carboplatin AUC5 every 3 weeks: 25 (24.5%) Carboplatin AUC2 and paclitaxel 60 mg/m^2 weekly: 5 (4.9%) Carboplatin AUC2 weekly: 3 (2.9%) Response to NAC (RECIST): • Complete: 35 (34.3%) • Partial: 33 (32.4%) • Stable: 18 17.6%) • Progressive: 13 (12.7%) • Missing: 3 (2.9%) **Optimal cytoreduction** defined as residual disease no greater than 1 cm (RD ≤ 1 cm) (n = 57; 67.1%) • NMRD (described in study as RD0): 37/85 (43.5%) • SVRD: 20 (23.5%) • LVRD (RD > 1): 28 (32.9%)

Cioffi 2018 *(Continued)*

Outcomes	Overall survival defined as interval from the date of initial diagnosis to the date of death or last follow-up Median overall survival: 25 months Multivariate Cox PH model for overall survival adjusted for age, number of chemotherapy courses, debulking surgery, ASA score, hypoalbuminaemia (defined as albuminaemia < 32 g/L), FIGO stage, presence of ascites, high tumour dissemination and Charlson comorbidity score: • SVRD < 1 cm (including NMRD) (vs LVRD > 1): HR 0.29 (95% CI 0.127 - 0.662), P = 0.003 Progression-free survival defined as interval from the date of initial diagnosis to the date of first recurrence, death or last follow-up. Median progression-free survival: 11 months Multivariate Cox PH model for PFS adjusted for age, number of chemotherapy courses, debulking surgery, ASA score, hypoalbuminaemia (defined as albuminaemia < 32 g/L), FIGO stage, presence of ascites ≥ 500 mL, high tumour dissemination and Charlson comorbidity score: • SVRD < 1 cm (including NMRD) (vs LVRD > 1 cm): HR 0.43 (95% CI 0.205 to 0.935), P = 0.03
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome. Overall survival defined as interval from the date of initial diagnosis to the date of death or last follow-up. 5. Adjustment for other prognostic factors (a-g): low risk HR for OS was adjusted for residual disease, age, number of neoadjuvant chemotherapy courses, debulking surgery, ASA score, hypoalbuminaemia (defined as albuminaemia < 32 g/L), FIGO stage, presence of ascites ≥ 500 mL, high tumour dissemination and Charlson comorbidity score 6. Statistical analysis and reporting (a-d): unclear risk No conceptual framework; although appears all variables were used in the multivariate models **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome. Progression-free survival defined as interval from the date of initial diagnosis to the date of first recurrence, death, or last follow-up. 5. Adjustment for other prognostic factors (a-g): low risk

Cioffi 2018 *(Continued)*

	HR for PFS was adjusted for residual disease, age, number of neoadjuvant chemotherapy courses, debulking surgery, ASA score, hypoalbuminaemia (defined as albuminaemia < 32 g/L), FIGO stage, presence of ascites ≥ 500 mL, high tumour dissemination and Charlson comorbidity score. 6. Statistical analysis and reporting (a-d): unclear risk No conceptual framework; although appears all variables were used in the multivariate models
Notes	ASA score: 1: 5 (4.9%); 2: 36 (35.3%); 3: 51 (50%); 4: 7 (6.9%) BMI (mean): 24.4 (≥ 70 years) and 25.5 (< 70 years) Charlson comorbidity score ≥1: 27 (26.5%) Procedures before NAC: diagnostic laparoscopy: 78 (27.7%); clinical exam/imaging: 196 (69.5%); unknown: 8 (2.8%)

Cuylan 2018

Study characteristics	
Methods	Retrospective study
Participants	218 women with stage III non-serous EOC Median age of women was 54 (range: 18 to 78) years Stage, n (%): • IIIA1: 55 (25.5%) • IIIA2: 14 (6.4%) • IIIB: 34 (15.6%) • IIIC: 115 (52.8%) 55 (25.2%) women underwent maximal CRS, 163 (74.8%) had optimal debulking Histopathology, n (%): endometrioid 64 (29.4%), mucinous 61 (28%), mixed 39 (17.9%), clear 54 (24.8%) Ascites, n (%): present 122 (56%), absent 96 (44%) Serum CA 125 (median, IU/mL): ≥ 240 IU/mL 109 (50%), < 240 IU/mL 109 (50%) Grade 1: 31 (14.2%), Grade 2: 57 (26.1%), Grade 3: 76 (34.9%) Turkey
Residual disease details	Speciality of surgeon: gynaecologic oncologist All women underwent maximal or optimal primary CRS followed by 6 cycles of carboplatin plus paclitaxel chemotherapy Residual disease was noted as follows: • NMRD after primary CRS: 55 (25.2%) • 'Optimal' cytoreduction, defined as SMRD (≤ 1 cm): 163 (74.8%)
Outcomes	HR for prognostic factors for OS: • Age 51 to 69 years vs ≤ 50 years (HR 1.73, 95% CI 1.23 to 2.66) • Age ≤ 50 vs ≥ 70 years (HR 2.6, 95% CI 1.215 to 5.591)

Cuylan 2018 *(Continued)*

- NMRD (HR 0.31, 95% CI 0.166 to 0.615)

HR for prognostic factors for PFS:

- Bilaterality (HR 1.44, 95% CI 1.01 to 2.056)
- Age (HR 2.25, 95% CI 1.176 to 4.323)
- NMRD (HR 0.34, 95% CI 0.202 to 0.58)

Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): unclear risk Valid and reliable measurement of RD. Multicentre design may introduce heterogeneity in measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): low risk HR for OS was adjusted for age, maximal cytoreduction and stage in multivariable Cox model 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria in multivariate model **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): low risk HR for PFS was adjusted for age, maximal cytoreduction and stage in multivariable Cox model 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria in multivariate model
Notes	Median duration of follow-up was 31.5 (range: 1 to 20) months 5-year PFS rate was 34.8% 5-year OS rate was 44.2%, median OS was 47 months (95% CI 36.12 to 57.88) A univariate analysis showed an OS rate of 81.2% the maximal CRS group Status: alive 109 (50%); dead 109 (50%)

Davidson 2019

Study characteristics	
Methods	Multicentre retrospective and single-centre prospective cohort Prospective data collection was to explore minimally-invasive surgery following NACT
Participants	All participants received NACT followed by interval debulking surgery for an advanced ovarian, fallopian tube or primary peritoneal cancer At Duke, information on women receiving NACT was collected retrospectively between January 2000 and September 2013 and prospectively (with subject informed consent after October 2013). At the Ohio State University and the University of Oklahoma, subjects were identified retrospectively. Women at all 3 institutions were included if they were diagnosed prior to 30 June 2017 to allow for at least 12 months of post-diagnosis follow-up. N = 282 participants with advanced ovarian, fallopian tube or primary peritoneal cancer Median age: 63.9 (range: 34.1 to 84.8) Race: Caucasian – 229 (81.2%) FIGO: IIIC – 114 (40.4%); IV – 101 (35.8%); presumed AOC – 57 (20.2%); unknown stage – 10 (3.5%) Histology: serous – 227 (80.5%); undifferentiated – 4 (1.5%); endometrioid – 1 (0.4%); mixed – 5 (1.8%); clear cell – 5 (1.8%); NOS – 21 (7.5%); unknown – 15 (5.3%) Ascites: 88 (31.2%)
Residual disease details	Carboplatin and paclitaxel: 87.2% Median NACT cycles: 4 (range: 2 to 10) Indication for NACT: disease volume - 80 (28.4%); comorbidities - 19 (6.7%); both - 29 (10.3%) Median surgery duration, minutes: 194 (range: 45 to 459) Determination of resectability: diagnostic laparoscopy – 78 (27.7%); clinical exam/imaging – 196 (69.5%); unknown – 8 (2.8%) Surgical approach at IDS: laparoscopy only – 27 (9.6%); laparoscopy converted to laparotomy – 26 (9.2%); exploratory laparotomy only – 221 (78.4%) Median surgical complexity score: 2 Surgical complexity score: • Low (0 to 3): 193 (68.4%) • Moderate (4 to 7): 80 (28.4%) • Complex (8 to 9): 9 (3.2%) Intraoperative complications: 23 women (8.7%). Bowel injuries (including serosal injuries) (n = 16); bladder (n = 6); vascular injuries (n = 6). Postoperative complications were seen in 62 women (22%) prior to hospital discharge and included: • Ileus/small bowel obstruction: 26 (9.2%) • Pulmonary issues: 12 (4.3%) • Altered mental state: 10 (3.6%) • Wound cellulitis/haematoma, UTI and cardiac concerns: 5 (1.8%) • Re-operation: 1 (0.4%) 32 (11.3%) experienced complications after discharge and within 30 days of surgery

Davidson 2019 *(Continued)*

	18 (6.4%) re-admitted. Data for reasons for re-admission available for n = 7: infectious complications (n = 3), gastrointestinal dysmotility (n = 3), acute renal failure related to urinary retention (n = 1). 2 required re-operation during re-admission. 1 underwent re-operation in outpatient setting for wound debridement. **Optimal cytoreduction** defined using two methods: NMRD (described in study as RD0) (n = 165/271; 60.9%) or SVRD ≤ 1 (n = 228/271; 84.1%). The latter definition is used in multivariable analysis. • NMRD: 165 (60.9%) • SVRD: 63 (23.2%) • LVRD 1 cm to 2 cm: 6 (2.2%) • LVRD > 2 cm: 37 (13.7%) • Missing (n = 11)
Outcomes	Disease-specific overall survival (DSS) defined as time from completion of adjuvant chemotherapy to death due to cancer Median disease-specific overall survival (DSS): 24.8 months Median DSS in RD ≤ 1: 25 months Median DSS in RD > 1: 23.5 months Multivariable Cox PH for DSS adjusted for ASA score, age, SCS and major morbidity: • LVRD > 1 cm (vs SVRD ≤ 1 cm): HR 1.7 (95% CI 1.1 to 2.8), P = 0.03 No deaths within 30 days of IDS
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): unclear risk Adequate cut-off for residual disease used. Multicentre design may introduce heterogeneity in measurement of RD. **Outcome level assessment:** **Outcome: disease-specific survival** 4. Outcome measurement (a-c): high risk Overall survival not used as outcome. Rather, disease-specific survival was used. Disease-specific survival (DSS) defined as time from completion of adjuvant chemotherapy to death due to cancer. 5. Adjustment for other prognostic factors (a-g): high risk Age arbitrarily categorised; ASA score dichotomised. Model predicting DSS adjusted for ASA score, age, SCS, presence of major morbidity. Few of these were deemed important prognostic factors. 6. Statistical analysis and reporting (a-d): unclear risk No conceptual framework; data driven based on P values of univariate associations **Outcome: progression-free survival**

Davidson 2019 *(Continued)*

	Not reported
Notes	—

Eisenkop 2003

Study characteristics	
Methods	This is a prospective study of women with FIGO stage IIIC ovarian cancer treated with primary cytoreductive surgery followed by platinum-based chemotherapy between 1990 and 2002 at a single North American institution
Participants	408 consecutive women presenting with stage IIIC epithelial ovarian cancer form the study group The median age at study entry was 62.8 years (range: 24 to 91) All women presented with FIGO stage IIIC epithelial ovarian cancer: 408 (100%) Tumour cell type: serous: 239 (58.5%), unspecified adenocarcinoma: 98 (24%), endometrioid: 32 (8%), clear cell: 10 (2.5%), mucinous: 18 (4.5%), mixed: 9 (2%), transitional cell: 2 (0.5%) Tumour grade: 1: 21 (5%), 2: 82 (20%), 3: 304 (75%), unspecified: 1 woman Volume of ascites: none: 20 (5%), ≤ 1000 mL: 114(28%), > 1000 mL: 249(61%), not recorded: 24(6%) GOG performance score: 0: 17 (4%), 1: 88 (21.5%), 2: 177 (43.5%), 3: 59 (14.5%), 4: 2 (0.5%), unspecified: 65 (16%) Preoperative tumour volume: Location of the largest metastases: omentum and adjacent structures: 228 (56%), pelvis: 102 (25%), retroperitoneal lymph nodes: 34 (8%), diaphragm: 12 (3%), other (large bowel, small bowel, mesentery, etc): 32 (8%) Largest metastatic disease: < 10 cm: 104 (26%), > 10 cm: 302 (74%)
Residual disease details	Residual disease was noted as follows: 1. NMRD: 351 (86%) 2. SVRD: 41 (10%) 3. LVRD (> 1 cm): 16 (6%) Surgery was undertaken by a gynaecological oncologist and disease was assessed intraoperatively in each of the following 5 regions: the left and right upper abdominal quadrants, the pelvis, the retroperitoneum and the central abdomen. A specifically defined numerical rank of 0 to 3 was assigned to each of the 5 regions and the ranks for each of the 5 regions were summed to give a total score before cytoreduction. 'Optimal' cytoreduction was defined as complete cytoreduction with no visible residual disease. The authors have previously described in other publications how this can be achieved at different anatomical sites but recourse to bowel resection was routine as was pelvic and para-aortic nodal dissection. Postoperative chemotherapy was platinum-based: cisplatin (50 to 100 mg/m^2) or carboplatin (300 to 400 mg/m^2) given in combination therapy with either cyclophosphamide or paclitaxel every 3 weeks for a planned 6 to 8 cycles.
Outcomes	Overall survival: HR adjusted for sum of rankings (a numerical ranking system was devised to reflect the continuum of progressively extensive tumour involvement for 5 anatomic regions) using a Cox model: SVRD vs NMRD: HR 2.32 (95% CI 1.20 to 5.37)

Eisenkop 2003 *(Continued)*

	LVRD (> 1 cm) vs NMRD HR: 2.98 (95% CI 1.74 to 5.23) **Direct surgical morbidity and mortality** Postoperative mortality occurred in 10 (2.5%) women Other morbidity including surgically related systemic morbidity such as chest infection, thromboembolic disease and cardiovascular events have not been reported **Recovery** The median length of hospital stay was 10 days
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome. Survival was measured in months from the date of primary surgery to the time of death or last follow-up appointment using life table analysis. 5. Adjustment for other prognostic factors (a-g): high risk HR for OS was adjusted for residual disease and sum of rankings (a numerical ranking system was devised to reflect the continuum of progressively extensive tumour involvement for 5 anatomic regions) in a multivariable Cox model 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria in multivariate model **Outcome: progression-free survival** Not reported
Notes	The median follow-up interval was 32.8 months Survival was measured in months from the date of primary surgery to the time of death or last follow-up appointment using life table analysis. Survival outcomes were analysed based on the numerical ranking of disease in each anatomical region, the sum of the ranking and the cytoreductive outcome. The median survival was 58.2 months (24% to 91%) and the estimated 5-year survival was 49% **Ranking of disease load** 349 (85.5%) of women had ranking in all 5 designated regions. Ranking was not possible in the rest because lymph node dissection was deferred in 48 women (12%) or the pattern of spread was inconsistent with ranking criteria in 16 women (4%).

Eisenkop 2003 *(Continued)*

On univariate analysis, categorisation of the sum of ranking scores (0 to 5 vs 6 to 10, vs ≥ 11), as well as ranking in the left upper abdominal quadrant and in the central abdomen were statistically important determinants of survival.

Univariate analysis showed that any rank score over zero (any disease) in the left upper abdominal quadrant (P = 0.01) and in the central abdominal region (P = 0.04) adversely affected survival. An effect of the anatomical site of disease on survival was not confirmed on multivariate analysis.

On multivariate analysis, survival was most influenced by the completeness of cytoreduction (P = 0.001), and less influenced by the categorised sum of rankings (P = 0.05).

This study demonstrates that high rates of complete cytoreduction can be achieved within dedicated teams with suitable training. The independent effect of completeness of cytoreduction on survival is confirmed though the median length of follow-up in the report is modest.

Feng 2016

Study characteristics	
Methods	Retrospective study
Participants	625 women who underwent primary staging or debulking surgery for high-grade serous ovarian cancer (HGSC) Age at diagnosis, median (range), years: 56 (30 to 84) FIGO stage: early (I,II) - 58 (9.3%); advanced (III, IV) - 567 (90.7%) Performance status: 0 to 379 (60.6%); 1 to 202 (32.3%); 2 to 44 (7.0%) 132 (21.1%) underwent bowel resection; 91 (14.6%) underwent upper abdominal surgery; 104 (16.6%) underwent lymphadenectomy CA-125: < 500 U/mL - 144 (23.6%); ≥ 500 U/mL - 465 (76.5%) Ascites: no - 75 (12%); < 500 mL - 104 (16.7%); ≥ 500 mL - 445 (71.3%) China
Residual disease details	Speciality of surgeon not reported After primary cytoreduction, all women received platinum-based intravenous chemotherapy Chemotherapy regimen: • Paclitaxel + carboplatin - 518 (82.9%) • Other platinum and taxane agents - 91 (14.6%) • Platinum and other agents - 16 (2.6%) Majority (441, 70.6%) of women had completed 6 to 8 cycles at intervals of 3 weeks R0 was defined as NMRD after surgery and was noted as follows: • No - 209 (33.4%) • Yes - 416 (66.6%)
Outcomes	PFS was defined as the time interval from the date of primary surgery to the date of disease progression or recurrence Median PFS was 18 months; 2-year PFS was 38.4%; 5-year PFS was 21.4%

Feng 2016 *(Continued)*

	OS was defined as the time interval from the date of the primary surgery to the date of death or last follow-up 2-year OS was 82.5%; 5-year OS was 51.4% At the time of analysis, 355 (56.8%) women were still alive
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of OS. OS was defined as the time interval from the date of the primary surgery to the date of death or last follow-up 5. Adjustment for other prognostic factors (a-g): unclear risk Multivariate models for OS adjusted for age, FIGO stage and time to chemotherapy 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection strategy into multivariate model. Unclear on reasoning behind inclusion of other prognostic factors in Cox models. **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of PFS; PFS was defined as the time interval from the date of primary surgery to the date of disease progression or recurrence 5. Adjustment for other prognostic factors (a-g): unclear risk Multivariate models for PFS adjusted for age, FIGO stage and time to chemotherapy. 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection strategy into multivariate model. Unclear on reasoning behind inclusion of other prognostic factors in Cox models.
Notes	The median (range) follow-up time was 29 (3 to 100) months The median (range) of time to chemotherapy (TTC) was 15 (4 to 62) days. TTC was longer for women who underwent bowel resection (P < 0.001). There were no differences in PFS and OS between women initiating chemotherapy before and after 15 days (P = 0.604 and 0.826 respectively) or among 4 groups categorised by quartile values (< 10 days, 10 to 14 days, 15 to 20 days, or ≥ 21 days after surgery) (P = 0.471 and 0.516, respectively). The time interval between surgery and chemotherapy seemed to have no prognostic impact on women with HGSC within 6 weeks. Length of hospital stay not reported

Hofstetter 2013

Study characteristics	
Methods	Prospective multicentre study
Participants	191 women with stage IIIA to IV primary ovarian cancer. Stage IIIa: 3, IIIb: 8, IIIc: 147, IV: 33 ECOG performance status (only available for 183 women) 0: 113, 1: 60; 2/3: 10 Age < 57: 98, > 57: 93 Histological subtypes; serous: 182, mixed serous:1, serous/clear cell: 4, undifferentiated: 4 Tumour grade 1/2: 51, 3: 140
Residual disease details	All women underwent primary surgery. All women received postoperative intravenous or intraperitoneal platinum-based chemotherapy. Women that received neoadjuvant chemotherapy were excluded **Postoperative residual disease defined as** • NMRD (n = 121) • Macroscopic or 'suboptimal' if residual tumour lesions of any size or number (n = 70)
Outcomes	Median follow-up was 42 months 3-year OS: HR of NMRD vs macroscopic RD: 2.95 (95% CI 1.87 to 4.67) HR adjusted for interval between surgery and start of chemotherapy, tumour stage, age and extent of surgery **Morbidity** Intraoperative complications included bladder injury (2), ureteral injury (1), intestinal injury (1), vascular injury (2), other operative injury (1). 9 of 185 women required blood transfusions. Postoperative complications comprised surgical site complications (35), medical complications (42), infectious complications (22) and reoperation's (22). **Adjuvant chemotherapy** • Intravenous carboplatin/taxane 1 cycle (3), 3 cycles (3), 4 cycles (6), 5 cycles (9), 6 cycles (139), 7 cycles (5), 8 cycles (4), 9 cycles (1) • Intraperitoneal platinum/taxane (13) • 9 women had single agent carboplatin: 2 cycles (1), 3 cycles (1), 6 cycles (7) • 1 women received carboplatin/liposomal doxorubicin
Risk of bias (QUIPS)	1. Study participation (a-f): unclear risk Adequate number of participants and description of target population. Baseline characteristics, sampling frame and period/place study took place presented clearly. Though, inclusion criteria not detailed. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD

Hofstetter 2013 *(Continued)*

	Outcome level assessment: **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcomes 5. Adjustment for other prognostic factors (a-g): unclear risk Interval from primary surgery to chemotherapy (continuous) arbitrarily dichotomised along the median. Multivariate model predicting OS adjusted for interval from surgery to chemotherapy, FIGO stage, age and extent of surgery 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; variable selection strategy into multivariate model unclear. HRs for centre not included in the results for multivariate analysis. There were other factors that were also significant at univariate analysis but were not included in multivariate model. **Outcome: progression-free survival** Not reported
Notes	The median time interval from primary surgery to the start of platinum-based chemotherapy was 28 days (range: 4 to 128). Women who received the first cycle of chemotherapy less than 28 days after surgery had a significantly improved 3-year survival rate of 70% as opposed to 60% in women with a later start of cytotoxic treatment.

Iwase 2015

Study characteristics	
Methods	Single-centre retrospective analysis of medical records
Participants	N = 124 women with advanced EOC who received NACT-IDS therapy at the Cancer Institute Hospital (Tokyo, Japan) between 2000 and 2008. Median age: 58 (range: 29 to 83) FIGO: IIIB – 6 (4.8%); IIIC – 77 (62.1%); IV – 41 (33.1%) Histology: serous – 105 (84.6%); mixed adenocarcinoma or carcinosarcoma included serous component – 10 (8.1%); non-serous – 9 (7.3%) Median CA-125 at pre-NACT, U/mL: 1569.4 (range: 13.5 to 24821) Median CA-125 post-NACT, U/mL: 15.8 (range: 2.3 to 1965.1) Lymph node metastasis: positive – 49 (39.5%); negative – 41 (33.1%); not evaluated – 34 (27.4%)
Residual disease details	Strategy for NACT-IDS therapy consisted of intensive chemotherapy (6 or more cycles) aimed at complete resection during IDS and pathological complete response followed by maximum debulking surgery included systematic retroperitoneal lymphadenectomy in principle. After about 6 cycles of NACT, we then performed IDS unless the disease had progressed. After IDS, ACT was generally administered for about 3 cycles using the same regimen. However, some women did not receive 3 cycles of ACT due to having undergone intensive chemotherapy before surgery or having undergone highly invasive surgery. Conversely, more than 3 cycles of ACT were necessary in the case of some women for whom complete resection was not achieved. Method to diagnose: laparotomy - 62 (50%); non-laparotomy - 62 (50%)

Iwase 2015 *(Continued)*

	Median NACT cycles: 6 (range: 2 to 9) NACT regimen: ifosfamide, epirubicin and cisplatin (IEP) including cyclophosphamide, adriamycin and cisplatin (CAP) – 44 (35.5%); paclitaxel and carboplatin (TC) including docetaxel and carboplatin (DC) – 80 (64.5%); irinotecan (CPT) base – 3 (2.4%) Surgical procedure at IDS: exploratory laparotomy – 11 (8.9%); total abdominal hysterectomy, bilateral salpingo-oophorectomy, and omentectomy (TAH + BSO + OM) – 10 (8.1%); TAH + BSO + OM + excision of other organs – 17 (13.7%); TAH + BSO + OM + retroperitoneal lymphadenectomy – 48 (38.7%); TAH + BSO + OM + excision of other organs + retroperitoneal lymphadenectomy – 38 (30.6%) Median operative blood loss, mL: 1291 (range: 220 to 5640) Blood transfusion: 72 women (70.6%) Median adjuvant CT cycles: 3 (range: 1 to 8) ACT regimen: ifosfamide, epirubicin and cisplatin (IEP) including cyclophosphamide, adriamycin and cisplatin(CAP) – 25 (20.2%); paclitaxel and carboplatin (TC) including docetaxel and carboplatin (DC) – 65 (52.4%); docetaxel and cisplatin (DP) including docetaxel (DTX) – 22 (17.7%); others – 7 (5.6%) **'Optimal' cytoreduction** defined as SVRD < 1 cm (n = 113; 91.1%) • NMRD: 98 (79%) • SVRD (RD < 1): 15 (12.1%) • LVRD (RD ≥ 1): 11 (8.9%) * Note: in multivariable analysis, it is RD > 0 cm vs NMRD
Outcomes	2-year OS: NMRD (88.8%); SVRD (40%); LVRD (≥ 1 cm) (36.3%) 5-year OS: NMRD (43.4%); SVRD (0%); LVRD (≥ 1 cm) (0%) Multivariable Cox PH for overall survival adjusted for FIGO stage, histological subtype, NACT cycles, NACT regimen, systematic lymphadenectomy, excision of other organ(s), ascites cytology, lymph node metastasis: • RD > 0 cm (vs NMRD): HR 4.03 (95% CI 2.39 to 7.16), P < 0.001 2-year PFS: NMRD (39.8%); SVRD (< 1 cm) (13.3%); LVRD (≥ 1) (0%)
Risk of bias (QUIPS)	1. Study participation (a-f): unclear risk Number of participants below the minimum cutoff of n = 100 for this meta-analysis. Adequate description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Definition of OS not provided but it usually has a standard definition 5. Adjustment for other prognostic factors (a-g): low risk

Iwase 2015 *(Continued)*

	Adjustment for large number of important PFs (FIGO stage, histological subtype, NACT cycles, NACT regimen, systematic lymphadenectomy, excision of other organ(s), ascites cytology, lymph node metastasis) 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; criteria for variable selection for univariate and multivariate Cox PH for OS unspecified **Outcome: progression-free survival** Progression-free survival mentioned in methods but not reported in results
Notes	Median follow-up, months: 39.5 (range: 5 to 142) Exclusion criteria: synchronous or metachronous (within 5 years) malignancies other than carcinoma in situ, missing data because women were referred to a different institution for initial treatment, received only palliative therapy after exploratory laparotomy, stage III disease without macroscopic peritoneal dissemination (e.g. pT1N1, pT2N1, pT3aN0 and pT3aN1), and received PDS-ACT therapy as initial treatment. Finally, excluding women who were not able to undergo IDS because of disease progression during NACT.

Kaban 2017

Study characteristics	
Methods	Single-centre retrospective analysis of medical records
Participants	N = 203 women diagnosed with stage IIIC to IV ovarian, fallopian tube or primary peritoneal cancer (according to postoperative pathology reports) who underwent treatment with interval surgery after NACT at the Istanbul University Gynecological Oncology Department between January 2002 and December 2012. Median age: 59 (range: 28 to 84) FIGO staging not reported Histology: serous – 171 (84.2%); undifferentiated – 1 (0.4%); endometrioid – 2 (0.9%); carcinosarcoma – 7 (3.4%); mixed – 2 (0.9%); clear cell – 4 (1.9%); mesothelioma – 2 (0.9%); Brenner tumour – 1 (0.4%); missing – 10 (4.9%) Visible tumour in diaphragm/liver: 29 (14.3%) Presence of tumour in omentum: macroscopic – 144 (70.9%); tumour-free – 44 (21.6%); no macroscopic – 14 (6.9%); missing – 1 Median lymph node count 10 (range: 2 to 24) Nodal metastasis: 3
Residual disease details	NAC consisted of a carbo-platinum (area under the curves 5 to 6) and paclitaxel (135 to 175 mg/m2) regimen every 3 weeks Median NACT cycles: 6 (range: 1 to 10) Pelvic +/- para-aortic lymphadenectomy performed in n = 25 women (12.3%) Extra-surgical procedure: bowel resection (n = 4); splenectomy (n = 1)

Kaban 2017 *(Continued)*

	Intraperitoneal port placement: 13 (6.4%) After surgery, all women continued chemotherapy with 2 to 6 additional cycles **'Optimal' cytoreduction** defined as SVRD: • SVRD (RD ≤ 1): 165 (81.3%) • LVRD (RD > 1): 36 (17.9%) • Missing (n = 2)
Outcomes	Overall survival (OS) was defined as the time from initial treatment to death or to the last follow-up examination. 5-year OS: 33.4% Median OS: 37.5 months Median OS in RD ≤ 1 cm: 40.6 months Median OS in RD > 1 cm: 21.3 months Multivariable Cox PH for OS adjusted for age, lymphadenectomy, macroscopic tumour in omentum, number of chemotherapy cycles: • LVRD (> 1 cm) (vs SVRD): HR 1.629 (95% CI 1.024 to 2.593), P = 0.039
Risk of bias (QUIPS)	1. Study participation (a-f): unclear risk Adequate number of participants and description of target population. Baseline characteristics, sampling frame and period/place study took place presented clearly. Though, inclusion criteria not detailed. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD; 201 (99%) with available RD data **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of OS which was defined as the time from initial treatment to death or to the last follow-up examination. 5. Adjustment for other prognostic factors (a-g): unclear risk Number of chemotherapy cycles dichotomised along arbitrary cut-off. Model predicting OS adjusted for age, lymphadenectomy, macroscopic tumour in omentum, number of chemotherapy cycles. Inclusion of other important PFs in model may alter results. 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear how variables selected for multivariate models. But age was included even though it was not significant at univariate, suggesting some assessment of clinical judgment in selection of important PFs. **Outcome: progression-free survival** Not reported

Kaban 2017 *(Continued)*

Notes	Median follow-up, months: 34.5 (range: 1 to 124)

Kahl 2017

Study characteristics	
Methods	Retrospective, multicentre cohort study
Participants	793 women with FIGO stage IIIB to IV Median age, years (range) (% < 55 years): 60 (19 to 88) ECOG performance status (PS): 0 to 683 (86.1%); > 0 to 110 (13.9%) FIGO stages, n (%): • IIIB - 110 (13.9%) • Stage IIIC - 318 (40.1%) • Stage IV - 365 (46.0%) Ascites, mL: ≤ 500 to 450 (56.7%); > 500 to 343 (43.3%) Histology: high-grade serous - 660 (83.2%); others - 133 (16.8%) Surgical complexity score: low/intermediate (≤ 7) - 165 (20.8%); high (≤ 8) - 628 (79.2%) Lymph node dissection: systematic - 472 (59.5%); sampling - 111 (14%); no - 210 (26.5%) CDC: 0 to 2 - 593 (74.8%); 3 to 4 - 176 (22.1%); 5 - 24 (3.0%) Germany
Residual disease details	Procedure performed by accredited gynaecological oncologist All women underwent primary cytoreductive surgery followed by postoperative systemic therapy with platinum-based chemotherapy Residual disease was noted as follows, n (%): • NMRD: 482 (60.8%) • SVRD (1 mm to 10 mm): 226 (28.5%) • LVRD (> 10 mm): 85 (10.7%) Women were divided into 3 groups based on their age-adjusted Charlson Comorbidity Index (ACCI): low (0 to 1), intermediate (2 to 3), and high (≥ 4) Postoperative surgical complications were graded according to the Clavien-Dindo classification (CDC)
Outcomes	Multivariate analysis of prognostic factors for OS: Residual disease (versus NMRD): • SVRD (1 mm to 10 mm): HR 1.96 (95% CI 1.55 to 2.46) • LVRD (> 10 mm): HR 2.75 (95% CI 2.01 to 3.77) Multivariate analysis of prognostic factors for high complications (CDC 3 to 5): • Surgical complexity score: high (≤ 8): RR 1.70 (95% CI 1.01 to 2.85) • Blood loss: ≥ 500: RR 1.0 (95% CI 0.64 to 1.44)

Kahl 2017 *(Continued)*

	• Duration of surgery, minutes: ≥ 360: RR 1.84 (95% CI 1.24 to 2.72)
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): unclear risk Valid and reliable measurement of RD. Multicentre design may introduce heterogeneity in measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): unclear risk Ascites dichotomised along arbitrary cutoff of 500 mL. Multivariate model predicting OS adjusted for ACCI, ECOG, FIGO stage, histology and ascites. 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; criteria for variable selection into multivariate model is unclear. Dichotomisation of continuous variables also apparent. **Outcome: progression-free survival** Not reported
Notes	After a median follow-up was 47 months (interquartile range 18 to 87 months), 397 (50.1%) women had died. Significant differences between the 3 ACCI groups were detected for performance status (ECOG 0: 95.7% vs 84.2% vs 65.9%) and residual disease (NMRD 70.7% vs 55.3% vs 49.6%). Residual disease after debulking surgery was significantly more frequent in women with a high ACCI compared with women with an intermediate or low ACCI (50.4% vs 44.7% vs 29.3%) The mortality rate in the low-ACCI group was 1.2%, in the intermediate-ACCI group it was 2.3% and it was 9.8% for the high-ACCI group

Klar 2016

Study characteristics	
Methods	Retrospective analysis of primary trials
Participants	5055 participants with stages I to IV ovarian cancer from AGO Study groups were included in Klar 2016. A total of 4488/5130 (87.5%) were stage III/IV in the 4 reported trials that were included in Klar 2016 and n = 4850 were included in the RD analysis. AGO-OVAR 3 trial: n = 798

Klar 2016 *(Continued)*

	• FIGO stage IIIA to IV: 717/798 (89.85%) • Postoperative residual tumour size, n (%): unknown: 4; ≤ 1 cm: 488 (62.6%); > 1 cm: 291 (37.4%) AGO-OVAR 5 trial: n = 1308 • FIGO stage IIIA to IV: 1191/1308 (91.06%) • Postoperative residual tumour size, n (%): unknown: 122; ≤ 1 cm: 799 (67.4%); > 1 cm: 387 (32.6%) AGO-OVAR 7 trial: n = 1282 • FIGO stage IIIA to IV: 1156/1282 (90.17%) • Postoperative residual tumour size, n (%): unknown: 151; ≤ 1 cm: 773 (68.3%); > 1 cm: 358 (31.7%) AGO-OVAR 9 trial: n = 1716 • FIGO stage IIIA to IV: 1424/1742 (81.75%) • Postoperative residual tumour size, n (%): unknown: 156; ≤ 1 cm: 1.111 (70.1%); > 1 cm: 475 (29.9%) Total cohort characteristics: Overall mean age of all women was 57.4 years (standard deviation, 10.53) FIGO 1A to IIA: 184 (3.6%); FIGO IIB to IIIB: 1182 (23.4%); FIGO IIIC to IV: 3684 (72.9%) ECOG 0: 1999 (39.7%); ECOG 1: 2544 (50.5%); ECOG 2: 490 (9.7%); ECOG 3: 2 (0%); ECOG 4: 1 (0%) BMI: underweight: 330 (6.5%); normal weight: 2099 (41.5%); overweight: 2626 (51.9%) Residual tumour: NMRD: 1779 (36.7%); SVRD (1 mm to 10 mm): 1442 (29.7%); LVRD (> 10 mm): 1629 (33.6%) Grading: G1: 399 (8.3%); G2: 1572 (32.9%); G3: 2574 (53.8%); G4: 225 (4.7%); GX: 10 (0.2%) Histology: serous: 3656 (72.4%); endometrioid: 428 (8.5%); mucinous: 219 (4.3%); undifferentiated: 214 (4.2%); others: 533 (10.6%) Death: tumour related: 2686 (94.8%); therapy associated: 24 (0.8%); other: 124 (4.4%) Germany, Austria and France
Residual disease details	Speciality of surgeon not reported All women underwent surgical cytoreduction followed by chemotherapy regimens: AGO-OVAR 3 trial: comparison of the combination of carboplatin/paclitaxel with paclitaxel/cisplatin AGO-OVAR 5 trial: comparison of carboplatin/paclitaxel and epirubicin with carboplatin/paclitaxel AGO-OVAR 7 trial: comparison of carboplatin/paclitaxel followed by topotecan with carboplatin/paclitaxel AGO-OVAR 9 trial: comparison of carboplatin and paclitaxel with or without gemcitabine
Outcomes	The effect of young age on PFS and OS in a multivariate analysis including all potential confounders FIGO III to IV versus IIB to IIIB: • PFS (HR 1.55, 95% CI 1.40 to 1.71) • OS (HR 1.51, 95% CI 1.34 to 1.70) Residual tumour NMRD versus SVRD: • PFS (HR 0.47, 95% CI 0.43 to 0.52) • OS (HR 0.43, 95% CI 0.38 to 0.49)

Klar 2016 *(Continued)*

	Residual tumour LVRD (> 10 mm) versus SVRD (1 mm to 10 mm): • PFS (HR 1.22, 95% CI 1.12 to 1.33) • OS (HR 1.21, 95% CI 1.10 to 1.33)
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place of study presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): unclear risk Adequate cut-off for residual disease used. As data come from different trials, this may introduce heterogeneity in measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): unclear risk Age and BMI dichotomised. Tumour grading also dichotomised. Multivariate model for OS adjusted for ECOG, BMI, FIGO stage, tumour grading and histology. 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear on reasons why the particular specific set of variables were selected for univariate screening. Criteria for variable selection into multivariate models unclear. Dichotomisation of continuous variables apparent. **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): unclear risk Age and BMI dichotomised. Tumour grading also dichotomised. Multivariate model for PFS adjusted for ECOG, BMI, FIGO stage, tumour grading and histology. 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear on reasons why the particular specific set of variables were selected for univariate screening. Criteria for variable selection into multivariate models unclear. Dichotomisation of continuous variables apparent.
Notes	Follow-up times: • AGO-OVAR 3 trial: women were followed for nearly 50 months in the trial • AGO-OVAR 5 trial: median follow-up time for surviving women in both groups was 42 months (range 0 to 61 months) • AGO-OVAR 7 trial: median KM follow-up time was 54 months for both groups • AGO-OVAR 9 trial: median follow-up time was 49 months in both groups

Langstraat 2011

Study characteristics	
Methods	Retrospective review of medical records
Participants	Women with stage IIIC to IV primary ovarian cancer and managed with the intention of complete tumour cytoreduction (NMRD) followed by treatment with Taxol and platinum-based chemotherapy Women had to be 65 years of age and older Exclusion: women who received neoadjuvant chemotherapy, underwent initial surgical debulking at another facility or had borderline tumour histology or non-epithelial cancer. Women who required emergent/urgent surgical intervention due to a small bowel obstruction were included if the stated primary surgical goal was to achieve complete cytoreduction, otherwise they were excluded. N = 280 Mean age 73.5 years (range: 65 to 89); 33% 80 years or older The group of women was divided into 4 age groups: 65 to 69, 70 to 74, 75 to 79, over 80 for statistical analysis ASA 1 to 2: 96, 3 to 4: 181 Stage IIIC: 210, Stage IV: 67 Histological subtype; serous: 205, mucinous: 6, endometrioid: 17, clear cell: 6, other: 43 40% albumin > 3.0 g/dL Mean creatinine = 1.05 USA
Residual disease details	Type of surgeon not reported Postoperative residual disease was defined as: • NMRD 61 (21.8%) • SVRD (0 cm to 1 cm) 120 (42.8%) • LVRD (> 1 cm) 95 (35.5%) The surgical complexity score (SCS) was assigned based on the extent of surgical effort and is calculated based on the number and type of procedures the women underwent. High complexity is defined if the score is over 7, and low complexity if the score is 3 or less.
Outcomes	OS HR (LVRD (> 1 cm) vs NMRD) 4.51 (95% CI 2.92 to 7.17) HR (SVRD vs NMRD) 2.24 (95% CI 1.48 to 3.49) HRs adjusted for creatinine, surgical complexity score, FIGO stage and age group
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk

Langstraat 2011 *(Continued)*

	Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Definition of OS not provided but it usually has a standard definition. 5. Adjustment for other prognostic factors (a-g): unclear risk Ascites was dichotomised with arbitrary cutoff of 1000 mL. Age as defined as a continuous and categorical variable in univariate analysis. CA-125 dichotomised with arbitrary cutoff of 750 U/mL. Creatinine dichotomised arbitrarily. Multivariate model predicting OS adjusted for creatinine, surgical complexity score, FIGO stage and age. 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection strategy into multivariate model **Outcome: progression-free survival** Not reported
Notes	Mean follow-up was of 3.2 years (range 0 to 15.8 years) 30-day mortality was observed in 12 of 280 (4.3%) women Older women who underwent surgery had a poorer performance score, higher mean creatinine, lower mean albumin and were more likely to have stage III disease. Only 15% of women who underwent surgery in the oldest age group had stage IV disease, compared to 26% of the rest of the cohort. Survival benefit was most apparent with complete cytoreduction but this benefit decreased with increasing age (median survival 5 years versus age group 65 to 69 at 5.9 years. Despite the trend towards lower surgical complexity in the older women over age 80 years (45%), there was a significant increase in surgical morbidity, mortality and the inability to receive chemotherapy. Similar trend was seen in women aged > 75 years.

Lecointre 2020

Study characteristics	
Methods	Retrospective, multicentre cohort study in 9 referral centres of France, constituting the FRANCOGYN study group
Participants	501 women with histologically confirmed advanced epithelial ovarian cancer of stages III or IV according to the FIGO classification, diagnosed between January 2000 and June 2017. Participants were split into those with ≤ 4 NACT cycles and > 4 NACT cycles. Median age: ≤ 4 NACT cycles: 60.7 years; > 4 NACT cycles: 62.6 years BMI: < 25: 406 (81%); 25 to 30: 2 (1%); > 30: 93 (18%) White ethnicity: 246/284 (87%) Personal or familiar history of gynaecological cancer: 171 (34%) FIGO III: 409 (82%); FIGO IV: 92 (18%)

Lecointre 2020 *(Continued)*

	Serous histology: 274/478 (57%) Pre-operative CA-125, U/mL: > 500: 302 (60%); ≤ 500: 199 (40%) Charlson index ≥ 1: 103/298 (35%) Tumour grade 1 to 2: 65 (13%); tumour grade 3: 248 (87%)
Residual disease details	The type of surgery performed was classified as complete (R0) when all visible tumours were removed (NMRD (referred to RD0 in study)) at the end of the intervention, R1 when it was ≤ 2.5 mm, R2 when it was more than > 2.5 mm but less than 2.5 cm NMRD: 346/471 (73%); RD > 0 cm to 2.5cm: 125/471 (27%) 30 participants had missing RD data
Outcomes	Median OS: 54.2 months 5-year survival ≤ 4 cycles: 45.6%; > 4 cycles: 27.6% 10-year survival ≤ 4 cycles: 26 %; > 4 cycles: 11% In multivariate Cox model controlling for number of NACT cycles (≤ 4, > 4); age (cat); Charlson index; FIGO; lymph node status (N+ vs N0); response to NACT; residual disease (RD > 0 cm to 2.5 cm vs NMRD) (adjusted HR 2.04 (95% CI 1.53 to 2.72)) Median PFS: 22.9 months 5-year survival ≤ 4 cycles: 19.7%; > 4 cycles: 11.7% In multivariate Cox model controlling for number of NACT cycles (≤ 4, >4); age (cat); response to NACT; residual disease (RD > 0 cm to 2.5 cm vs NMRD) (adjusted HR 1.36 (95% CI 1.05 to 1.76))
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): unclear risk Valid and reliable measurement of RD. 471 (94%) have RD data. Multicentre design may introduce heterogeneity in measurement of RD. **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): high risk Multivariate Cox model for OS adjusted for number of NACT cycles (≤ 4, > 4); age (cat); Charlson index; FIGO; lymph node status (N+ vs N0); response to NACT; residual disease (RD > 0 cm to 2.5 cm vs RD 0 cm) Large missing data rate for Charlson index (40%) and response to NACT (24%) - no methods discussed to handle missing data therefore assumed complete case analysis. 6. Statistical analysis and reporting (a-d): unclear risk

Lecointre 2020 *(Continued)*

No conceptual framework; data driven based on P values of univariate associations. Unclear on reasons why the particular specific set of variables were selected for univariate screening. Although multivariate estimates for RD were presented in the text of results, they did not appear in the corresponding tables.

Outcome: progression-free survival

4. Outcome measurement (a-c): low risk

Valid and reliable measurement of outcome

5. Adjustment for other prognostic factors (a-g): high risk

Multivariate Cox model for PFS adjusted for number of NACT cycles (≤ 4, > 4); age (cat); response to NACT; residual disease (RD > 0 cm to 2.5 cm vs RD 0 cm)

Large missing data rate for response to NACT (24%) - no methods discussed to handle missing data therefore assumed complete case analysis.

6. Statistical analysis and reporting (a-d): unclear risk

No conceptual framework; data driven based on P values of univariate associations. Unclear on reasons why the particular specific set of variables were selected for univariate screening. Although multivariate estimates for RD were presented in the text of results, they did not appear in the corresponding tables.

Notes	Study reports n = 471 with RD data, but due to missing data from other variables in the multivariate model, the HR estimates for OS and PFS may not be based on complete case analysis and could be based on less, unless imputation was used (e.g. multiple imputation by chained equations). Median NACT cycles ≤4 cycles: median 4 (range 3 to 4); > 4 cycles: median 6 (range 5 to 8) NACT regime Platinum and taxane: 464 (93%); other platinum-based: 37 (7%) Response to NACT: Complete response: 73/380 (19%); partial: 307/380 (81%) Time from diagnosis to IDS, months ≤ 4 NACT cycles: 3.8 (range 3.1 to 4.7); > 4 cycles: 5.9 (range 5.1 to 7.7) Operating duration, minutes ≤ 4 cycles: 328 (range 300 to 375); > 4 cycles: 360 (range 293 to 450) Blood transfusion: Yes: 44/77 (57%); no: 33/77 (43%) Intraoperative complications: Yes: 57/387 (15%); no: 330/387 (85%)

Lecuru 2019

Study characteristics

Lecuru 2019 *(Continued)*

Methods	Secondary analysis of the CHIVA double-blind randomised phase II GINECO study. The CHIVA trial explored the role of nintedanib in combination with NACT vs placebo in combination with NACT.
Participants	N = 163 participants treated with NACT with FIGO stage IIIC to IV AOC considered as unresectable after laparoscopic (lap) evaluation 188 participants were originally enrolled into the trial. The decision to exclude 25 participants was not stated.
Residual disease details	Women were treated with 3 to 4 cycles of platinum-taxane NACT + oral nintedanib before interval debulking surgery (IDS). CT (up to 6 cycles in total) and nintedanib were pursued postoperatively. No definition of optimal cytoreduction provided. Complete surgical resection response (referred to in study as CC0) included as variable but no explicit definition.
Outcomes	Multivariable Cox PH model adjusted for ECOG, ascites, neutrophil/lymphocyte ratio, PCI at baseline, RECIST ORR, CC0 at IDS, PCR and treatment arm (nintedanib vs placebo): • Complete surgical response (CC0) was predictive of both PFS and OS in multivariable Cox PH models ($P < 0.01$)
Risk of bias (QUIPS)	1. Study participation (a-f): high risk Abstract only therefore insufficient information on study participation 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Definition of OS not provided but it usually has a standard definition. 5. Adjustment for other prognostic factors (a-g): high risk Not explicitly stated but implied that model predicting OS adjusted for ECOG, ascites, neutrophil/lymphocyte ratio, Peritoneal Cancer Index at baseline, response rate at end of NACT according to RECIST (RECIST ORR), pathological complete or near complete response rate and treatment arm 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; variable selection criteria undefined and magnitude of effect not reported, only P value **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Definition of PFS not provided but it usually has a standard definition. 5. Adjustment for other prognostic factors (a-g): high risk Not explicitly stated but implied that model predicting OS adjusted for ECOG, ascites, neutrophil/lymphocyte ratio, Peritoneal Cancer Index at baseline, response rate at end of NACT according to RECIST (RECIST ORR), pathological complete or near complete response rate and treatment arm

Lecuru 2019 *(Continued)*

	6. Statistical analysis and reporting (a-d): high risk No conceptual framework; variable selection criteria undefined and magnitude of effect not reported, only P value
Notes	Abstract only Refer to Ferron 2019 for trial results for all n = 188 participants From Ferron 2019: Women with FIGO stage IIIC to IV chemotherapy-naive AEOC considered as unresectable after laparoscopic evaluation were randomised (2:1) to be treated with 3 to 4 cycles (cy) of carboplatin (AUC 5 mg/mL/min) and paclitaxel (175 mg/m^2) (CP) before interval debulking surgery (IDS) followed by 2 to 3 cycles of CP for a total of 6 cycles, plus either 200 mg of nintedanib (arm A) or placebo (arm B) twice daily on days 2 to 21 q3 week at cycles 1 and 2, 5 and 6 and maintenance therapy for up to 2 years.

Liu 2020

Study characteristics	
Methods	Retrospective analysis of past medical data from First Affiliated Hospital of Third Military Medical University from January 2009 to December 2017 China
Participants	114 women with stage III to IV epithelial ovarian cancer diagnosed by biopsy or cytologic examination based on histological proofs who received neoadjuvant chemotherapy followed by laparoscopic conservative interval debulking surgery (NACT + LIDS) Mean age: 51.6 (SD 9.3) Mean BMI: 23.2 (SD 3.3) FIGO III: 94 (82%); FIGO IV: 10 (18%) Serous histology: 97 (85%) Tumour grade High: 92 (81%); medium: 4 (3%); low 3 (3%); unknown: 15 (13%) Lymph node status Positive: 56 (49%); negative: 58 (51%)
Residual disease details	NMRD (referred to in study as R0) disease was defined as all diseases that were cytoreduced by electronic devices. If these diseases were not resected using an en bloc approach, leaving SVRD (≤1 cm), authors considered it as optimal (R1). NMRD: 66 (58%) SVRD (< 1 cm): 48 (42%)
Outcomes	Median OS: 56 months Univariate association between RD and OS was reported ≥ SVRD (< 1 cm) vs NMRD: HR 9.589 (95% CI 3.911 to 23.507) No variable other than RD was included in the "multivariate" model. Therefore, this was not included in the analysis and this is noted in the interpretation of the results. Median DFS: 14 months After controlling for age (continuous), residual disease (SVRD vs NMRD): adjusted HR 6.022 (95% CI 3.632 to 9.986)

Liu 2020 *(Continued)*

Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcomes 5. Adjustment for other prognostic factors (a-g): high risk No variable other than RD was included in the "multivariate" model. Therefore, this was not included in the analysis and this is noted in the interpretation of the results. 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear how variables were selected into multivariate model and why the absence of key variables. Selection strategy led to multivariate Cox model for OS with RD as the only predictor. **Outcome: disease-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcomes 5. Adjustment for other prognostic factors (a-g): high risk Only adjustment for age in multivariate model for DFS 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear how variables were selected into multivariate model and why the absence of key variables.
Notes	Patients received IV paclitaxel and carboplatin/ cisplatin or IV docetaxel and cisplatin every 3 weeks Number of NACT cycles 2: 67 (59%); 3: 37 (32%); 10: (9%) Number of adjuvant chemotherapy cycles 3 to 4: 30 (26%); 5: 42 (37%); ≥ 6: 42 (37%)

Lorusso 2016

Study characteristics

Lorusso 2016 *(Continued)*

Methods	Multicentre, retrospective review of consecutive women who underwent NACT-IDS in 5 Italian centres
Participants	N = 193 participants with advanced-stage ovarian cancer
Residual disease details	3 NACT cycles: 77 (44%) 4 NACT cycles: 74 (38%) 5 NACT cycles or more: 43 (22%) Text suggests residual disease was treated as NMRD vs any macroscopic RD (> 0 cm)
Outcomes	5-year overall survival (OS) was 46% and 31% for women having 3 and 4+ cycles of NACT 10-year OS was 26% and 18% for women having 3 and 4+ cycles of NACT "A trend towards worse OS was observed for women with residual disease at IDS": HR 1.29 (95% CI 0.98 to 1.70), P = 0.06 Unknown number of covariates in model except for ECOG performance status. Residual disease variable presumed to be RD > 0 cm vs NMRD.
Risk of bias (QUIPS)	1. Study participation (a-f): high risk Abstract only therefore insufficient information on study participation 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Definition of OS not provided but it usually has a standard definition. 5. Adjustment for other prognostic factors (a-g): high risk Unclear on which variables were adjusted for but we know there is at least ECOG and number of NACT cycles 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear on reasons why the particular specific set of variables were selected for multivariate model **Outcome: progression-free survival** Not reported
Notes	Abstract only

Luger 2020

Study characteristics	
Methods	Retrospectively review of patients diagnosed between 2000 and 2016 Austria
Participants	178 stage III and IV ovarian cancer patients Median age at diagnoses was 64.6 years (interquartile range (IQR) 50.8 to 72.7) Only patients without surgically removed enlarged cardiophrenic lymph nodes (CPLN) were eligible for this study FIGO III: 91 (51%); FIGO IV: 87 (49%) Histology Serous: 157 (88%); mucinous: 3 (2%); endometrioid: 13 (7%); clear cell: 5 (3%) Tumour grade: 1: 17 (10%); 2: 82 (46%); 3: 79 (44%) Median follow-up duration: 49.6 months (IQR 32.9 to 66.3)
Residual disease details	All patients received primary upfront primary debulking surgery (PDS) by dedicated teams including at least one certified gynaecologic oncologist, and all received adjuvant platinum-based chemotherapy. The authors defined “No residual disease” as complete macroscopic tumour resection at the end of debulking surgery Residual disease groups: NMRD: 133 (75%) RD > 0 cm: 45 (25%)
Outcomes	Overall and progression-free survival
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Definition of OS not provided but it usually has a standard definition 5. Adjustment for other prognostic factors (a-g): low risk HR for OS was adjusted for age (> 64.6 years), CA-125, paraaortic nodes (positive), stage, residual disease, and CPLN dimension in multivariable Cox model 6. Statistical analysis and reporting (a-d): high risk

Luger 2020 *(Continued)*

	No conceptual framework; unclear of variable selection criteria in multivariate model **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Definition of PFS not provided but it usually has a standard definition 5. Adjustment for other prognostic factors (a-g): low risk HR for PFS was adjusted for age (> 64.6 years), CA-125, paraaortic nodes (positive), stage, residual disease, and CPLN dimension in multivariable Cox model 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria in multivariate model
Notes	Residual disease in multivariate model for: PFS: HR 2.44 (95% CI 1.23 to 4.84), P = 0.011; OS: HR 2.17 (95% CI 1.11 to 4.69), P = 0.028. The upper 95% CI for OS was entered into forest plots as 4.26 so slight margin of error in the reported statistic). Multivariate model was adjusted for age, CA-125, histologically positive paraaortic lymph nodes, FIGO stage (IIIA to IIIC vs FIGO IVA and IVB), cardiophrenic lymph node (CPLN) and residual disease. Recurrence was observed in 66.9% (n = 119) of patients and the median progression-free survival was 12.0 months (IQR 5.5 to 30.5). 80 patients (44.9%) died during a median time of follow-up of 49.6 months (IQR 32.89 to 66.26). Adjuvant chemotherapy: Carboplatin + paclitaxel: 150 (84%); carboplatin: 24 (14%); carboplatin + endoxan: 4 (2%) Platinum response: Refractory + resistant: 35 (20%); sensitive: 143 (80%) A systematic pelvic and paraaortic lymphadenectomy (removal of ≥ 20 retroperitoneal lymph nodes was performed in 84.2% of patients Systematic retroperitoneal lymphadenectomy (removal of ≥ 20 nodes): 150 (84.2%) Sampling retroperitoneal lymphadenectomy (removal of < 20 nodes): 8 (4%) Median number of removed nodes: 26 (IQR 7 to 37) 88 (68%) had exhibited histologically proven retroperitoneal lymph node metastases Intraperitoneal carcinomatosis radiologically evident in 151 (85%) Radiological diagnosis of upper abdominal spread in 72 (41%)

McGuire 1995

Study characteristics	
Methods	Retrospective analysis of a prospective randomised controlled trial comparing different chemotherapy dosing schedules. It aimed to determine the importance of chemotherapy dose intensity on survival, progression-free survival (PFS) and response. This was not a trial of surgery but the report allows a comparison of survival outcomes for subgroups women with stage III ovarian cancer who have had < 2 cm or ≥ 2 cm of residual disease following surgery and therefore is relevant to this review.
Participants	458 women with FIGO stage III and IV epithelial ovarian cancer were recruited. These were women who had more than 1 cm residual disease following initial surgery. 27 women were ineligible: incorrect stage (n = 5), incorrect primary tumour (n = 9), incorrect cell type (n = 7), history of prior malignancy (n = 3), prior chemotherapy (n = 1) and other (n = 2) Women with borderline ovarian tumours (low malignant potential) were excluded

McGuire 1995 *(Continued)*

	Recruitment was from December 1986 to April 1990 and all women had undergone a surgical procedure The median age at study entry was 60 years (range: 20 to 83) 305 (67%) and 153 (33%) women had FIGO stage III and IV disease, respectively Tumour cell type: serous 312 (68.1%), endometrioid: 64 (14%), mucinous; 12 (2.6%), clear cell: 12 (2.6%), other: 58 (12.7%) Tumour grade: 1: 26 (9%), 2: 114 (39%), 3: 152 (52%), not specified 2 (1%) GOG score: 0: 150 (32.8), 2: 213 (46.5%), 3: 95 (20.7%)
Residual disease details	Residual disease was noted as follows: 1. LVRD between 1 cm and 2 cm for women with stage III disease: 31 (6.8%) 2. LVRD greater than 2 cm for women with stage III disease: 274 (58.9%) 3. LVRD between 1 cm and 2 cm for women with stage IV disease: 54 (11.8%) 4. LVRD greater than 2 cm for women with stage IV disease: 99 (21.6%) **Definition of optimal surgery:** All women were 'suboptimally' cytoreduced with > 1 cm of residual disease **Chemotherapy:** 2 trial arms with women receiving either standard chemotherapy: cyclophosphamide 500 mg/m^2 and cisplatin 50 mg/m^2 intravenously every 3 weeks for 8 courses OR intense chemotherapy: cyclophosphamide 1000 mg/m^2 and cisplatin 100 mg/m^2 intravenously every 3 weeks for 4 courses. Dose modification was rigidly controlled to maintain intensity.
Outcomes	Overall survival and progression-free survival: HR adjusted for age, GOG performance status, histological sub-type, stage/residual disease and measurable disease using Cox model: III, ≥ 2 cm vs III, 1 to 2 cm: HR 1.91 IV, 1 cm to 2 cm vs III, 1 to 2 cm: HR 1.89 IV, ≥ 2 cm vs III, 1 to 2 cm: HR 2.29 Overall and progression-free survival (PFS) were measured from the date of randomisation. All eligible women were included in the analysis of outcomes. All causes of death were used to calculate survival, and the estimates were based on Kaplan-Meier procedures.
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk

McGuire 1995 *(Continued)*

	Valid and reliable measurement of outcome. OS was measured from the date of randomisation. 5. Adjustment for other prognostic factors (a-g): unclear risk Multivariate model for OS adjusted for age, GOG performance status, histological subtype and measurable disease 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria in multivariate model. Magnitude of effect not reported with confidence interval and only P value was available. **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome. PFS was measured from the date of randomisation. 5. Adjustment for other prognostic factors (a-g): unclear risk Multivariate model for PFS adjusted for age, GOG performance status, histological subtype and measurable disease 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria in multivariate model. Magnitude of effect not reported with confidence interval and only P value was available.
Notes	Mean and median length of follow-up were not reported. Since this trial was a trial of chemotherapeutic regimens, the randomisation did not aim to compare the effect of different degrees of surgical debulking. The findings borne out on multivariate analysis are similar to those in retrospective and cohort studies. The prospective nature of this study has, however, facilitated the collection of a fairly complete data set and gives this work some authority. Other variables in Cox model: Age (years): reference group: women aged less than 55 years (P = 0.47): 55 to 65: HR 1.08; > 65: HR 1.38 GOG performance status: reference group: GOG 0 (P = 0.009) 1: HR 1.26, 2: HR 1.56 Histological subtype: reference group: serous adenocarcinoma (P < 0.001): Endometrioid: HR 0.951, mucinous: HR 8.31, clear cell: HR 1.79, other: HR 0.84 Measurable disease: reference group: No: (P = 0.01) Yes: HR 1.43 From the study both advancing age and worsening performance status were associated with poorer survival. In addition, mucinous histology is associated with an 8.3 times greater death rate than serous histology (P < 0.001). The study shows residual disease after surgery impacts on survival. Even in 'suboptimal' cytoreduction (residual disease greater than 1 cm), women with stage III disease and residual disease diameter less than 2 cm exhibited lower death rates than either those with stage III diease and residual disease diameter of ≥ 2 cm, or those with stage IV disease.

Melamed 2017a

Study characteristics

Melamed 2017a *(Continued)*

Methods	Retrospective cohort study
Participants	307 women with stage IIIC to IV epithelial clear cell carcinoma were included in the analysis Age group: • < 40: 10 (3.3%) • 40 to 49: 59 (19.2%) • 50 to 59: 131 (42.7%) • 60 to 69: 82 (26.7%) • 70 to 79: 23 (7.5%) • 80+: 2 (0.7%) Median age was 56 years Race/ethnicity: • Asian: 25 (8.1%) • Black: 18 (5.9%) • Hispanic: 24 (7.8%) • White: 240 (78.2%) Stage: • IIIC: 241 (78.5%) • IV: 66 (21.5%) USA
Residual disease details	Speciality of surgeon not reported All women underwent primary cytoreductive surgery and adjuvant chemotherapy Residual disease status was classified as follows: • NMRD: 141 (45.9%) • SVRD (1 cm or less): 77 (25.1%) • LVRD measuring > 1 cm: 23 (7.5%) • Unknown: 66 (21.5%)
Outcomes	The primary outcome for OS was time from diagnosis to death from any cause, or to last contact, as recorded by the cancer registrar NMRD: (AHR 0.34, 95% CI 0.18 to 0.64) SVRD (≤ 1 cm): (AHR 0.94, 95% CI 0.50 to 1.75) LVRD (> 1 cm): (AHR referent)
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk

Melamed 2017a *(Continued)*

	Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome. OS was time from diagnosis to death from any cause, or to last contact, as recorded by the cancer registrar. 5. Adjustment for other prognostic factors (a-g): high risk Age arbitrarily categorised. Multivariate model predicting OS adjusted for age, race/ethnicity, stage, region, insurance status, treating facility type, hospital annual ovarian cancer volume and presence of comorbidities 6. Statistical analysis and reporting (a-d): unclear risk Authors reported that covariates were selected a priori but difficult to verify **Outcome: progression-free survival** Not reported
Notes	Analysis is a subgroup of women who were analysed from a study that identified 6013 women with stage IIIC and IV high-grade serous, 307 with clear cell and 140 with mucinous histology The median follow-up was 34.1 months

Melamed 2017b

Study characteristics	
Methods	Retrospective cohort study
Participants	6013 women with stage IIIC to IV epithelial high-grade serous ovarian cancer were included in the analysis Age group, n (%): • < 40: 117 (1.8%) • 40 to 49: 859 (13.3%) • 50 to 59: 1827 (28.3%) • 60 to 69: 2047 (31.7%) • 70 to 79: 1297 (20.1%) • 80+: 314.8%) Median age was 63 years Race/ethnicity, n (%): • Asian: 236 (3.7%) • Black: 467 (7.2%) • Hispanic: 377 (5.8%) • White: 5318 (82.3%) • Other/unknown: 62 (1.0%) Stage, n (%):

Melamed 2017b *(Continued)*

	• IIIC: 4954 (76.7%) • IV: 1506 (23.3%) USA
Residual disease details	Speciality of surgeon not reported All women underwent primary cytoreductive surgery and adjuvant chemotherapy Residual disease status was classified as follows: • NMRD: 2048 (34.1%) • SVRD measuring 1 cm or less: 1848 (30.7%) • LVRD measuring > 1 cm: 546 (9.1%) • Unknown: 1571 (26.1%)
Outcomes	The primary outcome for OS was time from diagnosis to death from any cause, or to last contact, as recorded by the cancer registrar NMRD: (AHR 0.58, 95% CI 0.49 to 0.69) SVRD (≤ 1 cm): (AHR 0.85, 95% CI 0.72 to 1.01) LVRD (> 1 cm): (AHR referent)
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome. OS was time from diagnosis to death from any cause, or to last contact, as recorded by the cancer registrar. 5. Adjustment for other prognostic factors (a-g): high risk Age arbitrarily categorised. Multivariate model predicting OS adjusted for age, race/ethnicity, stage, region, insurance status, treating facility type, hospital annual ovarian cancer volume and presence of comorbidities 6. Statistical analysis and reporting (a-d): unclear risk Authors reported that covariates were selected a priori but difficult to verify **Outcome: progression-free survival** Not reported
Notes	Analysis is a subgroup of women who were analysed from a study that identified 6013 women with stage IIIC and IV high-grade serous, 307 with clear cell and 140 with mucinous histology

Melamed 2017b *(Continued)*

	The median follow-up was 34.1 months

Paik 2018

Study characteristics	
Methods	Retrospective analysis of data obtained from electronic medical records
Participants	419 EOC women of stages IIIB, IIIC or IV with high-grade serous type histology were investigated 48 (11.5%) with a normal-sized ovary (less than 4 cm in the longest diameter, with a tumour size greater than 5 × 5 mm within the ovarian substance) Mean age of women was 54.5 ± 10.3 years Women with enlarged-ovarian tumour were younger (54.0 ± 10.3 vs 58.4 ± 9.2 years) than those in the normal-sized ovary group The mean size of ovary was 7.5 ± 3.9 cm for the whole group: • With enlarged-ovarian tumour (n = 371): 8.1 ± 3.8 cm • With normal-sized ovary (n = 48); 3.2 ± 1.1 cm FIGO stage IIIB: 15 (3.6%); stage IIIC: 335 (84.7%); stage IV: 49 (11.7%) Initial CA-125 (U/mL): 1922.4 ± 2968.9 ASA physical status: • I: 191 (45.6 • II: 178 (42.5) • III: 18 (4.3) • Unknown: 32 (7.6) Korea
Residual disease details	Speciality of surgeon not reported Women were treated with primary debulking surgery (PDS) with adjuvant chemotherapy for primary treatment Residual disease status after PDS (cm) was classified as follows, n(%): • NMRD: 107 (25.5%) • SVRD (< 1 cm): 147 (35.1%) • LVRD (≥ 1 cm): 165 (39.4%) For adjuvant chemotherapy, the first cycle of combination chemotherapy consisting of taxane/platinum was initiated routinely within 2 weeks of surgery Subsequent chemotherapy cycles were performed every 3 weeks for 6 cycles, but there could have been variation in the number of cycles depending on women situation Overall survival (OS) was defined as the time between initial diagnosis and women death or loss to follow-up Progression-free survival (PFS) was designated as the time between diagnosis and women recurrence/progression or loss to follow-up

Paik 2018 *(Continued)*

Outcomes	Multivariate Cox proportional hazards analysis of PFS and OS to adjust for risk-associated prognostic clinical features Residual disease status after PDS (cm): • NMRD: PFS and OS (HR 1) • SVRD (< 1 cm): PFS (HR 1.591, 95% CI 1.153 to 2.193); OS (HR 2.291, 95% CI 1.398 to 3.752) • LVRD (≥ 1 cm): PFS (HR 1.698, 95% CI 1.239 to 2.326); OS (HR 2.549, 95% CI 1.564 to 4.152)
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome. OS was defined as the time between initial diagnosis and women death or loss to follow-up. 5. Adjustment for other prognostic factors (a-g): unclear risk CA-125 arbitrarily dichotomised at cutoff of 35 mL. Multivariate model for OS adjusted for age, CA-125, FIGO stage and normal sized ovary 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear variable selection criteria into multivariate model **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome. PFS was designated as the time between diagnosis and women recurrence/progression or loss to follow-up. 5. Adjustment for other prognostic factors (a-g): unclear risk CA-125 arbitrarily dichotomised at cutoff of 35 mL. Multivariate model for PFS adjusted for age, CA-125, FIGO stage and normal sized ovary. 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear variable selection criteria into multivariate model
Notes	In total cohort with a median follow-up period of 43 months (range, 3 to 164 months), Inferior overall survival (OS) was shown in the normal-sized ovary group (median OS, 71.2 vs 41.4 months At the time of analysis, of the 419 enrolled women, 298 (71.1%) experienced a relapse, and 192 (45.8%) died after a median observation time of 43 months (range, 3 to 164 months)

Paik 2018 *(Continued)*

Other variables in cox model:

Age (continuous): PFS (HR 0.966, 95% CI 0.985 to 1.007); OS (HR 1.003, 95% CI 0.989 to 1.017)

CA-125 level (U/mL):

- < 35: PFS and OS (HR 1)
- ≥ 35: PFS (HR 2.167, 95% CI 1.020 to 4.601); OS (HR 4.437, 95% CI 1.077 to 17.549)

FIGO stage:

- IIIB: PFS and OS (HR 1)
- IIIC: PFS (HR 1.130, 95% CI 0.529 to 2.414); OS (HR 0.638, 95% CI 0.280 to 1.453)
- IV: PFS (HR 1.178, 95% CI 0.520 to 2.671); OS (HR 0.621, 95 % CI 0.249 to 1.550)

Normal-sized ovary:

- No: PFS and OS (HR 1)
- Yes: PFS (HR 1.180, 95% CI 0.839 to 1.660); OS (HR 1.593, 95% CI 1.097 to 2.314)

For primary surgical treatment, bilateral salpingo-oophorectomy, hysterectomy, peritoneal washing, retroperitoneal lymphadenectomy, omentectomy and tumourectomy of any metastatic lesions were performed routinely

Peiretti 2010

Study characteristics	
Methods	Retrospective study
Participants	259 with advanced epithelial ovarian and fallopian tube cancer met the inclusion criteria Median age was 58 years (range: 22 to 77 years) Primary site disease: ovary 256 (98%); fallopian tube 3 (2%) FIGO stages: IIIC: 199 (76%); IV: 60 (24%) Tumour grades: grade 1 to 2: 53 (21%); grade 3: 198 (76%); grade N/A: 8 (3%) Histological type: • Serous: 184 (71%) • Endometrioid: 39 (15%) • Clear cell: 8 (3%) • Mixed: 26 (10%) • Others: 2 (1%) Peritoneal carcinomatosis: yes: 188 (72%); no: 71 (28%) Location of largest mass: • Pelvis: 130 (50%) • Omentum: 110 (42%) • Upper abdomen: 14 (5%) • Retroperitoneal node: 1 (0.4%) • Other: 4 (1.6%) Intraoperative units blood transfused, n (%):

Peiretti 2010 *(Continued)*

	• None: 147 (56%) • 1 to 2: 67 (26%) • 3 to 4: 31 (12%) • > 5: 14 (5%) Postoperative units blood transfused, n (%): • None: 122 (47%) • 1 to 2: 113 (43%) • 3 to 4: 23 (8%) • > 5: 4 (2%) Size of largest mass (cm): ≤ 10: 98 (38%); > 10: 161 (62%) Median CA-125 (range): 913 U/mL (17 to 52,817) Median ascites (range): 1500 cc (100 to 15,000) Spain and Italy
Residual disease details	All these women underwent an attempt of maximal surgical cytoreduction unless there was unresectable disease as determined by the attending surgeon. Speciality of surgeon not reported. Postoperative platinum-based chemotherapy was administered in all women Residual tumour classed as: • NMRD: 115 (44%) • 1 mm to 5 mm: 50 (19%) • 6 mm to 10 mm: 33 13%) • 11 mm to 20 mm: 18 (7%) • > 20 mm: 43 (17%) Progression-free survival (PFS) was defined as the time interval from date of surgery to the date of the documented first recurrence of disease
Outcomes	At multivariate analysis, age greater than 60 years (P = 0.025), stage IV vs IIIC (P = 0.037) and any residual disease (P = 0.032) were shown to have an independent association with worse PFS Median estimated blood loss (range): 700 cc (50 to 6000) The median length of hospital stay was 9 days Median length of surgery (range): 270 minutes (70 to 480)
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival**

Peiretti 2010 *(Continued)*

	4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome. OS was defined as the time interval from date of surgery to the date of death or last follow-up 5. Adjustment for other prognostic factors (a-g): high risk Not reported in multivariate analyses. Only univariate results. 6. Statistical analysis and reporting (a-d): high risk Not reported in multivariate analyses **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome. PFS was defined as the time interval from date of surgery to the date of the documented first recurrence of disease. 5. Adjustment for other prognostic factors (a-g): high risk Age categorised. Multivariate model predicting PFS adjusted for age and FIGO stage. Unclear if ascites was included in multivariate model or not. 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; variable selection criteria for multivariate analyses unstated. Multivariate results (hazard ratios) PFS not displayed, only P values.
Notes	After a median follow-up of 29.8 months, PFS and overall median survival (OS) were 19.9 and 57.6 months respectively 92% of the women completed 5 or more cycles of platinum-based systematic chemotherapy At univariate analysis, factors significantly associated with decreased PFS included: age greater than median (N60 years), stage IV, presence of ascites N1000 cc, presence of diffuse peritoneal carcinomatosis and macroscopic residual disease

Peiretti 2012

Study characteristics	
Methods	Retrospective medical chart review
Participants	238 consecutive women who underwent rectosigmoid colectomy as part of cytoreductive surgery for ovarian cancer during the study interval were included Median age was 59.7 years (range: 22 to 85 years) FIGO stage IIC: 3 (1%); IIIA: 1(0.4%); IIIB: 2 (0.8%); IIIC: 174 (73%); IV: 58 (24%) Primary site disease: • Ovary: 230 (96%) • Fallopian tube: 4 (2%) • Peritoneal cancer: 4(2%) Tumour grade: • 1 to 2: 51 (22%) • 3: 184 (77%)

Peiretti 2012 *(Continued)*

	• N/A: 3 (1%) Histological subtype: • Serous: 200 (84%) • Endometrioid: 15 (6%) • Clear cell: 5 (3%) • Mixed: 18 (7%) Median ascites (range): 1500 cm^3 100 to 11,000) Italy (157) and USA (81)
Residual disease details	All operations were performed by gynaecologic oncologists Postoperative platinum-based chemotherapy was administered in all women • 62% underwent carbo-platinum and Taxol regimen • Doxorubicin liposomal, gemcitabine and topotecan were the other chemotherapeutic drugs used in association with platinum Complete cytoreduction was defined as no visible residual tumour at the completion of the primary operation. Reported categories for residual disease (mm) where as follows - no. of women (%): • NMRD: 99 (41%) • SVRD (1 mm to 10 mm): 106 (44%) • LVRD (> 10 mm): 32 (15%)
Outcomes	The risk factor significantly associated with decreased overall survival (OS) was the presence of any macroscopic residual disease at the end of surgery (P = 0.003) The median overall survival time from the time of surgery for all women was 55 months A statistically significant difference (P = 0.002) was observed in OS between the group with no macroscopic residual disease (median of 72 months) and the other women with any other gross residual disease (median of 42 months) Median estimated blood loss (range): 1000 cm^3 (200 to 8500) Intraoperative blood transfusion: 152 (64%) Postoperative blood transfusion: 150 (63%) Median length of hospitalisation (days): 10 (range: 4 to 24 days)
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. Sample consists of small subset (n = 3, 1%) of stage IIC participants. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): unclear risk Valid and reliable measurement of RD. Multicentre design may introduce heterogeneity in measurement of RD

Peiretti 2012 *(Continued)*

	Outcome level assessment: **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Definition of OS not provided but it usually has a standard definition 5. Adjustment for other prognostic factors (a-g): low risk Multivariate model predicting OS adjusted for age, stage, histology, grade and presence of ascites 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; variable selection criteria for multivariate analysis unstated **Outcome: progression-free survival** Not reported
Notes	Mean or median length of follow-up were not reported Among all groups of women 85% were able to complete at least 5 cycles of (platinum-based) systematic chemotherapy 50% of women recurred during the study period. Among them, 74% had a recurrence in the upper abdomen. 8% of the women presented with abdominal recurrence associated to pelvic disease. Only 5% of the women showed a relapse in the pelvis 14% of the women presented with distant metastases at the time of recurrence Both univariate and multivariate analyses including the following variables were performed: age, stage, histology, grade, presence of ascites and residual tumour at end of surgery, however no HR are presented in the study

Petrillo 2014

Study characteristics	
Methods	Single-centre retrospective of medical data (January 1995 to December 2010) retrieved from the electronic database of the Gynecologic Oncology Unit of the Catholic University of Rome and Campobasso
Participants	N = 322 women were admitted to the Gynecologic Oncology Unit of the Catholic University of Rome and Campobasso, with a diagnosis of advanced ovarian, tubal or peritoneal cancer. All these women were judged as having unresectable advanced disease after initial surgical exploration and submitted to NACT followed by IDS. ≤ 65 years: 226 (70.2%) > 65 years: 96 (29.8%) FIGO: IIIC – 251 (77.7%); IV – 72 (22.3%) Histology: serous – 264 (82%); other – 58 (18%) Tumour grade: G1 – 9 (2.7%); grade 2/3 – 313 (97.3%) Ascites: 247 (76.7%) Median CA-125 at diagnosis: 548 (range: 9 to 9999)

Petrillo 2014 *(Continued)*

	Carcinomatosis at diagnosis: 285 (88.5%) Within FIGO IV (n = 72) Presence of pleural effusion: 37 Metastasis in liver, spleen or lung: 34
Residual disease details	3 to 4 NACT cycles: 216 (82.3%) 6 NACT cycles: 57 (17.7%) NACT regimen: carboplatin alone – 51 (15.8%); carboplatin/paclitaxel or pegylated-liposomal doxorubicin (PLD) – 271 (84.2%) Pathological response to NACT: • Complete (cPR in cases with no macroscopic residual neoplastic cells in all the surgical specimens, including the adnexa): 21 (6.5%) • Microscopic response (without macroscopic lesions but with microscopic foci (maximum diameter ≤3 mm)): 104 (32.3%) • Macroscopic response (persistent macroscopic site of disease after NACT were classified as a macroscopic response): 197 (61.2%) Study did not provide a definition of optimal cytoreduction • NMRD: 236 (73.3%) • SVRD (0 cm to 1 cm): 36 (11.2%) • LVRD (RD > 1 cm): 50 (15.5%)
Outcomes	Overall survival defined as time elapsed between diagnosis and death or date of last follow-up (second half of 2012 in all women) Death from disease: 239 (74.2%) Median OS in those who had complete response (NMRD) from NACT: 72 months Median OS in those who had optimal response: 38 months Median OS in those who had suboptimal response: 29 months Multivariable Cox PH for OS adjusted for pathological response to NACT: • Residual tumour at IDS (RT = 0 vs RT ≤ 1 vs explorative laparotomy): X2 = 24.951, P = 0.001 Progression-free survival (PFS) calculated from the date of diagnosis to the date of first relapse or the date of the last follow-up (second half of 2012 in all women) Recurrences: 285 (88.2%) Median PFS in those who had complete response (NMRD) from NACT: 36 months Median PFS in those who had optimal response: 16 months Median PFS in those who had suboptimal response: 13 months Multivariable Cox PH for PFS adjusted for age, carcinomatosis at diagnosis, CA-125, pathological response to NACT: • Residual tumour at IDS (RT = 0 vs RT ≤ 1 vs explorative laparotomy): X2 = 39.716, P = 0.001 * No adjusted HR estimates provided for OS or PFS
Risk of bias (QUIPS)	1. Study participation (a-f): low risk

Petrillo 2014 *(Continued)*

	Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome; OS defined as time elapsed between diagnosis and death or date of last follow-up 5. Adjustment for other prognostic factors (a-g): high risk Unstated why explorative laparotomy is a category within the RD variable. Model predicting OS only adjusted for pathological response to NACT. 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; data driven based on P values of univariate associations. Results for multivariate analysis of OS not reported using hazard ratios. **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome; PFS calculated from the date of diagnosis to the date of first relapse or the date of the last follow-up 5. Adjustment for other prognostic factors (a-g): high risk Unstated why explorative laparotomy is a category within the RD variable. Model predicting PFS adjusted for Age, carcinomatosis at diagnosis, CA-125 and pathological response to NACT. 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; data driven based on P values of univariate associations. Results for multivariate analysis of PFS not reported using hazard ratios.
Notes	Median follow-up: 47 months (range: 3 to 181)

Phillips 2018

Study characteristics	
Methods	Single-centre retrospective study
Participants	N = 398 women undergoing interval debulking surgery (IDS) for stage 3 or 4 epithelial ovarian, tubal or peritoneal cancer (advanced ovarian cancer, AOC). All women were managed by subspecialty trained gynaecological oncologists at the Pan-Birmingham Gynaecological Cancer Centre (PBGCC), Birmingham, United Kingdom Mean age: 63.9 (95% CI 42.2 to 85.6)

Phillips 2018 *(Continued)*

	FIGO: III – 273 (68.6%); IV – 123 (31.4%) Histology: serous – 370 (93%); undifferentiated – 1 (0.3%); endometrioid – 1 (0.3%); carcinosarcoma – 12 (3%); mixed – 8 (2%); clear cell – 2 (0.5%); unknown – 4 (1%) Tumour grade: G1 – 13 (3.3%); G2 – 2 (0.5%); G3 – 374 (94%); unknown – 9 (2.3%) Disease site: ovary – 252 (63.3%); fallopian – 90 (22.6%); primary peritoneal: 56 (14.1%)
Residual disease details	≤ 4 NACT cycles: 231 (58%) • Group 1 (≤ 4 cycles) with 111 (48.1%) receiving standard treatment with 3 cycles of NACT and the remaining 120 (51.9%) receiving an additional cycle to facilitate timing of IDS ≥ 5 NACT cycles: 167 (42%) NACT regimen: • Carboplatin: 94 (23.6%) • Paclitaxel and carboplatin: 304 (76.4%) • Additional bevacizumab: 25 (6% Surgical complexity score: • Low (0 to 3): 263 (66.1%) • Inter (4 to 7): 89 (22.4%) • High (8+): 46 (11.6%) Median adjuvant CT after IDS: 3 cycles **'Optimal' cytoreduction** defined as SVRD or NMRD (RD 0 cm to 1 cm) (n = 310, 77.9%): • NMRD: 255 (64.1%) • RD greater than 0 cm but less than 1 cm (RD < 1): 55 (13.8%) • RD of 1 cm and above (RD ≥ 1): 88 (22.1%)
Outcomes	Median OS: 40.1 months Median OS in NMRD: 51.8 months Median OS in SVRD < 1 cm: 29.5 Median OS in LVRD ≥ 1 cm: 28.9 Multivariable Cox PH for OS adjusted for FIGO stage, chemotherapy regime (carbo/Taxol vs carboplatin): Within group 1 (≤ 4 cycles NACT; n = 231) SVRD < 1 cm (vs NMRD): HR 1.5723 (95% CI 0.928 to 2.664), P > 0.05; RD ≥ 1 (vs NMRD): HR 1.7709 (95% CI 1.069 to 2.933), P = 0.0264; SVRD < 1 cm (vs LVRD ≥ 1 cm): HR 0.8879 (95% CI 0.460 to 1.715), P > 0.05 Within group 2 (> 4 cycles NACT; n = 167) SVRD < 1 cm (vs NMRD): HR 2.781 (95% CI 1.663 to 4.650), P = 0.0001; LVRD ≥ 1 (vs NMRD): HR 2.6729 (95% CI 1.759 to 4.062), P < 0.00001; SVRD < 1 cm (vs LVRD ≥ 1 cm): HR 1.04 (95% CI 0.613 - 1.765), P > 0.05
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk

Phillips 2018 *(Continued)*

	Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Definition of OS not provided but it usually has a standard definition. 5. Adjustment for other prognostic factors (a-g): high risk Model predicting OS adjusted for FIGO stage, and chemotherapy regime (carbo/Taxol vs carboplatin) 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear how FIGO stage and chemotherapy regime were chosen to be in multivariate model **Outcome: progression-free survival** Not reported
Notes	Median BMI: 25

Polterauer 2012

Study characteristics	
Methods	Prospective, multicentre study (5 specialised European centres for gynaecologic oncology) Women enrolment between February 2005 and December 2008
Participants	226 women with epithelial ovarian cancer FIGO Stages IIA to IV in whom radical cytoreductive surgery was performed and standard chemotherapy with paclitaxel and carboplatin was applied. Women having received neoadjuvant chemotherapy followed by interval debulking were excluded Mean age 57.5 year (SD 11.9) FIGO stages II, III and IV: 15 (6.6%), 174 (76.9%) and 37 (16.4%); FIGO stages IIIC and IV: 198 women (87.6%) Histological type serous/other: 194/32 NMRD: 69.4% SVRD (≤ 1 cm): 87.2% (NB: this category also includes NMRD) Austria
Residual disease details	Residual disease was defined as: Any RD (SVRD (≤ 1 cm) or LVRD (> 1 cm) Complete debulking (NMRD)
Outcomes	3-year OS (unadjusted) with NMRD: 72.4%; minimal RD: 65.8%; gross RD: 45.2%

Polterauer 2012 *(Continued)*

	Subgroup analysis of stages IIIC and IV: 3-year OS (unadjusted) with NMRD 69.7% (SE 5.3%); any RD 53.6% (SE 8.3%) (P = 0.003) HR (apparently for 'Any RD' vs 'No RD', adjusted for FIGO-stage, histological grade, histological type and age) 1.4 (95% CI 1.0 to 2.1) "Multivariable survival analysis revealed residual tumour size (p=0.04) and older women age (p =0.02) as independent prognosticators for impaired overall survival. Complete cytoreduction was predictive for a higher rate of treatment response (p=0.001) and was associated with prolonged progression-free and overall survival (p<0.001 and p=0.001)." HR for PFS (apparently for 'Any RD' vs 'NMRD', adjusted for FIGO stage, histological grade, histological type and age) 1.6 (95% CI 1.3 to 2.1) Univariate survival analysis of categorical variables by the log-rank test. Multiple forward stepwise Cox regression analysis.
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): unclear risk Valid and reliable measurement of RD. Multicentre design may introduce heterogeneity in measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): low risk Cohort was recruited with objective to identify and verify clinical and molecular prognostic/predictive factors in ovarian cancer. Possible confounding prognostic factors would also have been included in study. Multivariate model for OS adjusted for FIGO stage, histological grade, histology subtype and age 6. Statistical analysis and reporting (a-d): unclear risk No conceptual framework; variable selection criteria for multivariate analysis unstated. **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): low risk Cohort was recruited with objective to identify and verify clinical and molecular prognostic/predictive factors in ovarian cancer. Possible confounding prognostic factors would also have been included in study. Multivariate models for PFS adjusted for FIGO stage, histological grade, histology subtype and age 6. Statistical analysis and reporting (a-d): unclear risk

Polterauer 2012 *(Continued)*

	No conceptual framework; variable selection criteria for multivariate analysis unstated
Notes	Source of funding: the European commission (FP6 Specific Targeted Research or Innovation Project) Declaration of interest: none declared Median follow-up: 25.0 months (range: 1 to 49) Retrospective non-randomised study. Blinding not reported (but not applicable). Adjusted HRs are derived from a prognostic model. No details on how modelling was performed, but this seems to have been done based on significance testing (and not on including putative confounders in the analysis, irrespective of statistical significance). Women and disease characteristics not reported according to debulking status. NB: possible overlap with Hofstetter 2013.

Shibutani 2020

Study characteristics	
Methods	The purpose of this study was to determine the optimal regimen of neoadjuvant chemotherapy (NAC) for advanced epithelial ovarian, fallopian tube and peritoneal cancers Retrospective study of data from the Hyogo Cancer Center between January 2006 and December 2015. Japan
Participants	171 patients with advanced epithelial ovarian, fallopian tube or primary peritoneal cancer who underwent dose-dense tri-weekly administration of paclitaxel and carboplatin (TC) or TC as NAC followed by IDS The median age of patients was 61 (range 35 to 79) years Performance status of patients: 0 for 47 patients (27%); 1 for 79 patients (46%); 2 for 38 patients (22%); and 3 for 7 patients (4%)
Residual disease details	Patients who underwent NAC followed by interval debulking surgery The median number of NAC cycles was 4 (range 2 to 10). The total number of cycles during the first treatment was 7 (range 4 to 16). Dose-dense paclitaxel and carboplatin (TC) was administered in 101 patients (59%); tri-weekly TC was administered 70 patients (41%) **Residual disease groups:** SVRD < 1 cm: 150 (88%) LVRD > 1 cm: 21 (12%)
Outcomes	Overall survival and progression-free survival
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement.

Shibutani 2020 *(Continued)*

	3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome. The overall survival was calculated from the date of the first chemotherapy to the date of death or last contact. 5. Adjustment for other prognostic factors (a-g): high risk Only univariate analysis of OS. Not included in multivariate analyses. 6. Statistical analysis and reporting (a-d): high risk No multivariate model predicting OS despite there being one for PFS **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome. Progression-free survival was calculated from the date of the first chemotherapy to the date of death or disease progression. 5. Adjustment for other prognostic factors (a-g): low risk HR for PFS was adjusted for age (< 61 vs ≥ 61), PS (0 to 1 vs 2 to 3), stage (III vs IV), disease (ovary vs others), histology, residual disease, NAC cycles and NAC regimens in multivariable Cox model 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; variable selection criteria for multivariate analysis unstated No multivariate model predicting OS despite there being one for PFS
Notes	The median observation period was 41 (range 4 to 138) months Median progression-free survival was 21 (95% CI 18 to 23) months and 15 (95% CI 13 to 17) months in the dose-dense TC and conventional TC group, respectively (HR 0.69, 95% CI 0.46 to 0.96; P = 0.02) The median overall survival was 59 (95% CI 46 to 72) and 40 (95% CI 32 to 57) months in the dose-dense TC group and conventional TC group (HR 0.72, 95% CI 0.48 to 1.06; P = 0.09) Multivariate analysis for progression-free survival demonstrated that dose-dense TC represented an independent prognostic factor (HR 0.70, 95% CI 0.50 to 0.99; P = 0.04). PFS multivariate prognostic factors were as follows: FIGO stage (HR 0.68, 95% CI 0.48 to 0.96 (table says 0.90); P = 0.03) and residual disease at IDS (HR 0.55, 95% CI 0.34 to 0.96 (table says 0.90 and this appears to be the correct estimate when log estimates are entered; P = 0.02). Also when reference is changed this estimate is: HR 1.82 (95% CI 1.12 to 2.97).

Shim 2016

Study characteristics	
Methods	Retrospective study
Participants	276 women with FIGO stage III or IV ovarian cancer consecutively treated

Shim 2016 *(Continued)*

	Median age at diagnosis was 54 years (range: 20 to 80 years) 258 (93.5%) women received postoperative platinum-based chemotherapy South Korea
Residual disease details	Speciality of surgeon not reported Surgery followed by platinum-taxane chemotherapy The 25%, 50% and 75% quartiles of intervals from surgery to start of chemotherapy were 18, 22 and 28 days, respectively
Outcomes	Time to chemotherapy (TTC) was analysed and correlated with outcome The following were significant prognostic factors for progression-free survival in multivariate analysis: • TTC (≤ 28 vs > 28 days; HR 1.578, 95% CI 1.057 to 2.355) • Complete debulking with NMRD (HR 0.419, 95% CI 0.274 to 0.640) • Preoperative albumin level (HR 0.549, 95% CI 0.382 to 0.791)
Risk of bias (QUIPS)	1. Study participation (a-f): high risk Abstract only therefore insufficient information on study participation 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** Not reported **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Definition of PFS not provided but it usually has a standard definition 5. Adjustment for other prognostic factors (a-g): high risk Time to chemotherapy arbitrarily categorised. Model predicting PFS adjusted for time to chemotherapy and preoperative albumin level. 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear on reasons why the particular specific set of variables were selected for multivariate model. PFS used as outcome but no overall survival.
Notes	Findings are from an abstract OS not reported Mean and median length of follow-up were not reported

Shim 2016 *(Continued)*

	Although delayed TTC (> 28 days) did not possess prognostic significance in women without postoperative residual disease (n = 94), it significantly correlated with progression-free survival in women with postoperative RD (n = 164, HR 1.893, 95% CI 1.209 to 2.962)

Stoeckle 2014

Study characteristics	
Methods	Single-centre retrospective study
Participants	N = 118 women diagnosed with primary ovarian carcinoma, epithelial cell type (stages IIIC with carcinomatosis and IV) who were treated by NACT + late IDS (after 6 cycles) in the taxane/platinum period (1998 to 2010) Median age: 64 (range: 37 to 88) FIGO: IIIC – 82 (69%); IV – 36 (31%) Histology: serous – 111 (94%); non-serous – 7 (6%) Had lymph node assessment: 105 (89%) Median node count: 32 (range: 4 to 81) Lymph node involvement • Positive: 56 (47%) • Negative: 49 (42%) • N/A: 13 (11%)
Residual disease details	All women had sampling biopsy. • Laparoscopy: 77 (65.3%) • Diagnostic laparotomy: 17 (14.4%) Median NACT cycles: 6 (range: 5 to 13) NACT regimen • Carboplatin – 4 (3.4%); • Paclitaxel and carboplatin – 114 (96.6%) All IDS performed by 2 surgeons (co-authors on paper) with experience in ovarian cancer surgery Resection categories (other than peritoneal stripping) • Salpingo-oophorectomy: 109 (92%) • Total abdominal hysterectomy: 109 (92%) • Omentectomy: 113 (96%) • Appendectomy: 102 (86%) • Pelvic lymph node dissection: 104 (88%) • Aortic lymph node dissection: 93 (79%) • Bowel surgery: 32 (27%) • Other organ resection (spleen, liver, small bowel etc.): 17 (14%) Number of resection categories Median: 6

Stoeckle 2014 *(Continued)*

	Range: 0 to 8 Standard surgery: 54 (46%) Extended surgery: 64 (54%) **'Optimal' cytoreduction** defined as RD < 1 cm (n = 111, 94%) • NMRD (referred to in study as RD0): 80 (68%) • SVRD (RD 0.1 cm to 1 cm): 31 (26%) • LVRD (RD ≥ 1 cm): 7 (6%) * In multivariable analysis, it is NMRD vs RD > 0 cm
Outcomes	Overall survival defined as time from date of initial diagnosis to date of death of any cause Median OS: 42 months Median OS in no macroscopic RD group (RD 0 cm): 50 months Median OS in RD > 0: 38 months Multivariable Cox PH for OS adjusted for tumour grade, WHO performance status, ASA, bowel surgery (yes/no), FIGO stage: RD > 0 cm vs NMRD: HR 2.2 (95% CI 1.2 to 4.0), P = 0.01 Progression-free survival (PFS) was calculated from the date of initial diagnosis to date of progression. Progression was defined as locoregional or metastatic recurrences after complete remission or progression of disease in women without complete remission. Median PFS: 17.2 months No multivariable analysis for PFS Death within 30 days of surgery: 2 (1.7%)
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome. Overall survival defined as time from date of initial diagnosis to date of death of any cause. 5. Adjustment for other prognostic factors (a-g): unclear risk Model predicting OS adjusted for tumour grade, WHO performance status, ASA, bowel surgery (yes/no), FIGO stage 6. Statistical analysis and reporting (a-d): unclear risk

Stoeckle 2014 *(Continued)*

	No conceptual framework; data driven based on P values of univariate associations **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome. Progression was defined as locoregional or metastatic recurrences after complete remission or progression of disease in women without complete remission. 5. Adjustment for other prognostic factors (a-g): high risk Model predicting PFS was not adjusted for any other prognostic factor 6. Statistical analysis and reporting (a-d): high risk Model predicting PFS was not adjusted for any other prognostic factor
Notes	Median follow-up: 37 months ASA score: • 1: 35 (30%) • 2 to 3: 83 (70%) WHO performance status • 0 to 1: 80 (68%) • 2 to 3: 38 (32%) At IDS, 96 (81%) presented with visible tumour. Median tumour size was 2 mm. Median length of hospital stay (all women): 10 (2 to 44) Median length of stay (women with complications): 16 (range: 7 to 44) Major morbidity was defined as a complication requiring a prolonged hospital stay (more than 10 days), re-hospitalisation or reoperation (by surgery or interventional imaging) needing correction by major medication (e.g. prolonged IV antibiotics or blood transfusion (5 packed red blood cells), or causing death during the first postoperative month 21 women (18%) had major complications, for a total of 24 major complications • Infection: 11 • Blood loss needing transfusion > 5 PRBC: 7 • Thromboembolic event: 2 • Cerebrovascular accident: 1 • Myocardial infarction: 1 • Bowel obstruction: 1 • Chylous ascites: 1 Rehospitalisation: 10 women Reoperation by surgery or imaging techniques: 8 women

Tewari 2016

Study characteristics	
Methods	Retrospective analysis

Tewari 2016 *(Continued)*

Participants	1718 women with newly diagnosed International Federation of Gynecology and Obstetrics stage III and IV ovarian, peritoneal or fallopian tube carcinoma were included in the analysis Median age (years): microscopic (58.5); optimal (60.1); suboptimal (60.2) Performance status - frequency (%): • Normal, asymptomatic: 848 (49.3%) • Symptomatic, ambulatory: 745 (43.4%) • Symptomatic, in bed < 50%: 125 (7.3%) Top-level FIGO stage: III: 1241 (72.2%); IV: 477 (27.8%) Histology: serous: 1477 (86%); mixed epithelial: 76 (4.4%); endometrioid: 56 (3.3%); clear-cell/mucinous: 60 (3.5%); other: 24 (1.4%) Ascites: no: 346 (20.1%); yes: 1372 (79.9%) Progression-free survival status: censored: 268 (15.6%); progression or death: 1450 (84.4%) Overall survival status: censored: 840 (48.9%); death: 878 (51.1%) USA
Residual disease details	Speciality of surgeon was not reported Primary cytoreductive surgery followed by platinum based chemotherapy Treatment arms: frequency (%) • I (standard chemotherapy): 580 (33.8%) • II (concurrent bevacizumab): 570 (33.2%) • III (extended bevacizumab): 568 (33%) Residual disease, n (%) • NMRD: 85 (4.9%) • SVRD (≤ 1 cm): 701 (40.8%) • LVRD (> 1 cm): 932 (54.2%)
Outcomes	Overall survival: HR adjusted for: TSIC = 15 days: ≤ 1 cm (AHR 1.41, 95% CI 0.77 to 2.58); > 1 cm (AHR 1.87, 95% CI 1.05 to 3.31) Residual = micro, 40 days: • Race/ethnicity = White (AHR 1.27, 95% CI 1.15 to 1.40) • Race/ethnicity = Asian (AHR 1.51, 95% CI 1.27 to 1.80) • Race/ethnicity = Black (AHR 1.18, 95% CI 1.00 to 1.40) • Race/ethnicity = Hispanic (AHR 1.18, 95% CI 0.97 to 1.43) • Race/ethnicity = other (AHR 1.41, 95% CI 1.15 to 1.74) Residual ≤ 1 cm, 40 days: • Race/ethnicity = Asian (AHR 1.17, 95% CI 1.01 to 1.35) Residual > 1 cm, 40 days • Race/ethnicity = Asian (AHR 1.24, 95% CI 1.07 to 1.44) Histology • Serous: (AHR 1 - referent) • Mixed epithelial: (AHR 1.33, 95% CI 0.97 to 1.84)

Tewari 2016 *(Continued)*

- Endometrioid: (AHR 0.70, 95% CI 0.44 to 1.11)
- Clear-cell/mucinous: (AHR 4.97, 95% CI 2.46 to 10.05)
- Other: (AHR 1.14, 95% CI 0.73 to 1.78)

Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Definition of OS not provided but it usually has a standard definition 5. Adjustment for other prognostic factors (a-g): low risk Arbitrary dichotomisation of time from surgery to chemotherapy. Multivariate model predicting OS adjusted for age, race, performance status, tumour grade, FIGO stage, histology, ascites, CA-125, time from surgery to chemotherapy and interaction terms 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria for multivariate analysis **Outcome: progression-free survival** Not reported
Notes	At 15 days, time to initiation of chemotherapy does not increase the risk of death for any women, whereas at 40 days most women have an increased risk of death. This represents a change-point in increasing time at which some women start to become affected negatively.

Tseng 2018

Study characteristics	
Methods	Retrospective cohort study
Participants	978 women with stage IIIB to IV ovarian, fallopian tube or primary peritoneal carcinoma Median age was 61 years (range: 19 to 95 years) FIGO stage - n (%): - IIIB: 33 (3%) - IIIC: 761 (78%) - IV: 184 (19%) Histology - n (%):

Tseng 2018 *(Continued)*

	• Serous: 869 (89%) • Other: 109 (11%) Estimated blood loss: 700 mL (range: 5 mL to 8000 mL) Median hospital length of stay was 8 days (range 1 to 22 days) USA
Residual disease details	Speciality of surgeon not reported All women underwent primary debulking surgery followed by intraperitoneal (IP) chemotherapy in (n = 949, 99%) Residual disease was classed as follows: • NMRD (defined as complete gross resection (CGR) in study) - 0 mm: 408 (42%) • SVRD (1 to 10 mm): 378 (39%) • LVRD (> 10 mm): 192 (20%)
Outcomes	Multivariable analysis of factors associated with PFS adjusted for PDS-year group Residual disease: • NMRD: (AHR: reference) • SVRD: (AHR 1.393, 95% CI 1.174 to 1.654) • LVRD: (AHR 1.921, 95% CI 1.547 to 2.386) Multivariable analysis of factors associated with OS adjusted for PDS-year group Residual disease: • NMRD: (AHR: reference) • SVRD: (AHR 1.36, 95% CI 1.118 to 1.653) • LVRD: (AHR 1.751, 95% CI 1.378 to 2.224) Median operative time was 280 minutes (range, 36 to 893 minutes) Median length of hospital stay (LOS) was 8 days (range: 1 to 22 days) 30-day all-cause mortality was 0.4% (4 deaths)
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome

Tseng 2018 *(Continued)*

	5. Adjustment for other prognostic factors (a-g): low risk Multivariate models for OS adjusted for age, albumin, FIGO stage, ASA score, histology, BRCA, tumour index, and postoperative intraperitoneal chemotherapy 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria for multivariate analysis **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): low risk Multivariate models for PFS adjusted for age, albumin, FIGO stage, ASA score, histology, BRCA, tumour index and postoperative intraperitoneal chemotherapy 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria for multivariate analysis
Notes	Median follow-up for the entire cohort was 77.7 months (range: 1.3 to 198 months)

Van Geene 1996

Study characteristics	
Methods	Prospective cohort study: the 2 groups were defined from data collected prospectively at laparotomy. All women with ovarian cancer referred to the departments of gynaecological oncology at 2 hospitals between 1981 and 1989 were entered into prospective surgical studies.
Participants	During the 8-year period in the study a total of 256 women with previously untreated primary EOC were referred for consideration of surgery and chemotherapy. 37 women with stage II disease were excluded from this analysis leaving 219 women with stage III to IV disease to form the basis of the study. Median age at study entry was 57 years (range: 24 to 75 years) 180 (82%) and 39 (18%) women had FIGO stage III and IV disease respectively Histological cell type was as follows: serous: 134 (61%), endometrioid: 34 (15%), mucinous: 32 (15%), clear cell: 7 (3%), undifferentiated: 12 (6%) 50 (25%) women had tumour grade classified as being well, 68 (34%) had grade as moderate, 75 (37%) had poor grade and in 9 (4%) women the grade was unknown 101 (46%) women had GOG performance status 0, 94 (43%) had status 1, 23 (10.5%) women had status 2 and for 1 (0.5%) woman their status was unknown Mode of spread was as follows: bulky: 100 (46%), spreading: 119 (54%) UK
Residual disease details	Reported categories for residual disease were as follows: 1. RD < 2 cm: 92 (42%) of which 15 were deemed to have had NMRD 2. LVRD (> 2 cm): 127 (58%)

Van Geene 1996 *(Continued)*

	All women received cis-platinum containing chemotherapy at the dose of 75 mg/m^2 up to a total of 6 courses depending on response and toxicity
Outcomes	Overall survival: HR adjusted for performance status and pattern of spread using Cox model: > 2 cm vs < 2 cm: HR 1.83, P < 0.0001 We requested the exact P value and 95% CI from the study authors but the data were no longer available. Table 4 is confusing as no macroscopic RD and less than 2cm RD was compared to > 2 cm. This was grouped in table 2.
Risk of bias (QUIPS)	1. Study participation (a-f): unclear risk There was insufficient information to permit judgement 2. Study attrition (a-e): unclear risk There was insufficient information to permit judgement 3. Prognostic factor measurement (a-f): unclear risk There was insufficient information to permit judgement **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): unclear risk There was insufficient information to permit judgement 5. Adjustment for other prognostic factors (a-g): unclear risk There was insufficient information to permit judgement 6. Statistical analysis and reporting (a-d): unclear risk There was insufficient information to permit judgement **Outcome: progression-free survival** Not reported
Notes	The 2 groups were defined from data collected prospectively at laparotomy. Women with small-volume (≤ 0.5 cm) but widespread disease (> 10 metastatic nodules) were assigned to the seedling group and women with large-volume disease (> 0.5 cm) spread outside the pelvis were assigned to the bulky disease group. Optimal debulking, i.e. residual disease less than 2 cm, was achieved in 92 (42%) of the women with similar rates between the 2 groups (P = 0.09). Complete macroscopic clearance was achieved in only 15 women, all of which were in the bulky spread group. Complete macroscopic clearance (NMRD) was achieved in only 15 women, all of which were in the bulky spread group.

Wimberger 2010

Study characteristics	
Methods	Retrospective data set review (retrieved from 3 prospective, randomised phase III trials: AGO-OVAR (OVAR-3/-5/-7))

Wimberger 2010 *(Continued)*

Participants	Cohort of women from three prospective, randomised phase III trials: AGO-OVAR (OVAR-3/-5/-7) in between 1995 and 2002 Previously untreated epithelial ovarian cancer FIGO stage IV, at least 18 years of age and required to have adequate haematologic, renal and hepatic function, defined as follows: absolute neutrophil count (ANC) of at least 1.5 × 109 cells/L, platelet count of at least 100 × 109 cells/L, serum creatinine and bilirubin of no more than 1.25 × upper normal limit N = 573, all FIGO stage IV disease: malignant pleural effusion = 214 (37.3%), parenchymal hepatic metastases = 146 (25.5%), other sites disease = 213 (37.2%) Median age was 59 years (range 19 to 83); age < 50 (17.6%), 50 to 65 (59.5%), > 65 (22.9%) ECOG performance status: 0 (28.2%), 1 (54.6%), 2 (17.2%) Histological subtypes; serous (68.2%), endometrioid: (6.9%), mucinous (16.0%) Peritoneal carcinomatosis: yes (87.8%), no (12.2) France and Germany
Residual disease details	Residual disease were defined as: • NMRD (12.3%) • SVRD (1 to 10 mm) (29.3%) • LVRD (> 10 mm) (58.4%) Women were randomly assigned to one of two treatment arms consisting of either carboplatin or cisplatin and paclitaxel, or a combination of carboplatin and paclitaxel versus the same combination with epirubicin or topotecan. All women were scheduled to receive at least 6 courses of platinum-taxane intravenously every 3 weeks.
Outcomes	**Women with stage IV** **Kaplan-Meier** Median OS (unadjusted) of NMRD 54.6 months, SVRD 25.8 months, LVRD > 1 cm 23.9 months Median PFS (unadjusted) of NMRD 19.1 months, SVRD 13.6 months, LVRD (> 1 cm) 11.3 months **Multivariant analysis for OS:** SVRD vs NMRD: HR 1.87 (95% CI 1.21 to 2.89) LVRD (> 1 cm) vs NMRD: HR 2.13 (95% CI 1.40 to 3.23) **Multivariate analysis for PFS:** SVRD vs NMRD: HR 1.51 (95% CI 1.05 to 2.19) LVRD (> 1 cm) vs NMRD: HR 1.82 (95% CI 0.28 to 2.59) HRs adjusted for age, performance status, histological type, presence of peritoneal carcinomatosis and multiple sites (Y/N)
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement.

Wimberger 2010 *(Continued)*

	3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): unclear risk Age arbitrarily categorised. Multivariate model for OS adjusted for age, ECOG, histology, peritoneal carcinomatosis and number of stage IV disease sites 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria for multivariate analysis **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): unclear risk Age arbitrarily categorised. Multivariate model for PFS adjusted for age, ECOG, histology, peritoneal carcinomatosis and number of stage IV disease sites 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear of variable selection criteria for multivariate analysis
Notes	All women with stage IV disease in 3 RCTs: OVAR-3 trial (1995 to 1997): 69 women received carboplatin-paclitaxel (7 women had complete resection) 64 women received cisplatin-paclitaxel (6 women had complete resection) OVCAR-5 trial (1997 to 1999: 112 carboplatin-paclitaxel (14 complete resection, 61 LVRD > 1 cm) 106 carboplatin-paclitaxel-epirubicin (12 complete resection, 63 LVRD > 1 cm) OVCAR-7 trial (1999 to 2002): 104 carboplatin-paclitaxel (15 complete resection) 118 carboplatin-paclitaxel-topotecan (15 complete resection) The difference in proportion of women with zero residual disease in all 3 trials is not statistically significant (OVAR-3, P = 0.88, OVAR-5 P = 0.79 and OVAR-7, P = 0.71). No significant trend difference in women recruited during the different time period. No relation between residual disease and the number of applied chemotherapy cycles. Therefore, all 3 trials were considered sufficiently similar to be combined for this study and analysis. Median OS was statistically reduced in FIGO stage IV 26.1 months compared to stage IIIC

Winter 2007

Study characteristics

Winter 2007 *(Continued)*

Methods	The current study was a retrospective review of data from women treated with platinum and paclitaxel combination chemotherapy on one of 6 prospective randomised clinical trials conducted by GOG: protocols 111, 114, 132, 152, 158 and 172 GOG 111: included LVRD (> 1 cm) stage III/IV EOC (eligible women = 123) GOG 114: included SVRD (< 1 cm) stage III EOC (eligible women = 226) GOG 132: included LVRD (> 1 cm) stage III/IV EOC (eligible women = 147) GOG 152: included LVRD (> 1 cm) stage III EOC (eligible women = 397) GOG 158: included LVRD (> 1 cm) stage III EOC (eligible women = 792) GOG 172: included SVRD (≤ 1 cm) stage III EOC (eligible women = 210)
Participants	Data from 1895 women with stage III invasive EOC who underwent primary surgical cytoreduction followed by paclitaxel/platinum chemotherapy, while participating in one of six GOG clinical trials, was analysed for the present study The median age was 57 years (range: 16 to 86 years) All 1895 women had FIGO stage III Histological cell type was as follows: serous: 1392 (73.5%), endometrioid: 166 (8.8%), mucinous: 34 (1.8%), mixed epithelial: 142 (7.5%), adenocarcinoma unspecified: 49 (2.6%), clear cell: 62 (3.3%), undifferentiated: 26 (1.4%), other: 24 (1.3%) 179 (9.5%) women had tumour grade 1, 719 (37.9%) had grade 2 and 997 (52.6%) women had tumour grade 3 Tumour grade details: 1: 179 (9.5%), 2: 719 (37.9%), 3: 997 (52.6%) Ethnicity details: White: 1669 (88.1%), African-American: 111 (5.9%), other: 115 (6.1%)
Residual disease details	Reported categories for residual disease were as follows: 1. NMRD: 437 (23.1%) 2. SVRD (0.1 cm to 1 cm): 791 (41.7%) 3. LVRD (> 1 cm): 667 (35.2%) Optimal was not defined, yet women were divided into 3 groups for analysis, based on RD status (as above). The following chemotherapy schedules were given in the 6 trials: • GOG 111: IV paclitaxel 135 mg/m^2, cisplatin 75 mg/m^2, 6 cycles • GOG 114: IV paclitaxel 135 mg/m^2, cisplatin 75 mg/m^2, 6 cycles • GOG 132: IV paclitaxel 135 mg/m^2, cisplatin 75 mg/m^2, 6 cycles • GOG 152: IV paclitaxel 135 mg/m^2, cisplatin 75 mg/m^2, 6 cycles ± interval debulking • GOG 158: IV paclitaxel 135 mg/m^2 (24 hours), cisplatin 75 mg/m^2, 6 cycles or IV paclitaxel 175 mg/m^2 (3 hours), carboplatin AUC 7.5, 6 cycles • GOG 172: IV paclitaxel 135 mg/m^2, cisplatin 75 mg/m^2, 6 cycles
Outcomes	Overall survival and progression-free survival: HR adjusted for age (discrete), race, GOG performance status, histology and tumour grade using Cox model: SVRD vs NMRD: HR 2.11 (95% CI 1.78 to 2.49), P < 0.001 and HR 1.96 (95% CI 1.70 to 2.26), P < 0.001 for OS and PFS respectively LVRD (> 1 cm) vs NMRD: HR 2.47 (95% CI 2.09 to 2.92), P < 0.001 and HR 2.36 (95% CI 2.04 to 2.73), P < 0.001 for OS and PFS respectively

Winter 2007 *(Continued)*

Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): unclear risk Age arbitrarily categorised. Model predicting OS adjusted for age, race, GOG performance status, histology, and tumour grade 6. Statistical analysis and reporting (a-d): unclear risk In methods, authors reported that all variables considered as potential prognostic factors were included in multivariate analyses, suggesting some conceptual framework **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): unclear risk Age arbitrarily categorised. Model predicting PFS adjusted for age, race, GOG performance status, histology, and tumour grade. 6. Statistical analysis and reporting (a-d): unclear risk In methods, authors reported that all variables considered as potential prognostic factors were included in multivariate analyses, suggesting some conceptual framework.
Notes	1505 recurrences and 1323 deaths were identified during a median follow-up period of 43 months: The median PFS was 17.1 months (95% CI 16.4 to 17.8 months) The median OS was 45.3 months (95% CI 43.0 to 47.7 months) **PFS for disease residual:** NMRD: N = 437, PFS was 33.0 months, 0.1 cm to 1.0 cm: N = 791, PFS) was 16.8 months, LVRD (> 1 cm): N = 667, PFS was 14.1 months, P < 0.001 **OS for disease residual:** NMRD: N = 437, OS was 71.9 months, SVRD: N = 791, OS was 42.4 months, LVRD (> 1.0 cm): N = 667, OS was 35.0 months, P < 0.001 Increasing age was associated with decreased PFS and OS. Median PFS and OS were shorter for women with a performance status (PS) of 1 or 2 when compared with those with a PS of 0. No difference in median PFS was evident between PS 1 and PS 2 women, whereas the difference in median OS between the

Winter 2007 *(Continued)*

	same groups was observed. Based on tumour histology, women with endometrioid histology had improved clinical outcomes compared with those with serous tumours. Women with mucinous or clear-cell tumours had decreased PFS and OS. Women with mucinous cell type had a median OS of only 15 months compared with 24, 45 and 56 months for clear-cell, serous and endometrioid cell types, respectively. Women with NMRD had the longest PFS and OS 33 and 72 months, respectively compared with women with any gross residual disease. The differences in median PFS and OS between the SVRD and LVRD (> 1 cm) groups were also evident, albeit small (3 months in median PFS and 7 months in median OS). Women with grade 2 or 3 tumours were associated with decreased PFS and OS. Race was not significantly associated with PFS or OS.

Winter 2008

Study characteristics	
Methods	Retrospective review of 4 RCTs. The current study was a retrospective review of data from women with stage IV EOC treated with platinum and paclitaxel combination chemotherapy on one of four prospective randomised clinical trials conducted by the GOG: protocols 111, 132, 152 and 162
Participants	360 women with stage IV invasive EOC who underwent primary surgical cytoreduction followed by paclitaxel/platinum chemotherapy while participating in one of four GOG clinical trials. The median age of women was 59 years (range: 24 to 86 years) 317 (88%) women were white, 28 (8%) were black and 15 (4%) were of other ethnic origin 97 (27%) had GOG performance status 0, 203 (56%) had status 1 and 60 (17%) had status 2 24 (7%) women had tumour grade 1, 112 (31%) grade 2 and 224 (62%) had grade 3 disease Histology was as follows: serous 268 (74.5%), endometrioid 28 (8%), mucinous 7 (2%), clear cell 12 (3%), adenocarcinoma unspecified 9 (2.5%), mixed epithelial 22 (6%), undifferentiated 9 (2.5%), other 5 (1.5%). The median residual tumour size was 3 cm (range 0.0 to 40.0) Stage IV disease site was as follows: distant: 45 (12.5%), parenchymal liver: 64 (17.75%), pleural effusion: 172 (47.75%), subcutaneous: 32 (9%), others: 3 (1%), multiple sites: 44 (12%)
Residual disease details	The maximum diameter of residual tumour that was used to define optimal cytoreduction: 1 cm (in original RCTs). All 4 RCTs included suboptimal disease (> 1 cm). Residual disease was noted as follows: • NMRD: 29 (8%) • SVRD of 0.1 cm to 1 cm: 78 women (22%) • LVRD of 1.1 cm to 2 cm: 50 women (14%) • LVRD of 2.1 cm to 3 cm: 40 women (11%) • LVRD of 3.1 cm to 4 cm: 30 women (8.25%) • LVRD of 4.1 cm to 5 cm: 44 women (12%) • LVRD of 5.1 cm to 6 cm: 30 women (8.25%) • LVRD larger than 6 cm: 59 women (16.5%) 'Optimal' cytoreduction was defined as RD < 1 cm and a sensitivity analysis was performed defining RD as < 2 cm

Winter 2008 *(Continued)*

	All women were treated with primary surgical cytoreduction and 6 cycles of a 24-hour infusion of intravenous paclitaxel 135 mg/m^2, followed by intravenous cisplatin 75 mg/m^2

Outcomes	• Overall survival: HR adjusted for several prognostic categories
	• Optimal: NMRD:
	  ◦ SVRD (< 1 cm) vs NMRD: HR 1.93 (95% CI 1.17 to 3.20)
	  ◦ 1 cm to 5cm vs NMRD: HR 1.83 (95% CI 1.14 to 2.94)
	  ◦ > 5 cm vs NMRD: HR 2.72 (95% CI 1.65 to 4.47)
	• Optimal: SVRD (≤ 1.0 cm):
	  ◦ LVRD (> 1 cm) HR 1.30 (95% CI 1.00 to 1.59)
	• Optimal: ≤ 2 cm RD:
	  ◦ LVRD (> 2 cm) HR 1.17 (95% CI 0.92 to 1.49)
	• Progression-free survival: HR adjusted for several prognostic categories
	• Optimal: NMRD:
	  ◦ SVRD (< 1 cm) vs NMRD: HR 1.99 (95% CI 1.24 to 3.18)
	  ◦ 1 cm to 5cm vs NMRD: HR 2.15 (95% CI 1.38 to 3.34)
	  ◦ > 5 cm vs NMRD: HR 2.96 (95% CI 1.86 to 4.71)
	• Optimal: SVRD (≤ 1 cm) RD:
	  ◦ LVRD (> 1 cm) HR: 1.49 (95% CI 1.16 to 1.92)
	• Optimal: ≤ 2.0 cm RD:
	  ◦ LVRD (> 2 cm) HR: 1.27 (95% CI 1.01 to 1.61)

Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): unclear risk Multivariate model for OS adjusted for histology and stage IV disease site 6. Statistical analysis and reporting (a-d): unclear risk In methods, authors reported that all variables considered as potential prognostic factors were included in multivariate analyses, suggesting some conceptual framework. However, age, race, GOG PS and tumour grade were excluded secondary at univariate analysis due to their P values falling above significance threshold **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome 5. Adjustment for other prognostic factors (a-g): unclear risk

Winter 2008 *(Continued)*

	Multivariate model for PFS adjusted for histology and stage IV disease site 6. Statistical analysis and reporting (a-d): unclear risk In methods, authors reported that all variables considered as potential prognostic factors were included in multivariate analyses, suggesting some conceptual framework. However, age, race, GOG PS and tumour grade were excluded secondary at univariate analysis due to their P values falling above significance threshold
Notes	The median length of follow-up was 28 months When evaluating the association of clinicopathologic factors with residual disease status, there was no difference between the RD groups and demographic, clinical and pathologic factors Stage IV site did not seem to have significant association with RD group distributions

Zhang 2018

Study characteristics	
Methods	Single-centre, retrospective study undertaken on women treated between January 2003 and December 2013, at the Department of Gynecology, Weifang Yidu Central Hospital, China
Participants	N = 200 women diagnosed with stage IIIC to IV invasive ovarian, fallopian tube or peritoneal high-grade serous carcinoma, who were treated with platinum-based NAC followed by IDS and adjuvant chemotherapy. Median age: 61 (range: 38 to 80) FIGO: IIIC – 169 (84.5%); IV – 31 (15.5%) Pre-operative ascites • < 500 mL: 116 (58%) • ≥ 500 mL: 84 (42%) Median CA-125 at diagnosis: 952 U/mL (range: 75 to 23,400) Median pre-operative CA-125: 572 (range: 43 to 986) Median CA-125 decreasing kinetics (ratio of the initial serum CA-125 level to the preoperative serum CA-125 level): 2.3 (range: 0.8 to 30.2) ≤ 3 tumour sites: 50 (25%) > 3 tumour sites: 150 (75%)
Residual disease details	Median NACT cycles: 3 (range: 1 to 8) NAC was administrated intraperitoneally for 90 (45%) women and intravenously for 110 (55%) women Median adjuvant CT cycles: 5 (range: 3 to 7) **'Optimal' cytoreduction** defined as RD < 1 cm (n = 156, 78%): • NMRD (referred to in study as RD0): 59 (29.5%) • SVRD (RD < 1 cm): 97 (48.5%) • LVRD between 1 cm to 2 cm inclusive: 8 (4%) • LVRD (> 2 cm): 30 (15%) • Unknown: 6 (3%)

Zhang 2018 *(Continued)*

Outcomes	Overall survival defined as interval between treatment initiation and death Median OS in participants with ascites regression: 32.1 Median OS in participants without ascites regression: 25.2 Multivariable Cox PH for OS adjusted for pre-operative ascites, number of tumour sites, CA-125 at diagnosis, CA-125 decreasing kinetics: • LVRD (> 1 cm (vs SVRD < 1 cm): HR 2.58, 95% CI 1.71 to 4.24), P < 0.01 Progression-free survival defined as interval between the beginning of treatment and documented disease progression or death from any cause in women with no evidence of progression Median PFS in participants with ascites regression: 22.3 Median PFS in participants without ascites regression: 18 Multivariable Cox PH for PFS adjusted for pre-operative ascites, number of tumour sites, number of NAC cycles, CA-125 at diagnosis, CA-125 decreasing kinetics: • LVRD (> 1 cm (vs SVRD): HR 2.43, 95% CI 1.44 to 4.08), P < 0.01
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): low risk Valid and reliable measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome. OS defined as interval between treatment initiation and death. 5. Adjustment for other prognostic factors (a-g): unclear risk Baseline CA-125 and preoperative CA-125 are likely to introduce multicollinearity. Model predicting OS adjusted for pre-operative ascites, number of tumour sites, CA-125 at diagnosis, CA-125 decreasing kinetics 6. Statistical analysis and reporting (a-d): unclear risk No conceptual framework; data driven based on P values of univariate associations **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome. PFS defined as interval between the beginning of treatment and documented disease progression or death from any cause in women with no evidence of progression. 5. Adjustment for other prognostic factors (a-g): unclear risk

Zhang 2018 *(Continued)*

	Baseline CA-125 and preoperative CA-125 are likely to introduce multicollinearity. Model predicting PFS adjusted for age, preoperative ascites, FIGO stage, tumour sites, baseline CA-125, preoperative CA-125, number of NACT cycles and CA-125 decreasing kinetics 6. Statistical analysis and reporting (a-d): unclear risk No conceptual framework; data driven based on P values of univariate associations
Notes	Median follow-up: 43.5 months Ascites regression defined as an ascites volume of less than 500 mL **Inclusion criteria** (i) Women histologically diagnosed as stage IIIc or IV invasive ovarian, fallopian tube or peritoneal high-grade serous carcinoma; (ii) women treated with platinum-based NAC followed by IDS and adjuvant chemotherapy; and (iii) women with an ascites volume of greater than or equal to 500 mL before NAC treatment as assessed by ultrasound examination **Exclusion criteria** (i) Fragile women who received slow-release evacuation procedure before NAC due to intolerable abdominal distension; (ii) women with extra-abdominal metastatic malignancy; and (iii) women whose preoperative serum cancer antigen 125 (CA-125) levels were less than or equal to 35 U/ mL **Treatment protocol** A NAC regimen consisting of carbo-platinum (area under the curves 5 to 6) and paclitaxel (135 to 175 mg/ m^2) was administered every 3 weeks. IDS was performed approximately 2 to 4 weeks after the NAC regimen. The adjuvant chemotherapy (at least 3 to 4 cycles) was the same as NAC. The standard IDS included bilateral/unilateral salpingo-oophorectomy, hysterectomy, appendectomy, pelvic/para-aortic lymphadenectomy and omentectomy. Extensive upper abdominal surgery was defined as splenectomy, diaphragm stripping and/or resection, distal pancreatectomy, cholecystectomy, partial liver resection and partial gastrectomy. Other surgery procedures, such as large/small bowel resection and peritoneal resection, were performed as necessary.

Zhu 2016

Study characteristics	
Methods	Multicentre, retrospective study
Participants	N = 672 women newly diagnosed with epithelial ovarian cancer between June 2008 and December 2015 at the Sun Yat-Sen University Cancer Center and Nan Fang Hospital of Southern Medical University, who were treated with NACT followed by IDS Median age: 55 (range: 30 to 70) FIGO: III – 564 (83.9%); IV – 108 (16.1%) Histology: serous – 484 (72%); non-serous – 188 (28%) Tumour grade: • G1 – 384 (57.1%) • G2/3 – 288 (42.9%) CA-125 at diagnosis, U/mL: • ≤ 35: 226 (33.6%)

Zhu 2016 *(Continued)*

	• > 35: 446 (66.4%) Comorbidity: • Chronic hepatitis B: 64 (9.5%) • Hypertension: 35 (5.2%) • Diabetes: 27 (4%) • Cardiovascular disease: 4 (0.6%) Chemosensitivity (RECIST complete/partial response): 444 (66.1%) Chemoresistance: 228 (33.9%)
Residual disease details	All participants given 3 cycles of NACT before IDS NACT regimen • Cisplatin plus paclitaxel: 298 (44.3 %) • Carboplatin plus paclitaxel: 250 (37.2 %) • Carboplatin plus docetaxel: 124 (18.5 %) Complete response to NACT (NMRD) in 61 (9.1%) **'Optimal' cytoreduction** was defined as RD ≤ 1 cm (n = 486; 72.3%)
Outcomes	Overall survival defined as interval between the date of diagnosis and the date of death from any cause or last follow-up 5-year OS: 36.7% Multivariable Cox PH for OS adjusted for FIGO stage, chemosensitivity, Glasgow prognostic score: • LVRD (> 1 cm) (vs SVRD): HR 1.332 (95% CI 1.057 to 1.679), P = 0.015 Progression-free survival defined as time from the date of diagnosis to the date of first relapse, progression, death from any cause or last follow-up 5-year PFS: 19.3% Multivariable Cox PH for PFS adjusted for FIGO stage, chemosensitivity, Glasgow prognostic score: • LVRD (> 1 cm) (vs SVRD): HR 1.268 (95% CI 1.051 to 1.589), P = 0.044
Risk of bias (QUIPS)	1. Study participation (a-f): low risk Adequate number of participants and description of target population. Baseline characteristics, eligibility criteria, sampling frame and period/place study took place presented clearly. 2. Study attrition (a-e): unclear risk Unclear if patients with incomplete follow-up were excluded before arriving at the stated sample size. Insufficient information to permit judgement. 3. Prognostic factor measurement (a-f): unclear risk Valid and reliable measurement of RD. Multicentre design may introduce heterogeneity in measurement of RD **Outcome level assessment:** **Outcome: overall survival** 4. Outcome measurement (a-c): low risk

Zhu 2016 *(Continued)*

	Valid and reliable measurement of outcome. OS defined as interval between the date of diagnosis and the date of death from any cause or last follow-up. 5. Adjustment for other prognostic factors (a-g): high risk Age arbitrarily dichotomised. Multivariate models for OS adjusted for FIGO stage, chemosensitivity and Glasgow Prognostic Score 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear on how variables were brought forward to multivariate model **Outcome: progression-free survival** 4. Outcome measurement (a-c): low risk Valid and reliable measurement of outcome. PFS defined as time from the date of diagnosis to the date of first relapse, progression, death from any cause or last follow-up 5. Adjustment for other prognostic factors (a-g): high risk Age arbitrarily dichotomised. Multivariate models for PFS adjusted for FIGO stage, chemosensitivity and Glasgow Prognostic Score. 6. Statistical analysis and reporting (a-d): high risk No conceptual framework; unclear on how variables were brought forward to multivariate model
Notes	Median follow-up: 38 months (range: 5 to 103) ECOG PS (Eastern Cooperative Oncology Group performance status) ≤ 1: 494 (73.5%); > 2: 178 (26.5%) Definition of Glasgow prognostic score: women in whom an elevated CRP level (> 10 mg/L) and hypoalbuminaemia (< 35 g/L) were both present were allocated a score of 2. Women with only one of these two biochemical abnormalities were given a score of 1. Women with neither of these abnormalities received a score of 0.

1. Study participation
   a. Adequate participation in the study by eligible persons
   b. Description of the target population or population of interest
   c. Description of the baseline study sample
   d. Adequate description of the sampling frame and recruitment
   e. Adequate description of the period and place of recruitment
   f. Adequate description of inclusion and exclusion criteria
2. Study attrition
   a. Adequate response rate for study participants
   b. Description of attempts to collect information on participants who dropped out
   c. Reasons for loss to follow-up are provided
   d. Adequate description of participants lost to follow-up
   e. There are no important differences between participants who completed the study and those who did not
3. Prognostic factor measurement
   a. A clear definition or description of the PF is provided
   b. Method of PF measurement is adequately valid and reliable
   c. Continuous variables are reported or appropriate cutpoints are used
   d. The method and setting of measurement of PF is the same for all study participants
   e. Adequate proportion of the study sample has complete data for the PF
   f. Appropriate methods of imputation are used for missing PF data

4. Outcome measurement
   a. A clear definition of the outcome is provided
   b. Method of outcome measurement used is adequately valid and reliable
   c. The method and setting of outcome measurement is the same for all study participants
5. Adjustment for other prognostic factors
   a. All other important PFs are measured
   b. Clear definitions of the important PFs measured are provided
   c. Measurement of all important PFs is adequately valid and reliable
   d. The method and setting of PF measurement are the same for all study participants
   e. Appropriate methods are used to deal with missing values of PFs, such as multiple imputation
   f. Important PFs are accounted for in the study design
   g. Important PFs are accounted for in the analysis
6. Statistical analysis and reporting
   a. Sufficient presentation of data to assess the adequacy of the analytic strategy
   b. Strategy for model building is appropriate and is based on a conceptual framework or model
   c. The selected statistical model is adequate for the design of the study
   d. There is no selective reporting of result

Overall risk of bias judgements were made per outcome for each included study

Abbreviations:

ACCI: age-adjusted Charlson Comorbidity Index; AHR: adjusted hazard ratio; AOC: advanced ovarian cancer; ASA: American Society of Anaesthesiologists; BMI: body mass index; CDC: Clavien-Dindo classification; CI: confidence interval; CPLN: cardiophrenic lymph nodes; CRP: c-reactive protein; CRS: cytoreductive surgery; DSS: disease-specific survival; ECOG: Eastern Cooperative Oncology Group; EOC: epithelial ovarian cancer; FIGO: International Federation of Gynecology and Obstetrics; GOG: Gynaecologic Oncology Group; HR: hazard ratio; ICU: intensive care unit; IDS: interval debulking surgery; IP: intraperitoneal; IQR: interquartile range; IV: intravenous; KM: Kaplan–Meier; LVRD: large-volume residual disease; NACT/ACT/CT: neoadjuvant chemotherapy/adjuvant chemotherapy/chemotherapy; NMRD: no macroscopic residual disease; NOS: not otherwise specified; OR: odds ratio; OS: overall survival; PCI: Peritoneal Cancer Index; PDS: primary debulking surgery; PH: proportional hazards; OS: overall survival; PF: prognostic factor; PFS: progression-free survival; RD: residual disease; RR: risk ratio; RT: residual tumour; SD: standard deviation; SE: standard error; SVRD: small-volume residual disease; TSIC: time from surgery to initiation of chemotherapy; UTI: urinary tract infection; WHO: World Health Organization

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Alberts 1993	No survival analysis by RD as all patients had suboptimal surgery (defined as more than 2 cm)
Alberts 1996	No multivariate analysis data
Alphs 2006	Included only 78 patients; 8 patients were early-stage and 9 patients received NAC
Altman 2012	No multivariate analysis data
Andersen Soegaard 2005	This study included only 83 patients, of which 66 received platinum-based chemotherapy. No multivariate analysis was performed.
Anuradha 2016	Scope of study focused on time interval between surgery and chemotherapy
Bailey 2006	Chemotherapy data are absent
Baker 1994	95% CI or SE (HR) are not reported and the HR point estimate for OS is 1.66 across all categories; it is not clear if the < 1 cm category was used as the reference group when compared to both 1 cm to 2 cm and > 2 cm residual disease
Barda 2004	27.3% of ovarian cancer received non-platinum chemotherapy
Benedetti-Panici 1996	Included only 66 patients and stage IIb. No survival data per RD. Also included NAC/IDS.

Study	Reason for exclusion
Bertelsen 1990	Study does not include a multivariate analysis
Bertelsen 1993	No survival data per residual disease
Bian 2016	No multivariate analysis data
Bonnefoi 1999	38 patients had NAC and 27 patients had non-platinum chemotherapy
Brinkhuis 1996a	No direct comparison by size of residual disease and there is no multivariate analysis
Brinkhuis 1996b	1 group of patients did not receive platinum chemotherapy except at progression. Survival data per RD is reported for all patients collectively.
Bristow 1999	Included only 84 patients
Cai 2007	Included 95 patients. We suspect that IDS cases were included.
Ceresoli 2018	Included only 56 patients at analysis, of which 28 treated with cytoreductive surgery + HIPEC and 28 treated with cytoreductive surgery alone.
Chekman 2015	Did not report outcomes for extent of residual disease by type of initial primary surgery
Clamp 2018	No multivariate analysis data
Colozza 1997	Included only 39 patients
Conte 1991	No survival data per residual disease
Conte 1996	There is no optimal group. No survival data per residual disease.
Crawford 2005	18% of the cases were stage IC and II
Creasman 1990	All cases were sub-optimal, defined as RD greater than 1 cm; no analysis by RD
Cummins 2019	Full text unavailable
Dao 2016	Included patients who had neoadjuvant chemotherapy
Del Campo 1994	Included only 91 patients
de Oliviera 1990	1 arm did not receive platinum-based chemotherapy
di Re 1996	14 patients had borderline tumours. Also included stage II cases. Before 1979, patients received non-platinum chemotherapy.
Elgamal 2019	Full text unavailable
Fagotti 2020	Did not report outcomes for extent of residual disease by type of initial primary surgery
Gao 2001	Only 31 cases
Gasimli 2016	Included selective group of women with cytoreduction of tumour to microscopic optimal disease (0 cm)
Geisler 2004	24 patients were stage I and II

Study	Reason for exclusion
Gershenson 1989	Included only 50 patients
Gershenson 1992	All patients were optimal, defined as RD less than 2 cm. No further analysis of survival by RD.
Gershenson 1995	Included only 51 patients
Greggi 2016	RD thresholds were not part of scope as the study focused on comparison of oncology specialist centres versus non-specialist centres
Grem 1991	Included only 43 patients
Hainsworth 1990	Included only 25 patients
Hakes 1992	Included only 78 patients
Hamid 2002	Only included 62 patients
Hardy 1991	Included only 30 stage IV patients
Heitz 2016	No multivariate analyses were reported
Hoskins 1992	All patients are optimal, i.e. less than 1 cm. Survival data is per preoperative disease volume rather than RD.
Hoskins 1996	Included only 29 patients
Hoskins 1997	No survival by residual disease
Itamochi 2002	Optimal surgery, i.e. size of RD, is not properly defined
Kaern 2005	Included only 31 stage III patients with no control group having RD more than 1 cm
Kehoe 2015	Comparisons of residual disease were based on type of intervention
Kessous 2017	No multivariate analysis data
Keyver-Paik 2016	No multivariate analyses were reported
Kirmani 1994	Included only 29 patients
Kristensen 1995	Included only 27 patients
Le 1997	Data for stage IIIC and IV subgroup was not reported and authors no longer had access to these data
Lee 2018	No multivariate analyses were reported and no response from corresponding author after request for adjusted estimates
Loizzi 2016	Included only 78 patients
Lorusso 1998	Included only 34 patients
Malik 1998	Included only 21 patients
Marchetti 1993	Included only 70 patients

Study	Reason for exclusion
McGuire 1996	No multivariate analyses were reported
Michaan 2018	Chemotherapy response score not same as optimal cytoreduction
Ngan 1989	Contained 65 patients only and 15 patients were excluded, so only 50 patients
Omura 1989	95% CIs and P values from Cox model in adjusted estimates are not reported. Cannot use Parmar's methods given the number of deaths and log rank P value as we need the unadjusted estimate.
Onda 2020	Did not report outcomes for extent of residual disease by type of initial primary surgery
Palmer 1992	Included only 70 patients
Piver 1991	43 patients did not receive platinum-based chemotherapy. No multivariate analysis.
Raspagliesi 2018	No multivariate analysis data
Redman 1986	Included 89 patients, 11 of whom initially did not receive platinum chemotherapy
Risum 2012	Only 17 women went through NACT-IDS
Rodriguez 2013	Comparisons were in terms of surgical procedures performed and could not be analysed by residual disease thresholds
Rose 2004	Reported on outcome after ''secondary'' debulking surgery. However, Winter 2007 included the results of GOG 152 by residual diease after primary cytoreductive surgery. This has been confirmed through personal communication with GOG statistician (Dr Mark Brady).
Ruscito 2016	Study did not distinguish between PDS and IDS
Rutten 2014	17% of sample made up of FIGO I and II
Salani 2007	Case-control study
Sessa 1991	No multivariate analysis performed
Shapiro 1998	Included only 26 patients
Shinozuka 1999	Some patients received preoperative chemotherapy
Sioulas 2017	Included women who received combination of intravenous/intraperitoneal chemotherapy and RD was not adequately reported in multivariate analyses
Skarlos 1996	Included patients with stage IIC disease
Smits 2015	Scope of study focused on obese and non-obese patients and included proportion of women who received neoadjuvant chemotherapy
Solmaz 2015	Did not report survival by residual disease
Son 2017	Included only 60 patients
Stewart 2015	Full text unavailable
Stewart 2016	No multivariate analysis

Study	Reason for exclusion
Strauss 1996	Included 42 patients only
Suidan 2015	Reported in abstract form only and unlikely that residual disease thresholds were assessed in appropriate multivariate analyses
Sun 2000	Patients who did not receive preoperative chemotherapy are only 76. Nature of chemotherapy received not clear.
Sutton 1989	Included only 56 patients
Takano 2006	Most patients had early-stage disease, which cannot be separated from late-staged cases
Takano 2007	Included early-stage disease (stage IC and II), which cannot be separated from late-staged cases
Tay 1996	Included 62 patients only. Did not include survival data per optimal versus suboptimal.
Taylor 1994	Included only 64 patients
Tingulstad 2003	6 patients did not receive chemotherapy and 6 patients received non-platinum chemotherapy
Todo 2003	Included patients who received NAC and IDS but did not report by extent of disease
Trhlík 2013	Full text unavailable
Uyar 2005	18 patients were stage I and II. No survival data per RD.
Vallejos 1997	Included only 30 patients
Van Der Burg 1996	Reported results per residual disease after NAC/IDS
Van Driel 2017	Non-platinum based chemotherapy was given to all the women
van Vliet 2015	Included patients with who received IDS
Vergote 2010	Comparisons of residual disease were based on type of intervention
Vergote 2018	Comparisons of residual disease were based on type of intervention
Vidal 2016	No multivariate analyses were reported
Wadler 1996	Survival reported per residual disease in all patients including 118 who received non-platinum chemotherapy
Wallace 2017	No multivariate analyses were reported
Warwick 1995	31 patients were stage II. No survival data per RD.
Willemse 1992	Included only 76 patients
Wils 1990	Included only 88 patients
Wimberger 2007	Multivariate analyses did not include residual disease and the study also included women with stage IIB and IIC disease. We attempted to contact the authors for further information but at time of submission of the review there had been no correspondence.

Study	Reason for exclusion
Yamamoto 2007	Included 67 ''selected'' patients with rare histological subtype
Zang 1999	Included only 71 patients and 31 of them received neoadjuvant chemotherapy
Zhang 2015	< 100 patients with advanced disease in study

CI: confidence interval; HIPEC: hyperthermic intraperitoneal chemotherapy; HR: hazard ratio; IDS: interval debulking surgery; NAC: neoadjuvant chemotherapy; OS: overall survival; PDS: Primary debulking surgery; RD: residual disease; SE: standard error

DATA AND ANALYSES

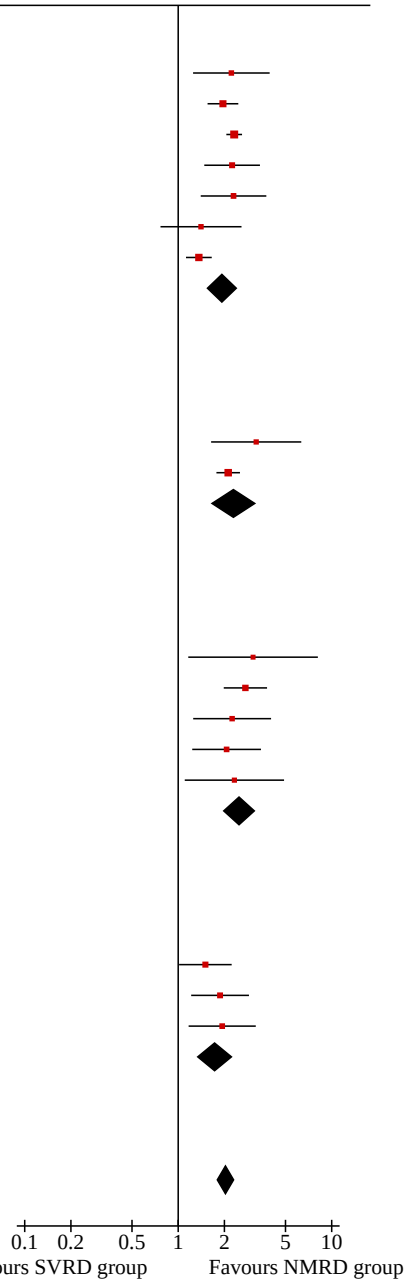
Comparison 1. PDS: SVRD (< 1 cm) versus NMRD

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.1 Overall survival	17		Hazard Ratio (IV, Random, 95% CI)	2.03 [1.80, 2.29]
1.1.1 Advanced stage (III/IV)	7		Hazard Ratio (IV, Random, 95% CI)	1.93 [1.55, 2.39]
1.1.2 Stage III	2		Hazard Ratio (IV, Random, 95% CI)	2.29 [1.66, 3.15]
1.1.3 Stage IIIC	5		Hazard Ratio (IV, Random, 95% CI)	2.49 [1.98, 3.13]
1.1.4 Stage IV	3		Hazard Ratio (IV, Random, 95% CI)	1.73 [1.34, 2.22]
1.2 Overall survival - sensitivity analysis using fixed-effect model	17		Hazard Ratio (IV, Fixed, 95% CI)	2.05 [1.91, 2.20]
1.2.1 Advanced stage (III/IV)	7		Hazard Ratio (IV, Fixed, 95% CI)	2.01 [1.84, 2.19]
1.2.2 Stage III	2		Hazard Ratio (IV, Fixed, 95% CI)	2.17 [1.83, 2.58]
1.2.3 Stage IIIC	5		Hazard Ratio (IV, Fixed, 95% CI)	2.49 [1.98, 3.13]
1.2.4 Stage IV	3		Hazard Ratio (IV, Fixed, 95% CI)	1.73 [1.34, 2.22]
1.3 Overall survival - sensitivity analysis excluding Klar 2016	16		Hazard Ratio (IV, Random, 95% CI)	1.99 [1.75, 2.27]
1.3.1 Advanced stage (III/IV)	6		Hazard Ratio (IV, Random, 95% CI)	1.81 [1.46, 2.25]
1.3.2 Stage III	2		Hazard Ratio (IV, Random, 95% CI)	2.29 [1.66, 3.15]
1.3.3 Stage IIIC	5		Hazard Ratio (IV, Random, 95% CI)	2.49 [1.98, 3.13]
1.3.4 Stage IV	3		Hazard Ratio (IV, Random, 95% CI)	1.73 [1.34, 2.22]
1.4 Progression-free survival	10		Hazard Ratio (IV, Random, 95% CI)	1.88 [1.63, 2.16]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.4.1 Advanced stage (III/IV)	5		Hazard Ratio (IV, Random, 95% CI)	1.82 [1.43, 2.32]
1.4.2 Stage III	2		Hazard Ratio (IV, Random, 95% CI)	2.21 [1.54, 3.18]
1.4.3 Stage IIIC	1		Hazard Ratio (IV, Random, 95% CI)	2.03 [1.25, 3.31]
1.4.4 Stage IV	2		Hazard Ratio (IV, Random, 95% CI)	1.68 [1.26, 2.24]
1.5 Progression-free survival - sensitivity analysis using fixed-effect model	10		Hazard Ratio (IV, Fixed, 95% CI)	1.93 [1.80, 2.06]
1.5.1 Advanced stage (III/IV)	5		Hazard Ratio (IV, Fixed, 95% CI)	1.92 [1.77, 2.08]
1.5.2 Stage III	2		Hazard Ratio (IV, Fixed, 95% CI)	2.01 [1.76, 2.31]
1.5.3 Stage IIIC	1		Hazard Ratio (IV, Fixed, 95% CI)	2.03 [1.25, 3.31]
1.5.4 Stage IV	2		Hazard Ratio (IV, Fixed, 95% CI)	1.68 [1.26, 2.24]
1.6 Progression-free survival - sensitivity analysis excluding Klar 2016	9		Hazard Ratio (IV, Random, 95% CI)	1.83 [1.56, 2.13]
1.6.1 Advanced stage (III/IV)	4		Hazard Ratio (IV, Random, 95% CI)	1.69 [1.33, 2.14]
1.6.2 Stage III	2		Hazard Ratio (IV, Random, 95% CI)	2.21 [1.54, 3.18]
1.6.3 Stage IIIC	1		Hazard Ratio (IV, Random, 95% CI)	2.03 [1.25, 3.31]
1.6.4 Stage IV	2		Hazard Ratio (IV, Random, 95% CI)	1.68 [1.26, 2.24]

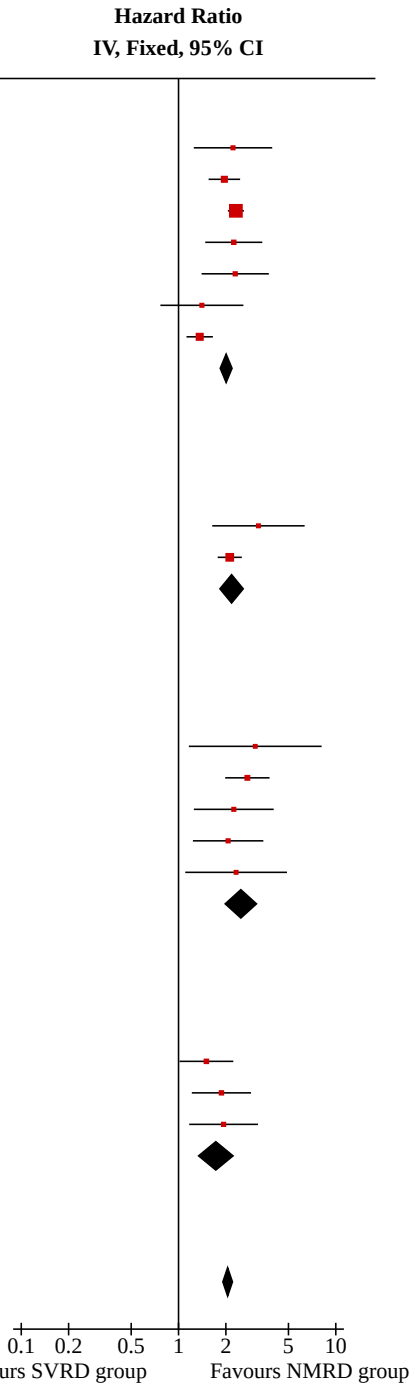
Analysis 1.1. Comparison 1: PDS: SVRD (< 1 cm) versus NMRD, Outcome 1: Overall survival

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
1.1.1 Advanced stage (III/IV)				
Chang 2012a	0.7975	0.293	3.5%	2.22 [1.25 , 3.94]
Kahl 2017	0.671	0.1176	10.0%	1.96 [1.55 , 2.46]
Klar 2016	0.84	0.06	13.6%	2.32 [2.06 , 2.61]
Langstraat 2011	0.8087	0.213	5.5%	2.24 [1.48 , 3.41]
Paik 2018	0.83	0.251	4.4%	2.29 [1.40 , 3.75]
Tewari 2016	0.342	0.31	3.2%	1.41 [0.77 , 2.58]
Tseng 2018	0.31	0.0985	11.2%	1.36 [1.12 , 1.65]
Subtotal (95% CI)			**51.5%**	**1.93 [1.55 , 2.39]**
Heterogeneity: Tau² = 0.05; Chi² = 23.15, df = 6 (P = 0.0007); I² = 74%				
Test for overall effect: Z = 5.94 (P < 0.00001)				
1.1.2 Stage III				
Cuylan 2018	1.17	0.345	2.7%	3.22 [1.64 , 6.34]
Winter 2007	0.75	0.09	11.8%	2.12 [1.77 , 2.53]
Subtotal (95% CI)			**14.4%**	**2.29 [1.66 , 3.15]**
Heterogeneity: Tau² = 0.02; Chi² = 1.39, df = 1 (P = 0.24); I² = 28%				
Test for overall effect: Z = 5.07 (P < 0.00001)				
1.1.3 Stage IIIC				
Aletti 2006	1.123	0.4959	1.4%	3.07 [1.16 , 8.13]
Bristow 2011	1.008	0.1657	7.4%	2.74 [1.98 , 3.79]
Chang 2012b	0.809	0.298	3.4%	2.25 [1.25 , 4.03]
Chi 2006	0.726	0.263	4.1%	2.07 [1.23 , 3.46]
Eisenkop 2003	0.843	0.38	2.3%	2.32 [1.10 , 4.89]
Subtotal (95% CI)			**18.6%**	**2.49 [1.98 , 3.13]**
Heterogeneity: Tau² = 0.00; Chi² = 1.17, df = 4 (P = 0.88); I² = 0%				
Test for overall effect: Z = 7.80 (P < 0.00001)				
1.1.4 Stage IV				
Ataseven 2016	0.408	0.201	5.9%	1.50 [1.01 , 2.23]
Wimberger 2010	0.628	0.2211	5.3%	1.87 [1.21 , 2.89]
Winter 2008	0.66	0.257	4.3%	1.93 [1.17 , 3.20]
Subtotal (95% CI)			**15.4%**	**1.73 [1.34 , 2.22]**
Heterogeneity: Tau² = 0.00; Chi² = 0.81, df = 2 (P = 0.67); I² = 0%				
Test for overall effect: Z = 4.24 (P < 0.0001)				
Total (95% CI)			**100.0%**	**2.03 [1.80 , 2.29]**
Heterogeneity: Tau² = 0.02; Chi² = 31.75, df = 16 (P = 0.01); I² = 50%				
Test for overall effect: Z = 11.39 (P < 0.00001)				
Test for subgroup differences: Chi² = 5.30, df = 3 (P = 0.15), I² = 43.4%				



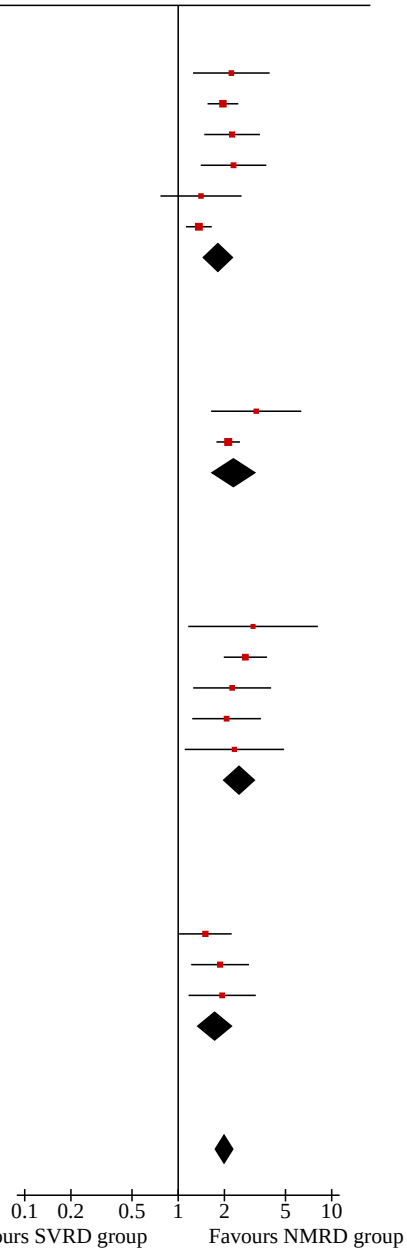
Analysis 1.2. Comparison 1: PDS: SVRD (< 1 cm) versus NMRD, Outcome 2: Overall survival - sensitivity analysis using fixed-effect model

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Fixed, 95% CI
1.2.1 Advanced stage (III/IV)				
Chang 2012a	0.7975	0.293	1.5%	2.22 [1.25 , 3.94]
Kahl 2017	0.671	0.1176	9.3%	1.96 [1.55 , 2.46]
Klar 2016	0.84	0.06	35.7%	2.32 [2.06 , 2.61]
Langstraat 2011	0.8087	0.213	2.8%	2.24 [1.48 , 3.41]
Paik 2018	0.83	0.251	2.0%	2.29 [1.40 , 3.75]
Tewari 2016	0.342	0.31	1.3%	1.41 [0.77 , 2.58]
Tseng 2018	0.31	0.0985	13.2%	1.36 [1.12 , 1.65]
Subtotal (95% CI)			**65.9%**	**2.01 [1.84 , 2.19]**
Heterogeneity: Chi² = 23.15, df = 6 (P = 0.0007); I² = 74%				
Test for overall effect: Z = 15.79 (P < 0.00001)				
1.2.2 Stage III				
Cuylan 2018	1.17	0.345	1.1%	3.22 [1.64 , 6.34]
Winter 2007	0.75	0.09	15.9%	2.12 [1.77 , 2.53]
Subtotal (95% CI)			**16.9%**	**2.17 [1.83 , 2.58]**
Heterogeneity: Chi² = 1.39, df = 1 (P = 0.24); I² = 28%				
Test for overall effect: Z = 8.92 (P < 0.00001)				
1.2.3 Stage IIIC				
Aletti 2006	1.123	0.4959	0.5%	3.07 [1.16 , 8.13]
Bristow 2011	1.008	0.1657	4.7%	2.74 [1.98 , 3.79]
Chang 2012b	0.809	0.298	1.4%	2.25 [1.25 , 4.03]
Chi 2006	0.726	0.263	1.9%	2.07 [1.23 , 3.46]
Eisenkop 2003	0.843	0.38	0.9%	2.32 [1.10 , 4.89]
Subtotal (95% CI)			**9.4%**	**2.49 [1.98 , 3.13]**
Heterogeneity: Chi² = 1.17, df = 4 (P = 0.88); I² = 0%				
Test for overall effect: Z = 7.80 (P < 0.00001)				
1.2.4 Stage IV				
Ataseven 2016	0.408	0.201	3.2%	1.50 [1.01 , 2.23]
Wimberger 2010	0.628	0.2211	2.6%	1.87 [1.21 , 2.89]
Winter 2008	0.66	0.257	1.9%	1.93 [1.17 , 3.20]
Subtotal (95% CI)			**7.8%**	**1.73 [1.34 , 2.22]**
Heterogeneity: Chi² = 0.81, df = 2 (P = 0.67); I² = 0%				
Test for overall effect: Z = 4.24 (P < 0.0001)				
Total (95% CI)			**100.0%**	**2.05 [1.91 , 2.20]**
Heterogeneity: Chi² = 31.75, df = 16 (P = 0.01); I² = 50%				
Test for overall effect: Z = 20.06 (P < 0.00001)				
Test for subgroup differences: Chi² = 5.23, df = 3 (P = 0.16), I² = 42.7%				



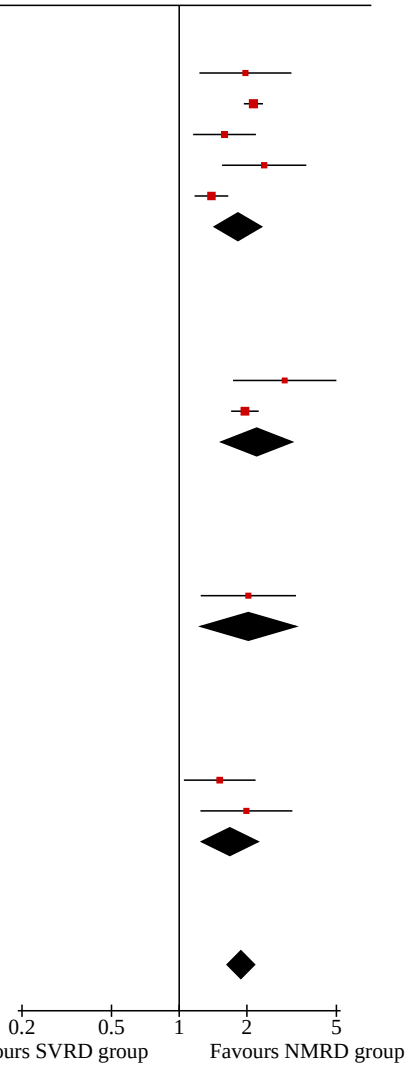
Analysis 1.3. Comparison 1: PDS: SVRD (< 1 cm) versus NMRD, Outcome 3: Overall survival - sensitivity analysis excluding Klar 2016

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
1.3.1 Advanced stage (III/IV)				
Chang 2012a	0.7975	0.293	4.0%	2.22 [1.25 , 3.94]
Kahl 2017	0.671	0.1176	11.6%	1.96 [1.55 , 2.46]
Langstraat 2011	0.8087	0.213	6.4%	2.24 [1.48 , 3.41]
Paik 2018	0.83	0.251	5.1%	2.29 [1.40 , 3.75]
Tewari 2016	0.342	0.31	3.7%	1.41 [0.77 , 2.58]
Tseng 2018	0.31	0.0985	13.1%	1.36 [1.12 , 1.65]
Subtotal (95% CI)			**43.9%**	**1.81 [1.46 , 2.25]**
Heterogeneity: Tau² = 0.04; Chi² = 10.77, df = 5 (P = 0.06); I² = 54%				
Test for overall effect: Z = 5.36 (P < 0.00001)				
1.3.2 Stage III				
Cuylan 2018	1.17	0.345	3.1%	3.22 [1.64 , 6.34]
Winter 2007	0.75	0.09	13.7%	2.12 [1.77 , 2.53]
Subtotal (95% CI)			**16.8%**	**2.29 [1.66 , 3.15]**
Heterogeneity: Tau² = 0.02; Chi² = 1.39, df = 1 (P = 0.24); I² = 28%				
Test for overall effect: Z = 5.07 (P < 0.00001)				
1.3.3 Stage IIIC				
Aletti 2006	1.123	0.4959	1.6%	3.07 [1.16 , 8.13]
Bristow 2011	1.008	0.1657	8.6%	2.74 [1.98 , 3.79]
Chang 2012b	0.809	0.298	3.9%	2.25 [1.25 , 4.03]
Chi 2006	0.726	0.263	4.7%	2.07 [1.23 , 3.46]
Eisenkop 2003	0.843	0.38	2.6%	2.32 [1.10 , 4.89]
Subtotal (95% CI)			**21.5%**	**2.49 [1.98 , 3.13]**
Heterogeneity: Tau² = 0.00; Chi² = 1.17, df = 4 (P = 0.88); I² = 0%				
Test for overall effect: Z = 7.80 (P < 0.00001)				
1.3.4 Stage IV				
Ataseven 2016	0.408	0.201	6.9%	1.50 [1.01 , 2.23]
Wimberger 2010	0.628	0.2211	6.1%	1.87 [1.21 , 2.89]
Winter 2008	0.66	0.257	4.9%	1.93 [1.17 , 3.20]
Subtotal (95% CI)			**17.8%**	**1.73 [1.34 , 2.22]**
Heterogeneity: Tau² = 0.00; Chi² = 0.81, df = 2 (P = 0.67); I² = 0%				
Test for overall effect: Z = 4.24 (P < 0.0001)				
Total (95% CI)			**100.0%**	**1.99 [1.75 , 2.27]**
Heterogeneity: Tau² = 0.02; Chi² = 25.43, df = 15 (P = 0.04); I² = 41%				
Test for overall effect: Z = 10.33 (P < 0.00001)				
Test for subgroup differences: Chi² = 6.22, df = 3 (P = 0.10), I² = 51.8%				



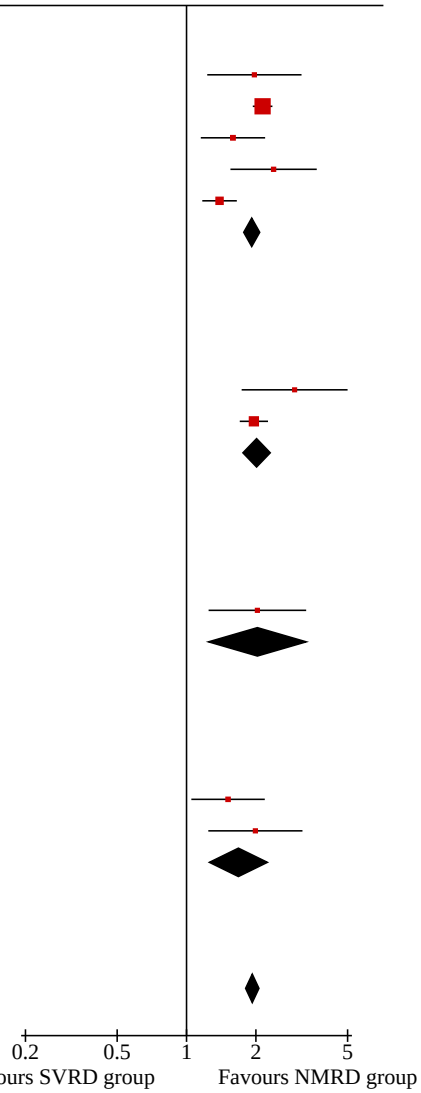
Analysis 1.4. Comparison 1: PDS: SVRD (< 1 cm) versus NMRD, Outcome 4: Progression-free survival

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
1.4.1 Advanced stage (III/IV)				
Chang 2012a	0.678	0.2403	6.2%	1.97 [1.23 , 3.15]
Klar 2016	0.76	0.05	18.6%	2.14 [1.94 , 2.36]
Paik 2018	0.464	0.164	9.9%	1.59 [1.15 , 2.19]
Shim 2016	0.87	0.22	7.0%	2.39 [1.55 , 3.67]
Tseng 2018	0.33	0.088	15.6%	1.39 [1.17 , 1.65]
Subtotal (95% CI)			**57.2%**	**1.82 [1.43 , 2.32]**
Heterogeneity: Tau² = 0.05; Chi² = 20.36, df = 4 (P = 0.0004); I² = 80%				
Test for overall effect: Z = 4.88 (P < 0.00001)				
1.4.2 Stage III				
Cuylan 2018	1.08	0.27	5.2%	2.94 [1.73 , 5.00]
Winter 2007	0.673	0.072	16.9%	1.96 [1.70 , 2.26]
Subtotal (95% CI)			**22.2%**	**2.21 [1.54 , 3.18]**
Heterogeneity: Tau² = 0.04; Chi² = 2.12, df = 1 (P = 0.15); I² = 53%				
Test for overall effect: Z = 4.27 (P < 0.0001)				
1.4.3 Stage IIIC				
Chang 2012b	0.708	0.2487	5.9%	2.03 [1.25 , 3.31]
Subtotal (95% CI)			**5.9%**	**2.03 [1.25 , 3.31]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 2.85 (P = 0.004)				
1.4.4 Stage IV				
Wimberger 2010	0.415	0.1872	8.5%	1.51 [1.05 , 2.19]
Winter 2008	0.688	0.24	6.2%	1.99 [1.24 , 3.18]
Subtotal (95% CI)			**14.7%**	**1.68 [1.26 , 2.24]**
Heterogeneity: Tau² = 0.00; Chi² = 0.80, df = 1 (P = 0.37); I² = 0%				
Test for overall effect: Z = 3.51 (P = 0.0004)				
Total (95% CI)			**100.0%**	**1.88 [1.63 , 2.16]**
Heterogeneity: Tau² = 0.03; Chi² = 24.62, df = 9 (P = 0.003); I² = 63%				
Test for overall effect: Z = 8.82 (P < 0.00001)				
Test for subgroup differences: Chi² = 1.49, df = 3 (P = 0.68), I² = 0%				

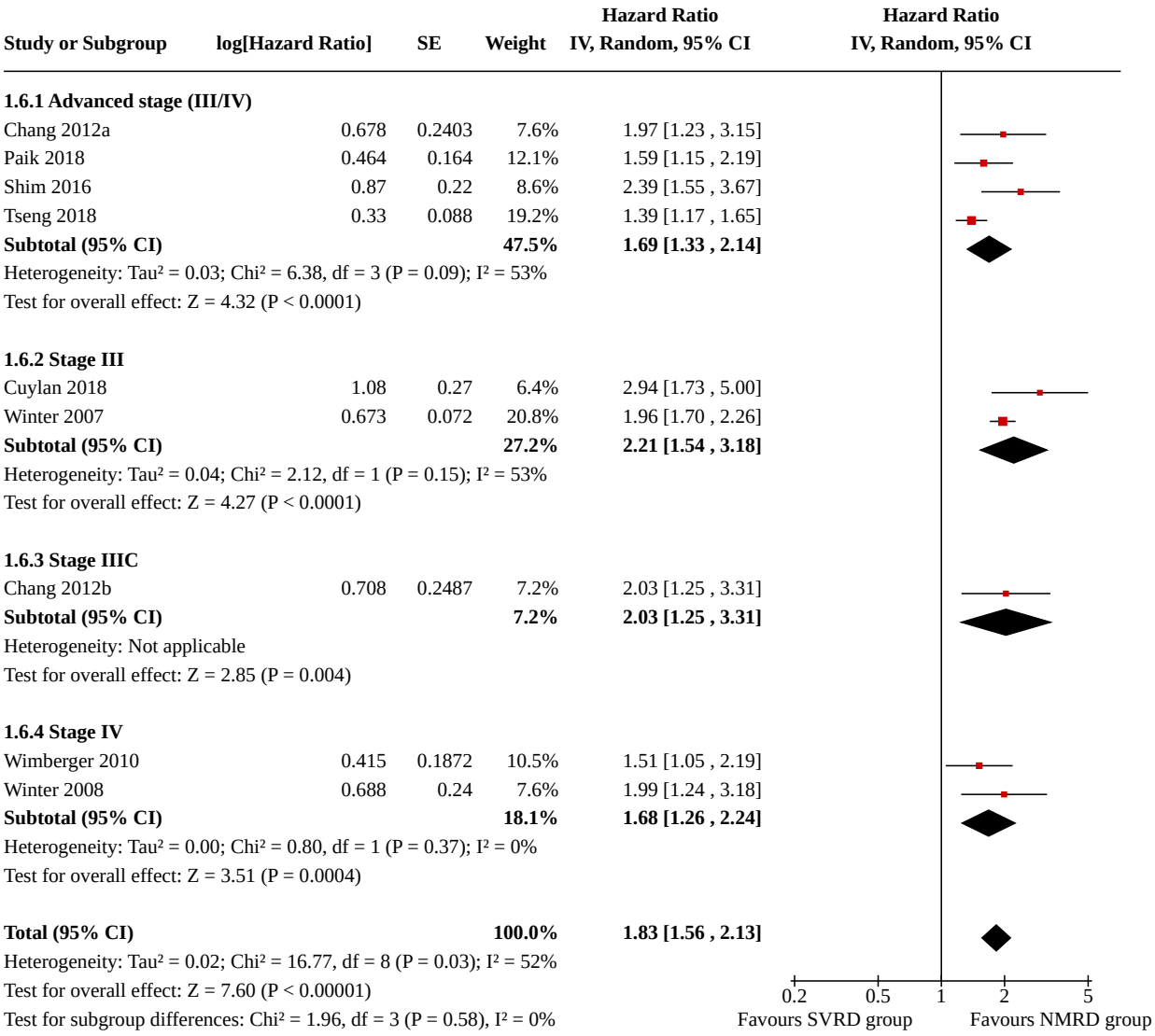


Analysis 1.5. Comparison 1: PDS: SVRD (< 1 cm) versus NMRD, Outcome 5: Progression-free survival - sensitivity analysis using fixed-effect model

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Fixed, 95% CI
1.5.1 Advanced stage (III/IV)				
Chang 2012a	0.678	0.2403	2.0%	1.97 [1.23 , 3.15]
Klar 2016	0.76	0.05	45.8%	2.14 [1.94 , 2.36]
Paik 2018	0.464	0.164	4.3%	1.59 [1.15 , 2.19]
Shim 2016	0.87	0.22	2.4%	2.39 [1.55 , 3.67]
Tseng 2018	0.33	0.088	14.8%	1.39 [1.17 , 1.65]
Subtotal (95% CI)			**69.2%**	**1.92 [1.77 , 2.08]**
Heterogeneity: Chi² = 20.36, df = 4 (P = 0.0004); I² = 80%				
Test for overall effect: Z = 16.01 (P < 0.00001)				
1.5.2 Stage III				
Cuylan 2018	1.08	0.27	1.6%	2.94 [1.73 , 5.00]
Winter 2007	0.673	0.072	22.1%	1.96 [1.70 , 2.26]
Subtotal (95% CI)			**23.7%**	**2.01 [1.76 , 2.31]**
Heterogeneity: Chi² = 2.12, df = 1 (P = 0.15); I² = 53%				
Test for overall effect: Z = 10.06 (P < 0.00001)				
1.5.3 Stage IIIC				
Chang 2012b	0.708	0.2487	1.9%	2.03 [1.25 , 3.31]
Subtotal (95% CI)			**1.9%**	**2.03 [1.25 , 3.31]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 2.85 (P = 0.004)				
1.5.4 Stage IV				
Wimberger 2010	0.415	0.1872	3.3%	1.51 [1.05 , 2.19]
Winter 2008	0.688	0.24	2.0%	1.99 [1.24 , 3.18]
Subtotal (95% CI)			**5.3%**	**1.68 [1.26 , 2.24]**
Heterogeneity: Chi² = 0.80, df = 1 (P = 0.37); I² = 0%				
Test for overall effect: Z = 3.51 (P = 0.0004)				
Total (95% CI)			**100.0%**	**1.93 [1.80 , 2.06]**
Heterogeneity: Chi² = 24.62, df = 9 (P = 0.003); I² = 63%				
Test for overall effect: Z = 19.41 (P < 0.00001)				
Test for subgroup differences: Chi² = 1.33, df = 3 (P = 0.72), I² = 0%				



Analysis 1.6. Comparison 1: PDS: SVRD (< 1 cm) versus NMRD, Outcome 6: Progression-free survival - sensitivity analysis excluding Klar 2016



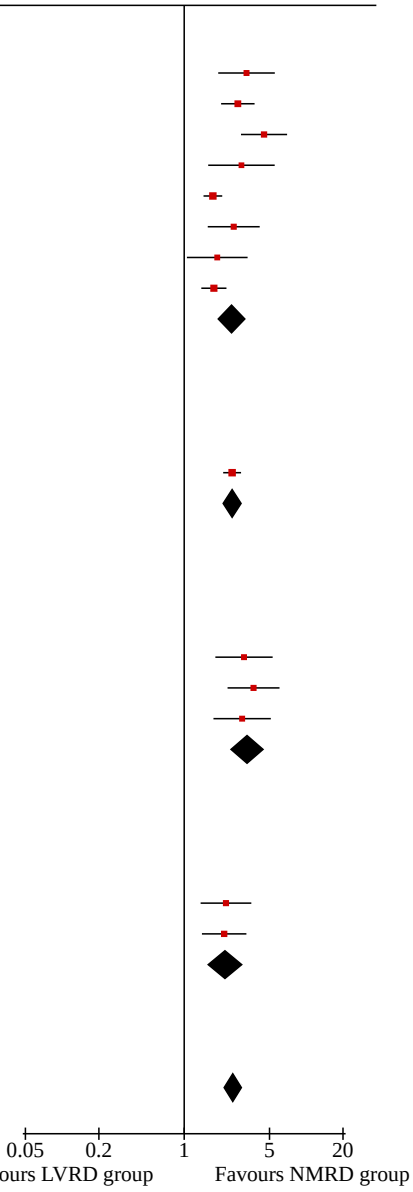
Comparison 2. PDS: LVRD (> 1 cm) versus NMRD

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.1 Overall survival	14		Hazard Ratio (IV, Random, 95% CI)	2.50 [2.13, 2.94]
2.1.1 Advanced stage (III/IV)	8		Hazard Ratio (IV, Random, 95% CI)	2.44 [1.90, 3.14]
2.1.2 Stage III	1		Hazard Ratio (IV, Random, 95% CI)	2.47 [2.09, 2.92]
2.1.3 Stage IIIC	3		Hazard Ratio (IV, Random, 95% CI)	3.27 [2.42, 4.42]
2.1.4 Stage IV	2		Hazard Ratio (IV, Random, 95% CI)	2.16 [1.57, 2.96]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.2 Overall survival - sensitivity analysis using fixed effects model	14		Hazard Ratio (IV, Fixed, 95% CI)	2.27 [2.09, 2.48]
2.2.1 Advanced stage (III/IV)	8		Hazard Ratio (IV, Fixed, 95% CI)	2.10 [1.87, 2.35]
2.2.2 Stage III	1		Hazard Ratio (IV, Fixed, 95% CI)	2.47 [2.09, 2.92]
2.2.3 Stage IIIC	3		Hazard Ratio (IV, Fixed, 95% CI)	3.27 [2.42, 4.42]
2.2.4 Stage IV	2		Hazard Ratio (IV, Fixed, 95% CI)	2.16 [1.57, 2.96]
2.3 Overall survival - sensitivity analysis excluding Melamed 2017b and Winter 2007	12		Hazard Ratio (IV, Random, 95% CI)	2.65 [2.20, 3.19]
2.3.1 Advanced stage (III/IV)	7		Hazard Ratio (IV, Random, 95% CI)	2.63 [1.99, 3.47]
2.3.2 Stage IIIC	3		Hazard Ratio (IV, Random, 95% CI)	3.27 [2.42, 4.42]
2.3.3 Stage IV	2		Hazard Ratio (IV, Random, 95% CI)	2.16 [1.57, 2.96]
2.4 Progression-free survival	6		Hazard Ratio (IV, Random, 95% CI)	2.10 [1.84, 2.40]
2.4.1 Advanced stage (III/IV)	3		Hazard Ratio (IV, Random, 95% CI)	1.92 [1.62, 2.27]
2.4.2 Stage III	1		Hazard Ratio (IV, Random, 95% CI)	2.36 [2.04, 2.73]
2.4.3 Stage IIIC	1		Hazard Ratio (IV, Random, 95% CI)	2.56 [1.54, 4.26]
2.4.4 Stage IV	1		Hazard Ratio (IV, Random, 95% CI)	1.82 [1.28, 2.59]

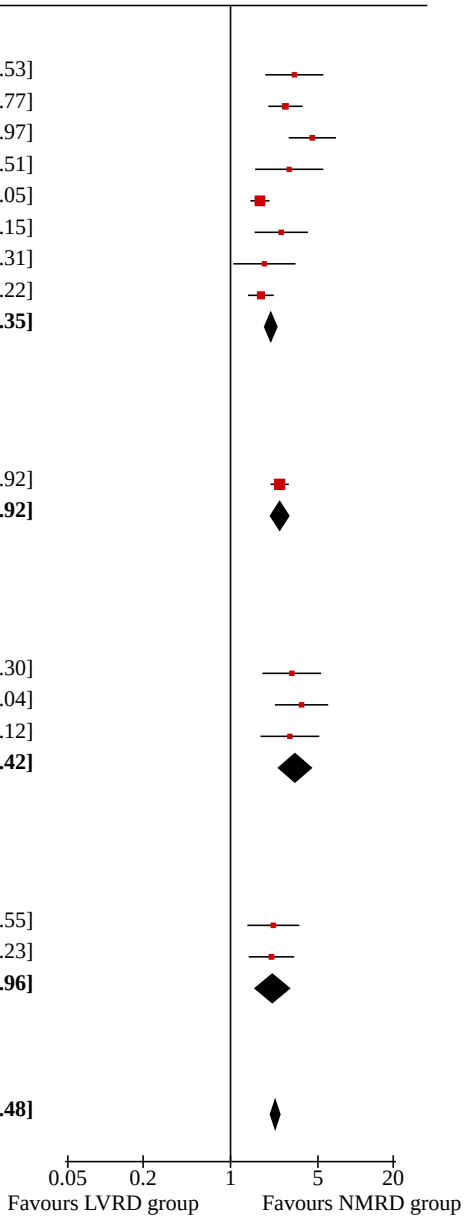
Analysis 2.1. Comparison 2: PDS: LVRD (> 1 cm) versus NMRD, Outcome 1: Overall survival

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
2.1.1 Advanced stage (III/IV)				
Chang 2012a	1.1756	0.2723	5.4%	3.24 [1.90 , 5.53]
Kahl 2017	1.012	0.161	8.9%	2.75 [2.01 , 3.77]
Langstraat 2011	1.5063	0.2218	6.8%	4.51 [2.92 , 6.97]
Melamed 2017a	1.08	0.32	4.4%	2.94 [1.57 , 5.51]
Melamed 2017b	0.541	0.09	11.5%	1.72 [1.44 , 2.05]
Paik 2018	0.9343	0.25	6.0%	2.55 [1.56 , 4.15]
Tewari 2016	0.6235	0.292	5.0%	1.87 [1.05 , 3.31]
Tseng 2018	0.56	0.121	10.4%	1.75 [1.38 , 2.22]
Subtotal (95% CI)			**58.4%**	**2.44 [1.90 , 3.14]**
Heterogeneity: Tau² = 0.09; Chi² = 26.34, df = 7 (P = 0.0004); I² = 73%				
Test for overall effect: Z = 7.00 (P < 0.00001)				
2.1.2 Stage III				
Winter 2007	0.904	0.086	11.6%	2.47 [2.09 , 2.92]
Subtotal (95% CI)			**11.6%**	**2.47 [2.09 , 2.92]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 10.51 (P < 0.00001)				
2.1.3 Stage IIIC				
Chang 2012b	1.1282	0.2757	5.4%	3.09 [1.80 , 5.30]
Chi 2006	1.308	0.25	6.0%	3.70 [2.27 , 6.04]
Eisenkop 2003	1.093	0.276	5.4%	2.98 [1.74 , 5.12]
Subtotal (95% CI)			**16.7%**	**3.27 [2.42 , 4.42]**
Heterogeneity: Tau² = 0.00; Chi² = 0.40, df = 2 (P = 0.82); I² = 0%				
Test for overall effect: Z = 7.71 (P < 0.00001)				
2.1.4 Stage IV				
Ataseven 2016	0.788	0.244	6.2%	2.20 [1.36 , 3.55]
Wimberger 2010	0.754	0.214	7.0%	2.13 [1.40 , 3.23]
Subtotal (95% CI)			**13.2%**	**2.16 [1.57 , 2.96]**
Heterogeneity: Tau² = 0.00; Chi² = 0.01, df = 1 (P = 0.92); I² = 0%				
Test for overall effect: Z = 4.78 (P < 0.00001)				
Total (95% CI)			**100.0%**	**2.50 [2.13 , 2.94]**
Heterogeneity: Tau² = 0.05; Chi² = 35.35, df = 13 (P = 0.0007); I² = 63%				
Test for overall effect: Z = 11.09 (P < 0.00001)				
Test for subgroup differences: Chi² = 3.93, df = 3 (P = 0.27), I² = 23.7%				



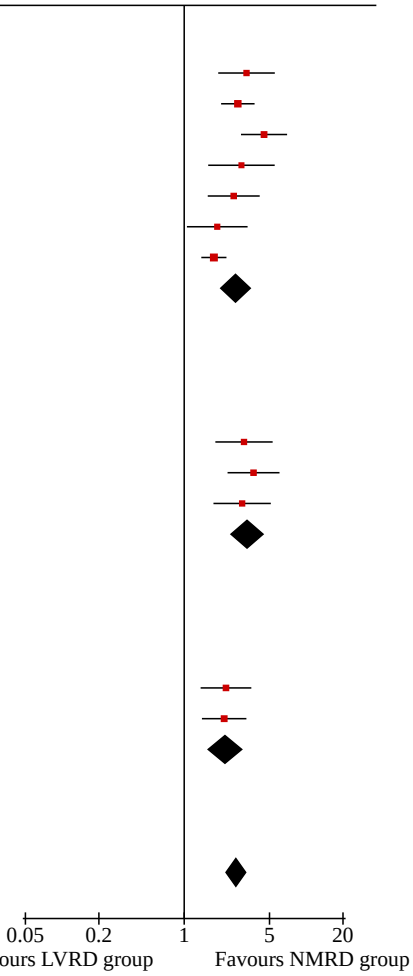
Analysis 2.2. Comparison 2: PDS: LVRD (> 1 cm) versus NMRD, Outcome 2: Overall survival - sensitivity analysis using fixed effects model

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Fixed, 95% CI
2.2.1 Advanced stage (III/IV)				
Chang 2012a	1.1756	0.2723	2.6%	3.24 [1.90 , 5.53]
Kahl 2017	1.012	0.161	7.5%	2.75 [2.01 , 3.77]
Langstraat 2011	1.5063	0.2218	3.9%	4.51 [2.92 , 6.97]
Melamed 2017a	1.08	0.32	1.9%	2.94 [1.57 , 5.51]
Melamed 2017b	0.541	0.09	23.8%	1.72 [1.44 , 2.05]
Paik 2018	0.9343	0.25	3.1%	2.55 [1.56 , 4.15]
Tewari 2016	0.6235	0.292	2.3%	1.87 [1.05 , 3.31]
Tseng 2018	0.56	0.121	13.2%	1.75 [1.38 , 2.22]
Subtotal (95% CI)			**58.3%**	**2.10 [1.87 , 2.35]**
Heterogeneity: Chi² = 26.34, df = 7 (P = 0.0004); I² = 73%				
Test for overall effect: Z = 12.86 (P < 0.00001)				
2.2.2 Stage III				
Winter 2007	0.904	0.086	26.1%	2.47 [2.09 , 2.92]
Subtotal (95% CI)			**26.1%**	**2.47 [2.09 , 2.92]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 10.51 (P < 0.00001)				
2.2.3 Stage IIIC				
Chang 2012b	1.1282	0.2757	2.5%	3.09 [1.80 , 5.30]
Chi 2006	1.308	0.25	3.1%	3.70 [2.27 , 6.04]
Eisenkop 2003	1.093	0.276	2.5%	2.98 [1.74 , 5.12]
Subtotal (95% CI)			**8.2%**	**3.27 [2.42 , 4.42]**
Heterogeneity: Chi² = 0.40, df = 2 (P = 0.82); I² = 0%				
Test for overall effect: Z = 7.71 (P < 0.00001)				
2.2.4 Stage IV				
Ataseven 2016	0.788	0.244	3.2%	2.20 [1.36 , 3.55]
Wimberger 2010	0.754	0.214	4.2%	2.13 [1.40 , 3.23]
Subtotal (95% CI)			**7.5%**	**2.16 [1.57 , 2.96]**
Heterogeneity: Chi² = 0.01, df = 1 (P = 0.92); I² = 0%				
Test for overall effect: Z = 4.78 (P < 0.00001)				
Total (95% CI)			**100.0%**	**2.27 [2.09 , 2.48]**
Heterogeneity: Chi² = 35.35, df = 13 (P = 0.0007); I² = 63%				
Test for overall effect: Z = 18.70 (P < 0.00001)				
Test for subgroup differences: Chi² = 8.60, df = 3 (P = 0.04), I² = 65.1%				

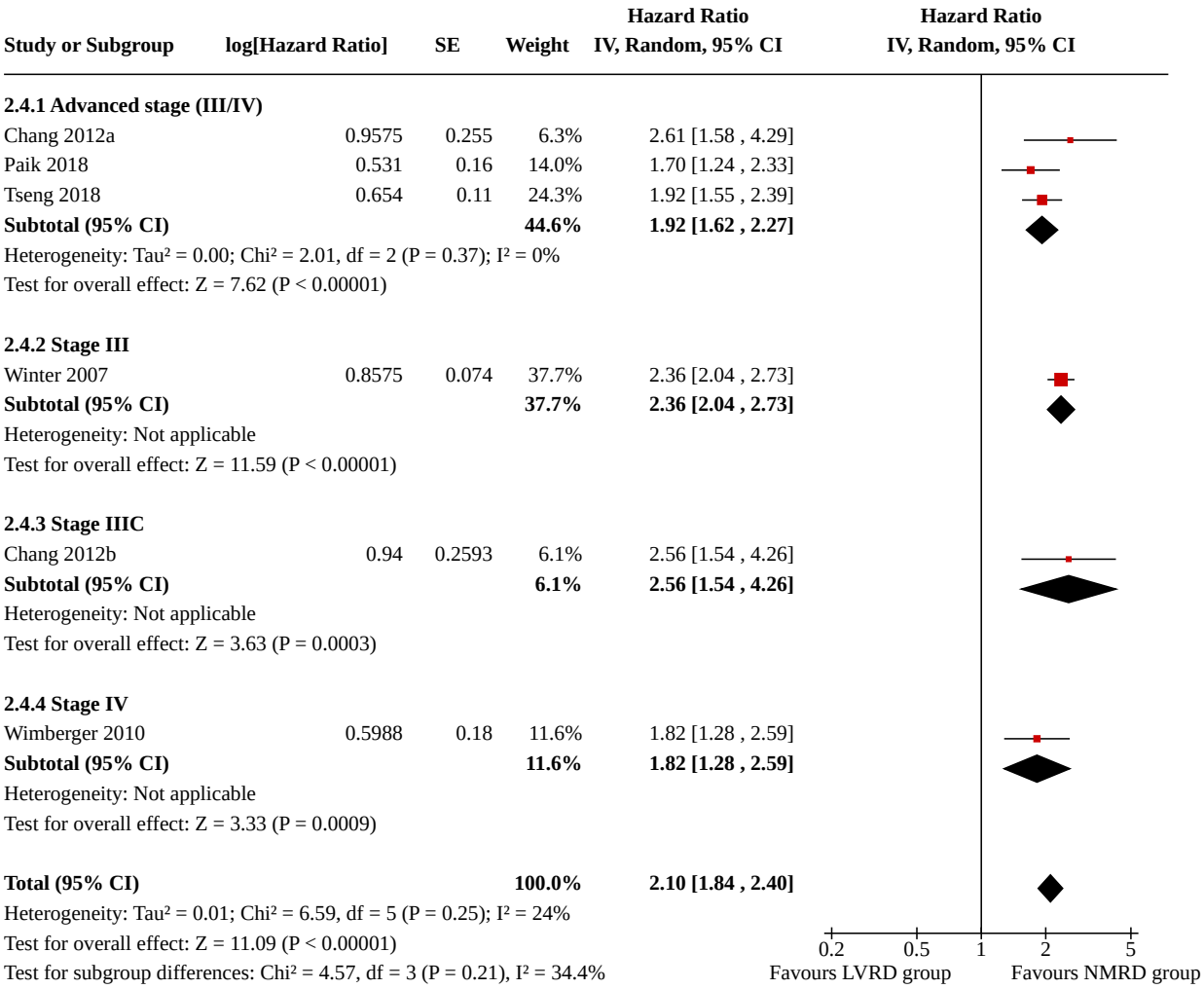


Analysis 2.3. Comparison 2: PDS: LVRD (> 1 cm) versus NMRD, Outcome 3: Overall survival - sensitivity analysis excluding Melamed 2017b and Winter 2007

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
2.3.1 Advanced stage (III/IV)				
Chang 2012a	1.1756	0.2723	7.1%	3.24 [1.90 , 5.53]
Kahl 2017	1.012	0.161	11.5%	2.75 [2.01 , 3.77]
Langstraat 2011	1.5063	0.2218	8.8%	4.51 [2.92 , 6.97]
Melamed 2017a	1.08	0.32	5.8%	2.94 [1.57 , 5.51]
Paik 2018	0.9343	0.25	7.8%	2.55 [1.56 , 4.15]
Tewari 2016	0.6235	0.292	6.5%	1.87 [1.05 , 3.31]
Tseng 2018	0.56	0.121	13.5%	1.75 [1.38 , 2.22]
Subtotal (95% CI)			**61.0%**	**2.63 [1.99 , 3.47]**
Heterogeneity: Tau² = 0.09; Chi² = 18.03, df = 6 (P = 0.006); I² = 67%				
Test for overall effect: Z = 6.83 (P < 0.00001)				
2.3.2 Stage IIIC				
Chang 2012b	1.1282	0.2757	7.0%	3.09 [1.80 , 5.30]
Chi 2006	1.308	0.25	7.8%	3.70 [2.27 , 6.04]
Eisenkop 2003	1.093	0.276	7.0%	2.98 [1.74 , 5.12]
Subtotal (95% CI)			**21.8%**	**3.27 [2.42 , 4.42]**
Heterogeneity: Tau² = 0.00; Chi² = 0.40, df = 2 (P = 0.82); I² = 0%				
Test for overall effect: Z = 7.71 (P < 0.00001)				
2.3.3 Stage IV				
Ataseven 2016	0.788	0.244	8.0%	2.20 [1.36 , 3.55]
Wimberger 2010	0.754	0.214	9.2%	2.13 [1.40 , 3.23]
Subtotal (95% CI)			**17.2%**	**2.16 [1.57 , 2.96]**
Heterogeneity: Tau² = 0.00; Chi² = 0.01, df = 1 (P = 0.92); I² = 0%				
Test for overall effect: Z = 4.78 (P < 0.00001)				
Total (95% CI)			**100.0%**	**2.65 [2.20 , 3.19]**
Heterogeneity: Tau² = 0.05; Chi² = 22.58, df = 11 (P = 0.02); I² = 51%				
Test for overall effect: Z = 10.33 (P < 0.00001)				
Test for subgroup differences: Chi² = 3.51, df = 2 (P = 0.17), I² = 43.1%				



Analysis 2.4. Comparison 2: PDS: LVRD (> 1 cm) versus NMRD, Outcome 4: Progression-free survival



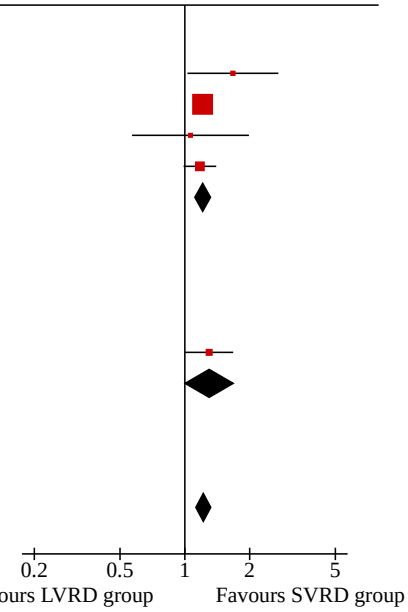
Comparison 3. PDS: LVRD (> 1 cm) versus SVRD (< 1 cm)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
3.1 Overall survival	5		Hazard Ratio (IV, Random, 95% CI)	1.22 [1.13, 1.32]
3.1.1 Advanced stage (III/IV)	4		Hazard Ratio (IV, Random, 95% CI)	1.21 [1.11, 1.32]
3.1.2 Stage IV	1		Hazard Ratio (IV, Random, 95% CI)	1.30 [1.00, 1.68]
3.2 Overall survival sensitivity analysis excluding Klar 2016	4		Hazard Ratio (IV, Random, 95% CI)	1.23 [1.08, 1.41]
3.2.1 Advanced stage (III/IV)	3		Hazard Ratio (IV, Random, 95% CI)	1.21 [1.03, 1.42]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
3.2.2 Stage IV	1		Hazard Ratio (IV, Random, 95% CI)	1.30 [1.00, 1.68]
3.3 Overall survival sensitivity analysis excluding 0 cm	3		Hazard Ratio (IV, Random, 95% CI)	1.20 [1.10, 1.30]
3.3.1 Advanced stage (III/IV)	3		Hazard Ratio (IV, Random, 95% CI)	1.20 [1.10, 1.30]
3.4 Overall survival sensitivity analysis including studies that included 0 cm	2		Hazard Ratio (IV, Random, 95% CI)	1.37 [1.09, 1.72]
3.4.1 Advanced stage (III/IV)	1		Hazard Ratio (IV, Random, 95% CI)	1.67 [1.03, 2.72]
3.4.2 Stage IV	1		Hazard Ratio (IV, Random, 95% CI)	1.30 [1.00, 1.68]
3.5 Progression-free survival	2		Hazard Ratio (IV, Random, 95% CI)	1.30 [1.08, 1.56]
3.5.1 Advanced stage (III/IV)	1		Hazard Ratio (IV, Random, 95% CI)	1.22 [1.12, 1.33]
3.5.2 Stage IV	1		Hazard Ratio (IV, Random, 95% CI)	1.49 [1.16, 1.92]

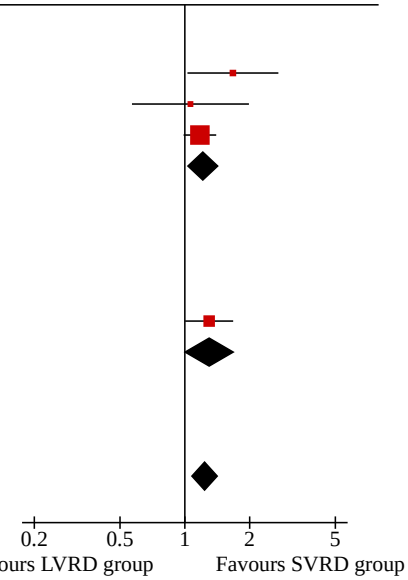
Analysis 3.1. Comparison 3: PDS: LVRD (> 1 cm) versus SVRD (< 1 cm), Outcome 1: Overall survival

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
3.1.1 Advanced stage (III/IV)				
Chan 2003	0.514	0.248	2.7%	1.67 [1.03 , 2.72]
Klar 2016	0.19	0.05	65.5%	1.21 [1.10 , 1.33]
Melamed 2017a	0.06	0.319	1.6%	1.06 [0.57 , 1.98]
Melamed 2017b	0.161	0.089	20.7%	1.17 [0.99 , 1.40]
Subtotal (95% CI)			**90.5%**	**1.21 [1.11 , 1.32]**
Heterogeneity: Tau² = 0.00; Chi² = 1.98, df = 3 (P = 0.58); I² = 0%				
Test for overall effect: Z = 4.48 (P < 0.00001)				
3.1.2 Stage IV				
Winter 2008	0.26	0.131	9.5%	1.30 [1.00 , 1.68]
Subtotal (95% CI)			**9.5%**	**1.30 [1.00 , 1.68]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 1.98 (P = 0.05)				
Total (95% CI)			**100.0%**	**1.22 [1.13 , 1.32]**
Heterogeneity: Tau² = 0.00; Chi² = 2.23, df = 4 (P = 0.69); I² = 0%				
Test for overall effect: Z = 4.87 (P < 0.00001)				
Test for subgroup differences: Chi² = 0.25, df = 1 (P = 0.61), I² = 0%				



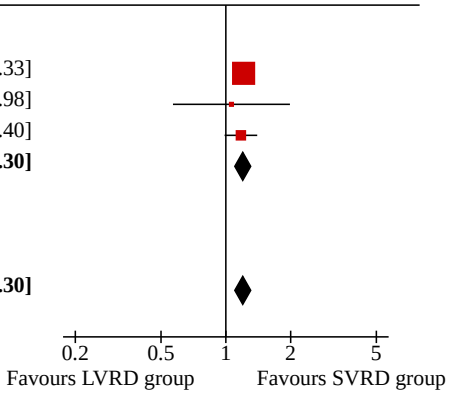
Analysis 3.2. Comparison 3: PDS: LVRD (> 1 cm) versus SVRD (< 1 cm), Outcome 2: Overall survival sensitivity analysis excluding Klar 2016

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
3.2.1 Advanced stage (III/IV)				
Chan 2003	0.514	0.248	7.7%	1.67 [1.03 , 2.72]
Melamed 2017a	0.06	0.319	4.7%	1.06 [0.57 , 1.98]
Melamed 2017b	0.161	0.089	59.9%	1.17 [0.99 , 1.40]
Subtotal (95% CI)			**72.3%**	**1.21 [1.03 , 1.42]**
Heterogeneity: Tau² = 0.00; Chi² = 1.98, df = 2 (P = 0.37); I² = 0%				
Test for overall effect: Z = 2.37 (P = 0.02)				
3.2.2 Stage IV				
Winter 2008	0.26	0.131	27.7%	1.30 [1.00 , 1.68]
Subtotal (95% CI)			**27.7%**	**1.30 [1.00 , 1.68]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 1.98 (P = 0.05)				
Total (95% CI)			**100.0%**	**1.23 [1.08 , 1.41]**
Heterogeneity: Tau² = 0.00; Chi² = 2.17, df = 3 (P = 0.54); I² = 0%				
Test for overall effect: Z = 3.06 (P = 0.002)				
Test for subgroup differences: Chi² = 0.19, df = 1 (P = 0.66), I² = 0%				



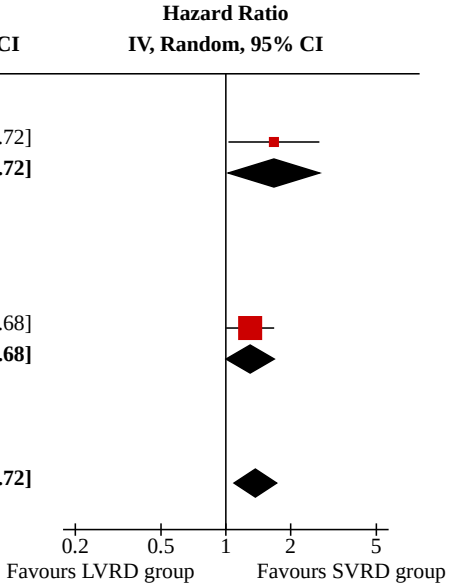
Analysis 3.3. Comparison 3: PDS: LVRD (> 1 cm) versus SVRD (< 1 cm), Outcome 3: Overall survival sensitivity analysis excluding 0 cm

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
3.3.1 Advanced stage (III/IV)				
Klar 2016	0.19	0.05	74.6%	1.21 [1.10 , 1.33]
Melamed 2017a	0.06	0.319	1.8%	1.06 [0.57 , 1.98]
Melamed 2017b	0.161	0.089	23.6%	1.17 [0.99 , 1.40]
Subtotal (95% CI)			**100.0%**	**1.20 [1.10 , 1.30]**
Heterogeneity: Tau² = 0.00; Chi² = 0.23, df = 2 (P = 0.89); I² = 0%				
Test for overall effect: Z = 4.19 (P < 0.0001)				
Total (95% CI)			**100.0%**	**1.20 [1.10 , 1.30]**
Heterogeneity: Tau² = 0.00; Chi² = 0.23, df = 2 (P = 0.89); I² = 0%				
Test for overall effect: Z = 4.19 (P < 0.0001)				
Test for subgroup differences: Not applicable				

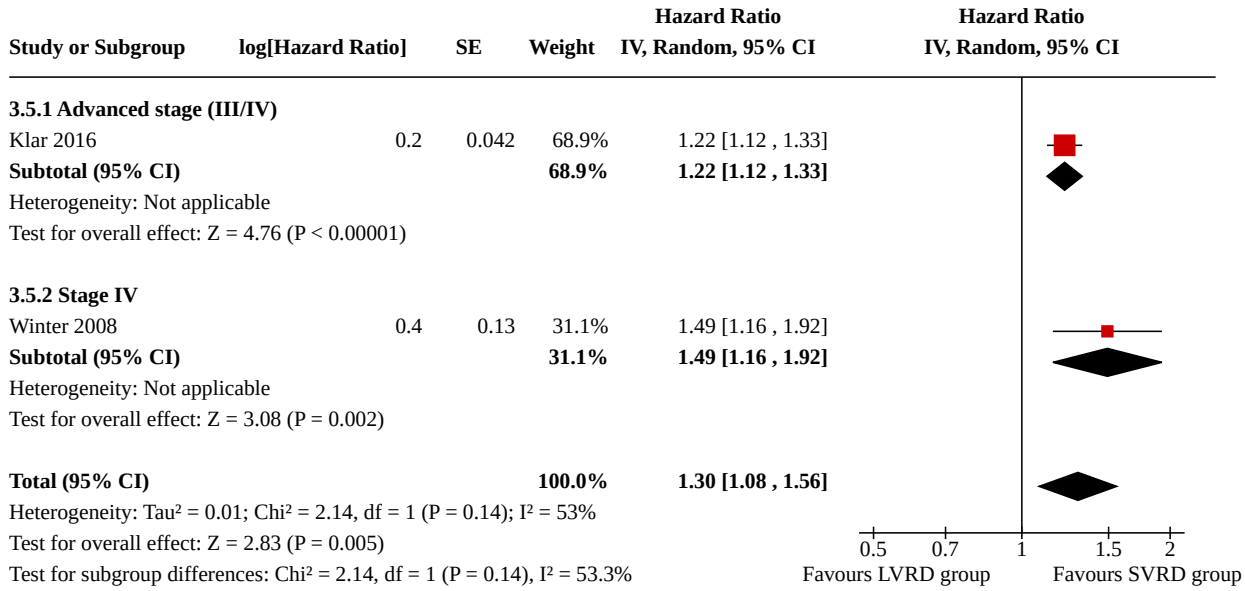


Analysis 3.4. Comparison 3: PDS: LVRD (> 1 cm) versus SVRD (< 1 cm), Outcome 4: Overall survival sensitivity analysis including studies that included 0 cm

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
3.4.1 Advanced stage (III/IV)				
Chan 2003	0.514	0.248	21.8%	1.67 [1.03 , 2.72]
Subtotal (95% CI)			**21.8%**	**1.67 [1.03 , 2.72]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 2.07 (P = 0.04)				
3.4.2 Stage IV				
Winter 2008	0.26	0.131	78.2%	1.30 [1.00 , 1.68]
Subtotal (95% CI)			**78.2%**	**1.30 [1.00 , 1.68]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 1.98 (P = 0.05)				
Total (95% CI)			**100.0%**	**1.37 [1.09 , 1.72]**
Heterogeneity: Tau² = 0.00; Chi² = 0.82, df = 1 (P = 0.37); I² = 0%				
Test for overall effect: Z = 2.72 (P = 0.006)				
Test for subgroup differences: Chi² = 0.82, df = 1 (P = 0.37), I² = 0%				



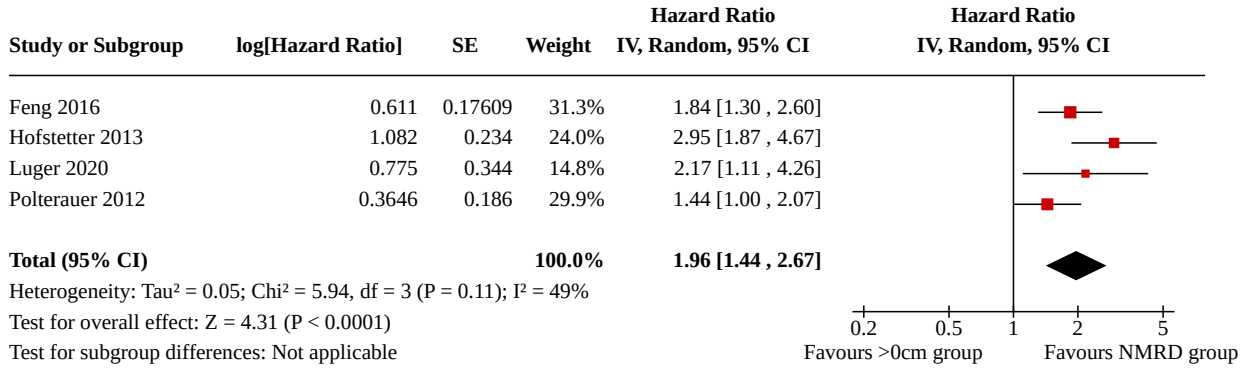
Analysis 3.5. Comparison 3: PDS: LVRD (> 1 cm) versus SVRD (< 1 cm), Outcome 5: Progression-free survival



Comparison 4. PDS: RD > 0 cm versus NMRD

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
4.1 Overall survival	4		Hazard Ratio (IV, Random, 95% CI)	1.96 [1.44, 2.67]
4.2 Progression-free survival	3		Hazard Ratio (IV, Random, 95% CI)	1.60 [1.36, 1.89]

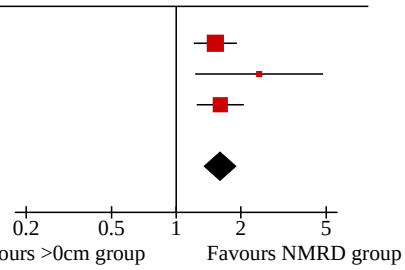
Analysis 4.1. Comparison 4: PDS: RD > 0 cm versus NMRD, Outcome 1: Overall survival



Analysis 4.2. Comparison 4: PDS: RD > 0 cm versus NMRD, Outcome 2: Progression-free survival

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
Feng 2016	0.42	0.118	51.3%	1.52 [1.21 , 1.92]
Luger 2020	0.89	0.35	5.8%	2.44 [1.23 , 4.84]
Polterauer 2012	0.47312	0.129	42.9%	1.60 [1.25 , 2.07]
Total (95% CI)			**100.0%**	**1.60 [1.36 , 1.89]**

Heterogeneity: Tau² = 0.00; Chi² = 1.62, df = 2 (P = 0.44); I² = 0%
Test for overall effect: Z = 5.56 (P < 0.00001)
Test for subgroup differences: Not applicable

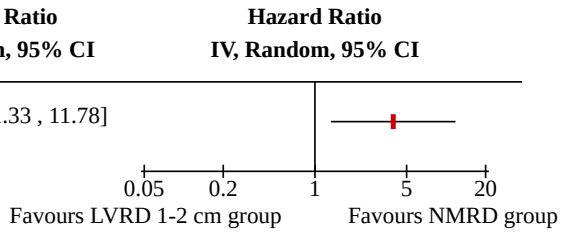


Comparison 5. PDS: LVRD 1 cm to 2 cm versus NMRD (stage IIIC)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
5.1 Overall survival	1		Hazard Ratio (IV, Random, 95% CI)	Subtotals only

Analysis 5.1. Comparison 5: PDS: LVRD 1 cm to 2 cm versus NMRD (stage IIIC), Outcome 1: Overall survival

Study or Subgroup	log[Hazard Ratio]	SE	Hazard Ratio IV, Random, 95% CI
Aletti 2006	1.3745	0.557	3.95 [1.33 , 11.78]

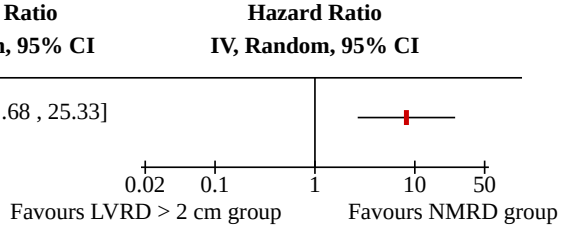


Comparison 6. PDS: LVRD (> 2 cm) versus NMRD (stage IIIC)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
6.1 Overall survival	1		Hazard Ratio (IV, Random, 95% CI)	Subtotals only

Analysis 6.1. Comparison 6: PDS: LVRD (> 2 cm) versus NMRD (stage IIIC), Outcome 1: Overall survival

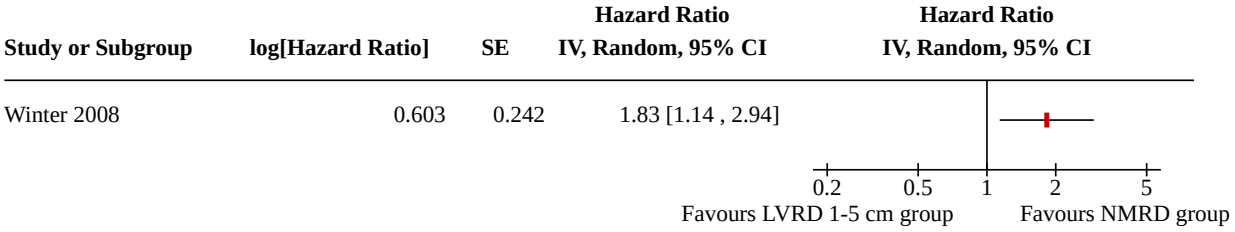
Study or Subgroup	log[Hazard Ratio]	SE	Hazard Ratio IV, Random, 95% CI
Aletti 2006	2.109	0.573	8.24 [2.68 , 25.33]



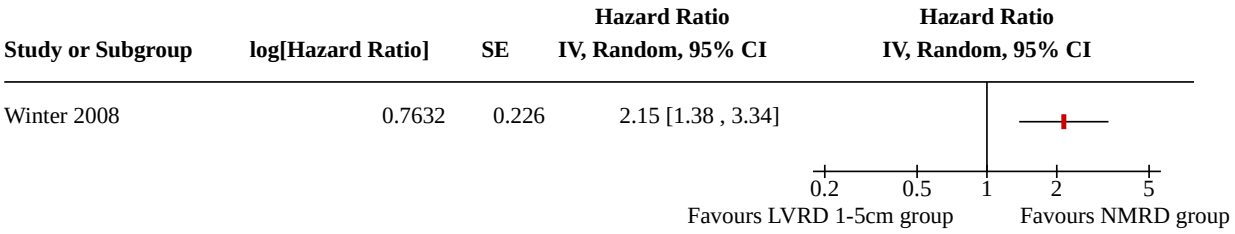
Comparison 7. PDS: LVRD 1 cm to 5 cm versus NMRD (stage IV disease)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
7.1 Overall survival	1		Hazard Ratio (IV, Random, 95% CI)	Subtotals only
7.2 Progression-free survival	1		Hazard Ratio (IV, Random, 95% CI)	Subtotals only

Analysis 7.1. Comparison 7: PDS: LVRD 1 cm to 5 cm versus NMRD (stage IV disease), Outcome 1: Overall survival



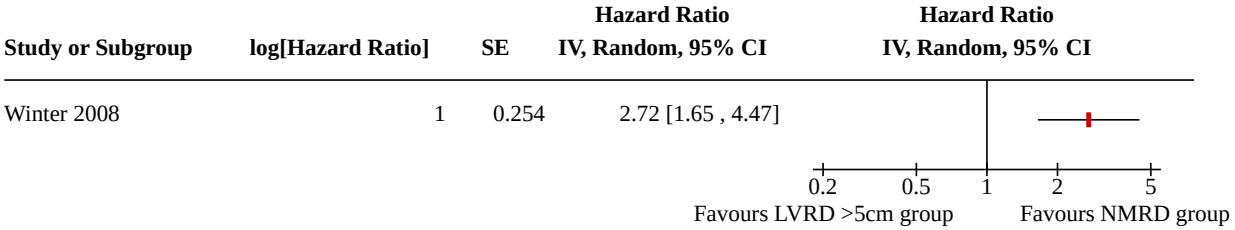
Analysis 7.2. Comparison 7: PDS: LVRD 1 cm to 5 cm versus NMRD (stage IV disease), Outcome 2: Progression-free survival



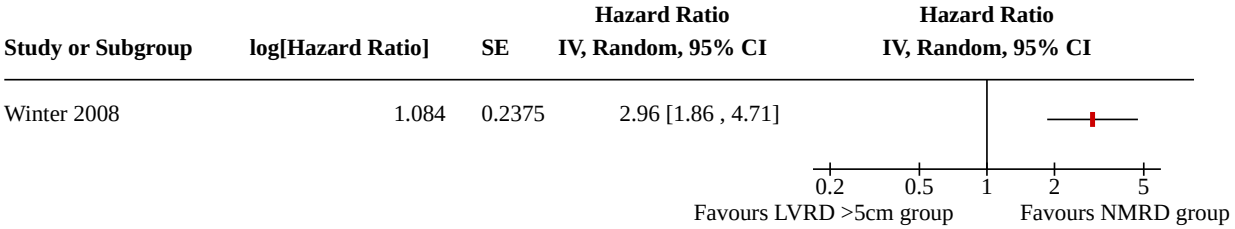
Comparison 8. PDS: LVRD (> 5 cm) versus NMRD (stage IV disease)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
8.1 Overall survival	1		Hazard Ratio (IV, Random, 95% CI)	Subtotals only
8.2 Progression-free survival	1		Hazard Ratio (IV, Random, 95% CI)	Subtotals only

Analysis 8.1. Comparison 8: PDS: LVRD (> 5 cm) versus NMRD (stage IV disease), Outcome 1: Overall survival



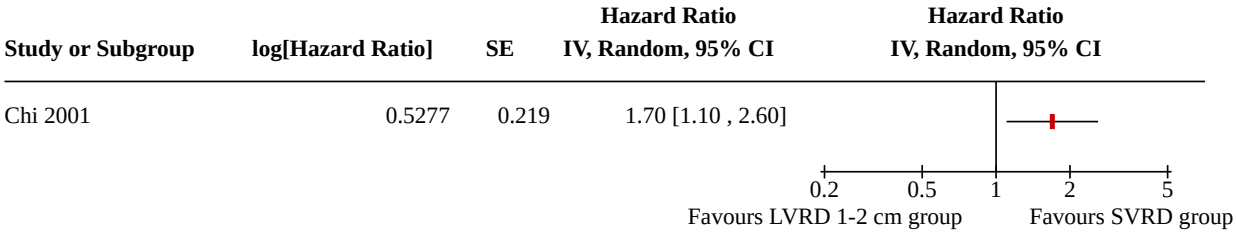
Analysis 8.2. Comparison 8: PDS: LVRD (> 5 cm) versus NMRD (stage IV disease), Outcome 2: Progression-free survival



Comparison 9. PDS: LVRD 1 cm to 2 cm versus SVRD (< 1 cm)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
9.1 Overall survival	1		Hazard Ratio (IV, Random, 95% CI)	Subtotals only

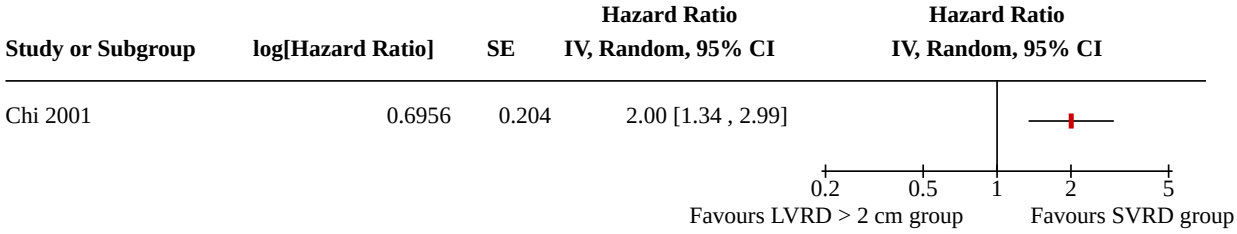
Analysis 9.1. Comparison 9: PDS: LVRD 1 cm to 2 cm versus SVRD (< 1 cm), Outcome 1: Overall survival



Comparison 10. PDS: LVRD (> 2 cm) versus SVRD (< 1 cm)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
10.1 Overall survival	1		Hazard Ratio (IV, Random, 95% CI)	Subtotals only

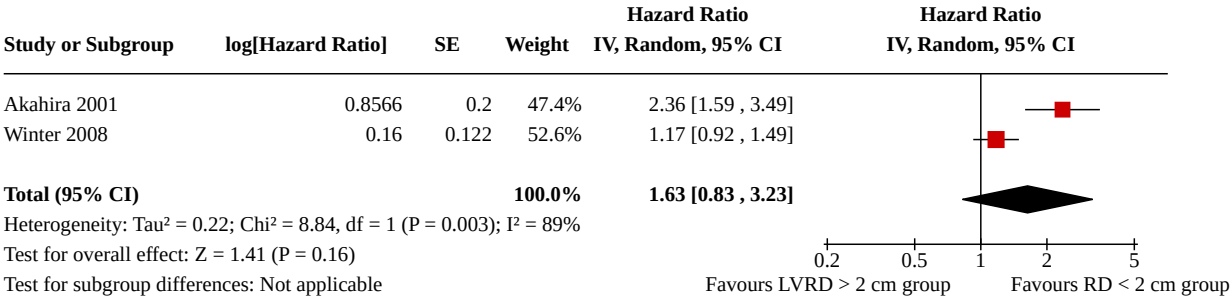
Analysis 10.1. Comparison 10: PDS: LVRD (> 2 cm) versus SVRD (< 1 cm), Outcome 1: Overall survival



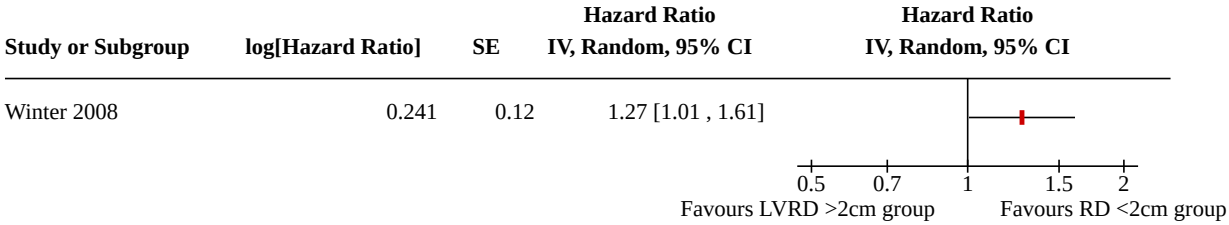
Comparison 11. PDS: LVRD (> 2 cm) versus RD < 2 cm (stage IV disease)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
11.1 Overall survival	2		Hazard Ratio (IV, Random, 95% CI)	1.63 [0.83, 3.23]
11.2 Progression-free survival	1		Hazard Ratio (IV, Random, 95% CI)	Subtotals only

Analysis 11.1. Comparison 11: PDS: LVRD (> 2 cm) versus RD < 2 cm (stage IV disease), Outcome 1: Overall survival



Analysis 11.2. Comparison 11: PDS: LVRD (> 2 cm) versus RD < 2 cm (stage IV disease), Outcome 2: Progression-free survival

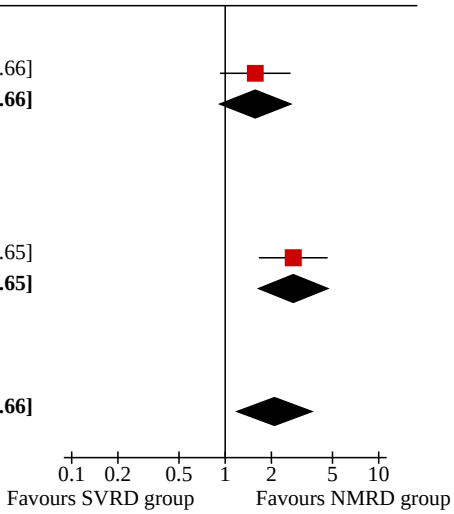


Comparison 12. IDS: SVRD (< 1 cm) versus NMRD

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
12.1 Overall survival	1		Hazard Ratio (IV, Random, 95% CI)	2.09 [1.20, 3.66]
12.1.1 ≤ 4 cycles of NACT	1		Hazard Ratio (IV, Random, 95% CI)	1.57 [0.93, 2.66]
12.1.2 > 4 cycles of NACT	1		Hazard Ratio (IV, Random, 95% CI)	2.78 [1.66, 4.65]
12.2 Progression-free survival	2		Hazard Ratio (IV, Random, 95% CI)	3.03 [0.81, 11.38]

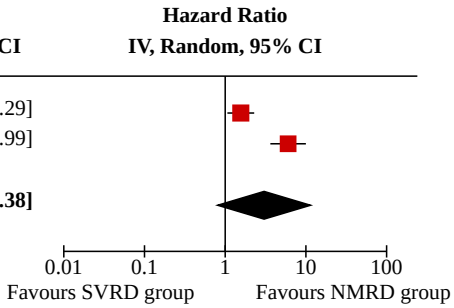
Analysis 12.1. Comparison 12: IDS: SVRD (< 1 cm) versus NMRD, Outcome 1: Overall survival

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
12.1.1 ≤ 4 cycles of NACT				
Phillips 2018	0.4513	0.2698	49.4%	1.57 [0.93 , 2.66]
Subtotal (95% CI)			**49.4%**	**1.57 [0.93 , 2.66]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 1.67 (P = 0.09)				
12.1.2 > 4 cycles of NACT				
Phillips 2018	1.021	0.263	50.6%	2.78 [1.66 , 4.65]
Subtotal (95% CI)			**50.6%**	**2.78 [1.66 , 4.65]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 3.88 (P = 0.0001)				
Total (95% CI)			**100.0%**	**2.09 [1.20 , 3.66]**
Heterogeneity: Tau² = 0.09; Chi² = 2.29, df = 1 (P = 0.13); I² = 56%				
Test for overall effect: Z = 2.60 (P = 0.009)				
Test for subgroup differences: Chi² = 2.29, df = 1 (P = 0.13), I² = 56.3%				



Analysis 12.2. Comparison 12: IDS: SVRD (< 1 cm) versus NMRD, Outcome 2: Progression-free survival

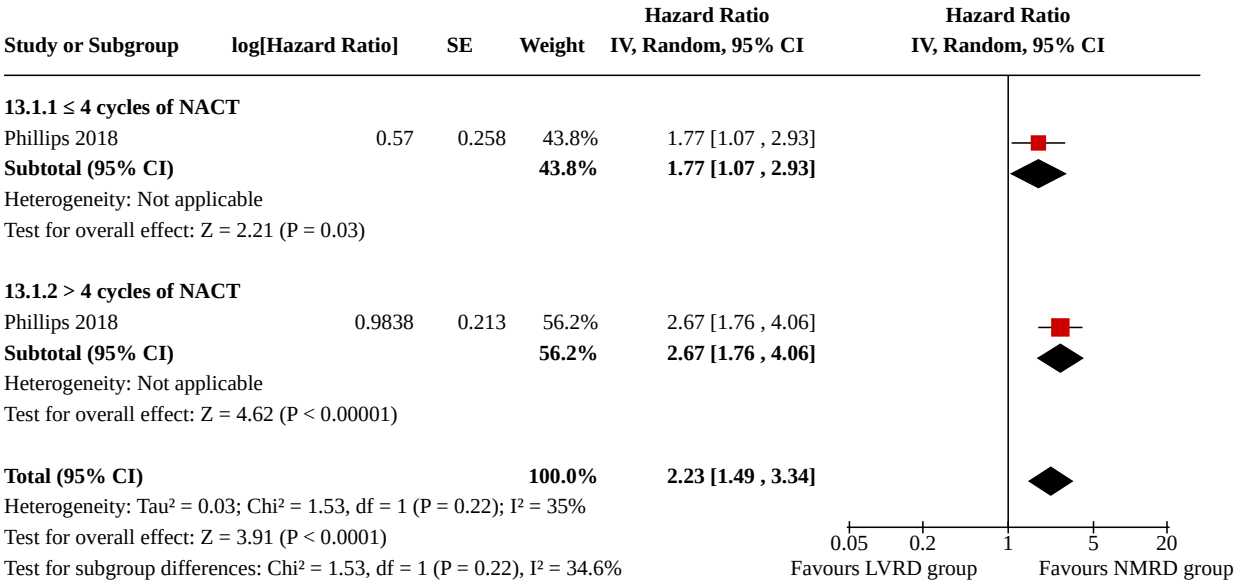
Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
Bixel 2020	0.445	0.195	50.8%	1.56 [1.06 , 2.29]
Liu 2020	1.795	0.2585	49.2%	6.02 [3.63 , 9.99]
Total (95% CI)			**100.0%**	**3.03 [0.81 , 11.38]**
Heterogeneity: Tau² = 0.86; Chi² = 17.38, df = 1 (P < 0.0001); I² = 94%				
Test for overall effect: Z = 1.64 (P = 0.10)				
Test for subgroup differences: Not applicable				



Comparison 13. IDS: LVRD (> 1 cm) versus NMRD

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
13.1 Overall survival	1		Hazard Ratio (IV, Random, 95% CI)	2.23 [1.49, 3.34]
13.1.1 ≤ 4 cycles of NACT	1		Hazard Ratio (IV, Random, 95% CI)	1.77 [1.07, 2.93]
13.1.2 > 4 cycles of NACT	1		Hazard Ratio (IV, Random, 95% CI)	2.67 [1.76, 4.06]

Analysis 13.1. Comparison 13: IDS: LVRD (> 1 cm) versus NMRD, Outcome 1: Overall survival



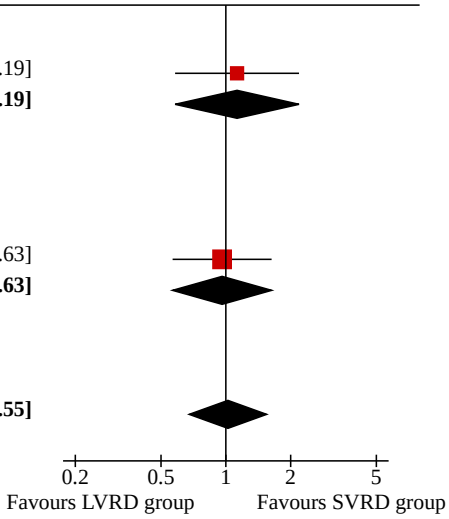
Comparison 14. IDS: LVRD (> 1 cm) versus SVRD (< 1 cm)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
14.1 Overall survival	1		Hazard Ratio (IV, Random, 95% CI)	1.02 [0.68, 1.55]
14.1.1 ≤ 4 cycles of NACT	1		Hazard Ratio (IV, Random, 95% CI)	1.13 [0.58, 2.19]
14.1.2 > 4 cycles of NACT	1		Hazard Ratio (IV, Random, 95% CI)	0.96 [0.57, 1.63]
14.2 Overall survival sensitivity analysis including 0 cm	6		Hazard Ratio (IV, Random, 95% CI)	1.60 [1.21, 2.11]
14.2.1 3 cycles of NACT	2		Hazard Ratio (IV, Random, 95% CI)	1.80 [0.94, 3.43]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
14.2.2 Median 4 cycles of NACT (disease-specific survival)	1		Hazard Ratio (IV, Random, 95% CI)	1.70 [1.06, 2.75]
14.2.3 Mean/median ~ 6 cycles of NACT	2		Hazard Ratio (IV, Random, 95% CI)	2.19 [1.07, 4.50]
14.2.4 ≤ 4 cycles of NACT	1		Hazard Ratio (IV, Random, 95% CI)	1.13 [0.58, 2.19]
14.2.5 > 4 cycles of NACT	1		Hazard Ratio (IV, Random, 95% CI)	0.96 [0.57, 1.63]
14.3 Overall survival sensitivity analysis excluding Phillips 2018	5		Hazard Ratio (IV, Random, 95% CI)	1.84 [1.34, 2.52]
14.3.1 3 cycles of NACT	2		Hazard Ratio (IV, Random, 95% CI)	1.80 [0.94, 3.43]
14.3.2 Median 4 cycles of NACT (disease-specific survival)	1		Hazard Ratio (IV, Random, 95% CI)	1.70 [1.06, 2.75]
14.3.3 Mean/median ~ 6 cycles of NACT	2		Hazard Ratio (IV, Random, 95% CI)	2.19 [1.07, 4.50]
14.4 Progression-free survival	4		Hazard Ratio (IV, Random, 95% CI)	1.76 [1.23, 2.52]
14.4.1 3 cycles of NACT	2		Hazard Ratio (IV, Random, 95% CI)	1.68 [0.90, 3.14]
14.4.2 Mean ~ 6 cycles of NACT	1		Hazard Ratio (IV, Random, 95% CI)	2.32 [1.08, 4.97]
14.4.3 All cycles	1		Hazard Ratio (IV, Random, 95% CI)	1.82 [1.12, 2.97]

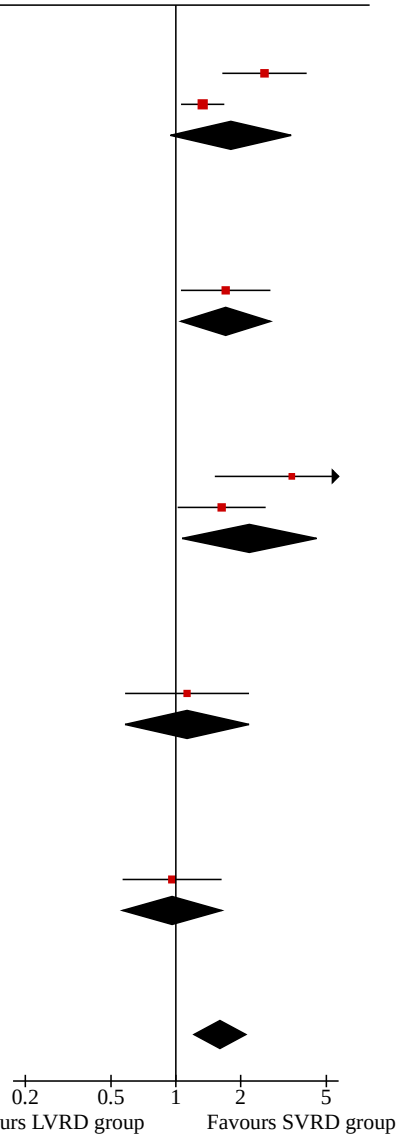
Analysis 14.1. Comparison 14: IDS: LVRD (> 1 cm) versus SVRD (< 1 cm), Outcome 1: Overall survival

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
14.1.1 ≤ 4 cycles of NACT				
Phillips 2018	0.12	0.338	39.0%	1.13 [0.58 , 2.19]
Subtotal (95% CI)			**39.0%**	**1.13 [0.58 , 2.19]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 0.36 (P = 0.72)				
14.1.2 > 4 cycles of NACT				
Phillips 2018	-0.04	0.27	61.0%	0.96 [0.57 , 1.63]
Subtotal (95% CI)			**61.0%**	**0.96 [0.57 , 1.63]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 0.15 (P = 0.88)				
Total (95% CI)			**100.0%**	**1.02 [0.68 , 1.55]**
Heterogeneity: Tau² = 0.00; Chi² = 0.14, df = 1 (P = 0.71); I² = 0%				
Test for overall effect: Z = 0.11 (P = 0.92)				
Test for subgroup differences: Chi² = 0.14, df = 1 (P = 0.71), I² = 0%				



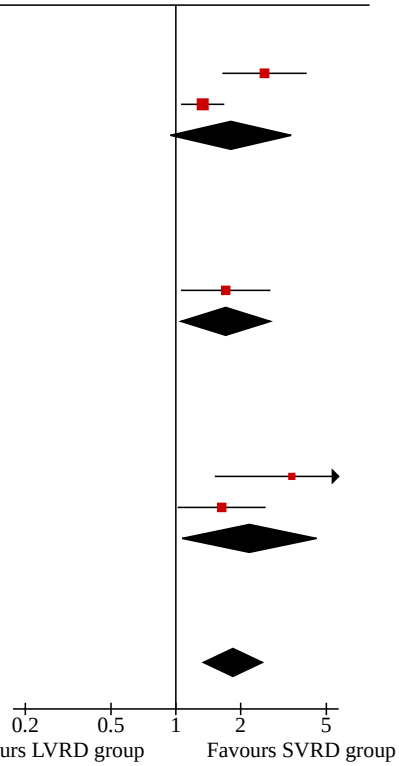
Analysis 14.2. Comparison 14: IDS: LVRD (> 1 cm) versus SVRD (< 1 cm), Outcome 2: Overall survival sensitivity analysis including 0 cm

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
14.2.1 3 cycles of NACT				
Zhang 2018	0.948	0.23	15.6%	2.58 [1.64 , 4.05]
Zhu 2016	0.287	0.118	22.5%	1.33 [1.06 , 1.68]
Subtotal (95% CI)			**38.1%**	**1.80 [0.94 , 3.43]**
Heterogeneity: Tau² = 0.19; Chi² = 6.54, df = 1 (P = 0.01); I² = 85%				
Test for overall effect: Z = 1.79 (P = 0.07)				
14.2.2 Median 4 cycles of NACT (disease-specific survival)				
Davidson 2019	0.533	0.244	14.8%	1.70 [1.06 , 2.75]
Subtotal (95% CI)			**14.8%**	**1.70 [1.06 , 2.75]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 2.18 (P = 0.03)				
14.2.3 Mean/median ~ 6 cycles of NACT				
Cioffi 2018	1.24	0.42	7.9%	3.46 [1.52 , 7.87]
Kaban 2017	0.49	0.24	15.1%	1.63 [1.02 , 2.61]
Subtotal (95% CI)			**23.0%**	**2.19 [1.07 , 4.50]**
Heterogeneity: Tau² = 0.16; Chi² = 2.40, df = 1 (P = 0.12); I² = 58%				
Test for overall effect: Z = 2.14 (P = 0.03)				
14.2.4 ≤ 4 cycles of NACT				
Phillips 2018	0.12	0.338	10.6%	1.13 [0.58 , 2.19]
Subtotal (95% CI)			**10.6%**	**1.13 [0.58 , 2.19]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 0.36 (P = 0.72)				
14.2.5 > 4 cycles of NACT				
Phillips 2018	-0.04	0.27	13.5%	0.96 [0.57 , 1.63]
Subtotal (95% CI)			**13.5%**	**0.96 [0.57 , 1.63]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 0.15 (P = 0.88)				
Total (95% CI)			**100.0%**	**1.60 [1.21 , 2.11]**
Heterogeneity: Tau² = 0.07; Chi² = 14.34, df = 6 (P = 0.03); I² = 58%				
Test for overall effect: Z = 3.33 (P = 0.0009)				
Test for subgroup differences: Chi² = 5.02, df = 4 (P = 0.28), I² = 20.4%				

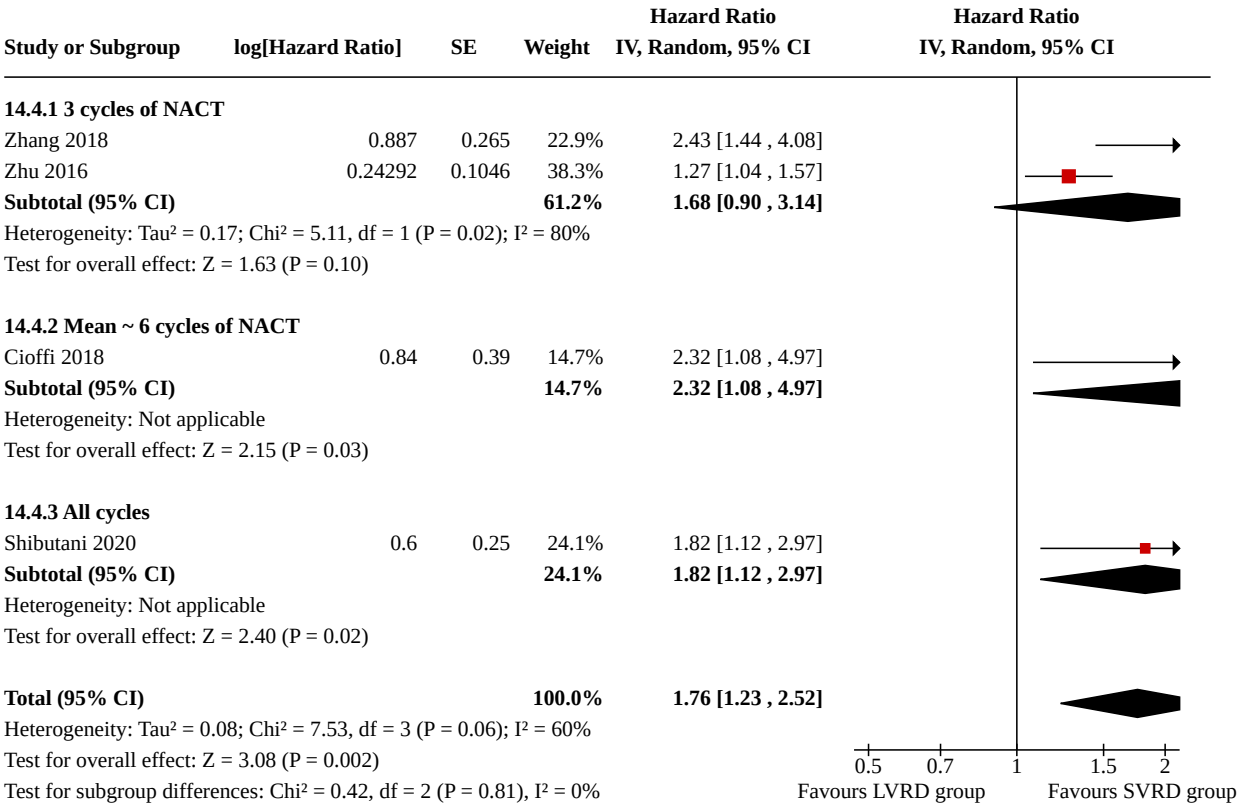


Analysis 14.3. Comparison 14: IDS: LVRD (> 1 cm) versus SVRD (< 1 cm), Outcome 3: Overall survival sensitivity analysis excluding Phillips 2018

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
14.3.1 3 cycles of NACT				
Zhang 2018	0.948	0.23	20.6%	2.58 [1.64 , 4.05]
Zhu 2016	0.287	0.118	29.7%	1.33 [1.06 , 1.68]
Subtotal (95% CI)			**50.3%**	**1.80 [0.94 , 3.43]**
Heterogeneity: Tau² = 0.19; Chi² = 6.54, df = 1 (P = 0.01); I² = 85%				
Test for overall effect: Z = 1.79 (P = 0.07)				
14.3.2 Median 4 cycles of NACT (disease-specific survival)				
Davidson 2019	0.533	0.244	19.5%	1.70 [1.06 , 2.75]
Subtotal (95% CI)			**19.5%**	**1.70 [1.06 , 2.75]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 2.18 (P = 0.03)				
14.3.3 Mean/median ~ 6 cycles of NACT				
Cioffi 2018	1.24	0.42	10.4%	3.46 [1.52 , 7.87]
Kaban 2017	0.49	0.24	19.8%	1.63 [1.02 , 2.61]
Subtotal (95% CI)			**30.2%**	**2.19 [1.07 , 4.50]**
Heterogeneity: Tau² = 0.16; Chi² = 2.40, df = 1 (P = 0.12); I² = 58%				
Test for overall effect: Z = 2.14 (P = 0.03)				
Total (95% CI)			**100.0%**	**1.84 [1.34 , 2.52]**
Heterogeneity: Tau² = 0.07; Chi² = 10.14, df = 4 (P = 0.04); I² = 61%				
Test for overall effect: Z = 3.79 (P = 0.0002)				
Test for subgroup differences: Chi² = 0.33, df = 2 (P = 0.85), I² = 0%				



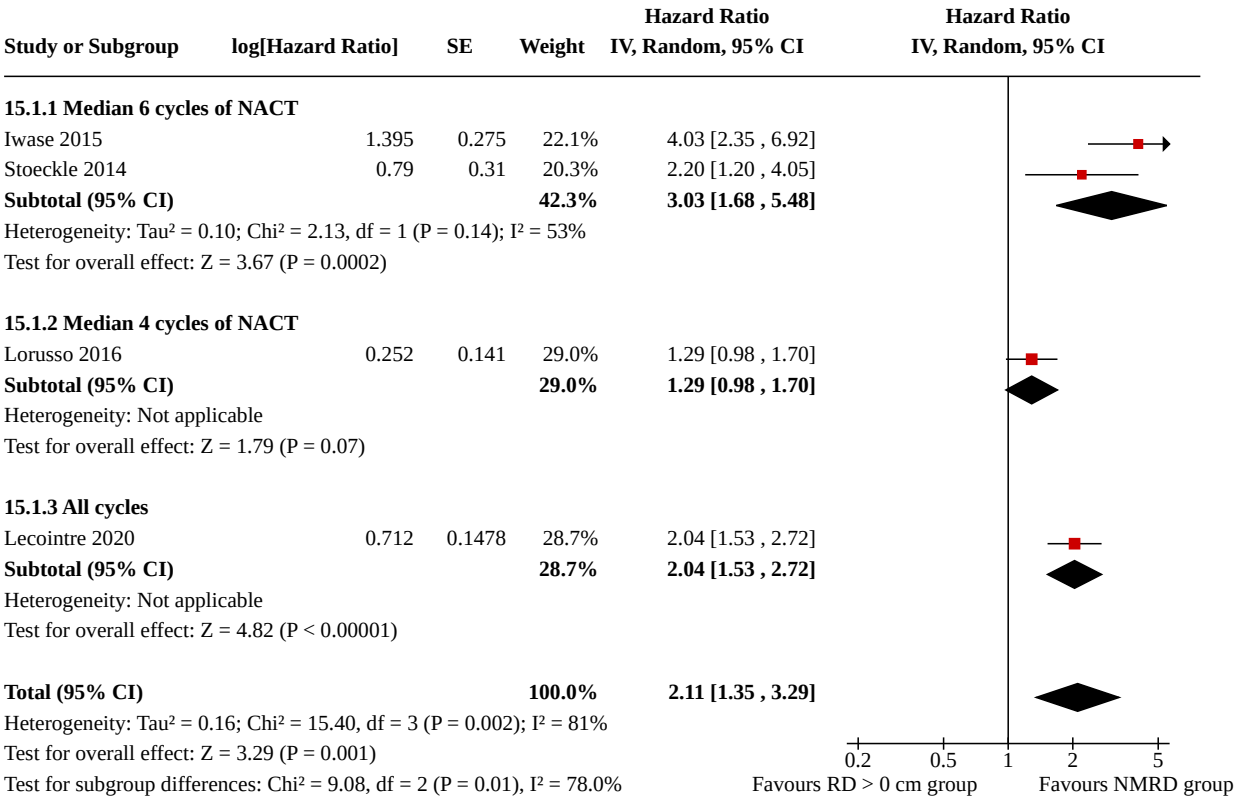
Analysis 14.4. Comparison 14: IDS: LVRD (> 1 cm) versus SVRD (< 1 cm), Outcome 4: Progression-free survival



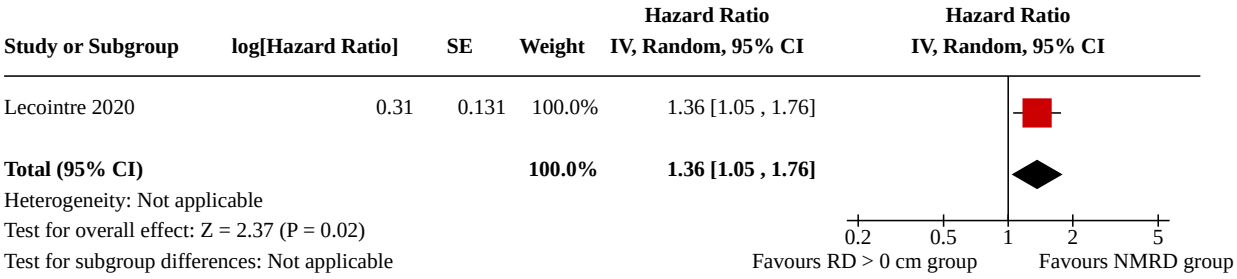
Comparison 15. IDS: RD > 0 cm versus NMRD

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
15.1 Overall survival	4		Hazard Ratio (IV, Random, 95% CI)	2.11 [1.35, 3.29]
15.1.1 Median 6 cycles of NACT	2		Hazard Ratio (IV, Random, 95% CI)	3.03 [1.68, 5.48]
15.1.2 Median 4 cycles of NACT	1		Hazard Ratio (IV, Random, 95% CI)	1.29 [0.98, 1.70]
15.1.3 All cycles	1		Hazard Ratio (IV, Random, 95% CI)	2.04 [1.53, 2.72]
15.2 Progression-free survival	1		Hazard Ratio (IV, Random, 95% CI)	1.36 [1.05, 1.76]

Analysis 15.1. Comparison 15: IDS: RD > 0 cm versus NMRD, Outcome 1: Overall survival



Analysis 15.2. Comparison 15: IDS: RD > 0 cm versus NMRD, Outcome 2: Progression-free survival



ADDITIONAL TABLES

Table 1. Summary of stage and residual disease in included upfront primary debulking surgery (PDS) studies

Study	No.	Stage		Optimal	Suboptimal	Median follow-up	Median age in years	Setting
	n	III n (%)	IV n (%)	n (%)	n (%)	Months	(Range)	
Akahira 2001	225	0 (0)	225 (100)	< 2: 70 (31)	> 2: 155 (69)	47.5 (13 to 112)	54 (26 to 85)	Japan
Aletti 2006	194	194 (100)	0 (0)	0: 46 (24) < 1: 85 (44)	1 to 2: 22 (11) > 2: 41 (21)	32.4 (0.2 to 126)	64 (24 to 87)	USA
Ataseven 2016	326	0 (0)	326 (100)	0: 157 (55) < 1: 88 (31)	> 1: 41 (14) NS: n = 40 exc.	34 (IQR: 12 to 70)	< 65: 205 (63) > 65: 121 (37)	Germany Austria
Bristow 2011	405	405 (100)	0 (0)	0: 209 (52) < 1: 196 (48)		33.0	59 (Range not reported)	USA
Chan 2003	104	84 (81)	20 (19)	< 1: 71 (68)	> 1: 33 (32)	33 (6 to 142)	Mean was 50.5 years and 61 years for younger and older women, respectively (Range: 22 and 85)	USA
Chang 2012a	203	189 (93)	14 (7)	0: 63 (31) < 1: 77 (38)	> 1: 63 (31)	43 (1 to 124)	54 (30 to 78)	South Korea
Chang 2012b	191	189 (100)	0 (0)	0: 61 (32) < 1: 67 (36)	> 1: 61 (32)	Not reported	54 (30 to 78)	South Korea
Chi 2001	282	216 (77)	66 (23)	< 1: 71 (25) 1 to 2: 73 (26)	> 2: 137 (49)	32 (1 to 139)	59 (22 to 87)	USA
Chi 2006	465	465	0	0: 67 (14)	> 1: 229 (49)	38	60	USA

Table 1. Summary of stage and residual disease in included upfront primary debulking surgery (PDS) studies *(Continued)*

		(100)	(0)	< 1: 169 (37)		(1 to 199)	(22 to 87)	
Cuylan 2018	218	218 (100)	0 (0)	0: 55 (25) < 1: 163 (75)		31.5	54 (18 to 78)	Turkey
Eisenkop 2003	408	408 (100)	0 (0)	0: 351 (86) < 1: 41 (10)	> 1: 16 (4)	32.8	62.8 (24 to 91)	USA
Feng 2016	625	n = 567 (91) stage III/IV		0: 209 (33)	> 0: 416 (67)	29 (3 to 100)	56 (30 to 84)	China
Hofstetter 2013	191	158 (83)	33 (17)	0: 121 (63)	> 0: 70 (37)	42	< 57: 98 > 57: 93	Europe
Kahl 2017	793	428 (54)	365 (46)	0: 482 (61) < 1: 226 (39)	> 1: 85	47 (IQR: 18 to 87)	60 (19 to 88)	Germany
Klar 2016	5055	4488/5130 (87.5) stage III/IV; n = 4850 in RD analysis		0: 1779 (37) < 1: 1442 (30)	> 1: 1629 (33)	0 to 144	Mean: 57.4 (SD 10.53)	Germany France Denmark
Langstraat 2011	280	210 (76)	67 (24)	0: 61 (22) < 1: 120 (43)	> 1: 95 (35)	3.2 years (0 to 15.8)	Mean: 73.5 (65 to 89)	USA
Luger 2020	178	91 (51)	87 (49)	0: 133 (75)	> 0: 45 (25)	49.6 (IQR 32.9 to 66.3)	64.6 years (IQR 50.8 to 72.7)	Austria
McGuire 1995	458	305 (67)	153 (33)	All sub-optimal	1 to 2 cm: 85 (18.6)	> 2 cm: 373 (81.4)	Not reported	USA
Melamed 2017a	307	241 (78)	66 (22)	0: 141 (59) < 1: 77 (32)	> 1: 23 (9) n = 66 missing	34.1	< 60: 200 (65) > 60: 107 (35)	USA
Melamed 2017b	6013	4954	1506	0: 2048 (46)	> 1: 546 (12)		< 60: 2803 (47)	

Table 1. Summary of stage and residual disease in included upfront primary debulking surgery (PDS) studies *(Continued)*

		(77)	(23)	< 1: 1848 (42)	1571 missing		> 60: 3210 (53)	
Paik 2018	419	370 (88)	49 (12)	0: 107 (26) < 1: 147 (35)	> 1: 165 (39)	43 (3 to 164)	Mean = 54.5 (SD 10.3)	South Korea
Peiretti 2010	259	199 (76)	60 (24)	0: 115 (44) < 1: 83 (32)	1 to 2: 18 (7) > 2: 43 (17)	29.8	58 (22 to 77)	Spain Italy
Peiretti 2012	238	180 (76)	58 (24)	0: 99 (41) < 1: 106 (44)	> 1: 32 (15)	Not reported	59.7 (22 to 85)	Italy USA
Polterauer 2012	226	II: 15 (7) III: 174 (77)	37 (16)	0: 157 (69)	> 0: 69 (31)	25.0 (1 to 49)	Mean: 57.5 (SD 11.9)	Europe
Shim 2016	276	III/IV (n = 276)		Not reported	Not reported	Not reported	54 (20 to 80)	South Korea
Tewari 2016	1718	1241 (72)	477 (28)	0: 85 (5) < 1: 701 (41)	> 1: 932 (54)	Not reported	58.5 to 60.2 for 0 to > 1 cm RD	USA
Tseng 2018	978	794 (81)	184 (19)	0: 408 (42) < 1: 378 (39)	> 1: 192 (19)	77.7 (1 to 198)	61 (19 to 95)	USA
Van Geene 1996	219	180 (82)	39 (18)	< 2 cm	< 2 cm: 92 (42)	> 2 cm: 127 (58)	57 (24 to 75)	UK
Wimberger 2010	573	573 (100)	0 (0)	0: 70 (12) < 1: 168 (29)	> 1: 335 (59)	Not reported	59 (19 to 83)	Germany France
Winter 2007	1895	1895 (100)	0 (0)	0: 437 (23) < 1: 791 (42)	> 1: 667 (35)	43	57 (16 to 86)	USA
Winter 2008	360	0	360	0 cm	0 cm: 29 (8) < 1 cm: 79 (22) Total: 108 (30)	28	59 (24 to 86)	USA

IQR: interquartile range;**RD:** residual disease; **SD:** standard deviation

Table 2. Summary of stage and residual disease in included interval debulking surgery (IDS) studies

Study	No.	Stage		Optimal	Suboptimal	Median follow-up	Median age in years	Setting
		III n (%)	IV n (%)	n (%)	n (%)	Months	(Range)	
Cioffi 2018	102	64 (63)	38 (37)	0: 37 (44) < 1: 20 (23)†	≥ 1: 28 (33)†	Not reported	Mean age ≥ 70 years: 74.5 (41%) < 70 years: 58.3 (59%)	Italy
Davidson 2019	282	IIIC: 114 (40) IV: 101 (36) Assumed AOC: 57 (20) Unknown: 10 (4)		0: 165 (59)‡ ≤ 1: 63 (22)‡	> 1 to 2: 6 (2)‡ > 2: 37 (13)‡	Not reported	63.9 (34.1 to 84.8)	USA
Iwase 2015	124	IIIB: 6 (5) IIIC: 77 (62)	41 (33)	< 1: 113 (91)	≥ 1: 11 (9)	39.5 (5 to 142)	58 (29 to 83)	Japan
Kaban 2017	203	Not reported		≤ 1: 165 (81)§	> 1: 36 (19)§	34.5 (1 to 124)	59 (28 to 84)	Turkey
Lecuru 2019	188	Not reported		Not reported		42.6	Not reported	France
Lorusso 2016	193	Not reported		Not reported		Not reported	Not reported	Italy
Petrillo 2014	322	251 (78)	72 (22)	No definition of optimal given 0: 236 (73) ≤ 1: 36 (11) > 1: 50 (16)		47 (3 to 181)	≤ 65: 226 (70%) > 65: 96 (30%)	Italy
Phillips 2018	398	273 (69)	123 (31)	0: 255 (64) < 1: 55 (14)	≥ 1: 88 (22)	Not reported	Mean: 63.9 (95% CI 42.2 to 85.6)	UK
Stoeckle 2014	118	82 (69)	36 (31)	0: 80 (68)	≥ 1: 7 (6)	37	64 (37 to 88)	France

Table 2. Summary of stage and residual disease in included interval debulking surgery (IDS) studies *(Continued)*

				< 1: 31 (26)				
Zhang 2018	200	169 (85)	31 (15)	0: 59 (30) < 1: 38 (19)	1 to 2: 8 (4) > 2: 30 (15)	43.5 (IQR 38.5 to 56.2)	61 (38 to 80)	China
Zhu 2016	672	564 (84)	108 (16)	≤ 1: 486 (72)	> 1: 186 (28)	38 (5 to 103)	55 (30 to 70)	China

†85/102 participants underwent debulking surgery following neoadjuvant chemotherapy.
‡Residual disease data available for n = 271/282.
§Residual disease data available for n = 201/203.
AOC: advanced ovarian cancer; **CI:** confidence interval; **IQR:** interquartile range

Table 3. Risk of bias assessment according to QUIPS (Quality in Prognostic Studies) for overall survival (OS) in primary debulking surgery (PDS) studies

Study	Study participation	Study attrition	Prognostic factor measurement	Outcome measurement	Adjustment for other prognostic factors	Statistical analysis and reporting
Akahira 2001	Low	Unclear	Unclear	Low	High	Unclear
Aletti 2006	Low	Unclear	Low	High	Low	Unclear
Ataseven 2016	Low	Unclear	Low	Low	Low	High
Bristow 2011	Low	Unclear	Low	Low	High	High
Chan 2003	Low	Unclear	Unclear	Low	Low	High
Chang 2012a	Low	Unclear	Low	Low	Low	High
Chang 2012b	Low	Unclear	Low	Low	Unclear	High
Chi 2001	Low	Unclear	Low	Low	Low	High
Chi 2006	Low	Unclear	Low	Low	Low	High
Cuylan 2018	Low	Unclear	Unclear	Low	Low	High
Eisenkop 2003	Low	Unclear	Low	Low	High	High
Feng 2016	Low	Unclear	Low	Low	Unclear	High
Hofstetter 2013	Unclear	Unclear	Low	Low	Unclear	High
Kahl 2017	Low	Unclear	Unclear	Low	Unclear	High
Klar 2016	Low	Unclear	Unclear	Low	Unclear	High
Langstraat 2011	Low	Unclear	Low	Low	Unclear	High
Luger 2020	Low	Unclear	Low	Low	Low	High
McGuire 1995	Low	Unclear	Low	Low	Unclear	High
Melamed 2017a	Low	Unclear	Low	Low	High	Unclear
Melamed 2017b	Low	Unclear	Low	Low	High	Unclear
Paik 2018	Low	Unclear	Low	Low	Unclear	High
Peiretti 2012	Low	Unclear	Unclear	Low	Low	High
Petrillo 2014	Low	Unclear	Low	Low	High	High
Polterauer 2012	Low	Unclear	Unclear	Low	Low	Unclear

Table 3. Risk of bias assessment according to QUIPS (Quality in Prognostic Studies) for overall survival (OS) in primary debulking surgery (PDS) studies *(Continued)*

Tewari 2016	Low	Unclear	Low	Low	Low	High
Tseng 2018	Low	Unclear	Low	Low	Low	High
Van Geene 1996	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear
Wimberger 2010	Low	Unclear	Low	Low	Unclear	High
Winter 2007	Low	Unclear	Low	Low	Unclear	Unclear
Winter 2008	Low	Unclear	Low	Low	Unclear	Unclear

Table 4. Risk of bias assessment according to QUIPS (Quality in Prognostic Studies) for overall survival (OS) in interval debulking surgery (IDS) studies

Study	Study participation	Study attrition	Prognostic factor measurement	Outcome measurement	Adjustment for other prognostic factors	Statistical analysis and reporting
Cioffi 2018	Low	Unclear	Low	Low	Low	Unclear
Davidson 2019	Low	Unclear	Unclear	High	High	Unclear
Iwase 2015	Unclear	Unclear	Low	Low	Low	High
Kaban 2017	Unclear	Unclear	Low	Low	Unclear	High
Lecointre 2020	Low	Unclear	Unclear	Low	High	Unclear
Lecuru 2019	High	Unclear	Low	Low	High	High
Liu 2020	Low	Unclear	Low	Low	High	High
Lorusso 2016	High	Unclear	Low	Low	High	High
Petrillo 2014	Low	Unclear	Low	Low	High	High
Phillips 2018	Low	Unclear	Low	Low	High	High
Stoeckle 2014	Low	Unclear	Low	Low	Unclear	Unclear
Zhang 2018	Low	Unclear	Low	Low	Unclear	Unclear
Zhu 2016	Low	Unclear	Unclear	Low	High	High

Table 5. Risk of bias assessment according to QUIPS (Quality in Prognostic Studies) for progression-free survival (PFS) in primary debulking surgery (PDS) studies

Study	Study participation	Study attrition	Prognostic factor measurement	Outcome measurement	Adjustment for other prognostic factors	Statistical analysis and reporting
Chang 2012a	Low	Unclear	Low	Unclear	Low	High
Chang 2012b	Low	Unclear	Low	Unclear	Unclear	High
Cuylan 2018	Low	Unclear	Unclear	Unclear	Low	High
Feng 2016	Low	Unclear	Low	Unclear	Unclear	High
Klar 2016	Low	Unclear	Unclear	Unclear	Unclear	High
Luger 2020	Low	Unclear	Low	Unclear	Low	High
McGuire 1995	Low	Unclear	Low	Unclear	Unclear	High
Paik 2018	Low	Unclear	Low	Unclear	Unclear	High
Peiretti 2010	Low	Unclear	Low	Unclear	High	High
Polterauer 2012	Low	Unclear	Unclear	Unclear	Low	Unclear
Shim 2016	High	Unclear	Low	Unclear	High	High
Tewari 2016	Low	Unclear	Low	Unclear	Low	High
Tseng 2018	Low	Unclear	Low	Unclear	Low	High
Wimberger 2010	Low	Unclear	Low	Unclear	Unclear	High
Winter 2007	Low	Unclear	Low	Unclear	Unclear	Unclear
Winter 2008	Low	Unclear	Low	Unclear	Unclear	Unclear

Table 6. Risk of bias assessment according to QUIPS (Quality in Prognostic Studies) for progression-free survival (PFS) in interval debulking surgery (IDS) studies

Study	Study participation	Study attrition	Prognostic factor measurement	Outcome measurement	Adjustment for other prognostic factors	Statistical analysis and reporting
Bixel 2020	Low	Unclear	Unclear	Unclear	High	High
Cioffi 2018	Low	Unclear	Low	Unclear	Low	Unclear
Lecointre 2020	Low	Unclear	Unclear	Unclear	High	Unclear

Table 6. Risk of bias assessment according to QUIPS (Quality in Prognostic Studies) for progression-free survival (PFS) in interval debulking surgery (IDS) studies *(Continued)*

Lecuru 2019	High	Unclear	Low	Unclear	High	High
Liu 2020	Low	Unclear	Low	Unclear	High	High
Petrillo 2014	Low	Unclear	Low	Unclear	High	High
Shibutani 2020	Low	Unclear	Low	Unclear	Low	High
Zhang 2018	Low	Unclear	Low	Unclear	Unclear	Unclear
Zhu 2016	Low	Unclear	Unclear	Unclear	High	High

APPENDICES

Appendix 1. MEDLINE search strategy

1. exp Ovarian Neoplasms/
2. (ovar* adj5 cancer*).mp.
3. (ovar* adj5 neoplas*).mp.
4. (ovar* adj5 carcinom*).mp.
5. (ovar* adj5 malignan*).mp.
6. (ovar* adj5 tumor*).mp.
7. (ovar* adj5 tumour*).mp.
8. 1 or 2 or 3 or 4 or 5 or 6 or 7
9. exp Surgical Procedures, Operative/
10. surg*.mp.
11. "surgery".fs.
12. 9 or 10 or 11
13. debulk*.mp.
14. cytoreduc*.mp.
15. 13 or 14
16. 8 and 12 and 15
17. "randomized controlled trial".pt.
18. "controlled clinical trial".pt.
19. randomized.ab.
20. randomly.ab.
21. trial.ab.
22. groups.ab.
23. exp Cohort Studies/
24. cohort*.mp.
25. (case adj series).mp.
26. 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25
27. 16 and 26
28. Animals/
29. Humans/
30. 28 not (28 and 29)
31. 27 not 30

Appendix 2. EMBASE search strategy

1. exp Ovary Tumor/
2. (ovar* adj5 (cancer* or neoplas* or carcinom* or malignan* or tumor* or tumour*)).mp.
3. 1 or 2
4. exp Surgery/
5. surg*.mp.
6. su.fs.
7. 4 or 5 or 6
8. (debulk* or cytoreduc*).mp.
9. 3 and 7 and 8
10. exp Controlled Clinical Trial/
11. crossover procedure/
12. double-blind procedure/
13. randomized controlled trial/
14. single-blind procedure/
15. random*.mp.
16. factorial*.mp.
17. (crossover* or cross over* or cross-over*).mp.
18. placebo*.mp.
19. (double* adj blind*).mp.
20. (singl* adj blind*).mp.
21. assign*.mp.
22. allocat*.mp.
23. volunteer*.mp.
24. exp cohort analysis/
25. cohort*.mp.
26. series.mp.
27. 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26
28. 9 and 27

key: mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name, fs=floating subheading

Appendix 3. CENTRAL search strategy

1. MeSH descriptor Ovarian Neoplasms explode all trees
2. ovar* near/5 cancer*
3. ovar* near/5 neoplas*
4. ovar* near/5 carcinom*
5. ovar* near/5 malignan*
6. ovar* near/5 tumor*
7. ovar* near/5 tumour*
8. (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7)
9. MeSH descriptor Surgical Procedures, Operative explode all trees
10. surg*
11. Any MeSH descriptor with qualifier: SU
12. (#9 OR #10 OR #11)
13. debulk*
14. cytoreduc*
15. (#13 OR #14)
16. (#8 AND #12 AND #15)

Appendix 4. Risk of bias and applicability assessment

Risk of bias and applicability assessment tool to assess risk of bias and applicability of prognostic factor studies (Riley 2019). Signalling questions and risk of bias ratings are listed in bullet points.

Domain 1: Participant selection

Risk of bias:

- Adequate participation in the study by eligible persons
- Description of the target population or population of interest
- Description of the baseline study sample
- Adequate description of the sampling frame and recruitment
- Adequate description of the period and place of recruitment
- Adequate description of inclusion and exclusion criteria

Risk of bias ratings:

- High: the relationship between the PF and outcome is very likely to be different for participants and eligible non-participants
- Moderate: the relationship between the PF and outcome may be different for participants and eligible non-participants
- Low: the relationship between the PF and outcome is unlikely to be different for participants and eligible non-participants

Applicability:

Are there concerns that the included women do not match the review question?

Domain 2: Study attrition

Risk of bias:

- Adequate response rate for study participants
- Description of attempts to collect information on participants who dropped out
- Reasons for loss to follow-up are provided
- Adequate description of participants lost to follow-up
- There are no important differences between participants who completed the study and those who did not

Risk of bias ratings:

- High: the relationship between the PF and outcome is very likely to be different for completing and non-completing participants
- Moderate: the relationship between the PF and outcome may be different for completing and non-completing participants
- Low: the relationship between the PF and outcome is unlikely to be different for completing and non-completing participants

Domain 3: Prognostic factor measurement

Risk of bias:

- A clear definition or description of the PF is provided
- Method of PF measurement is adequately valid and reliable
- Continuous variables are reported or appropriate cutpoints are used
- The method and setting of measurement of PF is the same for all study participants
- Adequate proportion of the study sample has complete data for the PF
- Appropriate methods of imputation are used for missing PF data

Risk of bias ratings:

- High: the measurement of the PF is very likely to be different for different levels of the outcome of interest
- Moderate: the measurement of the PF may be different for different levels of the outcome of interest
- Low: the measurement of the PF is unlikely to be different for different levels of the outcome of interest

Applicability:

Are there concerns that residual disease, the way that it is measured, or the way that it is interpreted, differ from the review question?

Domain 4: Outcome measurement

Risk of bias:

- A clear definition of the outcome is provided
- Method of outcome measurement used is adequately valid and reliable
- The method and setting of outcome measurement is the same for all study participants

Risk of bias ratings:

- High: the measurement of the outcome is very likely to be different related to the baseline level of the PF
- Moderate: the measurement of the outcome may be different related to the baseline level of the PF
- Low: the measurement of the outcome is unlikely to be different related to the baseline level of the PF

Applicability:

Are there concerns that outcome does not match the review question or that follow-up was not of sufficient duration?

Domain 5: Adjustment for other prognostic factors

Risk of bias:

- All other important PFs are measured
- Clear definitions of the important PFs measured are provided
- Measurement of all important PFs is adequately valid and reliable
- The method and setting of PF measurement are the same for all study participants
- Appropriate methods are used to deal with missing values of PFs, such as multiple imputation
- Important PFs are accounted for in the study design
- Important PFs are accounted for in the analysis

Applicability:

Did the prognostic factors adjusted for match the review question?

Risk of bias ratings:

- High: the observed effect of the PF on the outcome is very likely to be distorted by another factor related to PF and outcome
- Moderate: the observed effect of the PF on outcome may be distorted by another factor related to PF and outcome
- Low: the observed effect of the PF on outcome is unlikely to be distorted by another factor related to PF and outcome

Domain 6: Statistical analysis and reporting

Risk of bias:

- Sufficient presentation of data to assess the adequacy of the analytic strategy
- Strategy for model building is appropriate and is based on a conceptual framework or model
- The selected statistical model is adequate for the design of the study
- There is no selective reporting of results

Risk of bias ratings:

- High: the reported results are very likely to be spurious or biased related to analysis or reporting
- Moderate: the reported results may be spurious or biased related to analysis or reporting
- Low: the reported results are unlikely to be spurious or biased related to analysis or reporting

Appendix 5. Domains to be considered when judging the strength of the body of evidence

We considered the following domains when we assessed the strength of the body of evidence, based on the GRADE approach (Guyatt 2008):

- Risk of bias: Based on the results of the risk of bias assessments, we downgraded confidence in the evidence base if most evidence was from studies that we judged to be at high risk of bias.
- Indirectness: We downgraded confidence in the evidence base if we had concerns that the study sample, the prognostic factor, the outcome and/or the other factors in the models in the primary studies did not reflect the review question.

- Inconsistency: We downgraded confidence in the evidence base if there was unexplained heterogeneity or variability in results across studies.
- Imprecision: We downgraded confidence in the evidence base if the estimate of the effect size from a meta-analysis was not precise or, if no meta-analysis was performed, if the estimate of the size of effect from individual studies was not precise.
- Publication bias: Studies showing no association are likely to be unpublished, unless part of a larger study that specifically aimed to compare tests. We downgraded our confidence in the evidence base if we had reason to suspect publication bias from our assessments of reporting bias.
- Size of effect: We upgraded our confidence in the evidence base if the size of effect was moderate or large. If a meta-analysis was not possible, we upgraded if the size of effect was moderate or large for most included studies.

Appendix 6. Factors included in multivariate analysis in upfront primary debulking (PDS) studies

Citation	Factors included in multivariable (multivariate) analysis
Akahira 2001	Residual disease, histology and performance status
Aletti 2006	Residual disease, age, American Society of Anesthesiology (ASA) score, histological grade, operative time and aggressive surgery
Ataseven 2016	Age, performance status, residual tumour, tumour stage and ascites
Bristow 2011	Race, tumour grade 3, non-serous histology, ASA score > 3, surgical complexity score, serum albumin < 3.0 g/dL, platinum-based therapy, residual disease and perioperative morbidity
Chan 2003	Residual disease, age (older versus younger), stage (IV versus III) and performance status (1 to 2 versus 0)
Chang 2012a	Stage (IV), surgical procedure, residual disease and age
Chang 2012b	Residual disease, type of surgery, performance of lymphadenectomy and age
Chi 2001	Residual disease, age, stage (IIIC and IV versus IIIA/IIIB) and ascites (yes versus no)
Chi 2006	Residual disease, age and ascites
Cuylan 2018	Age, maximal cytoreduction and stage
Eisenkop 2003	Residual disease and sum of rankings
Feng 2016	Age, FIGO stage, residual disease and TTC
Hofstetter 2013	Interval from surgery to start of chemotherapy (≤ 28 versus < 28 days), stage (III versus IV), residual disease, age and extent of surgery
Kahl 2017	ACCI, ECOG PS, FIGO stage, surgical complexity score, blood loss, residual disease and duration of surgery
Klar 2016	Age, ECOG, BMI, stage, grading, residual tumour and histology
Langstraat 2011	Creatinine > 1.2 mg/dL, surgical complexity score, residual disease, stage IV disease and age
Luger 2020	Age (cat), CA-125, paraaortic nodes, FIGO, cardiophrenic lymph nodes dimension, residual disease
McGuire 1995	Residual disease, age, GOG performance status, histological subtype, stage or residual disease and measurable disease

(Continued)

Melamed 2017a	Age, race/ethnicity, stage, region, insurance status, treating facility type, hospital annual ovarian cancer volume, residual disease and presence of comorbidities
Melamed 2017b	
Paik 2018	Age, CA-125 level (U/mL), FIGO stage, residual disease and normal-sized ovary
Peiretti 2010	Age, stage IV vs IIIC and any residual disease
Peiretti 2012	Age, stage, histology, grade, presence of ascites and residual tumour at end of surgery
Polterauer 2012	Tumour stage, residual tumour, histological grade, histological type and age
Shim 2016	Not reported (abstract)
Tewari 2016	Age, race/ethnicity, performance status, grade, stage, histology,ascites, CA 125 (µg/ml), tumour residual and time from surgery to initiation of chemotherapy
Tseng 2018	Age, albumin, stage, ASA score, histology, BRCA status, OR Tumour Index, residual disease and postop IP chemotherapy
Van Geene 1996	Residual disease, performance status and pattern of spread
Wimberger 2010	Age, performance status, histology, residual tumour size, peritoneal carcinomatosis and stage IV disease site
Winter 2007	Residual disease, age (discrete), race, GOG performance status, histology and tumour grade
Winter 2008	Residual disease, histology and stage IV disease site

ACCI: age-adjusted Charlson Comorbidity Index; ASA: American Society of Anaesthesiologists; BMI: body mass index; BRCA: breast cancer gene; ECOG: Eastern Cooperative Oncology Group; FIGO: International Federation of Gynecology and Obstetrics; GOG: Gynaecologic Oncology Group; IP: intraperitoneal; PS: performance score; TTC: time to chemotherapy

Appendix 7. Factors included in multivariate analysis for each study on neoadjuvant chemotherapy (NACT) and interval debulking surgery (IDS)

Citation	Factors included in multivariable (multivariate) analysis
Bixel 2020	Residual disease, NACT cycles, route of chemotherapy administration (intraperitoneal or intravenous), maintenance therapy (yes/no)
Cioffi 2018	Residual disease, age, number of neoadjuvant chemotherapy courses, debulking surgery, ASA score, hypoalbuminaemia (defined as albuminaemia < 32 g/L), FIGO stage, presence of ascites ≥ 500 mL, high tumour dissemination and Charlson comorbidity score
Davidson 2019	Residual disease, ASA score, age, SCS and major morbidity
Iwase 2015	Residual disease, FIGO stage, histological subtype, NACT cycles, NACT regimen, systematic lymphadenectomy, excision of other organ(s), ascites cytology, lymph node metastasis
Kaban 2017	Residual disease, age, lymphadenectomy, macroscopic tumour in omentum, number of chemotherapy cycles

(Continued)

Lecointre 2020	Residual disease, number of NACT cycles (≤4, > 4), age (cat), Charlson index, FIGO, lymph node status (N+ vs N0), response to NACT
Lecuru 2019	Complete cytoreduction, ECOG, ascites, neutrophil/lymphocyte ratio, PCI at baseline, RECIST ORR (response rate at end of NACT), pCR and treatment arm (nintedanib vs placebo)
Liu 2020	Residual disease, age (cont)
Lorusso 2016	Residual disease, ECOG and number of NACT cycles†
Petrillo 2014	Residual disease, age, carcinomatosis at diagnosis, CA-125, pathological response to NACT
Phillips 2018	Residual disease, FIGO stage, chemotherapy regime (carbo/taxol vs carboplatin)
Shibutani 2020	Residual disease, age (cat), performance status, FIGO, disease type, histology, NACT cycles, NACT regimens
Stoeckle 2014	Residual disease, tumour grade, WHO performance status, ASA, bowel surgery (yes/no), FIGO stage
Zhang 2018	Residual disease, Pre-operative ascites, number of tumour sites, number of NAC cycles, CA-125 at diagnosis, CA-125 decreasing kinetics
Zhu 2016	Residual disease, FIGO stage, chemosensitivity, Glasgow prognostic score

†Full list of variables in multivariate analysis not explicitly mentioned.

ECOG: Eastern Cooperative Oncology Group; FIGO: International Federation of Gynecology and Obstetrics; NACT: neoadjuvant chemotherapy; PCI: Peritoneal Cancer Index; WHO: World Health Organization

HISTORY

Protocol first published: Issue 9, 2021

CONTRIBUTIONS OF AUTHORS

AE, BWR, KG and RN drafted the clinical and discussion sections of the review; AB, SH and PK data extracted items for inclusion in the review; AB drafted the methodological, results and discussion sections of the review. DC and LV reconciled the methodological and results sections of the review and contributed to the discussion. All authors agreed the final version.

DECLARATIONS OF INTEREST

- Andrew Bryant: none known
- Ahmed Elattar: none known
- Patience Kunonga: none known
- Brett A Winter-Roach: none known
- Shaun Hiu: none known
- Dawn Craig: none known
- Luke Vale: none known
- Ketankumar Gajjar: none known
- Raj Naik: none known

SOURCES OF SUPPORT

Internal sources

- None, Other

External sources

- National Institute for Health Research (NIHR), via Cochrane infrastructure funding to Cochrane Gynaecological, Neuro-oncology and Orphan Cancers, UK

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Three studies included a small proportion of women with early-stage (predominantly stage II) or unknown disease. Although not stringently part of our initial inclusion criteria, we included a study if the proportion with unknown or early-stage disease in the entire cohort was small. The proportion of women with early or unknown stage of disease in Feng 2016 (9.3%), Polterauer 2012 (6.6%) and Klar 2016 (12.5%) was not going to affect the applicability of the results.

The definitions of RD < 1 cm and RD > 1 cm were changed from near-optimal and suboptimal in the published protocol to small-volume residual disease (SVRD) and large-volume residual disease (LVRD), respectively. It was felt that this would make it easier to read for the non-clinical reader, as a combination of numbers and letters is more challenging and Cochrane Reviews have a large lay audience.

Appendix 6: Publication 4: Residual disease threshold after primary surgical treatment for advanced epithelial ovarian cancer (EOC), Part 1: A systematic review and network meta-analysis

American Journal of Therapeutics 30, e36–e55 (2023)

OPEN

Residual Disease Threshold After Primary Surgical Treatment for Advanced Epithelial Ovarian Cancer, Part 1: A Systematic Review and Network Meta-Analysis

Andrew Bryant, MSc,[1]* Eugenie Johnson, MSc,[1] Michael Grayling, PhD,[1] Shaun Hiu, MSc,[1] Ahmed Elattar, MD,[2] Ketankumar Gajjar, MD,[3] Dawn Craig, MSc,[1] Luke Vale, PhD,[1] and Raj Naik, MD[4]

Background: We present a systematic review and network meta-analysis (NMA) that is the precursor underpinning the Bayesian analyses that adjust for publication bias, presented in the same edition in AJT. The review assesses optimal cytoreduction for women undergoing primary advanced epithelial ovarian cancer (EOC) surgery.

Areas of Uncertainty: To assess the impact of residual disease (RD) after primary debulking surgery in women with advanced EOC. This review explores the impact of leaving varying levels of primary debulking surgery.

Data Sources: We conducted a systematic review and random-effects NMA for overall survival (OS) to incorporate direct and indirect estimates of RD thresholds, including concurrent comparative, retrospective studies of ≥100 adult women (18+ years) with surgically staged advanced EOC (FIGO stage III/IV) who had confirmed histological diagnoses of ovarian cancer. Pairwise meta-analyses of all directly compared RD thresholds was previously performed before conducting this NMA, and the statistical heterogeneity of studies within each comparison was evaluated using recommended methods.

Therapeutic Advances: Twenty-five studies (n = 20,927) were included. Analyses demonstrated the prognostic importance of complete cytoreduction to no macroscopic residual disease (NMRD), with a hazard ratio for OS of 2.0 (95% confidence interval, 1.8–2.2) for <1 cm RD threshold versus NMRD. NMRD was associated with prolonged survival across all RD thresholds. Leaving NMRD was predicted to provide longest survival (probability of being best = 99%). The results were robust to sensitivity analysis including only those studies that adjusted for extent of disease at primary surgery (hazard ratio 2.3, 95% confidence interval, 1.9–2.6). The overall certainty of evidence was moderate and statistical adjustment of effect estimates in included studies minimized bias.

Conclusions: The results confirm a strong association between complete cytoreduction to NMRD and improved OS. The NMA approach forms part of the methods guidance underpinning policy making in many jurisdictions. Our analyses present an extension to the previous work in this area.

Keywords: network meta-analysis, advanced epithelial ovarian cancer, complete cytoreduction, optimal cytoreduction, primary surgery

[1]*Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom;* [2]*Pan-Birmingham Gynaecological Oncology Cancer Centre, Birmingham, United Kingdom;* [3]*Nottingham City hospital, Obstetrics and Gynaecology, Nottingham, United Kingdom; and* [4]*Northern Gynaecological Oncology Centre, Gateshead, United Kingdom.*

The authors have no conflicts of interest to declare.

A. Bryant was lead author and drafted methodological, results, and discussion sections. E. Johnson contributed to methodological sections, formatting, and submission. M. Grayling and S. Hiu contributed to methodological and results sections. D. Craig and L. Vale added methodological expertise and reviewed the research. K. Gajjar and A. Elattar drafted clinical sections. R. Naik drafted clinical sections and was lead clinical expert.

**Address for correspondence: Andrew Bryant, MSc, Population Health Sciences Institute, Newcastle University, 4th Floor, Ridley Building 1, Queen Victoria Road, Newcastle upon Tyne NE1 7RU, England. E-mail: andy.bryant@ncl.ac.uk*

INTRODUCTION

Ovarian cancer is the seventh most common cancer among women up to 75 years of age and is a leading cause of death in women with gynecological malignancies.[1] Age older than 40 years, more than 90% of ovarian cancers originate from the surface (epithelial) cells of the ovary, termed epithelial tumors; the risk increases with age.[2,3] Around 70% of women with ovarian cancer are diagnosed at an advanced stage [International Federation of Gynaecology and Obstetrics (FIGO) stages III and IV].[4] That is, they have widespread tumor dissemination within the abdominal cavity, with the tumor potentially spreading to the liver, lungs, or distant organs.[5] As such, their prognosis is often poor.

Surgery and platinum-based chemotherapy are the mainstay of treatment in advanced epithelial ovarian cancer (EOC). The aim of primary surgery was to achieve "optimal cytoreduction," as the amount of residual disease (RD) (tumor remaining after surgery) is one of the most important prognostic factors for survival,[6–12] along with sensitivity to chemotherapy. The term "optimal cytoreduction" has been variably defined as referring to a maximal diameter of any residual tumor of between 0 and 1 cm, with RD greater than 1 cm being branded suboptimal.[7] "Complete cytoreduction" is achieved when there is no macroscopic residual disease (NMRD) (no visible tumor) left after surgery. A recently published National Ovarian Cancer Audit feasibility pilot report, by a British Gynaecological Cancer Society action group, highlights the need for more attempts at cytoreductive surgery in the United Kingdom.[13] In addition, some centers may not have the expertise to achieve complete cytoreduction, potentially leading to some patients not achieving optimal results for their individual surgery. The results from the Ovarian Cancer Audit feasibility pilot shows that on average only 51% of women with stage 2–4 and unstaged ovarian cancer receive surgery in England.[13] There are large disparities between surgeons and centers in their optimal and complete cytoreduction rates.[14–17] The development of these skills requires a shift in the surgeon's approach to surgery but, given that the additional procedures can be learned over a relatively short period, this could lead to increases in optimal or complete cytoreduction rates with no significant increases in perioperative morbidity.[15] It has previously been shown that optimal cytoreduction rates of up to 88% for primary laparotomy in advanced-stage ovarian cancer by gynecological oncologists working as a team can be achieved without any increase in morbidity.[16] Recent scientific and clinical studies relating to vascular epithelial growth factor receptors and BRCA/HRD status have opened up new avenues of treatment with biological agents, including vascular epithelial growth factor receptor inhibitors[18,19] and PARP inhibitors first line[20–23] and in relapsed setting[24,25] now becoming standard management practice. Thus, redefining the role and impact of complete cytoreduction in the overall survival (OS) outcomes of women with advanced EOC.

However, without reliable guidelines based on adequate empirical evidence, polarized views will continue to exist. Reliable quantification is important in its own right,[26] especially because there is still some resistance to incorporating statistical evidence into practice in many areas.[27] Although few refute the general conclusions of previous evidence suggesting that survival is better where there is complete cytoreduction compared with less-than-complete cytoreduction,[10,28–30] limitations in study design and in the conduct of previous analyses have not taken into account potential biases. Our review necessitated the inclusion of studies that reported adjusted analyses to attempt to minimize confounding bias. For example, if significantly more elderly women were included in a study where they were cytoreduced to NMRD than younger women with suboptimal RD thresholds, then there may be a confounding effect where suboptimal may be seen to have a better survival outcome. This is due to younger aged women being independently associated with prolonged survival, and therefore, NMRD may falsely seem to be associated as having worse survival than suboptimal RD.

Having the most up-to-date and reliable evidence is crucial to the development of clinical guidelines, and thus, it is of paramount importance that optimal analytical methods are used to appraise the available evidence.[31] Network meta-analysis (NMA)[32,33] is an extension to a standard pairwise meta-analysis that can incorporate and synthesize multiple treatments, or in this case RD thresholds, allowing for direct and indirect comparisons between groups that have previously not been compared in published studies. The use of NMA for guideline development is now common practice, with the method being well established within national health technology assessment agencies.[34] Furthermore, the World Health Organization and National Institute for Health and Care Excellence (NICE) have included recommendations on NMA within their clinical guidelines.[35,36] However, current guidelines related to optimal cytoreduction for women undergoing primary EOC surgery are not based on the highest level of evidence. A NMA on the back of the recent comprehensive systematic review (SR) in this area should provide robust evidence to policy makers

in the field.[31,37] The NMA reported in this SR is the precursor underpinning the Bayesian analyses that adjust for publication bias.[38] The Bayesian analyses are presented as the second part of this research and the publication is included in the same edition.

METHODS

Aim

To assess the impact of RD after primary debulking surgery in women with advanced EOC. This review explores the impact of leaving varying levels of RD after primary debulking surgery.

Eligibility criteria

We included retrospective prognostic studies that included adult women (older than 18 years) with surgically staged advanced EOC (FIGO stage III/IV) who had confirmed histological diagnoses of ovarian cancer. The population of interest was women who had received primary cytoreductive surgery followed by adjuvant platinum-based chemotherapy.[7]

The impact on survival of optimal and suboptimal cytoreduction for primary advanced disease was assessed using several RD thresholds reported in the literature. Included studies reported OS for comparisons of RD thresholds after surgery and used statistical adjustment for important baseline characteristics using multivariable analyses (eg, age, stage, and grade), to minimize confounding bias.[32,39] Owing to the nature of these retrospective studies, women were more likely to be allocated surgery by surgeon's preference. Consequently, there may be instances where a higher proportion of younger women, who are in better general health (measured using a performance status score[40]) for level of function and capability of self-care) undergo more aggressive surgery. These women may experience a better outcome than older women but this may be due to their better overall general health rather than the extent of resection. Therefore, adjusting for confounders is important to minimize effect distortion based on baseline imbalances. We included studies with a sample of at least 100 women. Smaller studies would have been restricted for the nature and extent of the adjusted analyses, due to the limited average number of participants per explanatory variable. Exclusion criteria included women with other concurrent malignancies, those who received chemotherapy before surgery (neoadjuvant), or intraperitoneal chemotherapy. This was to avoid the distortion of results to purify the data set and avoid the distorting effects of multitherapeutic interventions. Those with concurrent malignancies are not representative of EOC, and their inclusion would dilute external validity.

Search strategy

Electronic databases were searched from 1950 up to September 2021. Full reporting details are summarized in a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow chart (Figure 1) and in the published review.[7]

Study selection and data management

We followed the methodology as reported in Bryant et al,[7] in accordance with Cochrane guidelines.[32] At least 2 review authors were independently involved in the screening process and subsequently abstracted data.[7]

Risk of bias

At least 2 review authors independently assessed risk of bias. Although the included studies were a combination of RCTs, prospective, and retrospective designs, the comparison of RD was retrospective in nature. We therefore assessed risk of bias (and appraised quality) in the prognostic assessment of residual diease in included studies using the QUality In Prognosis Studies (QUIPS) tool. QUIPS is a tool designed to assess the risk of bias in prognostic factor studies.[41]

Data synthesis

The NMA synthesized studies according to guidance from the Cochrane Handbook for Systematic Reviews of Interventions,[32] NICE technical documents, and technology appraisal guidelines[42] and was reported according to the PRISMA extension for NMAs.[7,43,44] Although NMAs are typically used to synthesize only evidence from RCTs, the highly restrictive eligibility criteria applied to studies included in the SR underpinning the NMA permitted us to include retrospective studies, on the grounds that the women recruited into the studies being reviewed are comparable and could have been given surgery resulting in any of the RD thresholds considered in the network.[37]

The NMA used contrast based data and was conducted using a frequentist framework in Stata IC (version 15).[45–47] The analysis adjusted for multiarm trials and used the augmented approach.[47] Within the network, RD thresholds are depicted as nodes, with lines representing comparisons. All data sets and code in Stata are available on request from the corresponding author.

We did not anticipate design inconsistency to be a concern because our inclusion criteria limited heterogeneity in patient populations, primary disease, and

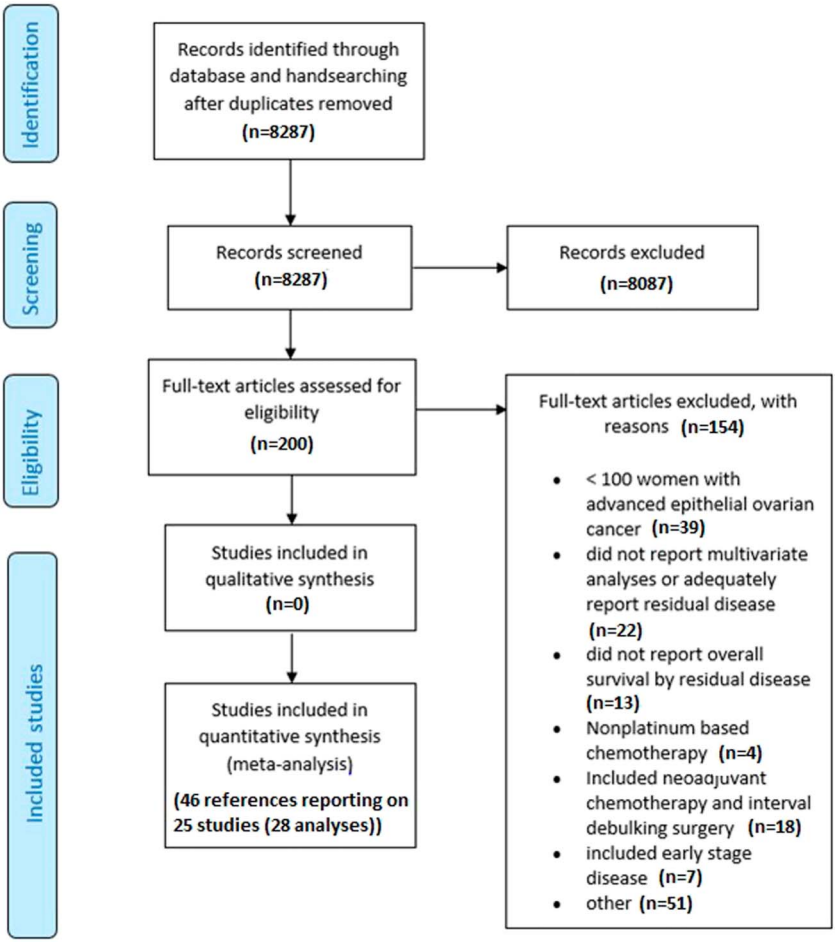


FIGURE 1. PRISMA flow diagram.

outcomes. There was no reason to suspect effect estimates would differ substantially in comparisons of thresholds across studies.

We conducted a network meta-regression for age, stage of disease, and histology to determine the similarity of studies for inclusion in the NMA. We presented the results of the network meta-regression using effect sizes reported as hazard ratios and 95% confidence intervals (CIs) because this is more useful than presenting a single global statistic in this case. All the RD thresholds are relative to the NMRD (0 cm) reference threshold. A meta-regression has been argued to have low power and be at risk of confounding[48,49] so we additionally checked summary and descriptive characteristics of studies to see whether there were any clear systematic differences between studies.

Transitivity in a NMA essentially necessitates that the underlying assumption of any indirect comparisons is that we can learn about the true relative effect of say RD <1 cm versus RD >1 cm through NMRD by combining the true relative effects of NMRD versus RD <1 cm and NMRD versus RD >1 cm. This means that we can compare RD <1 cm and RD >1 cm through NMRD. Therefore, the transitivity assumption underlying the NMA was evaluated by examining characteristics across studies; there were few concerns about potential effect modifiers across treatment comparisons as the distribution of key clinical characteristics, such as age, seemed similar across studies. Consistency, measured in agreement of direct and indirect evidence, was assessed by node-splitting analysis[46,47,50,51] and a formal global test for inconsistency.[46,47,51]

We presented the results of the NMA using effect sizes reported as hazard ratios and 95% CIs alongside results of the pairwise analyses reported in the SR underpinning the NMA. All the thresholds are relative to the NMRD reference threshold. We did not impute missing outcome data.

We also present plots showing the relative rank of all RD thresholds in OS (rankograms), which rank RD

Table 1. Characteristics of included studies in the NMA.

Study	Stage n (%)		RD (cm)		Median	RD reported in all models: covariates used in multivariable cox regression model	Median age in yr (range) or n (%) as reported	Country
	III	IV	Optimal n (%)	Suboptimal n (%)	F-U in mo (range)			
Akahira 2001[55]	0 (0)	225 (100)	<2: 70 (31)	>2: 155 (69)	47.5 (13–112)	Histology and performance status	54 (26–85)	Japan
Aletti 2006[14,56–58]	194 (100)	0 (0)	0: 46 (24) <1: 85 (44)	1–2: 22 (11) >2: 41 (21)	32.4 (0.2–126)	Age, ASA, histology, operative time, and aggressive surgery	64 (24–87)	USA
Ataseven 2016[59]	0 (0)	326 (100)	0: 157 (55) <1: 88 (31)	>1: 41 (14) NS: n = 40 exc	34 (IQR: 12–70)	Age, performance status, stage, and ascites	<65: 205 (63) >65: 121 (37)	Germany Austria
Bristow 2011[60]	405 (100)	0 (0)	0: 209 (52) <1: 196 (48)		33.0	Race, grade, histology, ASA, SCS, albumin, platinum therapy, and operative morbidity	59 Range not reported	USA
Chan 2003[61]	84 (81)	20 (19)	<1: 71 (68)	>1: 33 (32)	33 (6–142)	Age, stage, and performance status	Mean = 50.5 and 61 years for younger and older women, respectively, (range: 22 and 85).	USA
Chang 2012[62]	189 (93)	14 (7)	0: 63 (31) <1: 77 (38)	>1: 63 (31)	43 (1–124)	Age, stage, and type of surgery	54 (30–78)	South Korea
Chang 2012[63]	189 (100)	0 (0)	0: 61 (32) <1: 67 (36)	>1: 61 (32)	Not reported	Age, radical surgery, and lymphadenectomy	54 (30–78)	South Korea
Chi 2001[64]	216 (77)	66 (23)	<1: 71 (25) 1-2: 73 (26)	>2: 137 (49)	32 (1–139)	Age, stage, and ascites	59 (22–87)	USA
Chi 2006[65]	465 (100)	0 (0)	0: 67 (14) <1: 169 (37)	>1: 229 (49)	38 (1–199)	Age and ascites	60 (22–87)	USA
Cuylan 2018[67]	218 (100)	0 (0)	0: 55 (25) <1: 163 (75)		31.5	Age, stage, omental, peritoneal, and bilaterality present	54 (18–78)	Turkey
Eisenkop 2003[69]	408 (100)	0 (0)	0: 351 (86) <1: 41 (10)	>1: 16 (4)	32.8	Sum of rankings	62.8 (24–91)	USA
Feng 2016[70]	n = 567 (91) stage III/IV		0: 209 (33)	>0: 416 (67)	29 (3–100)	Age, stage, and time to chemotherapy	56 (30–84)	China
Hofstetter 2013[71]	158 (83)	33 (17)	0: 121 (63)	>0: 70 (37)	42	TSIC, stage, age, and extent of surgery	<57: 98 >57: 93	Europe

(*Continued on next page*)

Table 1. (*Continued*) Characteristics of included studies in the NMA.

Study	Stage n (%)		RD (cm)		Median	RD reported in all models: covariates used in multivariable cox regression model	Median age in yr (range) or n (%) as reported	Country
	III	IV	Optimal n (%)	Suboptimal n (%)	F-U in mo (range)			
Kahl 2017[72]	428 (54)	365 (46)	0: 482 (61) <1: 226 (39)	>1: 85	47 (IQR: 18–87)	Age adjusted CCI, performance status, stage, RD, histology, ascites, and SCS*	60 (19–88)	Germany
Klar 2016[73–78]	4488/5130 (87.5) stage III/IV; n = 4850 in RD analysis		0: 1779 (37) <1: 1442 (30)	>1: 1629 (33)	0–144	Age, ECOG status, BMI, stage, grade, and histology	Mean 57.4 (SD 10.53)	Germany France Denmark
Langstraat 2011[79]	210 (76)	67 (24)	0: 61 (22) <1: 120 (43)	>1: 95 (35)	3.2 years (0–15.8)	Age, creatinine, SCS, and stage	Mean: 73.5 (65–89)	USA
Luger 2020[80]	91 (51)	87 (49)	0: 133 (75)	>0: 45 (25)	49.6 (IQR: 33–66)	Age, CA-125, histologically positive paraaortic lymph nodes, FIGO, and CPLN.	64.6	Austria
Melamed 2017 [81]	241 (78)	66 (22)	0: 141 (59) <1: 77 (32)	>1: 23 (9) n = 66 missing	34.1	Age, ethnicity, stage, region, insurance status, facility type, hospital annual ovarian cancer volume, and comorbidities	<60: 200 (65) >60: 107 (35)	USA
Melamed 2017[81]	4954 (77)	1506 (23)	0: 2048 (46) <1: 1848 (42)	>1: 546 (12) 1571 missing			<60: 2803 (47) >60: 3210 (53)	
Paik 2018[82]	370 (88)	49 (12)	0: 107 (26) <1: 147 (35)	>1: 165 (39)	43 (3–164)	Age, CA-125, stage, and normal-sized ovary	Mean 54.5 (SD 10.3)	South Korea
Polterauer 2012[83]	II: 15 (7) III: 174 (77)	37 (16)	0: 157 (69)	>0: 69 (31)	25.0 (1–49)	Age, stage, grade, and histology	Mean 57.5 (SD 11.9)	Europe
Tewari 2016[84]	1241 (72)	477 (28)	0: 85 (5) <1: 701 (41)	>1: 932 (54)	Not reported	Age, ethnicity, performance status, grade, stage, histology, ascites, CA-125, and TSIC	58.5–60.2 for 0 to >1 cm RD	USA
Tseng 2018[85]	794 (81)	184 (19)	0: 408 (42) <1: 378 (39)	>1: 192 (19)	77.7 (1–198)	Age, albumin, stage, ASA score, histology, BRCA, OR tumor index, RD, and postop IP chemo	61 (19–95)	USA
Wimberger 2010[12]	573 (100)	0 (0)	0: 70 (12) <1: 168 (29)	>1: 335 (59)	Not reported	Age, performance status, histology, peritoneal carcinomatosis, and multiple sites	59 (19–83)	Germany, France

(*Continued on next page*)

Table 1. (*Continued*) Characteristics of included studies in the NMA.

Study	Stage n (%)		RD (cm)		Median	RD reported in all models: covariates used in multivariable cox regression model	Median age in yr (range) or n (%) as reported	Country
	III	IV	Optimal n (%)	Suboptimal n (%)	F-U in mo (range)			
Winter 2007[86–92]	1895 (100)	0 (0)	0: 437 (23) <1: 791 (42)	>1: 667 (35)	43	Age, race, performance status, histology, and grade	57 (16–86)	USA
Winter 2008[88,89,91,93,94]	360 (100)	0 (0)	0: 29 (8) <1: 78 (22)	1-5: 164 (46) >5: 89 (25)	28	Histology and stage IV disease site	59 (24–86)	USA
Winter 2008			<1: 78 (24)	>1: 253 (76)				
Winter 2008			<2: 50 (20)	>2: 203 (80)				

*SCS was added to multivariate analysis and was obtained through personal correspondence with Mr Beyhan Ataseven and included in the sensitivity analysis depicted in Table 5.

F-U, follow-up; NS, no surgery group excluded; OT, operative time; PS, performance status; ASA, American Society of Anaesthesiology score; SCS, surgical complexity score; omental, omental involvement; peritoneal, peritoneal involvement; CCI, Charlson comorbidity index; ECOG, Eastern Cooperative Oncology Group; BMI, body mass index; CA-125, cancer antigen 125 protein; TSIC, time from surgery to initiation of chemotherapy; BRCA, breast cancer mutation status; OR tumor index, scoring system to reflect extent of disease; IP, intraperitoneal; sum of rankings (numerical ranking system of progressively extensive tumor involvement for 5 anatomic regions); CPLN, cardiophrenic lymph node.

thresholds from having the highest probability (ranked 1) to the lowest probability (ranked 9) of maximizing OS. In addition, we report the "probability of being best" RD threshold, which assigns a probability that each RD threshold results in most prolonged survival relative to all others. Cumulative ranking probabilities using the surface under the cumulative ranking curve (SUCRA) were also calculated.[52] SUCRA presents a single value associated with each RD threshold. A value of 100% indicates the RD threshold is certain to be the most effective in the network (top ranked), while 0% indicates it is certain to be the least effective (in bottom rank). SUCRA was estimated through 10,000 repetitions in Stata using the network rank command.[45]

Sensitivity analysis

Because it was hypothesized that women with more extensive disease may have a poorer prognosis despite the outcome of their surgery, a sensitivity analysis including only studies that adequately adjusted for extent of disease at primary surgery was performed.

Certainty of the evidence

Guidance on the use of GRADE for prognostic factor studies has not yet been published,[53,54] but we appraised the quality and certainty of the evidence following existing guidelines for interventional SRs.[54] We based our judgment on the strength of the body of evidence based on the domains used by the GRADE Working Group (GRADE Working Group[54]). We interpreted our results in light of this graded evidence.

RESULTS

Study selection and characteristics

The flow of literature are shown in in the PRISMA diagram (Figure 1). The search strategy identified 8606 unique references, of which 200 progressed to full-text screening. At this stage, 154 were excluded, leaving 46 references[12,14,55–94] reporting on 25 primary studies[12,14,55,59–65,67,69–73,79–85,92,94] that met our inclusion criteria. Searches of the gray literature did not identify any additional relevant studies (Figure 1).

The 25 included studies assessed a total of 20,927 women, with the most having stage III disease. Three studies included a small proportion of women with early or unknown stage disease (range 3.6%–12.5%).[70,73,83] The analyses in Klar et al[73] included 1182 women with stage IIB-IIIB and 3684 women with stage IIIC-IV disease. This study contributed heavily to the analyses but results remained robust to its exclusion in a sensitivity analysis. See Table 1 for a full list of patient and study characteristics.

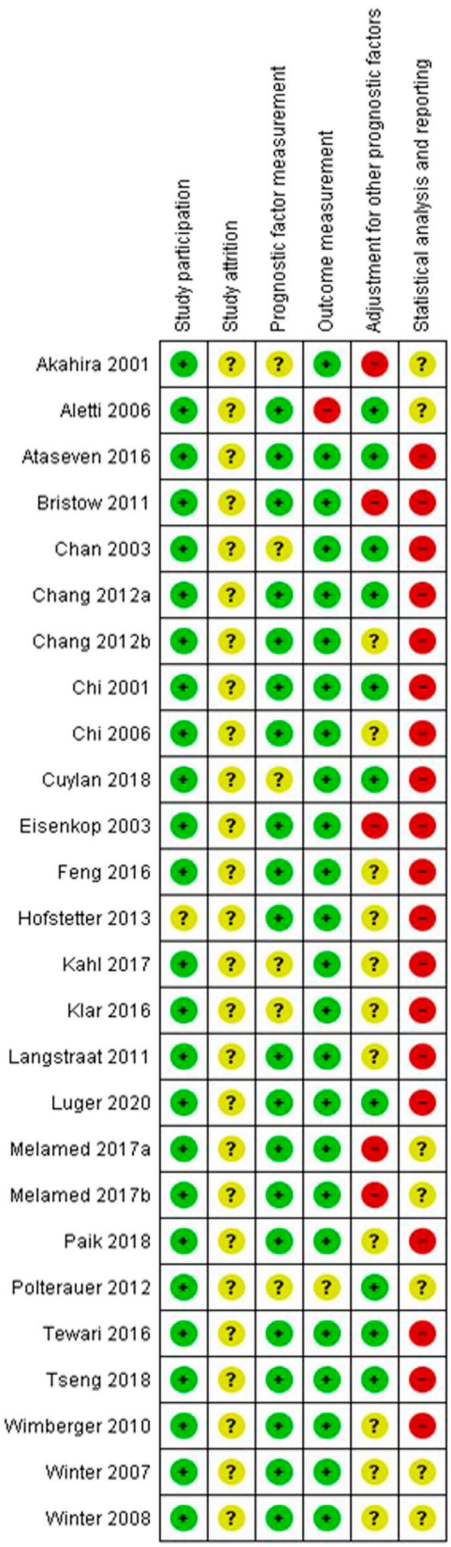


FIGURE 2. Risk of bias in included studies.

Table 2. Network meta-regression exploring age, FIGO stage, and histology.

RD*	Age†			FIGO stage‡			Histology§		
Ref¶ (0 cm)	HR‖	95% CI**	*P*††	HR‖	95% CI**	*P*††	HR‖	95% CI**	*P*††
<1 cm	0.98	0.96 to 1.01	0.24	1.00	1.00 to 1.01	0.13	0.99	0.99 to 1.00	0.02‡‡
>0 cm	1.03	0.93 to 1.13	0.62	1.00	0.97 to 1.03	0.86	0.96	0.93 to 1.00	0.07
1–2 cm	0.97	0.78 to 1.22	0.82	1.00	0.95 to 1.05	0.96	1.01	0.85 to 1.21	0.89
<2 cm	1.25	0.97 to 1.62	0.09	1.03	0.98 to 1.08	0.25	0.97	0.81 to 1.17	0.75
>1 cm	1.00	0.97 to 1.03	0.89	1.00	1.00 to 1.01	0.2	0.99	0.99 to 1.00	0.02‡‡
>2 cm	1.09	0.87 to 1.37	0.46	1.02	0.97 to 1.08	0.37	0.92	0.77 to 1.10	0.38

*RD thresholds of 1–5 cm and >5 cm were dropped due to detection of collinearity.
†Median age reported in this study was used except when not reported and mean was used.
‡Percentage of women in this study with International Federation of Gynecology and Obstetrics (FIGO) stage III EOC.
§Percentage of women in this study with serous histology.
¶Ref, reference: RD = 0 cm was used as the reference group.
‖HR, hazard ratio.
**CI, confidence interval.
††*P*: significance probability. This is the probability of the observed data or data more extreme, given the null hypothesis is true.
‡‡*P* was statistically significant but the HR point estimates and 95% CI's clearly show this is very unlikely to equate to any meaningful clinically significant differences in the percentage of women with serous histology across studies.

Risk of bias

The risk of bias assessments across all studies is shown in Figure 2. In general, most studies were at low to unclear risk of bias across domains but tended to be either at high or unclear risk for the statistical analysis and reporting domain. However, all included studies reported adjusted statistics to potentially minimize confounding bias. Owing to the restrictive inclusion criteria and attempts to minimize biases across the spectrum, studies were not necessarily at overall high risk of bias because they satisfied several of the criteria used to assess risk of bias.

Effects of interventions

The network meta-regression (Table 2) summarizes most covariates (age, stage, and histology) were not statistically significant ($P > 0.05$) in each of the RD comparisons. Although some covariates were statistically significant ($P < 0.05$) in a small number of comparisons, these differences were clearly not clinically meaningful. On examination of summary and descriptive characteristics (Table 1), there were no clear systematic differences between studies. We also checked the consistency assumption after completion of the NMA. There was no evidence of inconsistency in the network (see below).

Before data analysis, it is important to understand the geometry of the network.[95] The network plot shows which RD thresholds have been compared directly in studies and which can only be informed indirectly. The network geometry is depicted using the network diagram in Figure 3 and shows the range of RD thresholds and comparisons after optimal cytoreductive surgery for advanced EOC.[96] The RD thresholds presented in the NMA include complete cytoreduction to 0 cm (NMRD), 0.1–1 cm (0 cm < RD ≤1 cm, labelled as <1 cm for consistency with the published literature), >0 cm, 1–2 cm, >1 cm, 0.1–2 cm (labelled as <2 cm), >2 cm, 1–5 cm, and >5 cm. The nodes of some of the thresholds overlap, for example, >1 cm node overlaps with the 1–2 cm and >2 cm node, but these were all categorized as separate and unique nodes and interpreted accordingly and reflect the nature of data reported. Of note the 1–2 cm and <2 cm nodes included very sparse data so in that respect are less informative. Nodes where there were more comparative data available were for RD thresholds of 0 cm and <1 cm (indicated by the thick edge joining these 2 nodes in Figure 3). The comparisons of <1 cm and >1 cm included the 0 cm group, but this was deemed to have a negligible impact on the results and did not affect risk of bias profiles, certainty of the evidence or distort results because this was only applicable to 3 small studies.[61,64,94]

Table 3 summarizes the results of the NMA with a comparison of direct and indirect effect sizes of optimal and suboptimal RD thresholds. The results seem consistent across all split RD comparisons (sides), and there was no evidence of inconsistency in the network ($P = 0.48$).

The results in Table 4 and Figure 4 demonstrate prolonged survival if primary cytoreductive surgery

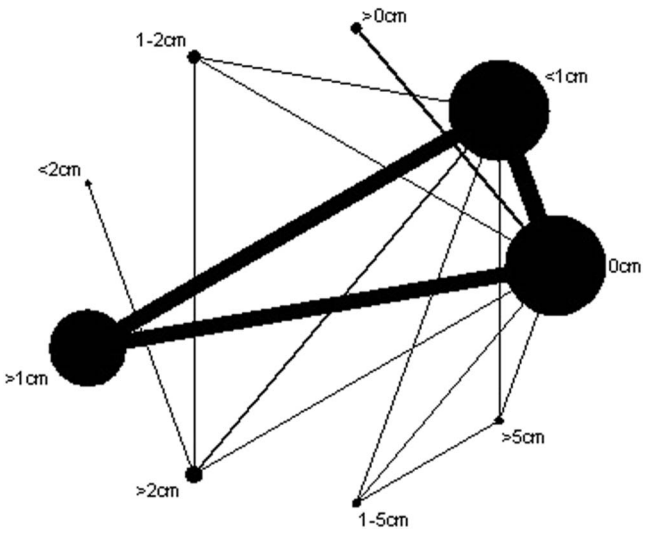


FIGURE 3. Network diagram showing RD comparisons after primary cytoreductive surgery for advanced EOC.

debulked to NMRD compared with any other RD threshold. Complete cytoreduction to NMRD was overwhelmingly the best ranked threshold because it was consistently ranked first (see Table 4 and Figure 5) with a very high probability of being the best RD threshold (SUCRA and *P*-best of 99.9% and 99%, respectively).

Table 4 also summarizes the benefit of incorporating reliable indirect estimates as an additional comparison between 0 cm versus <2 cm, while estimates for comparisons with sparse numbers are now more precise. A full breakdown of results is provided in in the detailed forest plots, which show results of all available comparisons (see Figure 6) and as a league table giving specific effect estimates for each and every comparison (see Table 5). There was no evidence of publication bias (see Figure 7).

Sensitivity analysis

A sensitivity analysis incorporating the results of 8 studies including an adequate adjustment for extent of disease at primary surgery increased the magnitude of effect estimates showed significantly prolonged survival in those with cytoreduction to NMRD (see Table 6). The results of this NMA also seem to be consistent across all sides in the network, and there was no evidence of overall inconsistency ($P = 0.31$). Other key probability and ranking statistics continued to provide strong evidence that NMRD (0 cm) is the best threshold ($P = 99.4\%$) and the SUCRA value remained very high (99.9%). Adjustment for extent of disease included: type (aggressive vs. standard) and extent of surgery; surgical complexity score; and progressively extensive tumor involvement in anatomic regions.

Table 3. Inconsistency test between direct and indirect RD threshold after primary surgery for advanced EOC comparisons in NMA.

Side	Direct		Indirect		Difference		
cm	Coefficient	SE	Coefficient	SE	Coefficient	SE	*P*
0 to <1*	0.688	0.063	0.570	0.353	0.118	0.358	0.741
0 to >0	No indirect estimate						
0 to 1–2	1.383	0.583	1.169	0.276	0.214	0.640	0.739
0 to >1	0.913	0.067	1.320	0.236	−0.406	0.245	0.097
0 to >2	2.119	0.597	1.349	0.265	0.770	0.648	0.235
0 to 1–5*	0.603	0.301	0.655	0.557	−0.052	0.639	0.936
0 to >5*	1.000	0.311	1.052	0.557	−0.052	0.639	0.936
0 to 1–2*	0.474	0.251	1.360	0.937	−0.886	0.957	0.354
0 to >1*	0.279	0.061	−0.356	0.335	0.635	0.340	0.062
<1 to >2*	0.743	0.243	1.629	0.939	−0.886	0.957	0.354
<1 to 1–5*	−0.057	0.308	−0.109	0.547	0.052	0.639	0.936
<1 to >5*	0.340	0.312	0.288	0.555	0.052	0.639	0.936
1–2 to >2*	0.265	0.250	−1.779	490.334	2.044	490.334	0.997
<2 to >2*	0.433	0.168	2.509	469.566	−2.075	469.566	0.996
1–5 to >5	No indirect estimate						

*All the evidence about these contrasts comes from the studies which directly compare them.
cm, centimeter; Coefficient, log hazard ratio; SE, standard error of log hazard ratio; *P*, significance probability (*P*) observed from the Z score.

Table 4. Results of NMA and pairwise analysis of optimal RD threshold after primary cytoreductive surgery for advanced EOC.

RD threshold versus 0 cm (reference)	NMA	Pairwise				
	HR (95% CI)	HR (95% CI)	n studies (participants)	Mean rank	*P* (best) %	SUCRA %
0 cm	Reference			1	99	99.9
<1 cm	1.98 (1.76–2.24)	2.03 (1.80–2.29)	17 (9404)	3.4	0	70.2
>0 cm	1.95 (1.48–2.58)	1.96 (1.44–2.67)	4 (1220)	3.4	0	70.6
1–2 cm	3.34 (2.04–5.47)	3.95 (1.33–11.78)	1 (68)	7.3	0	21.8
<2 cm	2.82 (1.58–5.04)	No direct estimate		6.0	0	36.9
>1 cm	2.57 (2.26–2.93)	2.50 (2.13–2.94)	14 (7988)	5.8	0	40.0
>2 cm	4.36 (2.69–7.04)	8.24 (2.68–25.33)	1 (87)	8.7	0	3.4
1–5 cm	1.85 (1.11–3.08)	1.83 (1.14–2.94)	1 (193)	3.2	1	72.0
>5 cm	2.75 (1.62–4.67)	2.72 (1.65–4.47)	1 (118)	6.2	0	35.3

HR, hazard ratio; *P* (best), probability that RD threshold is the best.

DISCUSSION

We identified 25 studies meeting our inclusion criteria. These studies assessed survival after primary cytoreductive surgery followed by adjuvant platinum-based chemotherapy in women with advanced EOC. The Sundar et al.[13] underpinning the NMA and the results of our updated analysis provides more precise and reliable estimates than seen in previous studies and reviews in this area,[6,8–12,97] which should enable more informed decisions to be made. Although the findings do not enable us to determine whether the survival benefit is a direct effect of the surgical intervention, they may encourage the surgical community to strive toward improving rates of complete cytoreduction and perhaps more centers adopting a more aggressive approach to attempt to improve rates of complete cytoreduction. Factors such as training, high-dependency unit support, patient selection and developing inter-surgical collaborations such as colo-rectal, upper gastrointestinal, hepato-biliary and vascular specialties could be important to help achieve this. RD and complete cytoreduction rates should be part of routinely collected cancer data and should be a quality indicator for advanced ovarian cancer surgery along with other indicators recommended by the British Gynaecological

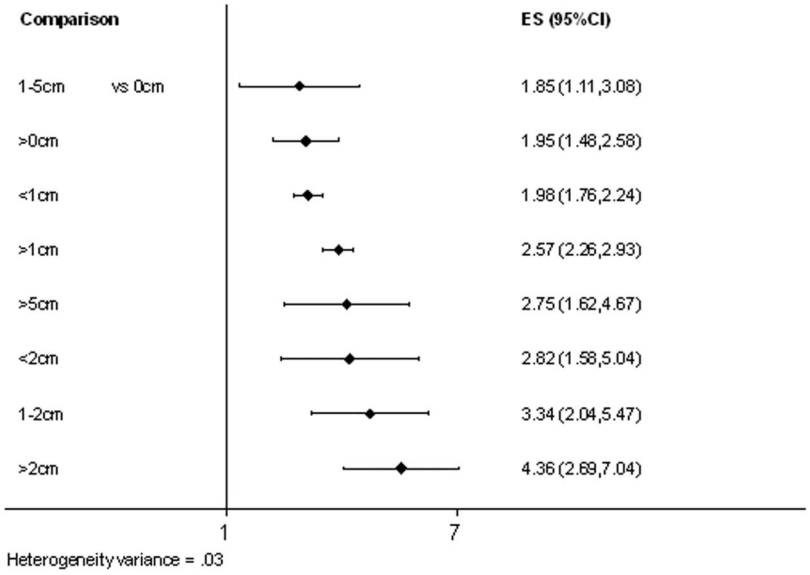


FIGURE 4. Forest plot showing RD thresholds versus complete cytoreduction (0 cm) after primary cytoreductive surgery for advanced EOC.

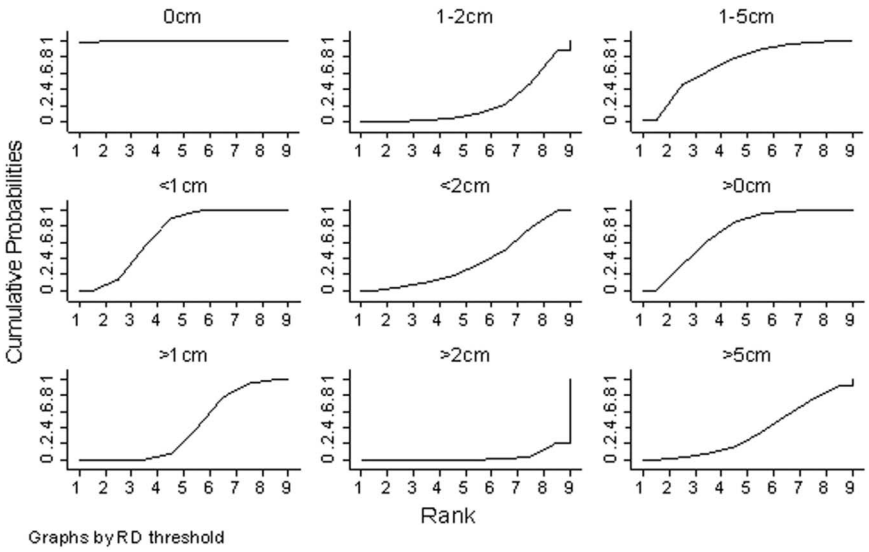


FIGURE 5. Rankograms showing ranks of RD thresholds for probability of being best at prolonging OS.

Cancer Society ovarian cancer action group and ESGO (European Society of Gynaecological Oncologists).

Pairwise analyses and NMAs clearly showed the prognostic importance of complete cytoreduction, with OS significantly prolonged in this RD threshold.[32,33] There should always be concern around "small study" biases, such as publication biases,[32,39] in meta-analyses. Although there was no evidence of publication bias (Figure 7), the results should still be interpreted with some caution. In addition, the nature of model selection procedures in the included studies may have meant study authors with nonstatistically significant *P* values may not have included RD in their final model.[98] However, including only studies that reported adjusted analyses should mean we have examined the best available evidence. Furthermore, a sensitivity analysis of the results of 8 studies that included an adequate adjustment for extent of disease at primary surgery strengthened the main conclusions. We emphasize the importance of this adjustment in this area; the results of the sensitivity analysis are key to proponents of aggressive surgery, as it was hypothesized that women with more extensive disease may have had poor prognosis despite the outcome of their surgery. However, the benefit of achieving complete cytoreduction became more evident after statistical adjustment for extent of disease. Nonetheless, we do suggest that all caveats should be discussed with patients before their primary surgery, especially in cases where there is likely to be a large trade-off between complete cytoreduction to NMRD and morbidity/quality of life.[99–101]

When compared with NMRD, all RD thresholds above this level resulted in shorter patient survival. When we compared different definitions of optimal and suboptimal cytoreduction, we observed the same survival patterns in those with greater removal of disease. However, these were attenuated compared with complete cytoreduction. Consequently, a key question is how much extra effort should be made to minimize RD if complete cytoreduction to NMRD is not possible. Although our findings do not enable us to determine whether the survival benefit is a direct effect of the surgical intervention, they do suggest that every effort should be made to reduce the tumor to microscopic disease. Where this is not considered achievable, attempts should be made to obtain near-optimal cytoreduction, defined as RD $<$ 1 cm. From the magnitude of effect sizes in comparisons of 0 cm versus larger amounts of RD (where there were sufficient evidence available for a give RD threshold), it seems that if RD cannot be limited to an optimal level then the surgeon could potentially prioritize their focus on morbidity and quality of life (QoL). The results of the SOCQER-2[102] study commissioned by NICE, assessed QoL in women undergoing standard or extensive surgery after primary surgery in advanced EOC. This study found no important differences in global QoL scores measured across 6 weeks, 6 months, and 12 months postsurgery in varying complexities of surgery. Patients who underwent low-complexity surgery were associated with higher rates of RD and lower survival compared with those with a similar disease burden undergoing surgery of intermediate complexity. Postoperative RD was associated with poorer OS, particularly in patients undergoing low-complexity surgery.

Table 5. League table giving specific effect estimates for all RD comparisons: HR (and 95% CI HRs) for OS.

	0 cm	<1 cm	>0 cm	1–2 cm	<2 cm	>1 cm	>2 cm	1–5 cm	>5 cm
0 cm	Reference	1.98* (1.76–2.24)	1.95 (1.48–2.58)	3.34 (2.04–5.47)	2.82 (1.58–5.04)	2.57 (2.26–2.93)	4.36 (2.69–7.04)	1.85 (1.11–3.08)	2.75 (1.62–4.67)
<1 cm	0.50* (0.45–0.57)	Reference	0.98 (0.73–1.33)	1.69 (1.04–2.73)	1.42 (0.81–2.52)	1.30 (1.15–1.46)	2.20 (1.38–3.51)	0.93 (0.56–1.56)	1.39 (0.82–2.35)
>0 cm	0.51 (0.39–0.68)	1.02 (0.75–1.38)	Reference	1.71 (0.97–3.02)	1.45 (0.76–2.76)	1.32 (0.97–1.79)	2.23 (1.28–3.89)	0.95 (0.53–1.70)	1.41 (0.78–2.56)
1–2 cm	0.30 (0.18–0.49)	0.59 (0.37–0.96)	0.58 (0.33–1.03)	Reference	0.84 (0.47–1.52)	0.77 (0.47–1.26)	1.30 (0.80–2.13)	0.55 (0.27–1.11)	0.82 (0.40–1.68)
<2 cm	0.35 (0.20–0.63)	0.70 (0.40–1.24)	0.69 (0.36–1.31)	1.18 (0.66–2.13)	Reference	0.91 (0.51–1.63)	1.54 (1.11–2.14)	0.65 (0.30–1.41)	0.97 (0.45–2.12)
>1 cm	0.39 (0.34–0.44)	0.77 (0.68–0.87)	0.76 (0.56–1.03)	1.30 (0.79–2.13)	1.10 (0.61–1.97)	Reference	1.69 (1.05–2.74)	0.72 (0.43–1.21)	1.07 (0.63–1.83)
>2 cm	0.23 (0.14–0.37)	0.46 (0.29–0.73)	0.45 (0.26–0.78)	0.77 (0.47–1.25)	0.65 (0.47–0.90)	0.59 (0.36–0.96)	Reference	0.42 (0.21–0.85)	0.63 (0.31–1.28)
1–5 cm	0.54 (0.32–0.90)	1.07 (0.64–1.79)	1.05 (0.59–1.89)	1.81 (0.90–3.65)	1.53 (0.71–3.28)	1.39 (0.83–2.34)	2.36 (1.18–4.71)	Reference	1.49 (0.82–2.70)
>5 cm	0.36 (0.21–0.62)	0.72 (0.42–1.22)	0.71 (0.39–1.29)	1.22 (0.60–2.48)	1.03 (0.47–2.23)	0.93 (0.55–1.60)	1.58 (0.78–3.20)	0.67 (0.37–1.22)	Reference

*Upper diagonal represents 0 cm versus <1 cm comparisons and inverse (<1 cm vs. 0 cm) in lower diagonal and repeated for other comparisons accordingly.

The overall certainty of the identified available evidence is moderate.[54] The evidence was primarily downgraded by one level from high certainty to moderate because the statistical analysis and reporting domain in the QUIPS tool[41] was assessed as being at high or unclear risk of bias in all included studies. Many study authors reported that statistically significant variables from the univariate analysis were included in the multivariable model, but gave no further details about any conceptual framework. The problem with this method is that there are variables that may not be important in a univariate association but are important in the full model. It is often more appropriate to include all pertinent variables that are plausibly important, potentially using data reduction methods to combine closely related variables.[103] This was the most serious bias from the QUIPS domains that could influence the effect estimates. The results are consistent and seem to be reliable and precise in conclusions drawn. Some comparisons were sparse with wide CIs, but even the lower 95% CI would be clinically significant as a point estimate in many cases, indicating a gain in OS. Consequently, further research is unlikely to change our confidence in the existing estimates of effect.[54] The exact reasons for performing one type of surgery over another were not well documented, and it was likely that women in generally poor health would be subjected to less aggressive surgery and thus would be more likely to have larger RD. This would most likely result in poorer survival. For this reason, we applied strict inclusion criteria and included studies that used statistical adjustment. However, it is generally accepted that the major reason for not achieving complete cytoreduction in most of the cases is not actually related to patient factors but is more associated with a deficiency in surgical skill and/or a lack of willingness in the surgeon to embrace ultraradical surgery.

The evidence suggests the need to redefine the term "optimal cytoreduction" by the Gynecological Cancer InterGroup, from its definition of <1 cm RD to NMRD.[10,104–106] We suggest retaining 3 categories of RD classification but redefining to optimal, "near optimal," and "suboptimal" cytoreduction rather than complete, optimal, and suboptimal for RD of 0 cm, <1 cm, and >1 cm, respectively. Similar suggestions using the terms complete, minimal, and gross have been previously published.[107]

Implications for research

Part 2 of this research is presented in the same edition and focused on adjustments for publication bias using expert elicitation.[38] This research aimed to conduct a series of sensitivity analyses to adjust the results of the

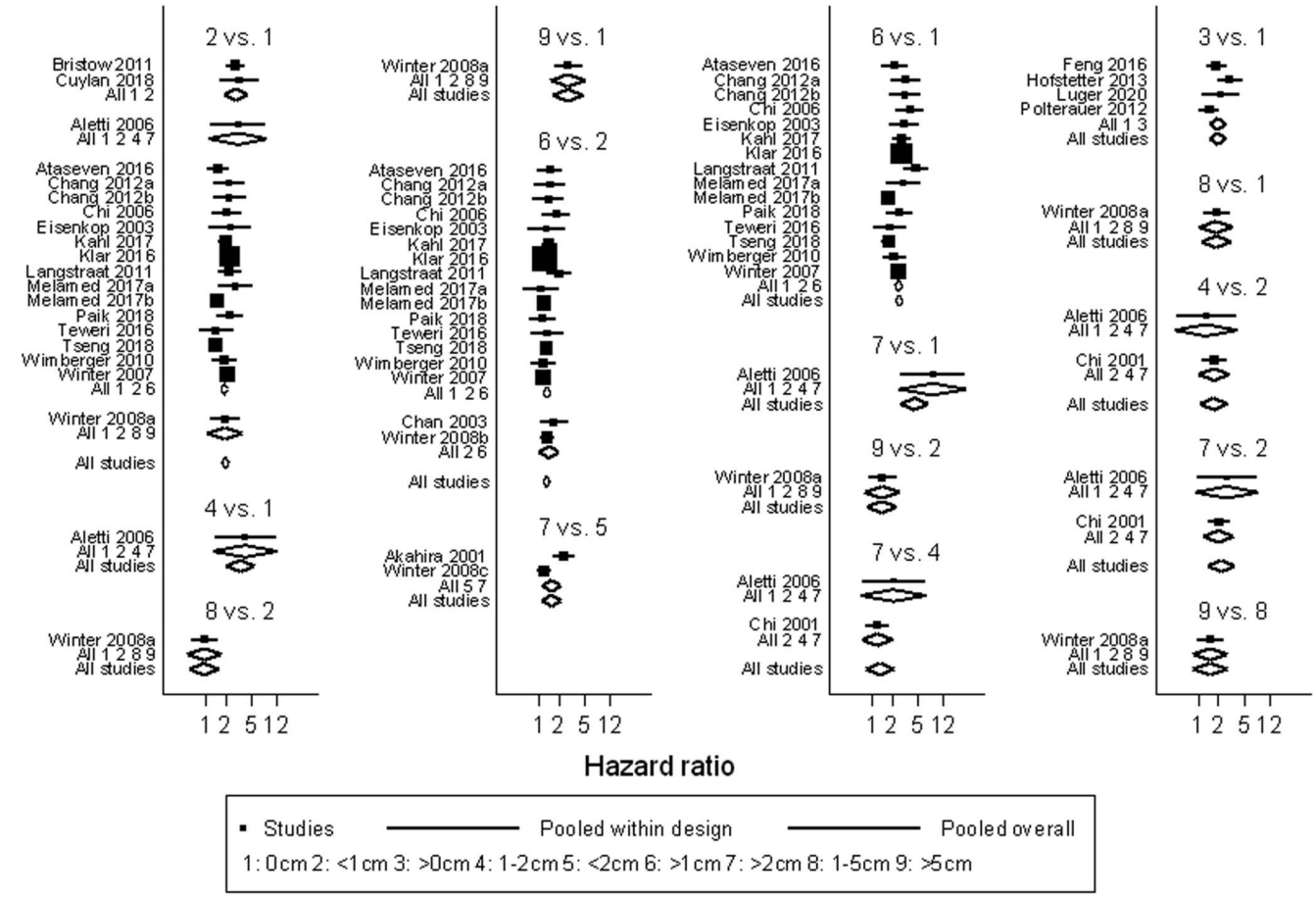


FIGURE 6. Forest plots showing results of all available RD comparisons and global test of consistency.

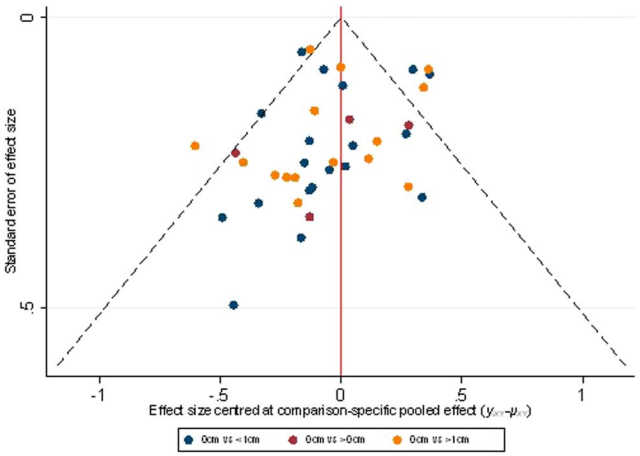


FIGURE 7. Funnel plot showing studies including comparisons of RD <1 cm, >0 cm, and >1 cm with complete cytoreduction (RD 0 cm).

NMA for publication bias, to confirm or refute the existing conclusions. The next piece of research in this area should focus on developing a model for accurately predicting important outcomes such as survival and quality of life based on remaining RD after primary surgery, and the effects of ultraradical surgery, so women can plan for their future and make informed decisions on subsequent treatment. This could be achieved by first conducting a review of prognostic studies to identify all studies reporting prognostic models for OS, as well as disease recurrence in women with advanced EOC following primary surgical debulking and also determine the performance of these models for predicting the risk stratification of women with this disease.[108] Measures should include discrimination, the area under the (curve) receiver operating characteristic curve, calibration, and overall model performance.[109,110] Given the fact that remaining RD after primary surgery is likely to remain the main predictor of survival in this area, a precise prediction

Table 6. Sensitivity analysis showing NMA of optimal RD threshold after primary cytoreductive surgery for advanced EOC including studies using adjustment for extent of disease.

RD threshold versus 0 cm (reference)*	HR (95% CI) NMAs	Mean rank	*P* (best) %	SUCRA %
0 cm	Reference	1	99.4	99.9
<1 cm	2.25 (1.93–2.63)	2.4	0	72.5
>0 cm	2.95 (1.87–4.67)	3.6	0	47.2
1–2 cm	3.32 (1.29–8.58)	3.9	0.7	41.9
>1 cm	3.41 (2.78–4.18)	4.3	0	33.5
>2 cm	6.89 (2.59–18.31)	5.8	0	5.0

*Comparisons involving <2 cm, 1–5 cm, and >5 cm RD thresholds were not reported.
HR, hazard ratio; *P* (best), probability that RD threshold is the best.

model would allow women and their families to plan for the future and aid future decisions on their subsequent further line treatment care pathway.

Because we have presented an updated and finalized analysis of impact of RD after primary surgery for advanced EOC, future research should also be conducted to determine whether increasing attempts at achieving complete cytoreduction have a direct effect on improving survival outcomes. This research should use methodologies and trial designs that reduce or eliminate confounding effects, such as the patient's performance status, disease spread, and tumor biology within the new paradigm of treatment with biological agents and genetic status. Despite the obvious challenges, this should be considered more than feasible because on average only around half of women with stage II–IV and unstaged ovarian cancer receive surgery in England.[13] Existing trials have shown conflicting results when further surgery was performed as an interval procedure after suboptimal cytoreduction at primary surgery.[111] Therefore, it seems best to increase attempts at optimizing to lower levels of RD at first surgery. Because there are large disparities between surgeons and centers in their optimal and complete cytoreduction rates,[14–17] it is worth considering randomizing patients to specialist centers providing more extensive surgery to achieve complete cytoreduction or to nonspecialist centers.[112] This may be best achieved by the conduct of a cluster randomized controlled trial. The increasing practice of offering neoadjuvant chemotherapy followed by interval debulking surgery should not complicate the performance of these trials, by including stratification for this factor within the study design.[113]

Another possible option is to randomize surgeons or hospitals to an intervention to develop their expertise and capability to perform more extensive ultraradical surgery, as additional training may be necessary.[114,115] There is a suggestion that maximal attempts to achieve complete cytoreduction are currently not being performed by most of the practising gynecological oncologists,[91] as previously indicated by low rates of complete cytoreduction to NMRD in many countries.[116,117] The development of these skills requires a shift in the surgeon's approach to surgery. Given that the additional procedures can be learned over a relatively short period, this could potentially lead to increases in optimal/complete cytoreduction rates with no significant increases in perioperative morbidity.[15] Similarly, it has been shown previously that optimal cytoreduction rates of up to 88% at primary laparotomy in advanced-stage ovarian cancer by gynecological oncologists working as a team can be achieved, without any increase in morbidity.[16]

CONCLUSIONS

Our results identified a strong association between achievement of complete cytoreduction and improved OS, highlighting a real need for clinical practice to follow Gynecological Cancer InterGroup recommendations. The NMA forms part of the methods guidance underpinning policy making in many jurisdictions. Part 2 of this research presents an extension to this work.[38]

ACKNOWLEDGMENTS

The authors thank the Cochrane Gynaecological and Neurological Cancer Review Group and the Biostatistics Research Group at Newcastle University for their excellent support. Both Luke Vale and Dawn Craig are funded by the NIHR Applied Research Collaboration for North East and North Cumbria. Luke Vale is also a member of the NIHR Newcastle In Vitro Diagnostics Cooperative.

REFERENCES

1. Ferlay J, Bray F, Siegel RL, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2015;65:87–108.
2. Kurman RJ, Carcangiu ML, Herrington CS. *WHO Classification of Tumours of Female Reproductive Organs*. 4th ed. Lyon, France: WHO Press; 2014.
3. Webb PM, Jordan SJ. Epidemiology of epithelial ovarian cancer. *Best Pract Res Clin Obstet Gynaecol.* 2017;41:3–14.
4. Berek JS, Kehoe ST, Kumar L, Friedlander M. Cancer of the ovary, fallopian tube, and peritoneum. *Int J Gynecol Obstet.* 2018;143(suppl 2):59–78.
5. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019;69:7–34.
6. Bristow RE, Tomacruz RS, Armstrong DK, et al. Survival effect of maximal cytoreductive surgery for advanced ovarian carcinoma during the platinum era: a meta-analysis. *J Clin Oncol.* 2002;20:1248–1259.
7. Bryant A, Hiu S, Kunonga PT, et al. Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery. *Cochrane Database Syst Rev.* 2022;9:CD015048.
8. Griffiths CT. Surgical resection of tumor bulk in the primary treatment of ovarian carcinoma. *Natl Cancer Inst Monogr.* 1975;42:101–104.
9. Hoskins WJ, McGuire WP, Brady MF, et al. The effect of diameter of largest residual disease on survival after primary cytoreductive surgery in patients with suboptimal residual epithelial ovarian carcinoma. *Am J Obstet Gynecol.* 1994;170:974–980.
10. du Bois A, Reuss A, Pujade-Lauraine E, et al. Role of surgical outcome as prognostic factor in advanced epithelial ovarian cancer: a combined exploratory analysis of 3 prospectively randomized phase 3 multicenter trials: by the Arbeitsgemeinschaft Gynaekologische Onkologie Studiengruppe Ovarialkarzinom (AGO-OVAR) and the Groupe dInvestigateurs Nationaux Pour les Etudes des Cancers de l'Ovaire (GINECO). *Cancer*. 2009;115:1234–1244.
11. Chang SJ, Hodeib M, Chang J, Bristow RE. Survival impact of complete cytoreduction to no gross residual disease for advanced-stage ovarian cancer: a meta-analysis. *Gynecol Oncol.* 2013;130:493–498.
12. Wimberger P, Wehling M, Lehmann N, et al. Influence of residual tumor on outcome in ovarian cancer patients with FIGO stage IV disease: an exploratory analysis of the AGO-OVAR (Arbeitsgemeinschaft Gynaekologische Onkologie Ovarian Cancer Study Group). *Ann Surg Oncol.* 2010;17:1642–1648.
13. Sundar S, Andreou A, Balega J, et al. *BGCS Call to Action—Response to Findings from National Ovarian Cancer Audit Feasibility Pilot*. 2021. Available at: https://www.bgcs.org.uk/wp-content/uploads/2021/05/OCAFP_BGCS-Call-to-action-21-05-2021-ref-14.00.pdf. Accessed October 1, 2022.
14. Aletti GD, Dowdy SC, Gostout BS, et al. Aggressive surgical effort and improved survival in advanced-stage ovarian cancer. *Obstet Gynecol.* 2006;107:77–85.
15. Chi DS, Eisenhauer EL, Zivanovic O, et al. Improved progression-free and overall survival in advanced ovarian cancer as a result of a change in surgical paradigm. *Gynecol Oncol.* 2009;114:26–31.
16. Naik R, Galaal K, Alagoda B, et al. Surgical training in gastrointestinal procedures within a UK gynaecological oncology subspecialty programme. BJOG 2010;117:26–31 Surgical training in gastrointestinal procedures within a UK gynaecological oncology subspecialty programme. *BJOG*. 2010;117:26–31.
17. van der Burg ME, van Lent M, Buyse M, et al. The effect of debulking surgery after induction chemotherapy on the prognosis in advanced epithelial ovarian cancer: gynecologic Cancer Cooperative Group of the European Organization for Research and Treatment of Cancer. *N Engl J Med.* 1995;332:629–634.
18. Burger RA, Brady MF, Bookman MA, et al. Incorporation of bevacizumab in the primary treatment of ovarian cancer. *N Engl J Med.* 2011;365:2473–2483.
19. Oza AM, Cook AD, Pfisterer J, et al. Standard chemotherapy with or without bevacizumab for women with newly diagnosed ovarian cancer (ICON7): overall survival results of a phase 3 randomised trial. *Lancet Oncol.* 2015;16:928–936.
20. Banerjee S, Moore KN, Colombo N, et al. Maintenance olaparib for patients with newly diagnosed advanced ovarian cancer and a BRCA mutation (SOLO1/GOG 3004): 5-year follow-up of a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* 2021;22:1721–1731.
21. Ray-Coquard I, Pautier P, Pignata S, et al. Olaparib plus bevacizumab as first-line maintenance in ovarian cancer. *N Engl J Med.* 2019;381:2416–2428.
22. Moore K, Colombo N, Scambia G, et al. Maintenance olaparib in patients with newly diagnosed advanced ovarian cancer. *N Engl J Med.* 2018;379:2495–2505.
23. González-Martín A, Pothuri B, Vergote I, et al. Niraparib in patients with newly diagnosed advanced ovarian cancer. *N Engl J Med.* 2019;381:2391–2402.
24. Poveda A, Floquet A, Ledermann JA, et al. Olaparib tablets as maintenance therapy in patients with platinum-sensitive relapsed ovarian cancer and a BRCA1/2 mutation (SOLO2/ENGOT-Ov21): a final analysis of a double-blind, randomised, placebo-controlled, phase 3 trial. *Lancet Oncol.* 2021;22:620–631.
25. Mirza MR, Monk BJ, Herrstedt J, et al. Niraparib maintenance therapy in platinum-sensitive, recurrent ovarian cancer. *N Engl J Med.* 2016;375:2154–2164.
26. NICE. *Contributing to Clinical Guidelines—A Guide for Patients and Carers*. National Institute for Health and Care Excellence. 2013. Available at: https://www.nice.org.uk/media/default/About/NICE-Communities/Public-involvement/Developing-NICE-guidance/Factsheet-1-contribute-to-developing-clinical-guidelines.pdf. Accessed October 1, 2022.

27. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the Biostatistics and results in the medical literature. *JAMA*. 2007;298:1010–1022.
28. Colombo N, Van Gorp T, Parma G, et al. Ovarian cancer. *Crit Rev Oncology/Hematology*. 2006;60:159–179.
29. Vergote I, De Wever I, Tjalma W, et al. Neoadjuvant chemotherapy or primary debulking surgery in advanced ovarian carcinoma: a retrospective analysis of 285 patients. *Gynecol Oncol*. 1998;71:431–436.
30. Vergote I, Trimbos BJ. Treatment of patients with early epithelial ovarian cancer. *Curr Opin Oncol*. 2003;15:452–455.
31. Leucht S, Chaimani A, Cipriani AS, et al. Network meta-analyses should be the highest level of evidence in treatment guidelines. *Eur Arch Psychiatry Clin Neurosci*. 2016;266:477–480.
32. Higgins JPT, Chandler J, Cumpston M, et al, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.0 Cochrane*. 2019.
33. Higgins JPT, Welton NJ. Network meta-analysis: a norm for comparative effectiveness? *The Lancet*. 2015; 386:628–630.
34. Laws A, Tao R, Wang S, et al. A comparison of national guidelines for network meta-analysis. *Value in Health*. 2019;22:1178–1186.
35. NICE. *Developing NICE Guidelines: The Manual Chapter 6*. National Institute for Health and Care Excellence (NICE); 2020.
36. Kanters S, Ford N, Druyts E, et al. Use of network meta-analysis in clinical guidelines. *Bull World Health Organ*. 2016;94:782–784.
37. Schmitz S, Adams R, Walsh C. Incorporating data from various trial designs into a mixed treatment comparison model. *Stat Med*. 2013;32:2935–2949.
38. Bryant A, Grayling M, Elattar A, et al. Residual disease after primary surgical treatment for advanced epithelial ovarian cancer; Part 2: network meta-analysis incorporating expert elicitation to adjust for publication bias. *Am J Ther*. 2022. doi: 10.1097/MJT.0000000000001548
39. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *BMJ*. 2012;344:d7762.
40. Azam F, Latif MF, Farooq A, et al. Performance status assessment by using ECOG (eastern cooperative oncology group) score for cancer patients by oncology healthcare professionals. *Case Rep Oncol*. 2019;12:728–736.
41. Riley RD, Moons KGM, Snell KIE, et al. A guide to systematic review and meta-analysis of prognostic factor studies. *BMJ*. 2019;364:k4597.
42. NICE. *Guide to the Methods of Technology Appraisal NICE*. 2013.
43. Moher DLA, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol*. 2009;62:1006–1012.
44. Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med*. 2015;162:777–784.
45. StataCorp. *Stata Statistical Software: Release 15*. 15th ed. College Station, TX: StataCorp LLC; 2017.
46. Shim S, Yoon BH, Shin IS, et al. Network meta-analysis: application and practice using Stata. *Epidemiol Health*. 2017;39:e2017047.
47. White IR. Network meta-analysis. *Stata J*. 2015;15:951–985.
48. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med*. 2002;21:1559–1573.
49. Borenstein M, Hedges LV, Higgins JPT, et al. Notes on subgroup Analyses and meta-regression. In: Borenstein M, Hedges LV, Higgins JPT, Rothstein HR, eds. *Introduction to Meta-Analysis*. 2009. Available at: https://doi.org/10.1002/9780470743386.ch21
50. Higgins JPT, Jackson D, Barrett JK, et al. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods*. 2012;3:98–110.
51. Dias S, Welton NJ, Caldwell DM, et al. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med*. 2010;29:932–944.
52. Mbuagbaw L, Rochwerg B, Jaeschke R, et al. Approaches to interpreting and choosing the best treatments in network meta-analyses. *Syst Rev*. 2017;6:79.
53. Foroutan F, Guyatt G, Zuk V, et al. GRADE Guidelines 28: use of GRADE for the assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks. *J Clin Epidemiol*. 2020;121:62–70.
54. Schünemann H, Guyatt G, Oxman A, eds. *The GRADE Working Group. GRADE Handbook for Grading Quality of Evidence and Strength of Recommendations. Cochrane Handbook*. 2013.
55. Akahira JI, Yoshikawa H, Shimizu Y, et al. Prognostic factors of stage IV epithelial ovarian cancer: a multicenter retrospective study. *Gynecol Oncol*. 2001;81:398–403.
56. Aletti GD, Dowdy SC, Podratz KC, Cliby WA. Relationship among surgical complexity, short-term morbidity, and overall survival in primary surgery for advanced ovarian cancer. *Am J Obstet Gynecol*. 2007;197:676.e1–676.e7.
57. Aletti GD, Dowdy SC, Podratz KC, Cliby WA. Surgical treatment of diaphragm disease correlates with improved survival in optimally debulked advanced stage ovarian cancer. *Gynecol Oncol*. 2006;100:283–287.
58. Aletti GD, Podratz KC, Jones MB, Cliby WA. Role of rectosigmoidectomy and stripping of pelvic peritoneum in outcomes of patients with advanced ovarian cancer. *J Am Coll Surgeons*. 2006;203:521–526.
59. Ataseven B, Grimm C, Harter P, et al. Prognostic impact of debulking surgery and residual tumor in patients with epithelial ovarian cancer FIGO stage IV. *Gynecol Oncol*. 2016;140:215–220.
60. Bristow RE, Ueda S, Gerardi MA, et al. Analysis of racial disparities in stage IIIC epithelial ovarian cancer

care and outcomes in a tertiary gynecologic oncology referral center. *Gynecol Oncol.* 2011;122:319–323.
61. Chan JK, Loizzi V, Lin YG, et al. Stages III and IV invasive epithelial ovarian carcinoma in younger versus older women: what prognostic factors are important? *Obstet Gynecol.* 2003;102:156–161.
62. Chang SJ, Bristow RE, Ryu HS. Impact of complete cytoreduction leaving no gross residual disease associated with radical cytoreductive surgical procedures on survival in advanced ovarian cancer. *Ann Surg Oncol.* 2012;19:4059–4067.
63. Chang SJ, Bristow RE, Ryu HS. Prognostic significance of systematic lymphadenectomy as part of primary debulking surgery in patients with advanced ovarian cancer. *Gynecol Oncol.* 2012;126:381–386.
64. Chi DS, Liao JB, Leon LF, et al. Identification of prognostic factors in advanced epithelial ovarian carcinoma. *Gynecol Oncol.* 2001;82:532–537.
65. Chi DS, Eisenhauer EL, Lang J, et al. What is the optimal goal of primary cytoreductive surgery for bulky stage IIIC epithelial ovarian carcinoma (EOC)? *Gynecol Oncol.* 2006;103:559–56464.
66. Eisenhauer EL, Abu-Rustum NR, Sonoda Y, et al. The effect of maximal surgical cytoreduction on sensitivity to platinum-taxane chemotherapy and subsequent survival in patients with advanced ovarian cancer. *Gynecol Oncol.* 2008;108:276–281.
67. Cuylan ZF, Meydanli MM, Sari ME, et al. Prognostic factors for maximally or optimally cytoreduced stage III nonserous epithelial ovarian carcinoma treated with carboplatin/paclitaxel chemotherapy. *J Obstet Gynaecol Res.* 2018;44:1284–1293.
68. Eisenkop SM, Friedman RL, Wang HJ. Complete cytoreductive surgery is feasible and maximizes survival in patients with advanced epithelial ovarian cancer: a prospective study. *Gynecol Oncol.* 1998;69:103–108.
69. Eisenkop SM, Spirtos NM, Friedman RL, et al. Relative influences of tumor volume before surgery and the cytoreductive outcome on survival for patients with advanced ovarian cancer: a prospective study. *Gynecol Oncol.* 2003;90:390–396.
70. Feng Z, Wen H, Bi R, et al. Prognostic impact of the time interval from primary surgery to intravenous chemotherapy in high grade serous ovarian cancer. *Gynecol Oncol.* 2016;141:466–470.
71. Hofstetter G, Concin N, Braicu I, et al. The time interval from surgery to start of chemotherapy significantly impacts prognosis in patients with advanced serous ovarian carcinoma—analysis of patient data in the prospective OVCAD study. *Gynecol Oncol.* 2013;131:15–20.
72. Kahl A, du Bois A, Harter P, et al. Prognostic value of the age-adjusted Charlson comorbidity index (ACCI) on short- and long-term outcome in patients with advanced primary epithelial ovarian cancer. *Ann Surg Oncol.* 2017;24:3692–3699.
73. Klar M, Hasenburg A, Hasanov M, et al. Prognostic factors in young ovarian cancer patients: an analysis of four prospective phase III intergroup trials of the AGO Study Group, GINECO and NSGO. *Eur J Cancer.* 2016;66:114–124.
74. Mahner S, Eulenburg C, Staehle A, et al. Prognostic impact of the time interval between surgery and chemotherapy in advanced ovarian cancer: analysis of prospective randomised phase III trials. *Eur J Cancer.* 2013; 49:142–149.
75. Pfisterer J, Weber B, Reuss A, et al. Randomized phase III trial of topotecan following carboplatin and paclitaxel in first-line treatment of advanced ovarian cancer: a gynecologic cancer intergroup trial of the AGO-OVAR and GINECO. *J Natl Cancer Inst.* 2006;98:1036–1045.
76. du Bois A, Herrstedt J, Hardy-Bessard AC, et al. Phase III trial of carboplatin plus paclitaxel with or without gemcitabine in first-line treatment of epithelial ovarian cancer. *J Clin Oncol.* 2010;28:4162–4169.
77. du Bois A, Meier W, Adams HP, et al. A randomized clinical trial of cisplatin/paclitaxel versus carboplatin/paclitaxel as first-line treatment of ovarian cancer. *Cancer-Spectrum Knowledge Environ.* 2003;95:1320–1329.
78. du Bois A, Weber B, Rochon J, et al. Addition of epirubicin as a third drug to carboplatinpaclitaxel in first-line treatment of advanced ovarian cancer: a prospectively randomized gynecologic cancer intergroup trial by the Arbeitsgemeinschaft Gynaekologische Onkologie Ovarian Cancer Study Group and the Groupe d'Investigateurs Nationaux pour l'Etude des Cancers Ovariens. *J Clin Oncol.* 2006;24:1127–1135.
79. Langstraat C, Aletti GD, Cliby WA. Morbidity, mortality and overall survival in elderly women undergoing primary surgical debulking for ovarian cancer: a delicate balance requiring individualization. *Gynecol Oncol.* 2011;123:187–191.
80. Luger AK, Steinkohl F, Aigner F, et al. Enlarged cardiophrenic lymph nodes predict disease involvement of the upper abdomen and the outcome of primary surgical debulking in advanced ovarian cancer. *Acta Obstetricia Gynecologica Scand.* 2020;99:1092–1099.
81. Melamed A, Manning-Geist B, Bregar AJ, et al. Associations between residual disease and survival in epithelial ovarian cancer by histologic type. *Gynecol Oncol.* 2017;147:250–256.
82. Paik ES, Kim JH, Kim TJ, et al. Prognostic significance of normal-sized ovary in advanced serous epithelial ovarian cancer. *J Gynecol Oncol.* 2018;29:e13–e.
83. Polterauer S, Vergote I, Concin N, et al. Prognostic value of residual tumor size in patients with epithelial ovarian cancer FIGO stages IIA-IV: analysis of the OVCAD data. *Int J Gynecol Cancer.* 2012;22:380–385.
84. Tewari KS, Java JJ, Eskander RN, et al. Early initiation of chemotherapy following complete resection of advanced ovarian cancer associated with improved survival: NRG Oncology/Gynecologic Oncology Group study. *Ann Oncol.* 2016;27:114–121.
85. Tseng JH, Cowan RA, Zhou Q, et al. Continuous improvement in primary debulking surgery for advanced ovarian cancer: do increased complete gross resection rates independently lead to increased

progression-free and overall survival? *Gynecol Oncol.* 2018;151:24–31.
86. Armstrong DK, Bundy BW, Wenzel L, et al. Intraperitoneal cisplatin and paclitaxel in ovarian cancer. *New Engl J Med.* 2006;354:34–43.
87. Markman M, Bundy BN, Alberts DS, et al. Phase III trial of standard-dose intravenous cisplatin plus paclitaxel versus moderately high-dose carboplatin followed by intravenous paclitaxel and intraperitoneal cisplatin in small-volume stage III ovarian carcinoma: an intergroup study of the gynecologic oncology group, southwestern oncology group, and eastern cooperative oncology group. *J Clin Oncol.* 2001;19:1001–1007.
88. McGuire WP, Hoskins WJ, Brady MF, et al. Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer. *New Engl J Med.* 1996;334:1–6.
89. Muggia FM, Braly PS, Brady MF, et al. Phase III randomized study of cisplatin versus paclitaxel versus cisplatin and paclitaxel in patients with suboptimal stage III or IV ovarian cancer: a gynecologic oncology group study. *J Clin Oncol.* 2000;18:106–115.
90. Ozols RF, Bundy BN, Greer BE, et al. Phase III trial of carboplatin and paclitaxel compared with cisplatin and paclitaxel in patients with optimally resected stage III ovarian cancer: a gynecologic oncology group study. *J Clin Oncol.* 2003;21:3194–3200.
91. Rose PG, Nerenstone S, Brady MF, et al. Secondary surgical cytoreduction for advanced ovarian carcinoma. *New Engl J Med.* 2004;351:2489–2497.
92. Winter WE, Maxwell GL III, Tian C, et al. Prognostic factors for stage III epithelial ovarian cancer: a gynecologic oncology group study. *J Clin Oncol.* 2007;25:3621–3627.
93. Spriggs DR, Brady MF, Vaccarello L, et al. Phase III randomized trial of intravenous cisplatin plus a 24- or 96-hour infusion of paclitaxel in epithelial ovarian cancer: a Gynecologic Oncology Group Study. *J Clin Oncol.* 2007;25:4466–4471.
94. Winter WE, Maxwell GL, Tian C, et al. Tumor residual after surgical cytoreduction in prediction of clinical outcome in stage IV epithelial ovarian cancer: a Gynecologic Oncology Group Study. *J Clin Oncol.* 2008;26:83–89.
95. Rouse B, Chaimani A, Li T. Network meta-analysis: an introduction for clinicians. *Intern Emerg Med.* 2017;12:103–111.
96. Chaimani A, Higgins JPT, Mavridis D, et al. Graphical tools for network meta-analysis in STATA. *PLoS ONE*. 2013;8:e76654.
97. Vergote I, Vlayen J, Heus P, et al. *KCE Reports 268. D/2016/10.273/49. Ovarian cancer: diagnosis, treatment and follow-up. Good Clinical Practice (GCP)*. Centre BHCK: Brussels, Belgian; 2016.
98. Williamson PR, Gamble C, Altman DG, et al. Outcome selection bias in meta-analysis. *Stat Methods Med Res.* 2005;14:515–524.
99. Brédart A, Bouleuc C, Dolbeault S. Doctor-patient communication and satisfaction with care in oncology. *Curr Opin Oncol.* 2005;17:351–354.
100. Frey MK, Philips SR, Jeffries J, et al. A qualitative study of ovarian cancer survivors' perceptions of endpoints and goals of care. *Gynecol Oncol.* 2014;135:261–265.
101. Wong BO, Clapp JT, Morris AM. Misinterpretation of surgeons' statements on cancer removal-the adverse effects of we got it all. *JAMA Oncol.* 2022.
102. Sundar S, Cummins C, Kumar S, et al. Quality of life from cytoreductive surgery in advanced ovarian cancer: investigating the association between disease burden and surgical complexity in the international, prospective, SOCQER-2 cohort study. *Int J Obstet Gynaecol.* 2022;129:1122–1132
103. Lo SK, Li IT, Tsou TS, See L. Non-significant in univariate but significant in multivariate analysis: a discussion with examples. *Changgeng Yi Xue Za Zhi.* 1995;18:95–101.
104. Stuart GC, Kitchener H, Bacon M, et al. 2010 gynecologic cancer InterGroup (GCIG) consensus statement on clinical trials in ovarian cancer: report from the fourth ovarian cancer consensus conference. *Int J Gynecol Cancer.* 2011;21:750–755.
105. Chang SJ, Bristow RE. Evolution of surgical treatment paradigms for advanced-stage ovarian cancer: redefining "optimal" residual disease. *Gynecol Oncol.* 2012;125:483–492.
106. Karam A, Ledermann JA, Kim JW, et al. Fifth ovarian cancer consensus conference of the gynecologic cancer InterGroup: first-line interventions. *Ann Oncol.* 2017;28:711–717.
107. Zapardiel I, Morrow CP. New terminology for cytoreduction in advanced ovarian cancer. *Lancet Oncol.* 2011;12:214.
108. EQUATOR. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement: Enhancing the QUAlity and Transparency Of health Research (EQUATOR). 2020. Available at: https://www.equator-network.org/reporting-guidelines/tripod-statement/. Accessed October 1, 2022.
109. Debray TP, Koffijberg H, Nieboer D, et al. Meta-analysis and aggregation of multiple published prediction models. *Stat Med.* 2014;33:2341–2362.
110. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology*. 2010;21:128–138.
111. Tangjitgamol S, Laopaiboon M, Lumbiganon P, Bryant A. Interval debulking surgery for advanced epithelial ovarian cancer. *Cochrane Database Syst Rev.* 2016;2016:CD006014.
112. Wimberger P, Lehmann N, Kimmig R, et al. Prognostic factors for complete debulking in advanced ovarian cancer and its impact on survival. An exploratory analysis of a prospectively randomized phase III study of the Arbeitsgemeinschaft Gynaekologische Onkologie Ovarian Cancer Study Group (AGO-OVAR). *Gynecol Oncol.* 2007;106:69–74.
113. Markman M. Concept of optimal surgical cytoreduction in advanced ovarian cancer: a brief critique and a call for action. *J Clin Oncol.* 2007;25:4168–4170.

114. Bristow RE, Palis BE, Chi DS, et al. The National Cancer Database report on advancedstage epithelial ovarian cancer: impact of hospital surgical case volume on overall survival and surgical treatment paradigm. *Gynecol Oncol.* 2010;118:262–267.
115. Eisenkop SM, Spirtos NM. What are the current surgical objectives, strategies and technical capabilities of gynaecologic oncologists treating advanced epithelial ovarian cancer? *Gynecol Oncol.* 2001;82:489–497.
116. Vergote I, Amant F, Kristensen GB, et al. European organization for research and treatment of cancer – gynaecological cancer group; NCIC clinical trials group. Neoadjuvant chemotherapy or primary surgery in stage IIIC or IV. *N Engl J.* 2010;363:943–953.
117. Crawford SC, Vasey PA, Paul J, et al. Does aggressive surgery only benefit patients with less advanced ovarian cancer? Results from an international comparison within the SCOTROC-1 trial. *J Clin Oncol.* 2005;23:5003–5011.

Appendix 7: Publication 5: Residual disease after primary surgery for advanced epithelial ovarian cancer: Expert elicitation exercise to explore opinions about potential impact of publication bias in a planned systematic review and meta-analysis

BMJ Open

Residual disease after primary surgery for advanced epithelial ovarian cancer: expert elicitation exercise to explore opinions about potential impact of publication bias in a planned systematic review and meta-analysis

Andrew Bryant ,[1] Michael Grayling,[1] Shaun Hiu,[1] Ketankumar Gajjar,[2] Eugenie Johnson,[1] Ahmed Elattar,[3] Luke Vale,[1] Dawn Craig,[1] Raj Naik[4]

To cite: Bryant A, Grayling M, Hiu S, *et al*. Residual disease after primary surgery for advanced epithelial ovarian cancer: expert elicitation exercise to explore opinions about potential impact of publication bias in a planned systematic review and meta-analysis. *BMJ Open* 2022;**12**:e060183. doi:10.1136/bmjopen-2021-060183

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-060183).

Received 14 December 2021
Accepted 03 August 2022

For numbered affiliations see end of article.

Correspondence to
Dr Andrew Bryant;
andy.bryant@ncl.ac.uk

ABSTRACT

Objectives We consider expert opinion and its incorporation into a planned meta-analysis as a way of adjusting for anticipated publication bias. We conduct an elicitation exercise among eligible British Gynaecological Cancer Society (BGCS) members with expertise in gynaecology.

Design Expert elicitation exercise.

Setting BGCS.

Participants Members of the BGCS with expertise in gynaecology.

Methods Experts were presented with details of a planned prospective systematic review and meta-analysis, assessing overall survival for the extent of excision of residual disease (RD) after primary surgery for advanced epithelial ovarian cancer. Participants were asked views on the likelihood of different studies (varied in the size of the study population and the RD thresholds being compared) not being published. Descriptive statistics were produced and opinions on total number of missing studies by sample size and magnitude of effect size estimated.

Results Eighteen expert respondents were included. Responders perceived publication bias to be a possibility for comparisons of RD <1 cm versus RD=0 cm, but more so for comparisons involving higher volume suboptimal RD thresholds. However, experts' perceived publication bias in comparisons of RD=0 cm versus suboptimal RD thresholds did not translate into many elicited missing studies in Part B of the elicitation exercise. The median number of missing studies estimated by responders for the main comparison of RD<1 cm versus RD=0 cm was 10 (IQR: 5–20), with the number of missing studies influenced by whether the effect size was equivocal. The median number of missing studies estimated for suboptimal RD versus RD=0 cm was lower.

Conclusions The results may raise awareness that a degree of scepticism is needed when reviewing studies comparing RD <1 cm versus RD=0 cm. There is also a belief among respondents that comparisons involving RD=0 cm and suboptimal thresholds (>1 cm) are likely to be impacted by publication bias, but this is unlikely to attenuate effect estimates in meta-analyses.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ In our elicitation exercise, designed in collaboration with senior gynae-oncologists, the number of respondents (n=18) was sufficient to provide a solid basis for meaningful conclusions to be drawn in an area of uncertainty.
⇒ Part A of the elicitation identifies areas where publication bias is of concern, but the questions asked do not provide an indication of the direction of any bias.
⇒ Therefore, in Part B of our elicitation exercise, we collected information that would enable any planned meta-analysis estimates to be adjusted for the anticipated impact of publication bias.
⇒ The approach adopted is inexpensive and easy to design and administer and did not rely on any contact with participants, who were able to complete at their own convenience.
⇒ However, answers given by the experts to open-ended questions were prone to an 'extreme answer bias'.

INTRODUCTION

Residual disease (RD) after upfront primary debulking surgery (PDS) for advanced epithelial ovarian cancer (EOC) is believed to be a key determinant of overall survival (OS). A recent prognostic factor systematic review protocol aims to demonstrate the superiority in terms of OS of the complete removal of RD in advanced EOC compared with leaving macroscopic disease (that is, the surgeon leaving some visible disease).[1]

However, much of the evidence in this area comes from small and/or retrospective studies. Relying on such studies to draw conclusions may be unsound. One reason for this relates to possible publication biases, which may be more pronounced for small,

retrospective evaluations. Publication bias can arise when the publication of research findings depends on the nature and direction of the results. It is more likely in smaller and retrospective studies than for larger randomised controlled trials.[2–6] Small studies might be underpowered and, furthermore, null findings might be due to deficiencies in the study design and conduct. Hence, including these studies might not lead to an appropriate adjustment of meta-analysis estimates. This is why we planned to include studies with a minimum sample size of 100 patients in the systematic review.

Therefore, given the nature of the evidence base, publication bias could be hypothesised to lead to a bias in favour of more complete removal of RD as described below.

Small and retrospective studies are also prone to other biases, particularly selection bias, (ie, systematic differences between groups in terms of baseline characteristics) compared with randomised trials.[7 8] Furthermore, all study designs may suffer from inadequacies of study conduct, such as deficiencies in blinding, high attrition and so on.[9 10] Again, these problems are potentially exacerbated for smaller retrospective studies.

As alluded to above, publication and other reporting biases[11 12] can have serious consequences to research and impact on summary of findings and recommendations in guidelines.[13 14] If it is suspected that publication bias is highly plausible, this may make the effect estimates obtained from meta-analyses uncertain and potentially unreliable. This is a concern when considering the results of the systematic review assessing OS for RD after PDS in advanced EOC.[1]

In this review, the data underpinning the estimates will be derived from the further analysis of data collected to address other research questions. Post hoc analyses of data collected to address other questions and secondary analyses of past medical records do not have to be prespecified anywhere, so there is a strong threat of data dredging. Therefore, the reporting of such data for individual studies may depend on the significance of their findings. For example, it is possible that only analyses producing 'significant' findings will be published. Thus, any meta-analysis may overestimate the effect of complete cytoreduction. This may be true even if many of the non-reported studies are small, as their cumulative impact on the meta-analysis may have a substantial overall effect.

Exploration of publication bias is an important part of a robust systematic review and should always be considered. At present, there is no consensus on a standard approach for identifying and adjusting for publication bias, although some methods, particularly around identification, do exist. Reduction of publication bias can be achieved by adherence to good review practice, such as a thorough search of grey literature.[15–17] Post hoc statistical approaches such as funnel plots,[18] trim and fill,[19 20] and file drawer number[21] could also be used. Furthermore, when there is evidence for publication bias or this bias is highly suspected, selection models[22 23] might be used to investigate how the results of a meta-analysis may be affected by publication bias. However, these usually require a large number of included studies in the analysis[12 24] and any adjustment generally requires an assumption of the underlying selection model.[12 22]

A potentially more practical approach is to incorporate external information into the meta-analysis. This external information could be gathered from various sources and incorporated using a Bayesian framework.[25–27] However, this approach would only be useful if the external information is obtained from a reliable source. This final point is the focus of our study, as we propose an approach that has hitherto received little attention in meta-analyses: the consideration of expert opinion and the incorporation of their views and opinions into the meta-analysis to inform the adjustment. We do this by conducting an elicitation exercise among eligible British Gynaecological Cancer Society (BGCS) members (based on a pertinent job title and expertise in gynaecology) to identify their expert opinions on the potential nature and extent of publication bias in a planned prospective systematic review and meta-analysis assessing OS in RD after PDS for advanced EOC.[1] The elicitation exercise relates to the conduct of the planned systematic review, where the findings from this exercise will be used to adjust the proposed meta-analyses for any perceived publication bias.

In the elicitation exercise, we ask participants to account for: (1) the sort of studies that have been conducted but not published; (2) the plausible magnitude and direction of any publication bias; and (3) possible explanations for why and how the publication bias occurs. These data could be used to adjust the results for publication bias in our planned meta-analysis assessing OS in RD thresholds after primary surgery for EOC.

METHODS

Case study

This research involved human participants outside of a study or trial setting. The elicitation exercise did not require ethical approval because it was sent to BGCS members and participation was optional. Information about any expert that participated in the elicitation exercise was kept confidential.

Participants were given details of a planned prospective systematic review and meta-analysis assessing OS for the extent of excision of RD (see online supplemental appendix 1). This will include data from studies or case series of 100 or more patients that include a concurrent comparison of different RD thresholds after primary surgical intervention in adult women with advanced EOC. The outcome of interest was OS for different categories of RD.

For the purposes of the case study, participants were told that bibliographic databases up to January 2020 were searched for pertinent data, so that they had a cut-off for their responses to each scenario. Participants were made aware that two review authors would independently

abstract data and assess risk of bias and, where possible, that the data would be synthesised in a meta-analysis. Full details of the methodology used in the review is provided in a Cochrane systematic review[1] and a summary of inclusion criteria is given in the elicitation exercise in online supplemental appendix 1.

The review objective is to assess the impact of RD after upfront and interval debulking surgery on survival outcomes. However, the focus of this paper and the elicitation exercise was OS in different RD thresholds after upfront primary surgery.

Design of elicitation exercise

The purpose of the elicitation exercise was to ask respondents for their opinions on the likelihood of studies not being published. Thereafter, we asked for their opinions on several different scenarios, all of which related to the likelihood of different studies not being published. These unpublished studies varied by both size of the study population and the impact of the RD threshold as a prognostic factor for OS.

The elicitation exercise was designed in consultation with four gynae-oncologists, to help ensure a sufficiently detailed level of explanation was provided regarding the purpose of the exercise, along with clear descriptions of the methodology and rationale. Visual examples were used to make what was being asked of respondents as transparent as possible.

Usually, expert opinions are elicited either directly using interview methods or via an elicitation exercise. In either case, opinions potentially need only be provided by as few as four experts.[28 29] However, it is advised to use more experts to give the results more generalisability and allow for the potential of a broader range of views.[30 31] Any widespread disagreement among experts can be reflected in the uncertainty of elicited estimates; all that is fundamental is that respondents have extensive knowledge and expertise in the area of interest.

The elicitation exercise consisted of three parts: A, B and C (online supplemental appendix 1 provides an example of the elicitation exercise). Part A adopted an existing method of elicitation,[30] while Part B used a de novo tool designed to provide a way of obtaining an estimate of the number of missing studies from a meta-analysis. Respondents also indicated the size of these missing studies, which can be used to calculate the magnitude of effect in the form of a HR with 95% CI. Parts A and B are described in more detail below. Part C was used to gauge the attitudes of the respondent cohort about reporting biases more generally and is not reported here.

To assist respondents in answering questions in Parts A and B, we provided brief guidance on the interpretation of commonly reported statistics from survival models (see introductory section of 'Expert elicitation' in online supplemental appendix 1).

Expert elicitation Part A

This part comprised one question (Q1) and attempted to assess publication bias by asking respondents about their views on the chance of publication for comparisons of different macroscopic RD thresholds (RD>0 cm) versus the reference comparator of complete cytoreduction (removal of tumour so that there was no visible disease with the naked eye, RD =0 cm). Specifically, for each comparison the sample size of the hypothetical study was varied between a minimum sample size (n=100) (which was part of the inclusion criteria in the planned review) and a maximum sample size (this maximum was based on observed sizes in the meta-analysis of included studies in an initial scope of the results up to January 2020).

Responders were then asked to assign a probability that a study reporting a given comparison with a given sample size would be published on a scale of 0 (no chance of publication) to 100 (certainly published). Other characteristics of the hypothetical study followed the inclusion criteria set out in the systematic review protocol by Bryant *et al*.[1] These have been summarised above and are reported in online supplemental appendix 1.

Expert elicitation Part B

Part B consisted of three broad questions and aimed to obtain the opinion of respondents on the estimated number of conducted-but-unpublished studies that might exist. For each question, participants were asked to consider a particular macroscopic RD threshold and compare it with RD =0 cm: Q2 (RD <1 cm vs RD =0 cm); Q3 (RD >1 cm vs RD =0 cm); and Q4 (RD >2 cm vs RD =0 cm). Subsequently, participants were asked on a Likert Scale from 1 (not likely at all) to 5 (extremely likely), the likelihood that relevant studies that either favoured macroscopic disease, or studies that found no statistically significant difference (p>0.05) in survival between macroscopic disease and RD =0 cm, would not be published.

Next, respondents were asked to give an estimate of how many studies of a certain size and magnitude of effect might be unpublished, along with a rationale for their answer. The sample size of unpublished studies was varied in increments of 100 from 100 to >500. The effect size, reported as the adjusted HR, was likewise varied in decrements of 0.1, between 1 and ≤0.5. In total, respondents were asked to think about the number of unpublished studies for 36 different hypothetical combinations of sample size and effect size. The questions were repeated for scenarios involving suboptimal RD thresholds (>1 cm and >2 cm) compared with RD =0 cm (See Q3-4 of elicitation exercise in online supplemental appendix 1).

The responses to the questions in Part B could be used to adjust the overall effect estimate from observed studies when data from unobserved studies are added.

Data collection and sampling

The elicitation exercise was vetted by the BGCS Survey panel; their helpful suggestions were incorporated and a link to the finalised elicitation exercise using Qualtrics was distributed to members via email by the BGCS administrator. BGCS have established guidelines for circulation of online surveys via the membership email

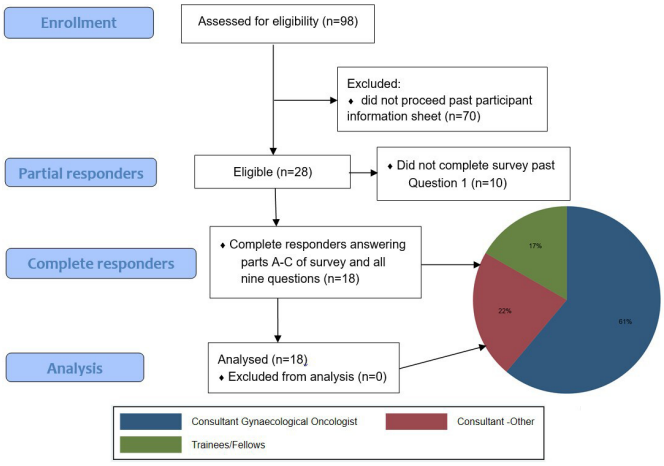


Figure 1 Elicitation exercise flow diagram.

directory, which were followed in our elicitation exercise and are available on request to the BGCS. The link to the elicitation exercise was open from 13 August 2020 to 26 October 2020 and two reminders were sent out. Study participation was voluntary and potential respondents were informed that the results of the elicitation would inform a publication. All acknowledgements are given with the consent of responders; all open-text responses provided have been anonymised and we have explicitly excluded cross-tabulation by job title, as this may have compromised the anonymisation.

Data analysis

The responses of the elicitation exercise are summarised using descriptive statistics. Further details are reported in online supplemental appendix 1. For the responses to Part B, we also provide in online supplemental appendix 2 an example of how the responses could be used to form an overall estimate of the total number of missing studies by sample size and magnitude of effect size for each question, reported as a HR and 95% CI. All analyses were conducted in StataIC V.15.[32]

Patient and public involvement

None.

RESULTS

Characteristics of respondents

The elicitation exercise was sent to all 455 BGCS members at the time, with over 80% being eligible to complete. A total of 98 BGCS members opened the link for the exercise and 28 proceeded past the participant information sheet. Of these, 18 respondents fully completed the elicitation exercise, and their responses are reported below. The remaining 10 participants did not adequately contribute to the exercise to be included in analysis (figure 1).

The distribution of expertise of completers of the exercise is also presented in figure 1. Most responders were consultant gynaecological oncologists (11/18; 61%) or subspecialist consultants (4/18; 22%). The median time to complete the exercise was 18 min (IQR 16–27 min) with a range of 8–61 min. The mean completion time was 23 min (SD 14 min).

Part A: Probability estimates that a study with minimum and maximum specified sample sizes is published for different macroscopic RD disease versus RD =0 cm

Table 1 shows the perceived probability that a study is published based on its sample size for the comparison of different RD thresholds (all compared with RD =0 cm). Responses suggest that publication bias may be quite likely in studies where the sample size was just 100. For example, responders suggest they thought there was less than a 60% chance that a comparison of RD <1 cm versus RD =0 cm would be reported for a study with a sample size of 100 participants.

Overall, there was widespread variation in the results, indicating that some responders thought the probability of publication was much higher than others (range 0%–100%). Responders appeared to indicate that the probability of publication was lowest for comparisons involving greater macroscopic disease volume (largest elicited median probability 20% (IQR 10%–75%) in macroscopic disease involving

Table 1 Summary statistics of responders' perceived chance (probability) of publication for studies of given sample size for residual disease thresholds compared with microscopic disease (0 cm)

Versus 0 cm	% for n minimum (n=100)			% for n maximum		
RD threshold	Mean (SD)	Median (IQR)	Observed range	Mean (SD)	Median (IQR)	Observed range
< 1 cm	57 (31.2)	55 (30–80)	0–100	95 (6.1)	99.5 (90–100)	80–100
> 0 cm	49 (33.6)	50 (20–80)	0–95	77 (25.5)	80 (70–99)	0–100
1–2 cm	48 (32.1)	50 (20–70)	0–100	58 (34.1)	72.5 (30–80)	0–100
< 2 cm	50 (36.6)	50 (10–85)	0–100	58 (36.6)	65 (20–90)	0–100
> 1 cm	49 (34.4)	45 (20–90)	0–100	85 (19.3)	95 (75–99)	40–100
> 2 cm	38 (36.6)	20 (10–75)	0–100	47 (37.9)	30 (15–80)	0–100
1–5 cm	29 (33.2)	10 (0–50)	0–95	42 (38.2)	27.5 (5–80)	0–100
> 5 cm	23 (34.4)	3.5 (0–50)	0–95	35 (40.7)	10 (0–80)	0–100

N maximum varies for different RD =0 cm versus RD threshold comparisons. For RD <1 cm and RD >1 cm, n=1000. For RD >0 cm, n=625, For RD 1–2 cm, n=210. The remainder are n=250.
RD, residual disease;

Table 2 Responders' perceived likelihood of publication bias in comparisons of near optimal (<1 cm) and suboptimal (>1/2 cm) versus complete cytoreduction (0 cm)

	RD <1 cm vs 0 cm		RD >1 cm vs 0 cm		RD >2 cm vs 0 cm	
Perceived likelihood of publication bias	N	%	N	%	N	%
Not likely at all (1)	1	5.5	10	55.5	15	83.5
Somewhat likely (2)	5	28	2	11	1	5.5
Quite likely (3)	8	44.5	3	17	0	0
Very likely (4)	2	11	2	11	1	5.5
Extremely likely (5)	2	11	1	5.5	1	5.5

RD, residual disease.

RD >2 cm vs RD =0 cm and as low as 3.5% (IQR: 0%–50%) for RD >5 cm vs RD =0 cm).

Respondents also indicated that there was potential for publication bias in some comparisons when studies had larger sample sizes. However, responders appeared to dismiss the threat of publication bias for comparisons of RD <1 cm versus RD =0 cm and RD >1 cm versus RD =0 cm. Mean and median probabilities were higher and close to 100%, indicating that respondents were highly certain that a study would be published. Comparisons involving higher volume suboptimal RD (greater macroscopic disease volume) versus RD =0 cm were considered to have a low probability of being published for larger studies (the largest elicited median probability was 30% (IQR: 15%–80%) in macroscopic disease involving RD >2 cm and the probability was much less for RD 1–5 cm and RD >5 cm). This was consistent with the results for smaller studies.

Part B: Perceived likelihood of publication bias and estimation of missing studies

Table 2 shows that most responders acknowledged that the likelihood of publication bias is 'somewhat' or 'quite' likely (72.5%) in the comparison of RD <1 cm with RD =0 cm, with only one responder (5.5%) thinking it was not likely at all. This view was completely reversed for comparisons involving suboptimal RD >1 cm with RD =0 cm, where most responders thought publication bias was 'not likely at all'.

The mean and median numbers of missing studies estimated by responders for comparison of RD <1 cm versus RD =0 cm was 17 (SD 16.5) and 10 (IQR 5–20), respectively (table 3). The average number of estimated missing studies was lower for the comparisons involving suboptimal macroscopic disease volume (RD thresholds that are >1 cm). The mean and median numbers of missing studies estimated by responders for the comparison of RD >1 cm versus RD =0 cm was 8.6 (SD 12.9) and 5 (IQR 0–10), respectively (table 3). The mean number of missing studies estimated by responders for the comparison of RD >2 cm versus RD =0 cm was 6 (SD 13.2) and median was 0.5 (IQR 0–5) (table 3).

Table 4 and the tables in online supplemental appendix 3 and 4 show that, in the opinion of respondents, the number of studies that might be missing may be influenced by the effect size of those missing studies detected. For example, for the comparison of RD <1 cm versus RD =0 cm, on average 9.4 of the 17 studies would be associated with an HR of 1. As the HR increased, fewer studies were felt to be missing such that, when the detected HR was 0.5, the average number of studies felt to be missing was less than 1. Considering all the studies that were felt to be missing by respondents, a weighted average HR was estimated. This weighted average HR of the effect size from the missing studies was 0.83 (95% CI 0.77 to 0.90) for the comparison of RD <1 cm compared with RD =0 cm. This HR was calculated based on a total of 3906 participants in the estimated missing studies and 2500 deaths given a 5-year survival rate of 36% (table 4).

Table 3 Summary statistics of responders' perceived likelihood of publication bias in comparisons of near optimal (<1 cm) and suboptimal (>1/2 cm) versus complete cytoreduction (0 cm)

	RD <1 cm vs 0 cm			RD >1 cm vs 0 cm			RD >2 cm vs 0 cm		
Summary statistics (Scale 1–5)	Mean (SD)	Median (IQR)	Range	Mean (SD)	Median (IQR)	Range	Mean (SD)	Median (IQR)	Range
Overall score of perceived likelihood of publication bias (n=18)	2.94 (1.1)	3 (2–3)	1–5	2 (1.3)	1 (1–3)	1–5	1.4 (1.1)	1 (1–1)	1–5
Total estimated missing studies (n)	17.8 (16.5)	10 (5–20)	0–50	8.6 (12.9)	5 (0–10)	0–50	6.2 (13.2)	0.5 (0–5)	0–50

RD, residual disease.

Table 4 Breakdown of distribution of size and magnitude of elicited unpublished studies of near-optimal RD <1 cm versus complete cytoreduction (0 cm)

n=321 (n=17.8)		Estimated effect size					
Assumed 5-year survival: 36%		HR=1	HR=0.9	HR=0.8	HR=0.7	HR=0.6	HR≤0.5
Size of studies missed that could have been included in the analysis	Sample size	RD <1 cm and 0 cm are the same	10% less chance of mortality favouring RD <1 cm	20% less chance of mortality favouring RD <1 cm	30% less chance of mortality favouring RD <1 cm	40% less chance of mortality favouring RD <1 cm	≥50% less chance of mortality favouring RD <1 cm
	n<100	Study excluded					
	n=100	122.08*	19.12	22.7	1.34	2.14	1.14
	n=200	25.08	11.12	12.62	4.38	2.18	2.18
	n=300	6.04	4.04	1.04	2.04	0	0
	n=400	10.37	9.37	9.37	9.37	9.37	9.37
	n=500	1.04	1.04	3.04	1.04	0	0
	n>500	5.08	4.04	4.04	3.04	1.04	1.04
Total studies† (mean)		169.7 (9.4)	48.7 (2.7)	52.8 (2.9)	21.2 (1.2)	14.7 (0.8)	13.7 (0.8)
Effective n‡ (mean)		26 879 (1493.3)	12 141 (674.5)	12 899 (716.6)	7790 (432.8)	5048 (280.4)	4948 (274.9)
Effective d§ (mean)		17 203 (956)	7770 (432)	8255 (459)	4986 (277)	3231 (179)	3167 (176)
SElogHR (√(4/d))¶		0.065	0.096	0.093	0.120	0.149	0.151
95% CI for HR**		0.88–1.14	0.75–1.09	0.67–0.96	0.55–0.89	0.45–0.80	0.37–0.67
Elicited estimate††		HR 0.83 (95% CI 0.77 to 0.90), logHR −0.19 SElogHR 0.04 (n=3906, d=2500)					

*Number of studies given in the breakdown were rescaled in three respondents to correspond to the total number estimated. Therefore, any non-integer numbers in the table are due to this rescaling.
†Absolute number of estimated missing studies elicited from responders with mean (simply absolute number divided by 18 (number of responders)) given in parentheses.
‡Absolute number of estimated missing participants elicited based on total studies with mean given in parentheses.
§Absolute number of deaths estimated from number of participants assuming 5-year survival rate of 36% with mean in ().
¶Approximation of the SE of the log HR using formula derived by Parmar,[46] namely the square root of 4 divided by mean number of deaths.
**95% CI for HR calculated using logHR±1.96 multiplied by SE of log HR then transforming back by taking the exponential.
††Elicited HR with 95% CI using mean responses for all aggregated effect sizes.
RD, residual disease.

Similarly, the mean number of missing studies estimated by responders for comparison of RD >1 cm versus RD =0 cm was 8.6 (table 3). The weighted average HR of the missing studies estimated HR 0.77 (95% CI 0.70 to 0.85); this was estimated using the same approach as described above, as reported in online supplemental appendix 3. The mean number of missing studies estimated by responders for comparison of RD >2 cm versus RD =0 cm was 6.2 (table 3). The weighted average HR was estimated to be 0.79 (95% CI 0.71 to 0.89; see online supplemental appendix 4 for more details of the data).

A further analysis of results by the strength of responders' opinions as to the likelihood of publication bias was conducted. Calculating an overall HR and 95% CI for missing studies based on responders in these likelihood of publication subgroups ('not likely at all', 'somewhat likely', 'quite likely; 'very or extremely likely') led to an estimated HR of 0.90 (95% CI 0.79 to 1.03) for comparison of RD <1 cm versus RD =0 cm (table 5). These analyses were not repeated for comparisons of RD >1 cm versus RD =0 cm and RD >2 cm versus RD =0 cm, as the opinions of responders shifted towards a general feeling that publication bias was 'not likely at all'. The range in the estimated number of conducted but unpublished studies according to RD <1 cm versus RD =0 cm is provided in table 5, but a breakdown of the range by study size and effect size is not presented but is available from the authors on request.

DISCUSSION

Principal findings

The elicitation exercise was likely to appeal to experts with polarised views of radical surgery and this was useful in getting representative opinion to inform priors.[26] It found that experts considered publication bias to be a possibility when assessing OS in the comparison of RD <1 cm versus RD =0 cm after PDS for EOC. This likelihood diminished considerably for the comparisons of suboptimal RD thresholds of >1 cm and >2 cm versus RD =0 cm, with most respondents (83.5%) believing it was not likely at all in comparison to RD >2 cm versus RD =0 cm. The

Table 5 Strength of responders' opinions as to likelihood of missing studies in RD <1 cm versus RD =0 cm and number of studies elicited

Strength of opinion of likelihood of missing studies	n	Estimated missing studies			Effect estimates*	
		Mean (SD)	Median (IQR)	Range	LogHR (SElogHR)	HR (95% CI)
'Not likely'	1	0	0	0	0 (0.25)†	1.0 (0.61 to 1.63)†
'Somewhat likely'	5	5.8 (2.4)	5 (5–5)	4–10	−0.098 (0.074)	0.91 (0.78 to 1.05)
'Quite likely'	8	17.8 (13.9)	12.5 (10–20)	7–50	−0.144 (0.054)	0.87 (0.78 to 0.96)
'Very/extremely likely'	4	37.5	40(25–50)	20–50	−0.078 (0.035)	0.92 (0.86 to 0.99)
All responders	18	17.8 (16.5)	10 (5–20)	0–50	−0.103 (0.066)	0.90 (0.79 to 1.03)

*Calculated using a simple weighted average of each responder.
†No studies were estimated from responder so for purposes of analysis and calculation of pooled estimate, one small and imprecise study was used.
RD, residual disease.

most striking finding was that experts were in large agreement about not needing to make any adjustments for publication bias in comparisons involving suboptimal cytoreduction versus complete resection, irrespective of role and surgical preference.

The average completion time of the elicitation exercise was quicker than the anticipated 30–60 min. This may have been due to some responders having an initial first look at the exercise before completing it during a later visit. This may help to explain the fastest completion time of 7.7 min. This hypothesis is consistent with how the exercise was designed, as we allowed up to 24 hours for completion following a first visit. In future work, we will consider a sensitivity analysis exploring the impact of excluding responses where completion times might be unrealistic.

Strengths and limitations

The elicitation exercise was designed in collaboration with senior gynae-oncologists. This is the main reason for the detailed level of explanation given, with visual examples, to ensure that potential respondents were clear about the tasks asked of them. This involved a trade-off between clarity of explanation and potentially dissuading some respondents from taking part. Our view was that getting data on a broader range of scenarios from a reduced number of respondents would be more valuable than getting data on a smaller number of scenarios from a greater number of respondents. This was not felt to be a major limitation as it has been argued that the opinions of only 4–16 experts are needed in expert elicitation exercises.[28–31] The sample size achieved (n=18) was comfortably above this.

Part A of the elicitation exercise was based on an existing elicitation approach.[30] This part was used to identify areas where publication bias is of concern. Part B built on this by exploring the potential direction of bias. Therefore, in Part B of our elicitation exercise we collected information that would enable meta-analysis estimates to be adjusted for the impact of publication bias. The approach, while practical to use, relies on accurate survival estimates being available as these are used to inform the study sizes. As noted above, it also requires that a sufficient number of experts provide an opinion (ie, 4–16).[28–31]

Answers given by the experts to open-ended questions were prone to an 'extreme answer bias'. Therefore, we made the instructions that accompanied the elicitation exercise quite extensive. We discussed this in detail when we designed the exercise, and we feel more biases would be introduced if a ceiling of the number of estimated studies had been applied. Further work is planned to explore the impact of extreme responses on the conclusions drawn.

It is questionable as to whether the information gathered from any expert elicitation exercise can be considered a reliable estimate of relative effect. Therefore, its incorporation in a meta-analysis for adjustment may lead to 'more precise' estimates as shown by a CI but these may not be considered more reliable (that is we have gained precision but may have introduced another bias). The results shown in tables 1 and 3 appear to show variability in the answers given by the 18 respondents. Therefore, a series of sensitivity analyses would need to be conducted in order to test how robust the overall conclusions are to variations in the value of the priors used.

Implications for researchers and policy makers

Numerous recommendations have been put forth to help prevent publication bias in a systematic review, such as preregistration,[33] openness to negative or null findings by journal reviewers and editors,[34] use of preprint services to ameliorate the file-drawer problem,[35] and encouraging publication regardless of journal impact—which is often conflated as a metric of research quality.[36] These may offer a solution and minimise publication bias. However, they are not without issues. This leaves a need for methods that can instead allow us to explore and characterise the impact of publication bias. Our proposed method of expert elicitation can assist in this exploration.

The elicitation exercise provided results that may facilitate adjusting estimated effect sizes obtained with a meta-analysis for publication bias. Responders estimated that

data for substantial numbers of participants might be missing (eg, the estimate was over 3900 for the comparison of RD <1 cm vs 0 cm); this could have an impact on the results of meta-analyses. In particular, the responses from the elicitation exercise could be used to form Bayesian priors for a meta-analysis; specifically, the prior could be used to adjust the observed effect estimates obtained from the meta-analysis to explore the expected impact of publication bias. The 'educated guesses' from respondents are the only substantial source of information in this area that may facilitate such adjustment. The use of this method may be particularly important in situations like the one presented, where there is broad agreement that there is selective reporting and that there are unpublished studies that would provide 'non-significant' or 'negative' results. Should the estimates derived from the elicitation be used to adjust the meta-analysis comparing RD <1 cm, RD > 1 cm and RD >2 cm with RD =0 cm, we would expect that this would dilute the point estimate of the HR from any meta-analysis that suggested a benefit in OS for women whose tumour was cytoreduced to RD =0 cm. However, in this particular instance, there would be increased precision around the point estimate.

Within the online supplemental appendix 2, we outline one way in which such a prior could be formed from the collected data. In this approach, the weight given to each adjustment varies for the comparison of the different RD thresholds versus RD =0 cm. For example, respondents estimated more missing studies which included a greater number of participants for the comparison of RD <1 cm versus RD =0 cm. Consequently, the comparison of RD <1 cm versus RD =0 cm would have more influence in any adjustments made in a meta-analysis. Whereas, for the comparison of RD >2 cm versus RD =0 cm the estimates from the meta-analysis would be less affected as the consensus among responders of the exercise was that there was far less concern about publication bias. Furthermore, our illustrative approach gives each responder the same weight so that they contribute equally to the prior elicitation. However, we note that it would be possible to explore giving different groups a different weight. This might be relevant if we believed that different groups have different views on the nature and extent of missing data.

In meta-analyses assessing OS in suboptimal RD after PDS for advanced EOC, the evidence is relatively sparse, especially for RD thresholds >2 cm compared with RD =0 cm. For example, in our provisional scope of the results (necessary to facilitate Part A of the elicitation exercise), there was only one study that directly compared RD >2 cm versus RD =0 cm, and three studies where some indirect evidence relevant to this comparison was available. These four studies included only 478 women who contributed data for the comparison of RD >2 cm versus RD =0 cm. In this circumstance, the impact of prior expectations on the nature and extent of publication bias is likely to considerably affect the estimate. However, as evidence accumulates, the weight given to a prior when making an adjustment to the meta-analysis result will be reduced.

Implications for clinicians

Publication bias can contribute to a false impression of the efficacy of a treatment effect or a prognostic factor within a body of literature.[37 38] In the context of our expert elicitation exercise, publication bias appears to be most prone in the comparison of RD <1 cm and RD =0 cm. This may be due to the difficulty in knowing for sure that surgery has completely removed all tumour, as there still may be macroscopic disease. The a priori expectation is that this would bias the effect estimates in favour of near-optimal cytoreduction (RD <1 cm). The likelihood of publication bias comparing suboptimal cytoreduction >1 cm versus RD =0 cm was perceived by experts to be very low. If the literature is positively biased towards a certain conclusion, then meta-analyses will reflect that trend. Although there are assistive methods to help identify and expose publication bias such as funnel plots,[18] they are by no means a full solution to the problem.

Research has shown that evidence from the literature is not the sole determining factor for clinical decision-making. Clinicians also have a preference for 'consensus-based decision-making' through relatively informal sources, such as their clinical colleagues and fellow academic experts. The opportunity to discuss and trade perspectives is treated as a valuable exchange to gather information and formulate one's judgement.[39 40] Therefore, expert elicitation could be used to explore the impact of areas of uncertainty when developing clinical guideline recommendations.

Implications for future research

An extension to our work could be to build on the idea of using individual patient data (IPD) in meta-analyses,[41] rather than using aggregate data. IPD can more easily incorporate a consistent selection of confounders to adjust for, which would reduce the impact of selective reporting of analyses and outcomes. An IPD analysis would also allow for comprehensive further exploration of confounders, which could include looking at possible interaction effects between confounders.[42]

Additionally, it may not necessarily be missing studies that are the sole cause of publication bias; a systematic review can also be prone to the selective reporting of outcomes and analyses within published studies.[9 43–45] This is an area that has been comprehensively critiqued and can be overcome to a large extent by conducting an IPD meta-analysis.[42] Knowing that selective reporting is highly likely in the area under consideration, participants of the exercise potentially factored this into their elicitation estimates as, effectively, it equates to a missing study.

CONCLUSION

Previous evidence from meta-analyses suggests that complete cytoreduction of EOC is associated with

increased OS. However, our elicitation exercise of 18 experts also suggests that there is the potential for some concern about the nature and extent of publication bias in this area. The concerns are such that the unpublished evidence may substantially reduce or even remove the suggested OS benefit from complete cytoreduction compared with RD <1 cm. The results may raise awareness that a degree of scepticism is needed when reviewing studies comparing RD <1 cm versus RD =0 cm, especially when such evidence comes from non-randomised and sometimes post hoc analyses. Expert elicitation can be used to explore the impact of areas of uncertainty when developing clinical guideline recommendations. However, there is a strong belief among respondents that complete cytoreduction has an improved survival outcome compared with RD >1 cm and that publication bias is not related to that perception.

Author affiliations
[1]Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK
[2]Obstetrics and Gynaecology, Nottingham City Hospital, Nottingham, UK
[3]Pan-Birmingham Gynaecological Oncology Cancer Centre, Sandwell and West Birmingham Hospitals NHS Trust, Birmingham, UK
[4]Northern Gynaecological Oncology Centre, Queen Elizabeth Hospital, Gateshead, UK

Twitter Andrew Bryant @AndrewBryant82

Acknowledgements The authors thank Debbie Lewis from BGCS for excellent support with the conduct of the elicitation exercise. The authors also thank the BGCS organisation and survey committee for assistance in improving the content of the elicitation exercise and the opportunity to disseminate to members to complete. Judgements for each of the scenarios were based on the personal opinions of the sample of British Gynaecological Cancer Society (BGCS) members and reflected their own experience in this area. They do not necessary represent the views of BGCS as an organisation.

Contributors AB is leading author and conceptualised the methodology aspect of the research, drafted the paper, carried out the statistical analyses and is responsible for the overall content as guarantor. MG is senior statistician on the paper and reviewed methodology, analyses and critically inputted into sections of the paper. SH is a statistician who critically assessed sections of the paper and offered expertise in survey design. KG and AE provided clinical expertise and critical review as senior gynae-oncologists and researchers. EJ critically reviewed the paper and offered research experience in evidence synthesis. LV and DC are senior research academics who rigorously reviewed the methods and results and had input into the discussion. RN is a gynae-oncologist who conceptualised the clinical aspect of the research and rigorously inputted into the methods and discussion. All authors reviewed and approved the final version.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This research involved human participants outside of a study or trial setting. The elicitation exercise did not require ethical approval because the elicitation exercise was sent to BGCS members and participation was optional. Information about any expert that participated in the elicitation exercise was kept confidential. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. An anonymised data set may be available on request and/or additional summary statistics provided.

ORCID iD
Andrew Bryant http://orcid.org/0000-0003-4351-8865

REFERENCES

1 Bryant A, Hiu S, Kunonga P. *Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery*. Cochrane Database of Systematic Reviews, 2021.
2 Abbasi K. Compulsory registration of clinical trials. *BMJ* 2004;329:637–8.
3 Moher D, Hopewell S, Schulz KF, *et al*. Consort 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869.
4 Moher D, Jones A, Lepage L, *et al*. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 2001;285:1992–5.
5 Song F, Parekh S, Hooper L, *et al*. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess* 2010;14
6 UKRI Medical Research Council. Mrc clinical trials review; 2019 [Accessed 01 Oct 2020].
7 Murad MH, Asi N, Alsawas M, *et al*. New evidence pyramid. *Evid Based Med* 2016;21:125–7.
8 O'Connor D, Green S, Higgins JPT. Chapter 5: Defining the Review Question and Developing Criteria for Including Studies. In: *Cochrane Handbook for systematic reviews of interventions version 5.1.0*. The Cochrane Collaboration, 2011. https://handbook-5-1.cochrane.org/
9 Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: *Cochrane Handbook for systematic reviews of interventions version 5.1.0*. The Cochrane Collaboration, 2011. https://handbook-5-1.cochrane.org/
10 Schulz KF, Grimes DA, Altman DG, *et al*. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ* 1996;312:742–4.
11 Burdett S, Stewart LA, Tierney JF. Publication bias and meta-analyses: a practical example. *Int J Technol Assess Health Care* 2003;19:129–34.
12 Sterne JAC, Egger M, Moher D. Chapter 10: Addressing reporting biases. In: *Cochrane Handbook for systematic reviews of interventions version 5.1.0*. The Cochrane Collaboration, 2011. https://handbook-5-1.cochrane.org/
13 Ekmekci PE. An increasing problem in publication ethics: publication bias and editors' role in avoiding it. *Med Health Care Philos* 2017;20:171–8.
14 Landewé RBM. Editorial: how publication bias may harm treatment guidelines. *Arthritis Rheumatol* 2014;66:2661–3.
15 Egger M, Jüni P, Bartlett C, *et al*. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? empirical study. *Health Technol Assess* 2003;7:1–82.
16 Hopewell S, McDonald S, Clarke M, *et al*. Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database Syst Rev* 2007;2:MR000010.
17 Mallett S, Hopewell S, Clarke M. Grey literature in systematic reviews: The first 1000 Cochrane systematic reviews. In: *4Th*

Symposium on systematic reviews: pushing the boundaries. Oxford, UK, 2002.
18 Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol* 2001;54:1046–55.
19 Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;56:455–63.
20 Duval S, Tweedie R. Practical estimates of the effect of publication bias in meta-analysis. *Australasian Epidemiologist* 1998;5:14–17.
21 Rosenthal R. The file drawer problem and tolerance for null results. *Psychol Bull* 1979;86:638–41.
22 Copas JB, Malley PF. A robust P-value for treatment effect in meta-analysis with publication bias. *Stat Med* 2008;27:4267–78.
23 Sutton AJ, Song F, Gilbody SM, *et al*. Modelling publication bias in meta-analysis: a review. *Stat Methods Med Res* 2000;9:421–45.
24 Egger M, Smith GD, Altman DG, eds. *Systematic revews in health care: Meta-analysis in context*. London, UK: BMJ Publishing Group, 2001.
25 Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian approaches to clinical trials and health-care evaluation*. Chichester, UK: John Wiley & Sons, 2004.
26 Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Stat Methods Med Res* 2001;10:277–303.
27 Sutton AJ, Abrams KR, Jones DR. *Methods for meta-analysis in medical research*. Chichester, UK: John Wiley & Sons, 2000.
28 Colson AR, Cooke RM. Expert elicitation: using the classical model to validate experts' judgments. *Rev Environ Econ Policy* 2018;12:113–32.
29 Cooke RM, Goossens LJH. Procedures guide for structured expert judgment. European Commission; 2000.
30 Mavridis D, Welton NJ, Sutton A, *et al*. A selection model for accounting for publication bias in a full network meta-analysis. *Stat Med* 2014;33:5399–412.
31 Wilson ECF, Usher-Smith JA, Emery J, *et al*. Expert elicitation of multinomial probabilities for decision-analytic modeling: an application to rates of disease progression in undiagnosed and untreated melanoma. *Value Health* 2018;21:669–76.
32 StataCorp. *Stata statistical software: release 15*. College Station, TX: StataCorp LLC, 2017.
33 Nosek BA, Ebersole CR, DeHaven AC, *et al*. The preregistration revolution. *Proc Natl Acad Sci U S A* 2018;115:2600–6.
34 Bespalov A, Steckler T, Skolnick P. Be positive about negatives-recommendations for the publication of negative (or null) results. *Eur Neuropsychopharmacol* 2019;29:1312–20.
35 Verma IM. Preprint servers facilitate scientific discourse. *Proc Natl Acad Sci U S A* 2017;114:12630.
36 Bohannon J. Hate Journal impact factors? new study gives you one more reason. *Science* 2016.
37 Montori VM, Smieja M, Guyatt GH. Publication bias: a brief review for clinicians. *Mayo Clinic Proceedings* 2000;75:1284–8.
38 Riley RD, Moons KGM, Snell KIE, *et al*. A guide to systematic review and meta-analysis of prognostic factor studies. *BMJ* 2019;6:k4597.
39 Gupta DM, Boland RJ, Aron DC. The physician's experience of changing clinical practice: a struggle to unlearn. *Implementation Science* 2017;12:28.
40 Kristensen N, Nymann C, Konradsen H. Implementing research results in clinical practice- the experiences of healthcare professionals. *BMC Health Serv Res* 2015;16:48.
41 Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221.
42 Riley RD, Debray TPA, Fisher D, *et al*. Individual participant data meta-analysis to examine interactions between treatment effect and participant-level covariates: statistical recommendations for conduct and planning. *Stat Med* 2020;39:2115–37.
43 Dwan K, Altman DG, Clarke M, *et al*. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. *PLoS Med* 2014;11:e1001666.
44 Williamson PR, Gamble C. Identification and impact of outcome selection bias in meta-analysis. *Stat Med* 2005;24:1547–61.
45 Williamson PR, Gamble C, Altman DG, *et al*. Outcome selection bias in meta-analysis. *Stat Methods Med Res* 2005;14:515–24.
46 Parmar MKB, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med* 1998;17:2815–34.

SUPPLEMENTARY MATERIAL

Appendix 1: Elicitation exercise on residual disease at primary surgery for advanced ovarian cancer sent to BGCS members

Job title

Please specify your job title

1. Sub-Specialist Consultant
2. Consultant gynae oncologist
3. Consultant gynaecologist –Unit Lead
4. Consultant gynaecologist –Other
5. Consultant Clinical Oncologist
6. Consultant Medical Oncologist
7. Consultant Histopathologist
8. Consultant Cytopathologist
9. Consultant Radiologist
10. Staff or Associate Specialist –Gynaecological Oncology
11. Staff or Associate Specialist –Other
12. Subspecialty Trainee Gynaecological Oncology
13. Specialty Registrar or Clinical/Research Fellow –O&G
14. Specialty Registrar or Clinical/Research Fellow –Clinical/Medical Oncology
15. Specialty Registrar or Clinical/Research Fellow -Radiology
16. Specialty Registrar or Clinical/Research Fellow –Palliative Care
17. Specialty Registrar or Clinical/Research Fellow –Other

Introduction

Participant Information Sheet

Bryant A, *et al*. *BMJ Open* 2022; 12:e060183. doi: 10.1136/bmjopen-2021-060183

Invitation

This is an invitation to complete a complex survey on residual disease at primary surgery for advanced ovarian cancer that will take up to 30 minutes, but as BGCS members you might consider the altruistic value of contributing towards an area of uncertainty within your field.

The nature of expert elicitation surveys are that they typically only need completion from relatively few experts but it is important that respondents have the necessary expertise and interest in the area. Elicitation surveys are often the only way of resolving issues of uncertainty.

The survey has been designed in consultation with several gynae-oncologists and that is the main reason for the detailed level of explanation given with visual examples so it is clear what is being asked of the respondent.

**Please use a computer or laptop to complete the survey as it is not mobile-friendly.*

Introduction to research problem

Residual disease at surgery for advanced ovarian cancer is one of the factors that influences survival. However, there is a lack of randomised controlled trials (RCTs) in upfront surgery for advanced ovarian cancer. This may be because some clinicians believe that tumour biology plays a greater role in predicting patient survival, undermining the importance of making every possible effort to obtain complete cytoreduction.

Available studies are retrospective in nature, looking at residual disease at surgery and patient survival after upfront surgery and chemotherapy. There is also huge variation in reporting and definitions. One consequence of this is the potential for publication bias due to selective or nonreporting of studies.

This presents challenges when conducting systematic reviews and meta-analyses. To overcome some of the challenges, we can think about what sort of studies have been conducted but not published. One way to do this is to ask for the opinions of experts such as yourself and incorporate your beliefs into our analyses. To do this we would like your opinions about a number of different scenarios describing the likelihood of different studies not being published.

Bryant A, *et al*. *BMJ Open* 2022; 12:e060183. doi: 10.1136/bmjopen-2021-060183

Impact of this survey

Meaningful and reliable conclusions will be drawn from this survey and it is the views from experts that is crucial to get informative, reliable and representative results. The adjustments for publication bias based on the survey results can potentially be transferred into other areas of Oncology so the survey will be extremely informative moving forward.

The results of the survey will be confidentially shared with all contributors and you will of course be acknowledged for your efforts. The results of the survey will be part of a publication on residual disease threshold after primary surgical treatment for advanced epithelial ovarian cancer (EOC), using your expert views to adjust for potential publication bias. This publication will be sent to BGCS members as soon as it is published.

How the survey works

The next sections describe the overall objective of the research this survey will inform, and a short summary of the methods used to address this. You will then be presented with the expert elicitation exercise, which will have three main parts. Expert elicitation is essentially a scientific consensus methodology. It allows for parametrisation (using your highly 'educated guesses'), for the respective questions and scenarios under consideration. The main purpose of this elicitation exercise is to quantify uncertainty.

Objectives

Objective of the type of research this survey will inform

1. *To evaluate the effects of residual disease on survival after primary cytoreductive surgery for women with advanced epithelial ovarian cancer (stages III and IV).*

To address this objective the following methods, briefly summarised next, will be used.

Please take some time to familiarise yourself with the methods.

Bryant A, *et al*. *BMJ Open* 2022; 12:e060183. doi: 10.1136/bmjopen-2021-060183

Types of studies
Data from randomised controlled trials (RCTs), prospective and retrospective cohort studies, and unselected case series of 100 or more patients that included concurrent comparison of different residual disease (RD) thresholds after primary surgical intervention.

Any data collected from RCTs were retrospective and taken from trials that randomised groups of women to various chemotherapy protocols after primary surgery and the surgical outcome was categorised as complete (microscopic or no visible disease), optimal, and suboptimal based on the maximum size of postoperative residual disease.

Case-control studies, studies that did not have concurrent comparison groups, and case series of fewer than 100 patients were excluded.

In order to minimise selection bias, we included only studies that used statistical adjustment for baseline case mix using multivariable analyses (for example age, stage, grade, extent of disease).

Types of participants
Adult women (over 18 years of age) with surgically staged advanced epithelial ovarian cancer (FIGO stage III/IV) who had confirmed histological diagnoses. Women with other concurrent malignancies were excluded.

Types of interventions
Intervention: primary optimal cytoreductive surgery followed by adjuvant platinum-based chemotherapy. We only included studies that defined optimal cytoreduction as surgery leading to residual tumours with a maximum diameter of any threshold up to 2 cm. Patients who received chemotherapy prior to surgery were excluded.

Comparison: women who had primary surgery resulting in residual disease which did not meet the criteria specified in the study as optimal, followed by adjuvant platinum-based chemotherapy.

Outcome
Overall survival was the outcome of interest and was defined as survival until death from all causes.

Searches
Electronic databases indicating the Cochrane Gynaecological Cancer Collaborative Review Group Trials Register, CENTRAL, MEDLINE and EMBASE were searched from 1950 up to January 2020. A comprehensive search of the grey literature was performed and extensive hand searches were carried out in pertinent areas. There were no language restrictions.

Expert elicitation

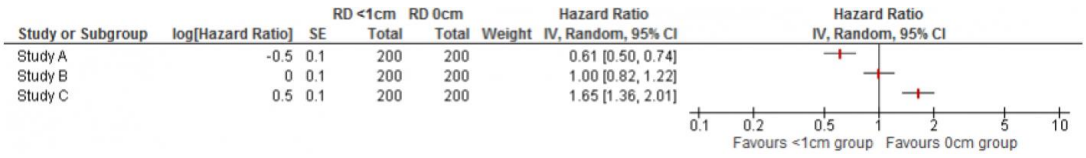
The expert elicitation exercise has three parts: A-C. Please answer these parts in order.

Before you do this please read the following text:

In subsequent tasks in the survey, you will be presented with statistics commonly reported in studies using survival models e.g., hazard ratios (HR). To help you familiarise with these statistics, kindly take a moment to consider the forest plot below for three studies A, B, and C and the associated interpretations in bullet points. *Do not worry about having to memorise the results, they are merely for illustrative purposes.*

- Study A shows statistically significant prolonged survival in the RD <1 cm threshold (or more risk of death in the RD 0cm threshold) than when residual disease was completely cytoreduced to 0cm.
- Study B shows no statistically significant difference in the risk of death between RD <1 cm and RD 0cm thresholds.
- Study C shows statistically significant prolonged survival in the 0cm threshold (or more risk of death in the RD <1cm threshold) than in the RD <1cm threshold.

Although it is possible for studies favouring RD <1cm (or other RD thresholds) over complete cytoreduction (0cm) to be published, it seems less likely because of the greater likelihood of reporting bias amongst studies reporting no statistical significance or ones favouring RD <1cm over RD of 0cm. This will be interpreted in light of any adjustment made.



Part A

Question 1

This section requires you to please provide estimates of the chance (probability) a study of a given sample size, for a certain comparison, is published.

The table below shows residual disease (RD) thresholds and sample sizes, which are all compared to the reference microscopic disease (RD 0cm). The studies mimic the inclusion criteria as outlined in the introduction. Please complete what in your opinion would be the chance that a study of a certain sample size comparing a specific RD threshold versus RD 0cm is published. Kindly do this for each of the 16 options below.

Kindly enter the percentage chance of being published for studies of given sample size and residual disease thresholds compared to microscopic disease (0cm). Kindly enter a value between 0 (no chance of publication) and 100 (certainly published).

A percentage of 0% indicates that you think there is no chance at all of publication and 100% means it is certain to be published. The value you should put for each option should lie between 0 and 100% likelihood of being published. Tossing an unbiased coin and getting a head would have 50% chance. There is no correct answer; your judgements for each option are your own personal opinions and reflect your experience in this area, but it is with these we hope to use in our analyses.

RD threshold (versus microscopic disease (RD 0cm)	Sample size (n) in comparison with microscopic disease (RD 0cm)	% chance of being published [value between 0 (no chance) and 100 (certain)]	n in comparison with RD 0cm	% chance of being published [value between 0 (no chance) and 100 (certain)]
LESS THAN 1 cm	100		1000	
GREATER THAN 0cm	100		625	
BETWEEN 1cm and 2cm	100		210	
LESS THAN 2cm	100		250	
GREATER THAN 1cm	100		1000	
GREATER THAN 2cm	100		250	
BETWEEN 1cm and 5cm	100		250	
GREATER THAN 5cm	100		250	

Part B

In lay terms, there is large literature suggesting a strong association with complete cytoreduction (0cm) and prolonged survival. However, due to the nature of studies looking at the association between complete cytoreduction and survival, whether there is selective reporting of studies is open to debate.

As experts in this area, it is assumed you will be very familiar with the literature and be aware of publications in ovarian cancer debulking journals on a regular basis. It is the studies that MAY have been conducted but not published in journals that you will not be aware of and we want you to consider how many of these there are likely to be.

In this part of the survey, we would like you to provide us with responses to questions that allows us to adjust the overall effect estimate when data from unobserved studies are added to the final analysis.

Question 2

2. Near optimal RD<1cm versus complete cytoreduction (0cm)

In this section, we would like you to provide us with responses to questions that allows us to adjust the overall effect estimate when data from unobserved studies are added to the final analysis. This adjustment will account for an absence (or not) of studies favouring near optimal RD <1cm or ones showing no statistically significant difference between RD <1cm and RD 0cm, based upon your own opinion and clinical experience in this area.

How likely is it that relevant studies reporting adequately sized analyses that did not favour complete cytoreduction (RD to 0cm) when compared to RD <1cm would not have been identified from the literature searches and therefore omitted from the meta-analysis? By this we mean how likely is it that studies that either favoured RD <1cm or studies that found no statistically significant difference (p>0.05) in survival between RD 0cm and RD <1cm) would not be published?

Bryant A, *et al*. *BMJ Open* 2022; 12:e060183. doi: 10.1136/bmjopen-2021-060183

- Studies reporting statistically significant prolonged survival in favour of RD LESS THAN 1cm (that is the effect size in the form of a hazard ratio is less than 1 and the upper 95% confidence interval does not cross 1)

OR

- Studies that reported no statistically significant difference in survival between RD LESS THAN 1cm and 0cm (that is the 95% confidence interval, reporting lower and upper estimates of hazard ratio, crosses 1)

	Not likely at all	Somewhat likely	Quite likely	Very likely	Extremely likely
Please indicate the strength of your opinion	○	○	○	○	○

Thinking about your response the question above and giving a realistic answer based on your own experience and awareness of previous analyses in this area, how many studies in total do you think will have been missed that should have been included?

Please give a brief reason for your answer

Please indicate in the table below where you think the number of studies you gave will be distributed. You're not expected to fill in all rows and columns and you may include multiple studies of the same size and magnitude. Assume for this scenario that the missing studies have an assumed 5 year survival of 36%

Assumed 5 year survival: 36%		RD <1cm and 0cm are the same i.e. HR = 1	10% less chance of mortality favouring RD <1cm i.e. HR = 0.9	20% less chance of mortality favouring RD <1cm i.e. HR = 0.8	30% less chance of mortality favouring RD <1cm i.e. HR = 0.7	40% less chance of mortality favouring RD <1cm i.e. HR = 0.6	>=50% less chance of mortality favouring RD <1cm i.e. i.e. HR ≤ 0.5
Size of studies missed that could have been included in the analysis	n=100						
	n=200						
	n=300						
	n=400						
	n=500						
	n>500						

Question 3

3. Sub-optimal RD>1cm versus complete cytoreduction (0cm)

In this section, we would like you to provide us with responses to questions that allows us to adjust the overall effect estimate when data from unobserved studies are added to the final analysis. This adjustment will account for an absence (or not) of studies favouring suboptimal RD >1cm or ones showing no statistically significant difference between RD >1cm and RD 0cm, based upon your own opinion and clinical experience in this area.

How likely is it that relevant studies reporting adequately sized analyses that did not favour complete cytoreduction (RD to 0cm) when compared to RD >1cm would not have been identified from the searches and therefore omitted from the meta-analysis? By this we mean how likely is it that studies that either favoured RD >1cm or studies that found no statistically significant difference (p>0.05) in survival between RD 0cm and RD >1cm) would not be published?

- Studies reporting statistically significant prolonged survival in favour of RD GREATER THAN 1cm (that is the effect size in the form of a hazard ratio is less than 1 and the upper 95% confidence interval does not cross 1)

OR

- Studies that reported no statistically significant difference in survival between RD GREATER THAN 1cm and 0cm (that is the 95% confidence interval, reporting lower and upper estimates of hazard ratio, crosses 1)

	Not likely at all	Somewhat likely	Quite likely	Very likely	Extremely likely
Please indicate the strength of your opinion	○	○	○	○	○

Thinking about your response the question above and giving a realistic answer based on your own experience and awareness of previous analyses in this area, how many studies in total do you think will have been missed that should have been included?

Please give a brief reason for your answer

Please indicate in the table below where you think the number of studies you gave will be distributed. You're not expected to fill in all rows and columns and you may include multiple studies of the same size and magnitude. Assume for this scenario that the missing studies have an assumed 5 year survival of 36%

Assumed 5 year survival: 36%		RD >1cm and 0cm are the same i.e. HR = 1	10% less chance of mortality favouring RD >1cm i.e. HR = 0.9	20% less chance of mortality favouring RD >1cm i.e. HR = 0.8	30% less chance of mortality favouring RD >1cm i.e. HR = 0.7	40% less chance of mortality favouring RD >1cm i.e. HR = 0.6	>=50% less chance of mortality favouring RD >1cm i.e. HR ≤ 0.5
Size of studies missed that could have been included in the analysis	n=100						
	n=200						
	n=300						
	n=400						
	n=500						
	n>500						

Question 4

4. Sub-optimal RD>2cm versus complete cytoreduction (0cm)

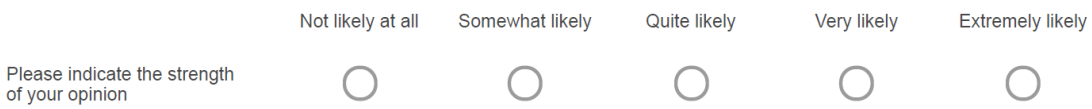
In this section, we would like you to provide us with responses to questions that allows us to adjust the overall effect estimate when data from unobserved studies are added to the final analysis. This adjustment will account for an absence (or not) of studies favouring suboptimal RD >2cm or ones showing no statistically significant difference between RD >1cm and RD 0cm, based upon your own opinion and clinical experience in this area.

How likely is it that relevant studies reporting adequately sized analyses that did not favour complete cytoreduction (RD to 0cm) when compared to RD >2cm would not have been identified from the searches and therefore omitted from the meta-analysis? By this we mean how likely is it that studies that either favoured RD >2cm or studies that found no statistically significant difference (p>0.05) in survival between RD 0cm and RD >2cm) would not be published?

- Studies reporting statistically significant prolonged survival in favour of RD GREATER THAN 2cm (that is the effect size in the form of a hazard ratio is less than 1 and the upper 95% confidence interval does not cross 1)

OR

- Studies that reported no statistically significant difference in survival between RD GREATER THAN 2cm and 0cm (that is the 95% confidence interval, reporting lower and upper estimates of hazard ratio, crosses 1)



Thinking about your response the question above and giving a realistic answer based on your own experience and awareness of previous analyses in this area, how many studies in total do you think will have been missed that should have been included?

Please give a brief reason for your answer

Please indicate in the table below where you think the number of studies you gave will be distributed. You're not expected to fill in all rows and columns and you may include multiple studies of the same size and magnitude. Assume for this scenario that the missing studies have an assumed 5 year survival of 36%

Assumed 5 year survival: 36%		RD >2cm and 0cm are the same i.e. HR = 1	10% less chance of mortality favouring RD >2cm i.e. HR = 0.9	20% less chance of mortality favouring RD >2cm i.e. HR = 0.8	30% less chance of mortality favouring RD >2cm i.e. HR = 0.7	40% less chance of mortality favouring RD >2cm i.e. HR = 0.6	>=50% less chance of mortality favouring RD >2cm i.e. HR ≤ 0.5
Size of studies missed that could have been included in the analysis	n=100						
	n=200						
	n=300						
	n=400						
	n=500						
	n>500						

Part C

Question 5

In a meta-analysis including non-randomised studies, often only univariate results are reported with no attempt made to adjust for potentially important baseline imbalances. This risks making the results biased.

On a scale of 0-100, to what extent do you think that the reason study authors only report univariate analyses is to maximise the magnitude in effect estimates to favour either an experimental or comparator group?

Question 6

In your opinion, how many attempted submissions should you make to journals to publish the results of your study?

Question 7

In your opinion, how many attempted submissions should you make to journals to publish the results of your study if it is not statistically significant (p>0.05)?

Question 8

What is lowest impact factor in a journal that you would consider submission of your work, regardless of the significance of your results?

	<1 e.g. Turkish Journal of Medical Sciences	1-5 e.g. BJOG	6-10 e.g. BMC Medicine	11-14 e.g PLoS Medicine	15-19 e.g. Annals of Internal Medicine	20-24 e.g. BMJ	25+ e.g. Lancet Oncology
Lowest impact factor	○	○	○	○	○	○	○

Question 9

To what extent do you think it is important to publish the results of a study even if the impact factor of the accepting journal is perceived to be very low?

Not important at all — Vitally important

0 10 20 30 40 50 60 70 80 90 100

Acknowledgment

(Optional) If you would like to be acknowledged for your contribution to the survey, kindly leave your name.

You may choose to provide your full name (e.g., Sam Smith) or an abbreviation (e.g., S Smith).

Data on your name will be kept separate from the data file containing the survey results so that your personal information cannot be traced back to your responses. Your name will only be used for purposes of acknowledgment and will not be used in analysis.

Which of the following mediums do you consent to being acknowledged in? You may choose all that apply.

- ☐ Journal publication
- ☐ Conference poster/oral presentation
- ☐ BGCS internal dissemination (e.g., newsletter)

Appendix 2: Statistical considerations and analysis

Part A

The scenario in part A of the elicitation exercise assumed that there is a population of studies which have been conducted assessing OS in RD thresholds after primary surgery for EOC. Then, it assumed that there are a finite number of published studies that have reported an estimated effect, with precision around that estimate (measured using standard error). In the presence of publication bias these studies are a non-random sample of all studies that have been conducted in this area. It is assumed that very large studies have a probability of being published very close to one, as journals tend to trust larger studies. Conversely, small studies have a diminished chance of publication. An additional consideration is that if effect size is correlated with the probability of a study being published, then this will introduce additional bias.(1)

Part B

The results of section B are a particularly novel aspect of this research, specifically as they could be used as prior information to inform adjustment of meta-analyses for publication bias. Here, we outline how this could be achieved.

First, we require that all calculations should give each expert responder the same weight, such that they contribute equally to the prior formation. The average study size for each effect size (HR) point estimate is then calculated; the sample size is dictated by the number of studies selected by each individual expert, with the average then calculated giving equal weight to each respondent. A normal prior is then formed for each HR, before these are combined in a weighted manner to a single elicited prior suitable for adjusting for publication bias.

We note that this is only one potential way to form a prior based on this elicited data and that a sensitivity analysis should certainly be conducted and potentially also other approaches considered. In our elicitation exercise, the choice of the number of missing

Bryant A, *et al*. *BMJ Open* 2022; 12:e060183. doi: 10.1136/bmjopen-2021-060183

studies was left open ended as to not lead experts to a choice and bias the results. Consequently, a sensitivity analysis could be conducted removing high estimates of unpublished studies if it was judged that unrealistic entries were unduly inflating an average.

Given an assumed 5 year survival rate of 36% (2-4) and a minimum sample size of n=100 to meet the criteria for inclusion in the network meta-analysis (NMA), then a minimum 64 events (deaths, d) would be required with 36 participants being alive and censored at the end of the study:

$$(d = 100\{1 - 0.36\})$$

Generalising this result, we assume that d can be related to n in general through the following formula.

$$n = \frac{d}{1 - (\text{5 year suvival rate})} = \frac{d}{0.64}.$$

The standard error of the log hazard ratio (SElogHR) can then be related to n by rearranging the following.

$$d = \frac{4}{\text{SE}(\log HR)^2},$$

$$\Rightarrow \text{SE}(\log HR) = \sqrt{\frac{4}{d}} = \sqrt{\frac{4}{0.64n}} = \sqrt{\frac{6.25}{n}}.$$

Next, we denote by m_{cij} the number of missing studies according to expert responder $c = 1, \dots, C$, with a HR of HR_j and a sample size of n_i, where:

$$n_1 = 100, n_2 = 200, n_3 = 300, n_4 = 400, n_5 = 500, n_6 = 625,$$

$$HR_1 = 1, HR_2 = 0.9, HR_3 = 0.8, HR_4 = 0.7, HR_5 = 0.6, HR_6 = 0.5.$$

We compute the average number of missing studies of type ij, across the responders, as:

$$m_{ij} = \frac{1}{C}\sum_{c=1}^{C} m_{cij}.$$

We use this to form an average sample size of missing studies with a HR of HR_j through:

$$m_j = \frac{\sum_i n_i m_{ij}}{\sum_i m_{ij}}.$$

With this, we assume that information from missing studies with a HR of HR_j can be categorised through the following distribution:

$$P_j \sim N\left(\log HR_j, \frac{6.25}{m_j}\right).$$

The P_j can then be combined in a weighted manner, giving more weight to those values of j with a larger value of m_j, via conflation. This gives a single elicited prior of:

$$P \sim N\left(\frac{\sum_j \frac{m_j \log HR_j}{6.25}}{\sum_j \frac{m_j}{6.25}}, \frac{1}{\sum_j \frac{m_j}{6.25}}\right) = N\left(\frac{\sum_j m_j \log HR_j}{\sum_j m_j}, \frac{6.25}{\sum_j m_j}\right).$$

This elicited estimate can then be used as prior information and be applied in a Bayesian analysis(5-7) that reflects the results of the expert opinion in the elicitation exercise.(1, 8).

Appendix 3: Breakdown of distribution of size and magnitude of elicited unpublished studies of sub-optimal RD >1cm versus complete cytoreduction (0cm)

N=154 (n=8.6)		**Estimated effect size**					
Assumed 5 year survival: 36%		**HR=1**	**HR=0.9**	**HR=0.8**	**HR=0.7**	**HR=0.6**	**HR≤0.5**
		RD <1cm and 0cm are the same	**10% less chance of mortality favouring RD <1cm**	**20% less chance of mortality favouring RD <1cm**	**30% less chance of mortality favouring RD <1cm**	**40% less chance of mortality favouring RD <1cm**	**>=50% less chance of mortality favouring RD <1cm**
Size of studies missed that could have been included in the analysis	**Sample size**						
	n<100	STUDY EXCLUDED					
	n=100	29.5	7.67	3.17	2.8	0.1	1.43
	n=200	14.5	6.67	3.17	2.8	0.1	1.43
	n=300	5	1.67	0	1.67	0	1.33
	n=400	9.66	8.33	8.33	8.33	8.33	9.66
	n=500	2.66	0	0	0	0	1.33
	n>500	6	2	0	0	0	6.33
Total studies[a] (mean)		67.3 (3.7)	26.3 (1.5)	14.7 (0.8)	15.6 (0.9)	8.5 (0.5)	21.5 (1.2)
Effective n[b] (mean)		16294 (905)	7184 (399)	4283 (238)	4673 (260)	3362 (187)	9313 (517)
Effective d[c] (mean)		10428 (579)	4598 (255)	2741 (152)	2991 (166)	2152 (120)	5960 (331)
SElogHR (√(4/d))[d]		0.083	0.125	0.162	0.155	0.183	0.110
95% CI for HR[e]		0.85-1.18	0.71-1.15	0.58-1.10	0.52-0.95	0.42-0.86	0.40-0.62
Elicited estimate[f]		HR=0.77 (95% CI 0.70 to 0.85), logHR=-0.26 SElogHR=0.05 (n=2500, d=1600)					

[a] Absolute number of estimated missing studies elicited from responders with mean (simply absolute number divided by 18 (number of responders)) given in parentheses

[b] Absolute number of estimated missing participants elicited based on total studies with mean given in parentheses

[c] Absolute number of deaths estimated from number of participants assuming 5 year survival rate of 36% with mean in ()

[d] Approximation of the standard error (SE) of the log hazard ratio (HR) using formula derived by Parmar(9), namely the square root of 4 divided by mean number of deaths

[e] 95% confidence interval for hazard ratio (HR) calculated using logHR ± 1.96 multiplied by standard error of log HR then transforming back by taking the exponential

[f] Elicited Hazard ratio with 95% confidence interval using mean responses for all aggregated effect sizes

[g] Number of studies given in the breakdown were rescaled in three respondents to correspond to the total number estimated

Appendix 4: Breakdown of distribution of size and magnitude of elicited unpublished studies of sub-optimal RD >2cm versus complete cytoreduction (0cm).

N=112 (6.2)		**Estimated effect size**					
Assumed 5 year survival: 36%		**HR=1**	**HR=0.9**	**HR=0.8**	**HR=0.7**	**HR=0.6**	**HR≤0.5**
		RD <1cm and 0cm are the same	**10% less chance of mortality favouring RD <1cm**	**20% less chance of mortality favouring RD <1cm**	**30% less chance of mortality favouring RD <1cm**	**40% less chance of mortality favouring RD <1cm**	**>=50% less chance of mortality favouring RD <1cm**
Size of studies missed that could have been included in the analysis	**Sample size**						
	n<100	STUDY EXCLUDED					
	n=100	14.67	7	5	0	0	0.67
	n=200	8.67	8	5	0	0	0.67
	n=300	0.67	0	0	0	0	0.67
	n=400	9	8.33	8.33	8.33	8.33	9
	n=500	1.33	0	0	0	0	0.67
	n>500	7	0	0	0	0	0.67
Total studies[a] (mean)		41.3 (2.3)	23.3 (1.3)	18.3 (1)	8.3 (0.5)	8.3 (0.5)	12.3 (0.7)
Effective n[b] (mean)		12042 (669)	5632 (313)	4832 (268)	3332 (185)	3332 (185)	4756 (264)
Effective d[c] (mean)		7707 (428)	3604 (200)	3092 (172)	2132 (118)	2132 (118)	3044 (169)
SElogHR (√(4/d))[d]		0.097	0.141	0.153	0.184	0.184	0.154
95% CI for HR[e]		0.83-1.21	0.68-1.19	0.59-1.08	0.49-1.00	0.42-0.86	0.37-0.68
Elicited estimate[f]		HR=0.79 (95% CI 0.71 to 0.89), logHR=-0.24 SElogHR=0.06 (n=1736, d=1111)					

[a] Absolute number of estimated missing studies elicited from responders with mean (simply absolute number divided by 18 (number of responders)) given in parentheses

[b] Absolute number of estimated missing participants elicited based on total studies with mean given in parentheses

[c] Absolute number of deaths estimated from number of participants assuming 5-year survival rate of 36% with mean in ()

[d] Approximation of the standard error (SE) of the log hazard ratio (HR) using formula derived by Parmar(9), namely the square root of 4 divided by mean number of deaths

[e] 95% confidence interval for hazard ratio (HR) calculated using logHR ± 1.96 multiplied by standard error of log HR then transforming back by taking the exponential

[f] Elicited Hazard ratio with 95% confidence interval using mean responses for all aggregated effect sizes

References

1. Mavridis D, Welton NJ, Sutton A, Salanti G. A selection model for accounting for publication bias in a full network meta-analysis. Stat Med. 2014;33(30):5399-412.
2. Ovarian cancer research alliance (OCRA). Stages of Ovarian Cancer. [Available from: https://ocrahope.org/patients/about-ovarian-cancer/staging/#:~:text=Most%20women%20diagnosed%20with%20Stage%20III%20ovarian%20cancer%20have%20a,survival%20rate%20of%20approximately%2039%25.
3. American Cancer Society. Survival Rates for Ovarian Cancer [Available from: https://www.cancer.org/cancer/ovarian-cancer/detection-diagnosis-staging/survival-rates.html.
4. Siegel RL, Miller, K.D. and Jemal, A. Cancer statistics. CA A Cancer J Clin. 2020;70:7-30.
5. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Chichester, UK: John Wiley & Sons; 2004.
6. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. Stat Methods Med Res. 2001;10(4):277-303.
7. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Methods for Meta-analysis in Medical Research. Chichester, UK: John Wiley & Sons; 2000.
8. Wilson ECF, Usher-Smith JA, Emery J, Corrie PG, Walter FM. Expert Elicitation of Multinomial Probabilities for Decision-Analytic Modeling: An Application to Rates of Disease Progression in Undiagnosed and Untreated Melanoma. Value in Health. 2018;21(6):669-76.
9. Parmar WKB, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Statistics in Medicine. 1998;17(24):2815-34.

Appendix 8: Expert elicitation exercise

Job title

Please specify your job title

1. Sub-Specialist Consultant
2. Consultant gynae oncologist
3. Consultant gynaecologist –Unit Lead
4. Consultant gynaecologist –Other
5. Consultant Clinical Oncologist
6. Consultant Medical Oncologist
7. Consultant Histopathologist
8. Consultant Cytopathologist
9. Consultant Radiologist
10. Staff or Associate Specialist –Gynaecological Oncology
11. Staff or Associate Specialist –Other
12. Subspecialty Trainee Gynaecological Oncology
13. Specialty Registrar or Clinical/Research Fellow –O&G
14. Specialty Registrar or Clinical/Research Fellow –Clinical/Medical Oncology
15. Specialty Registrar or Clinical/Research Fellow -Radiology
16. Specialty Registrar or Clinical/Research Fellow –Palliative Care
17. Specialty Registrar or Clinical/Research Fellow –Other

Introduction

Participant Information Sheet

Invitation

This is an invitation to complete a complex survey on residual disease at primary surgery for advanced ovarian cancer that will take up to 30 minutes, but as BGCS members you might consider the altruistic value of contributing towards an area of uncertainty within your field. The nature of expert elicitation surveys are that they typically only need completion from relatively few experts, but it is important that respondents have the necessary expertise and interest in the area. Elicitation surveys are often the only way of resolving issues of uncertainty.

The survey has been designed in consultation with several gynae-oncologists and that is the main reason for the detailed level of explanation given with visual examples, so it is clear what is being asked of the respondent.

**Please use a computer or laptop to complete the survey as it is not mobile-friendly.*

Introduction to research problem

Residual disease at surgery for advanced ovarian cancer is one of the factors that influences survival. However, there is a lack of randomised controlled trials (RCTs) in upfront surgery for advanced ovarian cancer. This may be because some clinicians believe that tumour biology plays a greater role in predicting patient survival, undermining the importance of making every possible effort to obtain complete cytoreduction.

Available studies are retrospective in nature, looking at residual disease at surgery and patient survival after upfront surgery and chemotherapy. There is also huge variation in reporting and definitions. One consequence of this is the potential for publication bias due to selective or nonreporting of studies.

This presents challenges when conducting systematic reviews and meta-analyses. To overcome some of the challenges, we can think about what sort of studies have been conducted but not published. One way to do this is to ask for the opinions of experts such as yourself and incorporate your beliefs into our analyses. To do this we would like your opinions about a number of different scenarios describing the likelihood of different studies not being published.

Impact of this survey

Meaningful and reliable conclusions will be drawn from this survey, and it is the views from

experts that is crucial to get informative, reliable, and representative results. The adjustments for publication bias based on the survey results can potentially be transferred into other areas of Oncology so the survey will be extremely informative moving forward.

The results of the survey will be confidentially shared with all contributors, and you will of course be acknowledged for your efforts. The results of the survey will be part of a publication on residual disease threshold after primary surgical treatment for advanced epithelial ovarian cancer (EOC), using your expert views to adjust for potential publication bias. This publication will be sent to BGCS members as soon as it is published.

How the survey works

The next sections describe the overall objective of the research this survey will inform, and a short summary of the methods used to address this. You will then be presented with the expert elicitation exercise, which will have three main parts. Expert elicitation is essentially a scientific consensus methodology. It allows for parametrisation (using your highly 'educated guesses'), for the respective questions and scenarios under consideration. The main purpose of this elicitation exercise is to quantify uncertainty.

Objectives

Objective of the type of research this survey will inform

1. *To evaluate the effects of residual disease on survival after primary cytoreductive surgery for women with advanced epithelial ovarian cancer (stages III and IV).*

To address this objective the following methods, briefly summarised next, will be used. Please take some time to familiarise yourself with the methods.

Expert elicitation

The expert elicitation exercise has three parts: A-C. Please answer these parts in order.

Before you do this, please read the following text:

In subsequent tasks in the survey, you will be presented with statistics commonly reported in studies using survival models e.g., hazard ratios (HR). To help you familiarise with these statistics, kindly take a moment to consider the forest plot below for three studies A, B, and C and the associated

interpretations in bullet points. *Do not worry about having to memorise the results, they are merely for illustrative purposes.*

- Study A shows statistically significant prolonged survival in the RD < 1 cm threshold (or more risk of death in the RD 0 cm threshold) than when residual disease was completely cytoreduced to 0 cm.
- Study B shows no statistically significant difference in the risk of death between RD < 1 cm and RD 0 cm thresholds.
- Study C shows statistically significant prolonged survival in the 0 cm threshold (or more risk of death in the RD < 1 cm threshold) than in the RD < 1 cm threshold.

Although it is possible for studies favouring RD < 1 cm (or other RD thresholds) over complete cytoreduction (0 cm) to be published, it seems less likely because of the greater likelihood of reporting bias amongst studies reporting no statistical significance or ones favouring RD < 1 cm over RD of 0 cm. This will be interpreted in light of any adjustment made.

Part A

Question 1

This section requires you to please provide estimates of the chance (probability) a study of a given sample size, for a certain comparison, is published.

The table below shows residual disease (RD) thresholds and sample sizes, which are all compared to the reference microscopic disease (RD 0 cm). The studies mimic the inclusion criteria as outlined in the introduction. Please complete what in your opinion would be the chance that a study of a certain sample size comparing a specific RD threshold versus RD 0 cm is published. Kindly do this for each of the 16 options below.

Kindly enter the percentage chance of being published for studies of given sample size and residual disease thresholds compared to microscopic disease (0 cm). Kindly enter a value between 0 (no chance of publication) and 100 (certainly published).

A percentage of 0% indicates that you think there is no chance at all of publication and 100%

means it is certain to be published. The value you should put for each option should lie between 0 and 100% likelihood of being published. Tossing an unbiased coin and getting a head would have 50% chance. There is no correct answer; your judgements for each option are your own personal opinions and reflect your experience in this area, but it is with these we hope to use in our analyses

RD threshold (versus microscopic disease (RD 0 cm)	Sample size (n) in comparison with microscopic disease (RD 0 cm)	% chance of being published [value between 0 (no chance) and 100 (certain)]	n in comparison with RD 0 cm	% chance of being published [value between 0 (no chance) and 100 (certain)]
LESS THAN 1 cm	100		1000	
GREATER THAN 0 cm	100		625	
BETWEEN 1cm and 2 cm	100		210	
LESS THAN 2 cm	100		250	
GREATER THAN 1 cm	100		1000	
GREATER THAN 2 cm	100		250	
BETWEEN 1 cm and 5 cm	100		250	
GREATER THAN 5 cm	100		250	

Part B

In lay terms, there is large literature suggesting a strong association with complete cytoreduction (0cm) and prolonged survival. However, due to the nature of studies looking at the association between complete cytoreduction and survival, whether there is selective reporting of studies is open to debate.

As experts in this area, it is assumed you will be very familiar with the literature and be aware of publications in ovarian cancer debulking journals on a regular basis. It is the studies that MAY have been conducted but not published in journals that you will not be aware of, and we want you to consider how many of these there are likely to be.

In this part of the survey, we would like you to provide us with responses to questions that allows us to adjust the overall effect estimate when data from unobserved studies are added to the final analysis.

Question 2

2. Near optimal RD<1cm versus complete cytoreduction (0cm)

In this section, we would like you to provide us with responses to questions that allows us to adjust the overall effect estimate when data from unobserved studies are added to the final analysis. This adjustment will account for an absence (or not) of studies favouring near optimal RD < 1 cm or ones showing no statistically significant difference between RD < 1 cm and RD 0 cm, based upon your own opinion and clinical experience in this area.

How likely is it that relevant studies reporting adequately sized analyses that did not favour complete cytoreduction (RD to 0 cm) when compared to RD < 1 cm would not have been identified from the literature searches and therefore omitted from the meta-analysis? By this we mean how likely is it that studies that either favoured RD < 1 cm or studies that found no statistically significant difference ($p > 0.05$) in survival between RD 0 cm and RD < 1 cm) would not be published?

- Studies reporting statistically significant prolonged survival in favour of RD LESS THAN 1 cm (that is the effect size in the form of a hazard ratio is less than 1 and the upper 95% confidence interval does not cross 1)

OR

- Studies that reported no statistically significant difference in survival between RD LESS THAN 1 cm and 0 cm (that is the 95% confidence interval, reporting lower and upper estimates of hazard ratio, crosses 1)

Thinking about your response the question above and giving a realistic answer based on your own experience and awareness of previous analyses in this area, how many studies in total do you think will have been missed that should have been included?

Please give a brief reason for your answer

Please indicate in the table below where you think the number of studies you gave will be distributed. You're not expected to fill in all rows and columns and you may include multiple studies of the same size and magnitude. Assume for this scenario that the missing studies have an assumed 5-year survival of 36%

Assumed 5-year survival: 36%		RD < 1 cm and 0 cm are the same i.e. HR = 1	10% less chance of mortality favouring RD < 1 cm i.e. HR = 0.9	20% less chance of mortality favouring RD < 1 cm i.e. HR = 0.8	30% less chance of mortality favouring RD < 1 cm i.e. HR = 0.7	40% less chance of mortality favouring RD < 1 cm i.e. HR = 0.6	≥ 50% less chance of mortality favouring RD < 1 cm i.e. HR ≤ 0.5
Size of studies missed that could have been included in the analysis	n=100						
	n=200						
	n=300						
	n=400						
	n=500						
	n>500						

Question 3

3. Sub-optimal RD > 1 cm versus complete cytoreduction (0 cm)

In this section, we would like you to provide us with responses to questions that allows us to adjust the overall effect estimate when data from unobserved studies are added to the final analysis. This adjustment will account for an absence (or not) of studies favouring suboptimal RD > 1 cm or ones showing no statistically significant difference between RD > 1 cm and RD 0 cm, based upon your own opinion and clinical experience in this area.

How likely is it that relevant studies reporting adequately sized analyses that did not favour complete cytoreduction (RD to 0 cm) when compared to RD > 1 cm would not have been identified from the searches and therefore omitted from the meta-analysis? By this we mean how likely is it

that studies that either favoured RD > 1 cm or studies that found no statistically significant difference (p > 0.05) in survival between RD 0 cm and RD > 1 cm) would not be published?

- Studies reporting statistically significant prolonged survival in favour of RD GREATER THAN 1 cm (that is the effect size in the form of a hazard ratio is less than 1 and the upper 95% confidence interval does not cross 1)

OR

- Studies that reported no statistically significant difference in survival between RD GREATER THAN 1 cm and 0 cm (that is the 95% confidence interval, reporting lower and upper estimates of hazard ratio, crosses 1)

Thinking about your response the question above and giving a realistic answer based on your own experience and awareness of previous analyses in this area, how many studies in total do you think will have been missed that should have been included?

Please give a brief reason for your answer

Please indicate in the table below where you think the number of studies you gave will be distributed. You're not expected to fill in all rows and columns and you may include multiple studies of the same size and magnitude. Assume for this scenario that the missing studies have an assumed 5-year survival of 36%

Assumed 5-year survival: 36%		RD > 1 cm and 0 cm are the same i.e. HR = 1	10% less chance of mortality favouring RD > 1cm i.e. HR = 0.9	20% less chance of mortality favouring RD > 1 cm i.e. HR = 0.8	30% less chance of mortality favouring RD > 1 cm i.e. HR = 0.7	40% less chance of mortality favouring RD > 1 cm i.e. HR = 0.6	≥ 50% less chance of mortality favouring RD > 1 cm i.e. HR ≤ 0.5
Size of studies missed that could have been included in the analysis	n=100						
	n=200						
	n=300						
	n=400						
	n=500						
	n>500						

Question 4

4. Sub-optimal RD > 2 cm versus complete cytoreduction (0 cm)

In this section, we would like you to provide us with responses to questions that allows us to adjust the overall effect estimate when data from unobserved studies are added to the final analysis. This adjustment will account for an absence (or not) of studies favouring suboptimal RD > 2 cm or ones showing no statistically significant difference between RD > 1 cm and RD 0 cm, based upon your own opinion and clinical experience in this area.

How likely is it that relevant studies reporting adequately sized analyses that did not favour complete cytoreduction (RD to 0 cm) when compared to RD > 2 cm would not have been identified

from the searches and therefore omitted from the meta-analysis? By this we mean how likely is it that studies that either favoured RD > 2 cm or studies that found no statistically significant difference (p > 0.05) in survival between RD 0 cm and RD > 2 cm) would not be published?

- Studies reporting statistically significant prolonged survival in favour of RD GREATER THAN 2 cm (that is the effect size in the form of a hazard ratio is less than 1 and the upper 95% confidence interval does not cross 1)

OR

- Studies that reported no statistically significant difference in survival between RD GREATER THAN 2 cm and 0 cm (that is the 95% confidence interval, reporting lower and upper estimates of hazard ratio, crosses 1)

Thinking about your response the question above and giving a realistic answer based on your own experience and awareness of previous analyses in this area, how many studies in total do you think will have been missed that should have been included?

Please give a brief reason for your answer

Please indicate in the table below where you think the number of studies you gave will be distributed. You're not expected to fill in all rows and columns and you may include multiple studies of the same size and magnitude. Assume for this scenario that the missing studies have an assumed 5-year survival of 36%

Assumed 5-year survival: 36%		RD > 2 cm and 0 cm are the same i.e. HR = 1	10% less chance of mortality favouring RD > 2 cm i.e. HR = 0.9	20% less chance of mortality favouring RD > 2 cm i.e. HR = 0.8	30% less chance of mortality favouring RD > 2 cm i.e. HR = 0.7	40% less chance of mortality favouring RD > 2 cm i.e. HR = 0.6	≥ 50% less chance of mortality favouring RD > 2 cm i.e. HR ≤ 0.5
Size of studies missed that could have been included in the analysis	n=100						
	n=200						
	n=300						
	n=400						
	n=500						
	n>500						

Part C

Question 5

In a meta-analysis including non-randomised studies, often only univariate results are reported with no attempt made to adjust for potentially important baseline imbalances. This risks making the results biased.

On a scale of 0-100, to what extent do you think that the reason study authors only report univariate analyses is to maximise the magnitude in effect estimates to favour either an experimental or comparator group?

Question 6

In your opinion, how many attempted submissions should you make to journals to publish the results of your study?

Question 7

In your opinion, how many attempted submissions should you make to journals to publish the results of your study if it is not statistically significant ($p > 0.05$)?

Question 8

What is lowest impact factor in a journal that you would consider submission of your work, regardless of the significance of your results?

Question 9

To what extent do you think it is important to publish the results of a study even if the impact factor of the accepting journal is perceived to be very low?

Acknowledgment

(Optional) If you would like to be acknowledged for your contribution to the survey, kindly leave your name.

You may choose to provide your full name (e.g., Sam Smith) or an abbreviation (e.g., S Smith).

Data on your name will be kept separate from the data file containing the survey results so that your personal information cannot be traced back to your responses. Your name will only be used for purposes of acknowledgment and will not be used in analysis.

Which of the following mediums do you consent to being acknowledged in? You may choose all that apply.

Appendix 9: Publication 6: Residual disease after primary surgery for advanced epithelial ovarian cancer, Part 2: Network meta-analysis incorporating results of expert elicitation to adjust for publication bias

American Journal of Therapeutics 30, e56–e71 (2023)

OPEN

Residual Disease After Primary Surgical Treatment for Advanced Epithelial Ovarian Cancer, Part 2: Network Meta-analysis Incorporating Expert Elicitation to Adjust for Publication Bias

Andrew Bryant,[1]* Michael Grayling,[1] Ahmed Elattar,[2] Ketankumar Gajjar,[3] Dawn Craig,[1] Luke Vale,[1] and Raj Naik[4]

Background: Previous work has identified a strong association between the achievements of macroscopic cytoreduction and improved overall survival (OS) after primary surgical treatment of advanced epithelial ovarian cancer. Despite the use of contemporary methodology, resulting in the most comprehensive currently available evidence to date in this area, opponents remain skeptical.

Areas of Uncertainty: We aimed to conduct sensitivity analyses to adjust for potential publication bias, to confirm or refute existing conclusions and recommendations, leveraging elicitation to incorporate expert opinion. We recommend our approach as an exemplar that should be adopted in other areas of research.

Data Sources: We conducted random-effects network meta-analyses in frequentist and Bayesian (using Markov Chain Montel Carlo simulation) frameworks comparing OS across residual disease thresholds in women with advanced epithelial ovarian cancer after primary cytoreductive surgery. Elicitation methods among experts in gynecology were used to derive priors for an extension to a previously reported Copas selection model and a novel approach using effect estimates calculated from the elicitation exercise, to attempt to adjust for publication bias and increase confidence in the certainty of the evidence.

Therapeutic Advances: Analyses using data from 25 studies (n = 20,927 women) all showed the prognostic importance of complete cytoreduction (0 cm) in both frameworks. Experts accepted publication bias was likely, but after adjustment for their opinions, published results overpowered the informative priors incorporated into the Bayesian sensitivity analyses. Effect estimates were attenuated but conclusions were robust in all analyses.

Conclusions: There remains a strong association between the achievement of complete cytoreduction and improved OS even after adjustment for publication bias using strong informative priors formed from an expert elicitation exercise. The concepts of the elicitation survey should be strongly considered for utilization in other meta-analyses.

Keywords: advanced epithelial ovarian cancer, residual disease, expert elicitation, publication bias, Bayesian network meta-analysis

[1]*Population Health Sciences Institute, Newcastle University, Newcastle Upon Tyne, United Kingdom;* [2]*Pan-Birmingham Gynaecological Oncology Cancer Centre, Birmingham, United Kingdom;* [3]*Nottingham City Hospital, Obstetrics and Gynaecology, Nottingham, United Kingdom; and* [4]*Northern Gynaecological Oncology Centre, Gateshead, United Kingdom.*

The authors have no conflicts of interest to declare.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.americantherapeutics.com).

A. Elattar, K. Gajjar, and R. Naik provided clinical expertise and contributed to the discussion sections of the paper; A. Bryant conceptualized the elicitation exercise and adjustment applied in Part B of the exercise and drafted the methodological, results and discussion sections of the paper. M. Grayling provided statistical expertise and applied and critically reviewed all aspects of methodology in the paper. D. Craig and L. Vale provided expert guidance and critical review of the paper. All authors agreed the final version.

**Address for correspondence: Population Health Sciences Institute, Newcastle University, 4th Floor, Idley Building 1, Queen Victoria Road, Newcastle upon Tyne NE1 7RU. E-mail: andy.bryant@ncl.ac.uk*

INTRODUCTION

Ovarian cancers remain a major concern to women worldwide.[1,2] In advanced disease, surgery and platinum-based chemotherapy are the standard treatment options. Traditionally, this included upfront primary debulking surgery (PDS) which is performed to remove as much visible disease as possible. This is because the amount of residual tumor is one of the most important prognostic factors for survival of epithelial ovarian cancer (EOC).[1] Chemotherapy followed by interval debulking surgery is an alternative primary treatment option for women diagnosed with advanced ovarian cancer, and evidence in this area is emerging. We focus our research on PDS because there is an established evidence base in which to apply our suggested methodology. A more extensive description of the aims of primary surgery in achieving "optimal cytoreduction" has been described in previous publications.[1,2] Evidence suggests that where there is "complete cytoreduction" (surgery that completely removes all visible tumour), survival is significantly improved compared with less-than-complete cytoreduction.[3–5] However, publication bias[6] leaves room for uncertainty as to the true value of complete cytoreduction, and opponents to the approach have raised concerns regarding the strength of the evidence base.

The Gynecological Cancer InterGroup defined "optimal" cytoreduction as having no macroscopic residual disease which is often reported in the literature as RD0 (residual disease (RD) = 0 cm), near-optimal RD (<1 cm), and suboptimal RD (>1 cm).[7] Although there is now less controversy about the prognostic importance of maximum cytoreduction, there remains divided opinion about the effects of any remaining RD after PDS and about what attempts should be made for maximal efforts at debulking. Different philosophies are evident within the surgical community, but there are also other important considerations, such as surgical skills, training, the woman's fitness for more radical treatment, morbidity, mortality, and quality of life. These are all considerations when assessing publication bias and the reliability of the effect estimates in published studies. There is also the issue about unreported studies that show "negative" results, which in this context may be a study showing no benefit of complete cytoreduction.

Indeed, publication bias is a well-known threat to the validity of meta-analyses.[6,8] Negative or statistically insignificant findings typically have less chance of being published; therefore, available studies tend to be a biased sample. This leads to an inflation of effect size estimates of unknown degree.[9] Consequently, it has been argued that attempting to correct for bias is typically better than incorporating no correction at all because publication bias is inevitable in most meta-analyses. This includes when no publication biases are detected, as available tests to ascertain the presence of publication bias typically have low power.[10] Ultimately, using adequate methods of bias correction can add confidence to the certainty of effect estimates in a meta-analysis.

Accordingly, this research had two main aims. First, to compare the results of a Bayesian network meta-analysis (NMA) using a noninformative prior[11] with ones attempting to adjust for selective reporting of outcomes and publication bias[6,12] by using expert elicitation methodology.[13–15] Elicitation was conducted using expert members of the British Gynaecological Cancer Society (BGCS). The adjustment for publication bias is a key component to this research because many skeptics refute conclusions in this area despite sound methodology being applied previously.[1,2,16–19] The use of novel NMA methodology in this area has been previously deployed,[20,21] but it is important to disseminate findings to the wider surgical community and not just proponents of aggressive surgery. This is only achievable by reporting effect estimates that are more likely to be closer to the true effects by removing a degree of bias. Secondly, and of paramount importance, is to encourage the use of this methodology in other areas of research, particularly where the magnitude of effects are disputed, affecting the certainty of the evidence. We promote the use of our methodology throughout the article and encourage others to attempt to implement the methods in their own research.

Specifically, to assess the potential effects of publication bias, we implement a modified version of the selection model described by Mavridis et al[22] (see also Chootrakool et al[23] and Mavridis et al).[24] This approach extends the popular Copas selection model for a conventional two-group meta-analysis[25–27] to the general NMA setting. It is, particularly, dependent on the specification of probabilities for the chance that "small" and "large" studies would be published, which we nominate using the results of an expert elicitation exercise. Although a small number of studies have previously performed this type of analysis in a NMA, they have specified these parameters somewhat arbitrarily (e.g., to reflect perceived levels of "low" and "high" publication bias). We are unaware of any previous work that has elicited these key parameters from experts.

We also use an alternative approach to adjusting for publication bias in a NMA, which to the best of our knowledge has not been considered previously, which leverages informative priors in an otherwise

conventional Bayesian NMA. In our case, the informative priors are formed based on the opinion of expert members of the BGCS.[28,29] We believe this approach would be easy to mimic for all oncology settings that use survival outcomes where an estimate of the control arm event rate can be reliably estimated.

METHODS

Search strategy and selection criteria

The NMAs reported in this article synthesized studies according to good research principles following the methods outlined by Bryant et al.[2] Bibliographic databases were searched from 1950 up to September 2021 (results of search are shown in Figure 1). We applied the same search and inclusion criteria as outlined by Bryant et al.[2] The population of interest was women who had received primary cytoreductive surgery followed by adjuvant platinum-based chemotherapy.[1] Included studies reported overall survival (OS) for comparisons of RD thresholds after surgery and used the same statistical adjustment constraints by Bryant et al[2] to minimize selection bias.[20,30] We sifted references identified from the search, extracted data on pertinent items, and assessed risk of bias in accordance with the Cochrane guidelines,[20] following on from the systematic review that underpins this analysis and the subsequent frequentist NMA.[1,2]

Expert elicitation exercise and statistical considerations

An expert elicitation exercise[31] was sent to members of the BGCS by the organizing committee. The elicitation exercise was conducted before the completion of the systematic review,[1] and the findings from this exercise used to adjust the meta-analyses for perceived publication bias. In the elicitation exercise, we asked participants to account for the sort of studies that have been conducted but not published, the plausible magnitude and direction of any publication bias and possible explanations for why and how the publication bias occurs. The survey consisted of two main parts, part A and part B, and is given in Supplementary Material. The results were used to perform the sensitivity analyses adjusting for publication bias, as described further below.

Data set and notation

The impact on OS of optimal and suboptimal cytoreduction for primary advanced disease was assessed using several RD thresholds that have been reported in the literature. Accordingly, our data set consists of the results of n studies, comparing a total of T RD thresholds (or arms; labeled 1,2,..). We use the terms, arms and RD thresholds interchangeably for the benefit of those mimicking our methods because it is likely that they will be applying the methodology to study arms in an RCT setting. We use the term design to refer to the set of RD thresholds compared in a given study, that is, a design is some subset of at least 2 RD thresholds in the network. Let $d = 1, \ldots, D$ index the designs used in our network, and n_d be the number of studies included in the network that used the dth design. Set also T_d as the number of RD thresholds in design d. Then, we have designs with T_d = 2,3,4. The designs in our data set are presented in Figure 2; we have n = 28, $T = 9$, and $D = 8$.

From a study of design d, the information used is: (a) $T_d - 1$ estimated effects (log hazard ratios, in our case) and their standard errors and; (b) $(T_d - 1)(T_d - 2)/2$ correlations between the $T_d - 1$ effects. We use subscript indices to identify this study and its design and superscript indices to denote the contrast being evaluated such that $y_{i,d}^{(a,b)}$ refers to the effect size for the ab comparison (where a and b are 2 RD thresholds) in the ith study that has the dth design. Similarly, we let $s_{i,d}^{(a,b)}$ denote the corresponding observed standard error (SE). Our data set, in this notation, is available on request.

Part A: Copas model approach

Part A of the elicitation exercise asked clinicians about their perceived probability of publication of individual studies relating to the standard error of their effect sizes. Part A was conducted to facilitate the conduct of a previously proposed method of adjusting for publication bias in a NMA.[24] We now describe this methodology.

Measurement model

Each observed effect $y_{i,d}^{(a,b)}$ in a two-threshold study (i.e., any study with $T_d = 2$) is modeled as a normal distribution:

$$y_{i,d}^{(a,b)} \sim N\left(\theta_{i,d}^{(a,b)}, \left(s_{i,d}^{(a,b)}\right)^2\right).$$

A random-effects model is assumed because publication bias is confounded with heterogeneity. Thus, it is assumed that the mean relative treatment effect is modeled as $\theta_{i,d}^{(a,b)} = \lambda^{(a,b)} + \delta_{i,d}^{(a,b)}$, where the random effects $\delta_{i,d}^{(a,b)}$ are normally distributed as $\delta_{i,d}^{(a,b)} \sim N(0, \tau^2)$.

In multithreshold study i of design d, the vector of $T_d - 1$ contrasts is modeled as a multivariate normal distribution. For example, if arms a, b, and c are included, then:

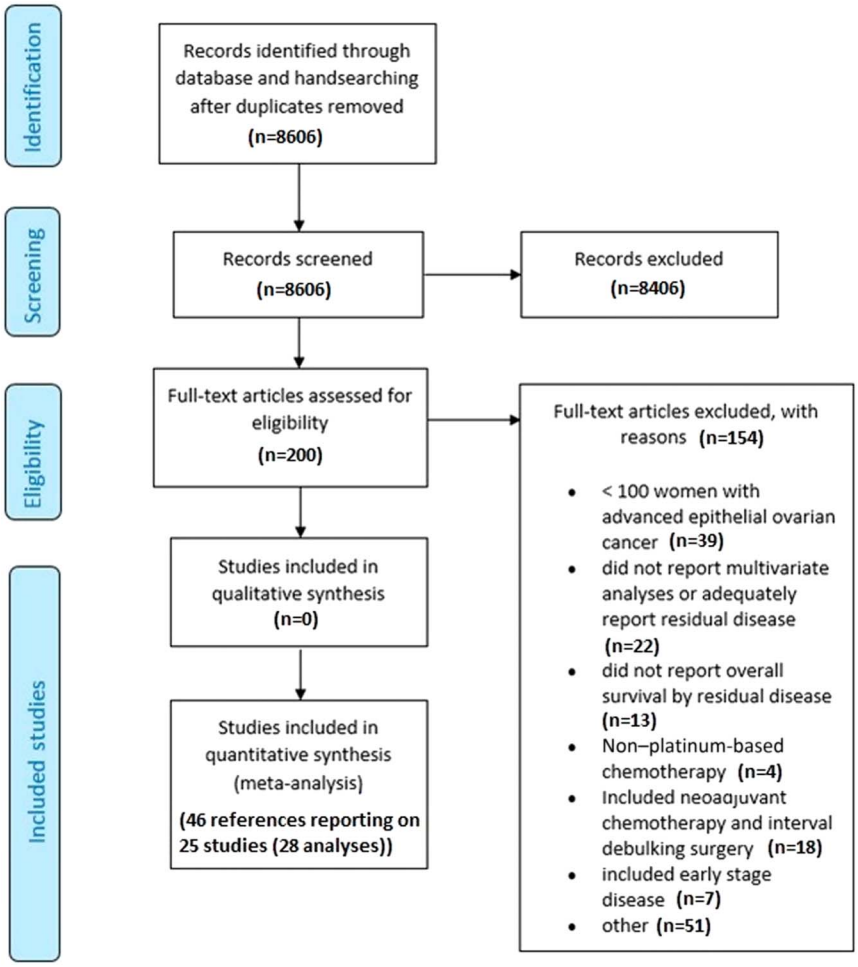
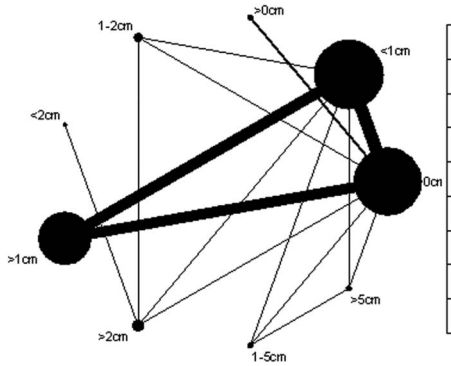


FIGURE 1. PRISMA flowchart.



Design (d)	Arms	T_d	n_d
1	1, 2, 4, 7	4	1
2	1, 2, 8, 9	4	1
3	1, 2, 6	3	15
4	2, 4, 7	3	1
5	1, 2	2	2
6	1, 3	2	4
7	2, 6	2	2
8	5, 7	2	2

FIGURE 2. Network diagram and summary of designs showing RD comparisons after primary cytoreductive surgery for advanced EOC. Arms 1–9 correspond to the following categories: 1 (0 cm), 2 (<1 cm), 3 (>0 cm), 4 (1–2cm), 5 (<2 cm), 6 (>1 cm), 7 (>2 cm), 8 (1–5cm), and 9 (>5 cm).

$$\begin{matrix} n y_{i,d}^{(a,b)} \\ n y_{i,d}^{(a,c)} \end{matrix} \Big| \begin{matrix} n\theta_{i,d}^{(a,b)} \\ n\theta_{i,d}^{(a,b)} \end{matrix} \Bigg) \sim N \Bigg(\begin{pmatrix} n\lambda^{(a,b)} \\ n\lambda^{(a,c)} \end{pmatrix} + \begin{matrix} n\delta_{i,d}^{(a,b)} \\ n\delta_{i,d}^{(a,c)} \end{matrix}, \begin{pmatrix} \left(s_{i,d}^{(a,b)}\right)^2 & cov\left(y_{i,d}^{(a,b)}, y_{i,d}^{(a,c)}\right) \\ cov\left(y_{i,d}^{(ab)}, y_{i,d}^{(a,c)}\right) & \left(s_{i,d}^{(a,c)}\right)^2 \end{pmatrix} \Bigg).$$

Assuming a common heterogeneity parameter across treatment comparisons, the random effects are:

$$\begin{matrix} n\delta_{i,d}^{(a,b)} \\ n\delta_{i,d}^{(a,c)} \end{matrix} \Bigg) \sim N \Bigg(\begin{pmatrix} n0 \\ n0 \end{pmatrix}, \tau^2 \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \Bigg).$$

Similarly, if arms *a*, *b*, *c*, and *e* are included, then:

$$\begin{pmatrix} n y_{i,d}^{(a,b)} & n\theta_{i,d}^{(a,b)} \\ n y_{i,d}^{(a,c)} & | \, n\theta_{i,d}^{(a,c)} \\ n y_{i,d}^{(a,e)} & n\theta_{i,d}^{(a,e)} \end{pmatrix} \sim N \Bigg(\begin{pmatrix} n\lambda^{(a,b)} \\ n\lambda^{(a,c)} \\ n\lambda^{(a,e)} \end{pmatrix} + \begin{pmatrix} n\delta_{i,d}^{(a,b)} \\ n\delta_{i,d}^{(a,c)} \\ n\delta_{i,d}^{(a,e)} \end{pmatrix}, \begin{pmatrix} \left(s_{i,d}^{(a,b)}\right)^2 & cov\left(y_{i,d}^{(a,b)}, y_{i,d}^{(a,c)}\right) & cov\left(y_{i,d}^{(a,b)}, y_{i,d}^{(a,e)}\right) \\ cov\left(y_{i,d}^{(a,b)}, y_{i,d}^{(a,c)}\right) & \left(s_{i,d}^{(a,c)}\right)^2 & cov\left(y_{i,d}^{(a,c)}, y_{i,d}^{(a,e)}\right) \\ cov\left(y_{i,d}^{(a,b)}, y_{i,d}^{(a,e)}\right) & cov\left(y_{i,d}^{(a,c)}, y_{i,d}^{(a,e)}\right) & \left(s_{i,d}^{(a,e)}\right)^2 \end{pmatrix} \Bigg),$$

$$\begin{pmatrix} n\delta_{i,d}^{(a,b)} \\ n\delta_{i,d}^{(a,c)} \\ n\delta_{i,d}^{(a,e)} \end{pmatrix} \sim N \Bigg(\begin{pmatrix} n0 \\ n0 \\ n0 \end{pmatrix}, \tau^2 \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix} \Bigg)$$

We assume a common between-study variance (heterogeneity) τ^2 across treatment comparisons; although arguably not realistic, this is common and often necessary in practice (given there are few studies per comparison).

Selection model

To model the probability each study is selected for publication, we assume that there is a latent variable underlying each study. This latent variable takes positive values if the specific study is published and negative values otherwise. Thus, there are as many latent variables as study designs, and each latent variable represents the propensity for publication given the design of that specific study. The propensity for publication for each design and study is denoted by $z_{i,d}$. It is modeled as a function of two parameters, α_d and β_d, and a function, $f(i,d)$, of the particular study and its design:

$$z_{i,d} = \alpha_d + \frac{\beta_d}{f(i,d)} = u_{i,d} + \xi_{i,d}.$$

Here, $\xi_{i,d} \sim N(0,1)$, and we constrain $\beta_d \geq 0$ because this will reflect the belief that larger studies are more likely to be published. In a two-threshold study i of design d that compares a and b, we set $f(i,d) = s_{i,d}^{(a,b)}$. Following Chootrakool et al (2011)[23] for a multithreshold trial, we use the average of the standard errors in this study. For example, in study i of design d that compares a, b, c, and e, we set:

$$f(i,d) = \frac{s_{i,d}^{(a,b)} + s_{i,d}^{(a,c)} + s_{i,d}^{(a,e)}}{3}.$$

With the above, the probability that study i with design d is published is equal to:

$$\mathbb{P}(z_{i,d} > 0) = \Phi\left(\alpha_d + \frac{\beta_d}{f(i,d)}\right) = \Phi(u_{i,d}).$$

This provides us with an interpretation of the parameters α_d and β_d. Informally, parameter α_d is the marginal probability that a study with design d is published, assuming it has infinite variance (not accounting for the approach taken to multithreshold studies). Parameter β_d is a discrimination parameter, discriminating the probabilities of publication between studies with difference variances.

Combined measurement and selection model

The measurement and selection models do not share common parameters but are connected through their residual terms. Specifically, we set $corr(y_{i,d}^{(ab)}, z_{i,d}) = \rho_d$ such that ρ_d controls how the effect size affects the probability of the study being published. Then, for two-threshold (thresholds *a* and *b*), three-threshold (thresholds *a*, *b*, and *c*), and four-threshold (thresholds *a*, *b*, *c*, and *e*) studies, the joint distribution for its effect sizes and propensity for publication are as follows:

$$\begin{pmatrix} n y_{i,d}^{(a,b)} \\ n z_{i,d} \end{pmatrix} \sim N \Bigg(\begin{pmatrix} n\theta_{i,d}^{(a,b)} \\ n u_{i,d} \end{pmatrix}, \begin{pmatrix} \left(s_{i,d}^{(a,b)}\right)^2 & \rho_d s_{i,d}^{(a,b)} \\ \rho_d s_{i,d}^{(a,b)} & 1 \end{pmatrix} \Bigg) I_{z_{i,d}>0},$$

$$\begin{pmatrix} n y_{i,d}^{(a,b)} \\ n y_{i,d}^{(a,c)} \\ n z_{i,d} \end{pmatrix} \sim N \Bigg(\begin{pmatrix} n\theta_{i,d}^{(a,b)} \\ n\theta_{i,d}^{(a,c)} \\ n u_{i,d} \end{pmatrix}, \begin{pmatrix} \left(s_{i,d}^{(a,b)}\right)^2 & cov\left(y_{i,d}^{(a,b)}, y_{i,d}^{(a,c)}\right) & \rho_d s_{i,d}^{(a,b)} \\ cov\left(y_{i,d}^{(a,b)}, y_{i,d}^{(a,c)}\right) & \left(s_{i,d}^{(a,c)}\right)^2 & \rho_d s_{i,d}^{(a,c)} \\ \rho_d s_{i,d}^{(a,b)} & \rho_d s_{i,d}^{(a,c)} & 1 \end{pmatrix} \Bigg) I_{z_{i,d}>0},$$

where I_X is the indicator variable for event X.

$$\begin{pmatrix} ny_{i,d}^{(a,b)} \\ ny_{i,d}^{(a,c)} \\ ny_{i,d}^{(a,e)} \\ nz_{i,d} \end{pmatrix} \sim N\left(\begin{pmatrix} n\theta_{i,d}^{(a,b)} \\ n\theta_{i,d}^{(a,c)} \\ n\theta_{i,d}^{(a,e)} \\ nu_{i,d} \end{pmatrix}, \begin{pmatrix} \left(s_{i,d}^{(a,b)}\right)^2 & cov\left(y_{i,d}^{(a,b)}, y_{i,d}^{(a,c)}\right) & cov\left(y_{i,d}^{(a,b)}, y_{i,d}^{(a,e)}\right) & \rho_d s_{i,d}^{(a,b)} \\ cov\left(y_{i,d}^{(a,b)}, y_{i,d}^{(a,c)}\right) & \left(s_{i,d}^{(a,c)}\right)^2 & cov\left(y_{i,d}^{(a,c)}, y_{i,d}^{(a,e)}\right) & \rho_d s_{i,d}^{(a,c)} \\ cov\left(y_{i,d}^{(a,b)}, y_{i,d}^{(a,e)}\right) & cov\left(y_{i,d}^{(a,c)}, y_{i,d}^{(a,e)}\right) & \left(s_{i,d}^{(a,e)}\right)^2 & \rho_d s_{i,d}^{(a,e)} \\ \rho_d s_{i,d}^{(a,b)} & \rho_d s_{i,d}^{(a,c)} & \rho_d s_{i,d}^{(a,e)} & 1 \end{pmatrix} \right) I_{z_{i,d}>0},$$

Prior distributions for model selection parameters

To fit the above model, prior distributions for the selection model parameters α_d and β_d are required. To form the priors, we need to specify lower and upper bounds, P_d^{low} and P_d^{high}, for the probability that a study of design d is published, where these extremes relate to small and large possible values of $f(i,d)$. P_d^{low} and P_d^{high} are modeled as random variables to reflect the uncertainty around them. Then, α_d and β_d are calculated using the inequalities:

$$P_d^{low} \leq \mathbb{P}(z_{id}>0|f(i,d)) \leq P_d^{high} \forall\, d.$$

Specifically, this gives:

$$\alpha_d + \frac{\beta_d}{\max\{f(i,d)\}} = \Phi^{-1}\left(P_d^{low}\right),$$

$$\alpha_d + \frac{\beta_d}{\min\{f(i,d)\}} = \Phi^{-1}\left(P_d^{high}\right).$$

Unlike Mavridis et al,[22] rather than setting $\min\{f(i,d)\}$ and $\max\{f(i,d)\}$ as the observed minimal and maximal values in the data set, we use the results of an elicitation exercise in which we asked experts about the probability studies of certain sizes would be published. For the population under investigation, we describe below why $\mathrm{SE}(\log HR) = \sqrt{6.25/n}$ is a reasonable assumption. Using this, on plugging in the minimal and maximal sample sizes from the elicitation exercise, the formulae for $f(i,d)$ gives:

$$\min\{f(i,d)\} = \begin{cases} 0.1 & : T_d = 2, \\ 0.122 & : T_d = 3, \\ 0.141 & : T_d = 4, \end{cases}$$

$$\max\{f(i,d)\} = \begin{cases} 0.25 & : T_d = 2, \\ 0.306 & : T_d = 3, \\ 0.354 & : T_d = 4. \end{cases}$$

All that then remains is to specify prior distributions $P_d^{low} \sim U(L_{1,d}, L_{2,d})$ and $P_d^{high} \sim U(U_{1,d}, U_{2,d})$. For those two-threshold designs that include the 0 cm arm, we are able to directly use the results from Part A of the survey. For those studies that did not contain the 0 cm arm, we calculate, similarly, swapping in their reference category for 0 cm (e.g., the probability of publication of <1 cm versus >2 cm would be taken as the elicited values for 0 cm versus >2 cm); the results are unlikely to be sensitive to this assumption because the number of studies that do not contain the reference category is small. For multi-threshold studies, we conservatively use the minimum probabilities across the various pairwise comparisons. We then consider three combinations of values for $L_{1,d}$, $L_{2,d}$, $U_{1,d}$, and $U_{2,d}$. We take them as the 0th and 50th (median) percentiles, the 25th (lower quartile) and 75th (upper quartile) percentiles, and the 50th and 100th percentiles of the elicited probabilities (with $L_{1,d}$ and $L_{2,d}$ set using the results for the smallest trial size we asked experts about and $U_{1,d}$ and $U_{2,d}$ set using the results for the largest trial size we asked experts about). We denote the elicited q^{th} percentile for the small study size by $P_{s,d,q}$ and similarly, $P_{l,d,q}$ for the large. The percentiles are then presented in Table 2.

Part B: alternative novel approach

Part B involved an alternative approach that asked clinicians to estimate the number of studies for key comparisons that they believed would be conducted but unpublished, and thus unidentified in the NMA. They were then subsequently asked to specify sample and effect sizes for each such missing study. The approach in Part B is a particularly novel aspect of this research because it can be used as prior information to inform adjustment of meta-analyses for publication bias in a way we believe to be previously unexplored. Here, we outline how this could be achieved.

We note that this is only one potential way to form a prior based on the elicited data and that a sensitivity analysis should certainly be conducted. For example, in our elicitation exercise, the choice of the number of miss-

ing studies was left open ended as to not lead experts to a choice and bias the results. Consequently, a sensitivity analysis could be conducted removing high estimates of unpublished studies if it was judged that unrealistic entries were unduly inflating an average.

Given an assumed 5-year survival rate of 36%[32–34] and a minimum sample size of n = 100 to meet the criteria for inclusion in the NMA, small studies might be underpowered and, furthermore, null findings might be due to deficiencies in the study design and conduct. Hence, including these studies might not lead to an appropriate adjustment of meta-analysis estimates. This is why we included studies with a minimum sample size of 100 patients in the systematic review,[1] and a minimum 64 events (deaths, d) are expected with 36 participants being alive and censored at the end of this study:

$$d = 100(1 - 0.36) = 64.$$

Generalizing this result, we assume that d can be related to n in general through the following formula:

$$n = \frac{d}{1 - (5 \text{ year suvival rate})} = \frac{d}{0.64}.$$

The standard error of the log hazard ratio $(\mathrm{SE}(\log HR))$ can then be related to n by rearranging the following formula:

$$d = \frac{4}{\mathrm{SE}(\log HR)^2},$$

$$\Rightarrow \mathrm{SE}(\log HR) = \sqrt{\frac{4}{d}} = \sqrt{\frac{4}{0.64n}} = \sqrt{\frac{6.25}{n}}.$$

Next, we denote by m_{cij} the number of missing studies according to expert responder $c = 1, \ldots, C$, with a HR of HR_j and a sample size of n_i, where:

$$n_1 = 100, n_2 = 200, n_3 = 300, n_4 = 400, n_5 = 500, n_6 = 625,$$

$$HR_1 = 1, HR_2 = 0.9, HR_3 = 0.8, HR_4 = 0.7, HR_5 = 0.6, HR_6 = 0.5.$$

We compute the average number of missing studies of type ij, across the responders, as:

$$m_{ij} = \frac{1}{C} \sum_{c=1}^{C} m_{cij}.$$

We use this to form an average sample size of missing studies with a HR of HR_j through:

$$m_j = \frac{\sum_i n_i m_{ij}}{\sum_i m_{ij}}.$$

With this, we assume that information from missing studies with a HR of HR_j can be categorized through the following distribution:

$$P_j \sim N\left(\log HR_j, \frac{6.25}{m_j}\right).$$

The P_j can then be combined in a weighted manner, giving more weight to those values of j with a larger value of m_j, through conflation. This gives an elicited prior of:

$$P \sim N\left(\frac{\sum_j \frac{m_j \log HR_j}{6.25}}{\sum_j \frac{m_j}{6.25}}, \frac{1}{\sum_j \frac{m_j}{6.25}}\right) = N\left(\frac{\sum_j m_j \log HR_j}{\sum_j m_j}, \frac{6.25}{\sum_j m_j}\right).$$

This elicited estimate can then be used as prior information and be applied in a Bayesian analysis[35–37] that reflects the results of the expert opinion in the elicitation exercise.[22,38]

Data analysis

We compare the results of a frequentist approach[2] with a NMA conducted within a Bayesian framework in WinBUGS 1.4.3 (MRC Biostatistics Unit, Cambridge, UK),[39,40] using two chains each with 100,000 simulations and a burn-in period of 30,000 simulations. The base case Bayesian analysis (analogous to the frequentist analysis) used vague noninformative priors and adjusted for multiarm trials using conditional distributions. Figure 2 shows a network diagram[41] of the thresholds (nodes) and comparisons (lines) available and a summary of designs in our network. Convergence of the model in the two chains was assessed using Brooks–Gelman–Rubin, trace and autocorrelation plots.[40]

Transitivity and design inconsistency were not deemed an issue because of restrictive inclusion criteria.[2] Consistency, which is measured in agreement of direct and indirect evidence, was assessed by comparing the individual data point's posterior mean deviance contributions for the consistency and inconsistency model.[42–44] Owing to the volume of sensitivity analyses, we did not conduct any further node splitting[42–44] because this was previously performed in the base case analysis.[2]

We present the results of the Bayesian NMA of optimal RD thresholds using effect sizes reported as posterior median HRs and 95% credible intervals (CrIs). All the thresholds are relative to the 0 cm macroscopic

RD reference threshold. We also present rankograms, which ranked RD thresholds from having highest probability of survival (ranked 1) to the lowest (ranked 9). In addition, we report the probability of being the best RD threshold and the surface under the cumulative ranking curves (SUCRAs).[45]

Sensitivity analyses (SA) form the crucial basis of this research; we perform a number of analyses that attempt to adjust the base case estimates for publication bias. We a priori focus on macroscopic RD to 0 cm, RD <1 cm, and suboptimal RD >1 cm. Other thresholds will add strength to the network but are not of direct interest. We use this approach in a complex situation that includes multiple RD thresholds (arms) and studies that included multiple thresholds (up to four in a study). In practice, it should be more straightforward following and applying the methodology to other analyses in different areas that have simpler networks and in a conventional intervention setting.

We repeated the base case Bayesian analysis above and used the elicitation exercise to use the Copas selection model (part A) and incorporate informative priors (part B) in place of the vague (noninformative) ones. For those wanting to restrict to a frequentist setting, in Part B, an analogous analysis in the frequentist framework is possible by including the elicited missing studies from the average experts' responses (artificially) in the observed studies in the NMA. However, we recommend applying the proposed Bayesian methods and formulating priors.

RESULTS

Summary of studies

The flow of the literature is presented in the PRISMA flowchart (Figure 1). The search strategy identified 8606 unique references, of which 200 progressed to full-text screening. Forty-six references, reporting on 25 primary studies which included 28 analyses (n = 20,927), met our inclusion criteria. Full details of searches along with a PRISMA flowchart, characteristics of included studies, and risk of bias assessments are provided by Bryant et al.[2]

The network diagram[41] and summary of designs in our network show the range of RD threshold comparisons after optimal cytoreductive surgery for advanced EOC (Figure 2). The most common RD thresholds were complete (0 cm) and near-optimal (<1 cm), while this was also the most widely reported comparison.

Base case analysis

The results of the base case Bayesian NMA were consistent with the frequentist analysis, and there was also no evidence of inconsistency in the network.[2] There were no issues with model convergence in WinBUGS,[39] as indicated by Brooks–Gelman–Rubin, trace and autocorrelation plots, with the number of simulations used adequate (see Appendix Figures 3–8, http://links.lww.com/AJT/A124, http://links.lww.com/AJT/A125, http://links.lww.com/AJT/A126, http://links.lww.com/AJT/A127, http://links.lww.com/AJT/A128, http://links.lww.com/AJT/A129, and http://links.lww.com/AJT/A130, respectively). The results of the base case analyses demonstrate prolonged OS if primary cytoreductive surgery achieved macroscopic RD to 0 cm compared with any other RD threshold (Table 1). Macroscopic RD to 0 cm was overwhelmingly the best ranked threshold because it was consistently ranked first (Table 1 and rankogram in Figure 9 in Supplementary Material, http://links.lww.com/AJT/A131), with a very high probability of being the best RD threshold (SUCRA and P best ranged from 98.4% to 99.9%). Sensitivity analyses using different random number seeds resulted in all Bayesian models being correct to one decimal place (data not shown). Low values in MC error terms in the model indicated reliability in estimates to good precision.[40]

Expert elicitation exercise

Eighteen expert members of the BGCS participated in the expert elicitation exercise. They were given the sample sizes (based on observed data for each RD threshold) and were asked in Part A of the elicitation exercise to state probabilities of publication for a study comparing different RD thresholds with complete cytoreduction (macroscopic RD to 0 cm). Table 2 presents the distribution of the elicited probabilities for each RD threshold, for the smallest and largest considered study sizes (full details of the expert clinician elicitation exercise are provided by Bryant et al[31]). In summary, responses suggest that publication bias may be quite likely in studies where the sample size was small. For example, the average response suggested that experts believed there was a 55% chance that a comparison of RD < 1 cm versus RD 0 cm would be reported for a study with a sample size of 100 participants. Responders seemed to indicate that the probability of publication was lowest for comparisons involving greater macroscopic disease volume [largest elicited median probability 20% (interquartile range 10–75) in macroscopic disease involving RD > 2 cm versus RD 0 cm and as low as 3.5% (interquartile range: 0–50) for RD > 5 cm vs. RD 0 cm]. However, respondents seemed to dismiss the threat of publication bias for comparisons of RD < 1 cm versus RD 0 cm and RD > 1 cm versus RD 0 cm in larger studies. Comparisons involving suboptimal

Table 1. Results of base case frequentist and Bayesian NMA of optimal RD threshold after primary cytoreductive surgery for advanced EOC.

	Frequentist				Bayesian			
RD	HR (95% CI)	Mean rank	P (best), %	SUCRA, %	HR (95% CrI)	Median rank	P (best), %	SUCRA, %
0 cm	Reference	1	99	99.9	Reference	1 (1–1)	98.42	99.8
<1 cm	1.98 (1.76–2.24)	3.4	0	70.2	1.99 (1.76–2.27)	3(2–5)	0	69.88
>0 cm	1.95 (1.48–2.58)	3.4	0	70.6	1.95 (1.46–2.63)	3(2–6)	0.005	70.43
1–2 cm	3.34 (2.04–5.47)	7.3	0	21.8	3.57 (2.14–5.99)	8(5–9)	0	18.58
<2 cm	2.82 (1.58–5.04)	6.0	0	36.9	2.89 (1.57–5.34)	7(2–8)	0.044	36.75
>1 cm	2.57 (2.26–2.93)	5.8	0	40.0	2.58 (2.26–2.97)	6(4–8)	0	40.91
>2 cm	4.36 (2.69–7.04)	8.7	0	3.4	4.47 (2.72–7.43)	9(7–9)	0	4.17
1–5cm	1.85 (1.11–3.08)	3.2	1	72.0	1.85 (1.06–3.22)	3(2–7)	1.498	71.93
>5 cm	2.75 (1.62–4.67)	6.2	0	35.3	2.75 (1.55–4.89)	6(2–9)	0.033	37.54

RD, residual disease; CI, confidence interval; P (best), probability that RD threshold is the best; CrI, credible interval; EOC, epithelial ovarian cancer; HR, hazard ratio; SUCRA, surface under the cumulative ranking curves.

RD (greater macroscopic disease volume) were considered to have a low probability of not being published for both the small and larger studies (but lower in smaller studies).

In part B of the elicitation exercise, the mean number of missing studies estimated by experts for comparison of RD < 1 cm versus RD 0 cm was 17.8. The average number of estimated missing studies was lower for the comparisons involving suboptimal macroscopic disease volume (RD thresholds that are > 1 cm).[31] In the comparison of RD < 1 cm versus RD 0 cm, on average, 9.4 of the 17.8 studies would be associated with a HR of 1. As the HR increased, fewer studies were believed to be missing such that, when the detected HR was 0.5, the average number of studies believed to be missing was less than 1 (Table 3). The weighted average HR of the effect size from the missing studies was 0.83 (95% CI 0.77–0.90) for the comparison of RD < 1 cm compared with RD 0 cm. This HR was calculated based on a total of 3906 participants in the estimated missing studies and 2500 deaths given a 5-year survival rate of 36% (Table 3). This corresponded to a log HR of −0.19 and SE log HR of 0.04; thus, we used ~N(−0.19, 0.04) as the distribution for our elicited prior for the <1 cm versus 0 cm comparison. Similarly, the mean number of missing

Table 2. The distribution of elicited probabilities for each RD threshold for the smallest and largest considered study sizes.

	Small study publication probabilities					Large study publication probabilities				
Design (*d*)	$P_{s,d,0}$	$P_{s,d,25}$	$P_{s,d,50}$	$P_{s,d,75}$	$P_{s,d,100}$	$P_{l,d,0}$	$P_{l,d,25}$	$P_{l,d,50}$	$P_{l,d,75}$	$P_{l,d,100}$
1	0	10	20	70	100	0	15	30	80	100
2	0	0	3.5	50	95	0	0	10	80	100
3	0	20	45	80	100	40	75	95	99	100
4	0	10	20	70	100	0	15	30	80	100
5	0	30	55	80	100	80	90	99.5	100	100
6	0	20	50	80	95	0	70	80	99	100
7	0	20	45	90	100	40	75	95	99	100
8	0	10	20	75	100	0	15	30	80	100

These are computed using the results of the elicitation exercise.
P,(*s,l*),*d*,(percentiles 0, 25, 50, 75, 100), probability that a small/large study is published with a specified design in a number of percentiles. Design (*d*) 1, arms 1,2,4,7; *d*(2), arms 1,2,8,9; *d*(3), arms 1,2,6; *d*(4), 2,4,7; *d*(5), arms 1,2; *d*(6), arms 1,3; *d*(7), arms 2.6; *d*(8), arms 5,7 where arms 1–9 correspond to the following categories: 1 (0 cm), 2 (<1 cm), 3 (>0 cm), 4 (1–2cm), 5 (<2 cm), 6 (>1 cm), 7 (>2 cm), 8 (1–5cm), and 9 (>5 cm).

Table 3. Breakdown of distribution of size and magnitude of elicited unpublished studies of near-optimal RD < 1 cm versus complete cytoreduction (0 cm).

N=321 (n=17.8)	Estimated effect size					
	HR = 1	HR = 0.9	HR = 0.8	HR = 0.7	HR = 0.6	HR ≤ 0.5
Assumed 5-year survival: 36%	RD <1 cm and 0 cm are the same	10% less chance of mortality favoring RD <1 cm	20% less chance of mortality favoring RD <1 cm	30% less chance of mortality favoring RD <1 cm	40% less chance of mortality favoring RD <1 cm	>=50% less chance of mortality favoring RD <1 cm
Sample size[h]						
n < 100			STUDY EXCLUDED			
n = 100	122.08[g]	19.12	22.7	1.34	2.14	1.14
n = 200	25.08	11.12	12.62	4.38	2.18	2.18
n = 300	6.04	4.04	1.04	2.04	0	0
n = 400	10.37	9.37	9.37	9.37	9.37	9.37
n = 500	1.04	1.04	3.04	1.04	0	0
n > 500	5.08	4.04	4.04	3.04	1.04	1.04
Total studies[a] (mean)	169.7 (9.4)	48.7 (2.7)	52.8 (2.9)	21.2 (1.2)	14.7 (0.8)	13.7 (0.8)
Effective *n*[b] (mean)	26,879 (1493.3)	12,141 (674.5)	12,899 (716.6)	7790 (432.8)	5048 (280.4)	4948 (274.9)
Effective *d*[c] (mean)	17,203 (956)	7770 (432)	8255 (459)	4986 (277)	3231 (179)	3167 (176)
SElogHR ($\sqrt{(4/d)}$)[d]	0.065	0.096	0.093	0.120	0.149	0.151
95% CI for HR[e]	0.88–1.14	0.75–1.09	0.67–0.96	0.55–0.89	0.45–0.80	0.37–0.67
Elicited estimate[f]	HR 0.83 (95% CI 0.77–0.90), logHR −0.19 SElogHR 0.04 (n = 3906, *d* = 2500)					
Elicited prior	~N(−0.19, 0.04)					

[a]Absolute number of estimated missing studies elicited from responders with mean (simply absolute number divided by 18 (number of responders)) given in parentheses ().
[b]Absolute number of estimated missing participants elicited based on total studies with mean given in parentheses.
[c]Absolute number of deaths estimated from the number of participants assuming a 5-year survival rate of 36% with mean in parentheses ().
[d]Approximation of the standard error (SE) of the log HR using formula derived by Parmar, namely the square root of 4 divided by the mean number of deaths.
[e]95% confidence interval for HR calculated using logHR ±1.96 multiplied by standard error of log HR then transforming back by taking the exponential.
[f]Elicited HR with 95% confidence interval using mean responses for all aggregated effect sizes.
[g]Number of studies given in the breakdown were rescaled in 3 respondents to correspond to the total number estimated. Therefore, any noninteger numbers in the table are due to this rescaling.
[h]Size of studies missed that could have been included in the analysis.

studies estimated in the elicitation exercise for comparison of RD > 1 cm versus RD 0 cm was 8.6.[31] The weighted average HR of the missing studies led to formulating ~N(−0.26, 0.05) as a prior. The mean number of missing studies estimated by responders for comparison of RD > 2 cm versus RD 0 cm was 6.2.[31] The weighted average HR of the missing studies led to formulating ~N(−0.24, 0.06) as a prior. However, there seemed to be widespread feeling among experts that publication bias was of much less concern

in suboptimal RD thresholds, and this is reflected in some of the sensitivity analyses (Table 5). A worked example surrounding derivation of priors based on these estimates is presented in Table 3 for the comparison of macroscopic RD with 0 cm and near-optimal cytoreduction to <1 cm.

Adjustment for publication bias

Tables 4 and 5 present the estimated effect sizes for RD thresholds for the sensitivity analyses incorporating an adjustment for publication bias. Models were constructed using responses from parts A and B of the expert elicitation exercise.

All analyses were based on 100,000 Markov Chain Monte Carlo simulations with a burn-in period of 30,000 draws, from two chains (as in Mavridis et al[22]). We present the median OS estimate for each RD group relative to the reference category (0 cm), along with its 95% CrI, SUCRA values, the median (and 95% CrI) rank for each group, and the estimated probability each group provides the best OS are also given.

Bayesian NMAs were fitted in a series of sensitivity analyses that used informative priors based on estimates obtained from the expert elicitation exercise (see above). We set out to explore a range of sensitivity analyses, from ones that best reflected the experts' views to more extreme scenarios that fully tested the robustness of the base case analysis presented in Table 1. Specifically, the main focus of our work was to examine the conclusions in the unadjusted analysis that identified three clear and distinct categories of RD groups after primary cytoreductive surgery, namely complete (0 cm), near-optimal (<1 cm), and suboptimal (>1 cm). Other reported RD thresholds contributed toward the network and added strength to the NMA, but clearly some comparisons such as when RD 1–2 cm is compared with macroscopic RD to 0 cm were not of paramount importance and would not necessarily be a widely reported and expected comparison. Therefore, it would not be appropriate to focus on publication bias in this example. Accordingly, sensitivity analyses focused on the main RD categories of complete RD to 0 cm, RD<1 cm, and suboptimal RD >1 cm.

Part A: Copas selection model

Table 4 presents the results of the selection model analyses. As would be expected, the introduction of increasing levels of publication bias adjustment typically results in greater reductions in the estimated OS benefit for the reference category. However, in almost all instances the results change little compared with the base case frequentist and Bayesian analyses (Table 1). The 0 cm category retains at least an 87.48% estimated chance of providing the best OS.

Part B: alternative novel approach

Sensitivity analysis (SA) 1 incorporated prior information using the estimates derived above (~N(−0.24, 0.06)) for RD <1 cm and RD > 0 cm, ~N(−0.26, 0.05) for RD >1 cm, and ~N(−0.24, 0.06) for RD>2 cm). In SA 2, informative priors were used for RD <1 cm, >0 cm, 1–2cm, <2 cm, and >1 cm and only RD < 1 cm and >0 cm in SA 3. SA 4 used informative priors for RD <1 cm, >1 cm, and >2 cm, and SA 5 incorporated priors for all RD thresholds. SA 5 was thus the most extreme sensitivity analysis.

SA 6 and 7 grouped RD < 2 cm into the RD < 1 cm threshold to reduce the number of RD groups to eight. This was due to the fact that RD < 2 cm was sparsely reported in the observed NMA comparisons because suboptimal RD is now clearly defined as >1 cm in the guidelines and the RD < 2 cm group was obtaining undue influence in the ranking statistics, which was wholly implausible. SA 6 incorporated prior information for RD < 1 cm, RD > 0 cm, >1 cm, and >2 cm. SA 7 incorporated prior information for RD < 1 cm, >0 cm, and >1 cm.

All sensitivity analyses were in line with the base case analysis and demonstrated prolonged OS if primary cytoreductive surgery achieved macroscopic RD to 0 cm compared with any other RD threshold (Table 5). However, the effect estimates were attenuated in comparisons involving macroscopic RD to 0 cm, although not to any suggestion of changing the existing conclusions. This was even the case in the most extreme sensitivity analysis (SA 5) that used all RD thresholds, including ones that would not have been expected to have been widely reported in reality. There remained three clear and distinct categories of RD thresholds after primary cytoreductive surgery. Complete macroscopic RD to 0 cm is still by far the best surgical option, with near-optimal (<1 cm) debulking a consolation if this is not possible. Suboptimal can therefore be defined as RD > 1 cm.

There were no issues with model convergence or other diagnostics in WinBUGS[39] in any of the sensitivity analyses, as previously indicated in the base case analysis.

DISCUSSION

There is a high level of uncertainty facing women undergoing treatment for advanced EOC, especially given differences in practice between surgeons in the United Kingdom and internationally. There are many

Table 4. Results of the selection model analyses are given, for the 3 considered sets of priors for the publication probabilities.

	$L_{1,d} = P_{s,d,0}, L_{2,d} = P_{s,d,50}, U_{1,d} = P_{l,d,0}, U_{2,d} = P_{l,d,50}$				$L_{1,d} = P_{s,d,25}, L_{2,d} = P_{s,d,75}, U_{1,d} = P_{l,d,25}, U_{2,d} = P_{l,d,75}$				$L_{1,d} = P_{s,d,50}, L_{2,d} = P_{s,d,100}, U_{1,d} = P_{l,d,50}, U_{2,d} = P_{l,d,100}$			
RD	HR (95% CrI)	Median rank (95% CrI)	P (best), (%)	SUCRA (%)	HR (95% CrI)	Median rank (95% CrI)	P (best), (%)	SUCRA (%)	HR (95% CrI)	Median rank (95% CrI)	P (best), (%)	SUCRA (%)
0 cm	—	1 (1–2)	96.47	99.55	—	1 (1–2)	87.48	98.38	—	1 (1–2)	96.53	99.56
<1 cm	1.93 (1.72–2.19)	3 (2–5)	0	68.58	1.93 (1.72–2.19)	4 (2–6)	0	63.11	1.95 (1.73–2.22)	3 (2–5)	0	68.95
>0 cm	1.89 (1.41–2.53)	3 (2–6)	0	69.45	1.92 (1.44–2.60)	4 (2–7)	0	62.61	1.94 (1.44–2.60)	3 (2–6)	0	68.40
1–2cm	3.28 (1.97–5.48)	8 (4–9)	0	19.91	3.36 (1.98–5.67)	8 (5–9)	0	17.49	3.42 (2.04–5.78)	8 (4–9)	0	19.06
<2 cm	2.72 (1.49–5.00)	6 (2–8)	0.07	36.89	2.80 (1.52–5.21)	7 (2–8)	0.06	32.26	2.80 (1.50–5.18)	7 (2–8)	0.06	36.80
>1 cm	2.49 (2.19–2.85)	6 (4–8)	0	39.43	2.51 (2.20–2.89)	6 (5–8)	0	34.92	2.53 (2.22–2.91)	6 (4–8)	0	39.65
>2 cm	4.10 (2.51–6.81)	9 (7–9)	0	5.13	4.28 (2.57–7.13)	9 (7–9)	0	3.52	4.30 (2.59–7.24)	9 (7–9)	0	4.79
1–5cm	1.75 (0.96–3.17)	3 (1–7)	3.28	71.99	1.45 (0.79–2.61)	2 (1–6)	10.99	80.27	1.77 (0.96–3.22)	3 (1–7)	3.26	72.22
>5 cm	2.58 (1.39–4.79)	6 (2–9)	0.18	39.07	1.98 (1.05–3.70)	4 (2–8)	1.47	57.45	2.59 (1.37–4.83)	6 (2–9)	0.15	40.56

$P_{(s,l),d}$(percentiles 50, 75, 100)=probability that a small/large study is published with a specified design in 50th, 75th, and 100th percentiles; RD: residual disease; P (best): probability that RD threshold is the best.

reviews and guidelines assessing the effect of remaining RD on OS after primary surgery. Unfortunately, many include low quality studies prone to selection and other biases due to poor design or inadequate conduct of statistical analyses. A NMA with stringent inclusion criteria that minimized selection bias was required to synthesize the evidence in this important clinical area. Because it is an area of great equipoise, to convince opponents and proponents of maximal efforts of surgical debulking alike, estimated effects need to report "fair" effect estimates. Making an adjustment for publication bias using elicited views of gynecological experts, we argue is the best approach to achieving this.

There is limited current guidance on methods for adjusting meta-analyses for publication bias, including strategies for choosing an informative prior.[46] Many systematic reviewers neglect to examine or discuss publication bias.[30,47] We are unaware of any literature on how often authors adjust for publication bias in their primary analyses or the methods they apply when adjustment is performed. However, the Copas model for NMAs has been infrequently cited, and inspection of the citations seems to indicate that nobody has elicited the parameters for its employment previously. More generally, elicitation within the context of meta-analyses is rare, likely because of its associated burden. Part B of our elicitation exercise, which elicits the average magnitude of the effect in missing studies is, to the best of our knowledge, novel. In conventional use of Bayesian methods, when prior information is scarce, it is advantageous to collaborate closely with experts. Prior information can be obtained systematically, and the information gathered can easily be formalized into prior distributions,[48] as we showed in our elicitation exercise. Often prior specifications should use available information because it can be the key to answering questions about populations that otherwise remain unanswered. The search for prior information may be intensive and time consuming, but the rewards are obvious because meta-analyses are almost all exclusively subject to some degree of reporting bias; therefore, we can improve the reliability of effect estimates by adjusting for publication bias. In our case, we used expert elicitation, but this could use some other systematic approach, with the key message that applying some kind of sensible adjustment for publication bias is better than doing nothing in most cases. However, incorporating prior information that disagrees with the information contained in data can lead to spurious conclusions, particularly if the prior is too informative. Obviously, there is no way of knowing this when estimating publication bias a priori, but substantial gains can be

Table 5. Results for the series of sensitivity analyses using elicited priors from part B of elicitation exercise.

	SA 1		SA 2		SA 3		SA 4		SA 5		SA 6		SA 7		
RD	HR (95% CrI)	SUCRA	HR (95% CrI)	SUCRA	HR (95% CrI)	SUCRA	HR (95% CrI)	SUCRA	HR (95% CrI)	SUCRA	HR (95% CrI)	SUCRA	HR (95% CrI)	SUCRA	Rank
0 cm	Ref	88.53	Ref	85.32	Ref	99.6	Ref	91.43	Ref	60.06	Ref	97.02	Ref	99.3	1
<1 cm	1.57 (1.23–1.86)	45.68	1.54 (1.22–1.83)	42.75	1.84 (1.60–2.08)	65.69	1.62 (1.32–1.88)	55.61	1.38 (1.07–1.69)	18.76	1.54 (1.20–1.82	60.01	1.75 (1.48–2.00)	63.18	2
>1 cm	2.00 (1.53–2.40)	19.3	1.97 (1.52–2.38)	18.5	2.45 (2.13–2.82)	40.38	2.07 (1.66–2.43)	23.89	1.78 (1.33–2.23)	0.8332	1.99 (1.51–2.38)	26.09	2.19 (1.83–2.52)	39.67	3

RD: Residual disease; SA: sensitivity analysis; SUCRA: Surface under cumulative ranking curve (using whole network of RD thresholds).
cm: Centimeters; Ref: Reference RD threshold is 0 cm; SD: standard deviation; ~N(mean, SD): Normal distribution with specified mean and SD; Rank: Based on highest median ranking between considered RD thresholds.
SA 1: ~N(−0.19,0.04) used to incorporate prior information for RD <1 cm and RD>0 cm, ~N(−0.26,0.05) for RD >1 cm, and ~N(−0.24,0.06) for RD>2 cm.
SA 2: ~N(−0.19,0.04) used to incorporate prior information for RD <1 cm and RD>0 cm, ~N(−0.26,0.05) for RD 1–2cm, <2 cm, and >1 cm.
SA 3: ~N(−0.19,0.04) used to incorporate prior information for RD <1 cm and RD>0 cm.
SA 4: ~N(−0.19,0.04) used to incorporate prior information for RD <1 cm, ~N(−0.26,0.05) for RD>1 cm, and ~N(−0.24,0.06) for RD>2 cm.
SA 5: ~N(−0.19,0.04) used to incorporate prior information for RD <1 cm and RD>0 cm, ~N(-0.26,0.05) for RD 1–2 cm, <2 cm and >1 cm, and ~N(−0.24,0.06) for RD>2 cm, 1–5 cm, and >5 cm.
SA 6: ~N(−0.19,0.04) used to incorporate prior information for RD <1 cm and RD>0 cm, ~N(−0.26,0.05) for RD >1 cm, and ~N(−0.24,0.06) for RD>2 cm where RD<2 cm grouped in RD<1 cm threshold.
SA 7: ~N(−0.19,0.04) used to incorporate prior information for RD <1 cm and RD>0 cm and ~N(−0.26,0.05) for RD >1 cm where RD<2 cm grouped in RD<1 cm threshold.

achieved when the inclusion of this information is appropriate. The use of well-specified informative priors can result in improved parameter estimation, where effect estimates will be more reliable.[49]

A good systematic review will eliminate some forms of reporting biases by following good research practice, at least by conforming to PRISMA guidelines.[50] However, many forms of assessment for publication bias, such as funnel plots and formal tests of asymmetry, as well as methods for addressing it, such as the trim and fill method, multiple imputation, and extensive searches of gray literature, are not adequate.[10,51] We also showed that the sophisticated selection models used in our analyses using results from part A of the elicitation exercise may also not have made the kind of adjustment for publication bias that reflected the opinions of the experts who participated. This was because the adjustment in part A was minimal. The novel methodology used in part B of the exercise where a prior was formulated from the average number of missing studies with their effect sizes may offer a simple and highly desirable approach. However, the adjustments in part A and B do not give different results leading to different conclusions, but Part B seemed to adjust effect estimates that seemed to more reflect the opinions of the experts. Consequently, there could be more scope for the results in Parts A and B to differ if this exercise was repeated in the future. In either approach, it is important to specify methods a priori as to not abuse the results by making post hoc adjustments. If the results from the two methods do differ in any future exercise in another research discipline, researchers can use our recommended sensitivity analyses to assess the impact on their conclusions. If there are widespread differences in results, then the confidence in any adjustment for publication bias will be low and it may be most appropriate to report the unadjusted effect estimates as the primary result.

In most areas where a meta-analysis is feasible, there will be experts in the area, so it is advisable to approach these or organizations like we did with the BGCS. Gynecological cancers are common in women, but other diseases may be more rare and not have the same kind of membership in such a society. Therefore, attempts to invite individual experts to participate in any elicitation exercise may be the only option. Thought should be given as to how experts could contribute to such exercises. When conducting a Bayesian analysis, it is important to always provide the origin of and reason behind the priors. We achieved this through our detailed elicitation exercise and a critique of the answers that each responder gave. We also provided the exact specifications of the priors.[48] We also strongly advise those replicating our methods to

conduct a series of sensitivity analyses as we did in Parts A and B of the elicitation exercise when we applied the results to priors. We clearly demonstrated the impact of various priors on the posterior estimates, ranging from noninformative to skeptical priors, to test the robustness of the conclusions.[48,52] A different research area may differ in the impact of such a wide range of priors, if for example, it was known evidence would be very limited in, say, a rare disease. Setting overly skeptical priors in this such a setting may not be sensible, but it is important to set out these justifications a priori if adjustments for publication bias are to have an impact. It is important to understand and interpret any differences between analyses with different priors.

Previous analyses[2] had shown a clear survival benefit of complete cytoreduction to no visible disease after primary cytoreductive surgery in women with advanced EOC. In a Bayesian framework, extreme value sensitivity analyses examined the plausibility of overturning conclusions in the base case analyses. There seemed to be little likelihood that the existing conclusions were not reliable. The selection model indicates that our findings are robust to large levels of publication bias. The elicited estimate used in Part B of the elicitation exercise as an adjustment for publication bias was also robust to the base case results, but seemed to be more representative of the strength of feeling in the experts' opinions. For example, the mean number of missing studies estimated by experts for comparison of RD $<$ 1 cm versus RD 0 cm was 17.8, corresponding to the derivation of an informative prior (N($-$0.24, 0.06)). Clearly, this had some impact on diluting the magnitude of effect to reflect the omission of unpublished studies in the base case meta-analysis and may compute more unbiased and representative estimates. Further research is now extremely unlikely to change our confidence in the existing estimates of effect. The estimates from the set of sensitivity analyses from the Copas selection model did not have the same desired impact. However, we believe the framework applied as an extension to NMA by incorporating multiarm studies was unique and will be of use in other settings.

Evidence from the literature is not the sole determining factor for clinical decision making. Clinicians also have a preference for 'consensus-based decision making.' This is often through relatively informal sources, such as conversations with clinical colleagues and fellow academic experts. Discussing and trading perspectives can be invaluable in gathering information to form judgments.[53,54] Empirical evidence that incorporates expert elicitation in areas of uncertainty may be of paramount importance to the development of clinical guidelines, enabling the disadvantages of contemporary statistical methodologies to be combined with previously implicit expert consensus. This NMA represents a major update and extension to previous analyses. The NMA adjusted for publication bias using elicited information on the three main comparisons of RD (namely, macroscopic RD to 0 cm, RD $<$ 1 cm, and RD $>$ 1 cm). The sensitivity analyses that use prior information and incorporate this into the effect estimates may help to remove any potential skepticism in the previous findings. The overall certainty of the evidence remains moderate despite a dilution of effect estimates in comparisons involving RD 0 cm. We believe our analyses, which have used advanced statistical methodology and expert opinion, offer the best and most comprehensive currently available evidence base to emphasize the reward for making every effort to perform aggressive surgery in women with advanced EOC, if at all feasible. Our findings should inform clinical guidelines and assist the shared decision-making process between patients, carers, and clinicians in routine practice on selecting the most appropriate choice of primary surgical approach for women with advanced EOC. if at all feasible. This work represents the best available evidence at this time.

Our analyses are also easily replicated in different oncology areas and other diseases. Although various different options for priors are available with various statistical approaches, we strongly recommend applying our methodology. Part B particularly is a very simple but highly effective and desirable approach and the use of incorporating the representative views of experts in their field, results will also be highly relevant, and most importantly, effect estimates should be reliable. Although it is important to test the robustness of the conclusions across all the specified sensitivity analyses, we intentionally used a more complicated exemplar as in practice most reviewers following the methods will be able to apply to more routine meta-analyses. In many cases, this will probably involve just two arm studies and potentially in just two comparison interventions. Previous experience with very straightforward exemplars then encountering difficulties when applying to real life data can be very demotivating, so we were conscious to ensure our methods can be repeated in almost all settings. To adopt the methods in Part B of the elicitation exercise, it merely requires a reliable survival estimate, for example, 5-year survival then follow the steps outlined in the paper. The number of experts required for the elicitation exercise and how many experts would constitute a representative number and the resources available to the research team.

REFERENCES

1. Bryant A, Hiu S, Kunonga P, et al. Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery. *Cochrane Database Syst Rev.* 2021:9. doi: 10.1002/14651858.CD015048.
2. Bryant A, Johnson E, Grayling M, et al. Residual disease threshold after primary surgical treatment for advanced epithelial ovarian cancer (EOC): a network meta-analysis. *BMC Syst Rev.* 2022. [submitted].
3. Colombo N, Van Gorp T, Parma G, et al. Ovarian cancer. *Crit Rev Oncol/Hematol*. 2006;60:159–179.
4. Vergote I, De Wever I, Tjalma W, et al. Neoadjuvant chemotherapy or primary debulking surgery in advanced ovarian carcinoma: a retrospective analysis of 285 patients. *Gynecol Oncol.* 1998;71:431–436.
5. Vergote I, Trimbos BJ. Treatment of patients with early epithelial ovarian cancer. *Curr Opin Oncol.* 2003;15:452–455.
6. Marks-Anglin A, Chen Y. A historical review of publication bias. *Res Syn Meth*. 2020;11:725–742.
7. Stuart GC, Kitchener H, Bacon M, et al. Gynecologic cancer InterGroup (GCIG) consensus statement on clinical trials in ovarian cancer: report from the fourth ovarian cancer consensus conference. *Int J Gynecol Cancer*. 2011;21:750–755.
8. Deeks J, Higgins JPT, Altman DG. Chapter 10: analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, et al, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.3 (Updated February 2022)*. Cochrane; 2022.
9. Ropovik I, Adamkovic M, Greger D. Neglect of publication bias compromises meta-analyses of educational research. *PLoS ONE*. 2021;16:e0252415.
10. Page MJ, Higgins JPT, Sterne JAC. Chapter 13: assessing risk of bias due to missing results in a synthesis. In: Higgins JPT, Thomas J, Chandler J, et al, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.3 (Updated February 2022)*. Cochrane; 2022.
11. Spiegelhalter DJ. Incorporating bayesian ideas into health-care evaluation. *Stat Sci.* 2004;19:156–174.
12. Dwan K, Gamble C, Williamson PR, et al. Reporting Bias Group. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS One*. 2013;8:e66844.
13. Iglesias CP, Thompson A, Rogowski WH, et al. Reporting guidelines for the use of expert judgement in model-based economic evaluations. *Pharmacoeconomics*. 2016;34:1161–1172.
14. Bojke L, et al. *Developing a Reference Protocol for Expert Elicitation in Healthcare Decision Making*. Health Technology Assessment Reports; 2019. [In Press].
15. Morgan MG. Use (and abuse) of expert elicitation in support of decision making for public policy. *Proc Natl Acad Sci USA*. 2014;111:7176–84.
16. Use Committee for Medicinal Products for Human Use (CHMP). In: Agency EM, ed. *Guideline on Adjustment for Baseline Covariates in Clinical Trials*; 2015.
17. Altman DG. *Covariate Imbalance, Adjustment for 2005*; 2005. https://doi.org/10.1002/0470011815.b2a01015.
18. Reeves BC, Deeks JJ, Higgins JPT, et al. Chapter 24: including non-randomized studies on intervention effects. In: Higgins JPT, Thomas J, Chandler J, et al, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 62 (Updated February 2021)*; 2021.
19. McKenzie JE, Brennan SE, Ryan RE, et al. Chapter 3: defining the criteria for including studies and how they will be grouped for the synthesis. In: Higgins JPT, Thomas J, Chandler J, et al, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 62 (Updated February 2021)*; 2021.
20. Higgins JPT, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.0 Cochrane*; 2019. (updated July 2019).
21. Higgins JPT, Welton NJ. Network meta-analysis: a norm for comparative effectiveness? *Lancet*. 2015;386:628–630.
22. Mavridis D, Welton NJ, Sutton A, et al. A selection model for accounting for publication bias in a full network meta-analysis. *Stat Med.* 2014;33:5399–5412.
23. Chootrakool H, Shi JQ, Yue R. Meta-analysis and sensitivity analysis for multi-arm trials with selection bias. *Stat Med.* 2011;30:1183–1198.
24. Mavridis D, Sutton A, Cipriani A, et al. A fully Bayesian application of the Copas selection model for publication bias extended to network meta-analysis. *Stat Med.* 2013;32:51–66.
25. Copas J, Shi JQ. Meta-analysis, funnel plots and sensitivity analysis. *Biostatistics*. 2000;1:247–262.
26. Copas JB. What works? Selectivity models and meta-analysis. *J Roy Stat Soc (Series A)*. 1999;162:95–109.
27. Copas JB, Shi JQ. A sensitivity analysis for publication bias in systematic reviews. *Stat Meth Med Res.* 2001;10:251–265.
28. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials (with discussion). *J Roy Statist Soc Ser A.* 1994;157:357–416.
29. Spiegelhalter DJ, Myles J, Jones D, et al. Bayesian methods in health technology assessment: a review. *Health Technol Assess Rep.* 2000;4:1–130.
30. Ahmed I, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *BMJ*. 2012;344:d7762.
31. Bryant A, Grayling M, Hiu S, et al. Residual disease after primary surgery for advanced epithelial ovarian cancer: expert elicitation exercise to explore opinions about potential impact of publication bias in a systematic review and meta-analysis. *BMJ Open*. 2022;0:e060183. doi: 10.1136/bmjopen-2021-060183.
32. (OCRA) Ocra. What is the survival rate for Stage 3 ovarian cancer? Available at: https://ocrahope.org/patients/about-ovarian-cancer/staging/#:~:text=Most%20women%20diagnosed%20with%20Stage%20III%20ovarian%20cancer%20have%20a,survival%20rate%20of%20approximately%2039%25. Accessed June 14, 2022.
33. American Cancer Society. *Survival Rates for Ovarian Cancer*; 2022.

34. Siegel RL, Miller KD, Jemal A. Cancer statistics. *CA A Cancer J Clin.* 2020;70:7–30.
35. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester, UK: John Wiley & Sons; 2004.
36. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Stat Methods Med Res.* 2001;10:277–303.
37. Sutton AJ, Abrams KR, Jones DR, et al. *Methods for Meta-Analysis in Medical Research*. Chichester, UK: John Wiley & Sons; 2000.
38. Wilson ECF, Usher-Smith JA, Emery J, et al. Expert elicitation of multinomial probabilities for decision-analytic modeling: an application to rates of disease progression in undiagnosed and untreated melanoma. *Value in Health*. 2018;21:669–676.
39. Lunn DJ, Best N, Spiegelhalter D. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput.* 2000;10:325–337.
40. Dias S, Welton NJ, Sutton AJ, et al. *NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials*. NICE Decision Support Unit; 2016.
41. Chaimani A, Higgins JPT, Mavridis D, et al. Graphical tools for network meta-analysis in STATA. *PLoS ONE*. 2013;8:e76654.
42. Dias S, Sutton AJ, Caldwell DM, et al. *NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials*. National Institute for Health and Care Excellence (NICE); 2014.
43. Higgins JP, Barrett JK, Lu G, et al. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods.* 2012;3:98–110.
44. Dias S, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med.* 2010;29:932–944.
45. Mbuagbaw L, Rochwerg B, Jaeschke R, et al. Approaches to interpreting and choosing the best treatments in network meta-analyses. *Syst Rev.* 2017;6:79.
46. Mueller KF, Meerpohl JJ, Briel M, et al. Methods for detecting, quantifying, and adjusting for dissemination bias in meta-analysis are described. *Rev Environ Econ Pol.* 2016;80:25–33.
47. Ross A, Cooper C, Gray H, et al. Assessment of publication bias and systematic review findings in top-ranked otolaryngology journals. *JAMA Otolaryngol Head Neck Surg.* 2019;145:187–188.
48. Zondervan-Zwijnenburg M, Peeters M, Depaoli S, et al. Where do priors come from? Applying guidelines to construct informative priors in small sample research. *Res Hum Develop.* 2017;20:305–320.
49. Quick H, Huynh T, Ramachandran G. A method for constructing informative priors for bayesian modeling of occupational hygiene data. *Ann Work Exposures Health.* 2017;61:67–75.
50. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
51. Page MJ, Sterne JAC, Higgins JPT, et al. Investigating and dealing with publication bias and other reporting biases in meta-analyses of health research: a review. *Res Synth Methods*. 2021;12:248–259.
52. McShane BB, Bockenholt U, Hansen KT. Adjusting for publication bias in meta-analysis: an evaluation of selection methods and some cautionary notes. *Perspect Psychol Sci.* 2016;11:730–749.
53. Gupta DM, Boland RJ, Aron DC. The physician's experience of changing clinical practice: a struggle to unlearn. *Implement Sci.* 2017;12:28.
54. Kristensen N, Nymann C, Konradsen H. Implementing research results in clinical practice- the experiences of healthcare professionals. *BMC Health Serv Res.* 2016;16:48.

Appendix 10: Statistical code used in the main Bayesian NMAs

R code for Part A in the sixth publication in Appendix 9

```
library(R2OpenBUGS)

copas model              <- "model{

 for (j in 1:D) {

  lower[j]              ~ dunif(l1[j], l2[j])

  upper[j]              ~ dunif(u1[j], u2[j])

 }

 for (i in 1:16) {

  S[i]                  <- (s[2*i - 1] + s[2*i])/2

 }

 for (i in 17:18) {

  S[i]                  <- (s[42 + 3*(i - 16) - 2] + s[42 + 3*(i - 16) - 1] + s[42 + 3*(i - 16)])/3

 }

 smin[1]                <- 0.122

 smax[1]                <- 0.306

 smin[2]                <- 0.122

 smax[2]                <- 0.306

 smin[3]                <- 0.1

 smax[3]                <- 0.25
```

```
  smin[4]              <- 0.1

  smax[4]              <- 0.25

  smin[5]              <- 0.1

  smax[5]              <- 0.25

  smin[6]              <- 0.1

  smax[6]              <- 0.25

  smin[7]              <- 0.141

  smax[7]              <- 0.354

  smin[8]              <- 0.141

  smax[8]              <- 0.354

  for (j in 1:D) {

    gu[j]              <- upper[j] - 0.5

    gl[j]              <- lower[j] - 0.5

    invNCDFU[j]        <- 5.531*(pow(upper[j]/(1 - upper[j]), 0.1193) - 1)*step(-gu[j]) - 5.531*(pow((1
- upper[j])/upper[j], 0.1193) - 1)*step(gu[j])

    invNCDFL[j]        <- 5.531*(pow(lower[j]/(1 - lower[j]), 0.1193) - 1)*step(-gl[j]) - 5.531*(pow((1 -
lower[j])/lower[j], 0.1193) - 1)*step(gl[j])

    beta[j]            <- (invNCDFL[j] - invNCDFU[j])/(1/smax[j] - 1/smin[j])

    alpha[j]           <- invNCDFU[j] - beta[j]/smin[j]

  }

  for (i in 1:16) {
```

```
  u[2*i - 1]            <- alpha[des[2*i - 1]] + beta[des[2*i - 1]]/S[i]

  u[2*i]                <- u[2*i - 1]

  z[2*i - 1]             ~ dnorm(u[2*i - 1], 1)T(0, )

  z[2*i]                <- z[2*i - 1]

  y[2*i - 1]             ~ dnorm(my[2*i - 1], w[2*i - 1])

  y[2*i]                 ~ dnorm(my[2*i], w[2*i])

  my[2*i - 1]            <- mu[2*i - 1] + R[2*i - 1]

  my[2*i]                <- mu[2*i] + R[2*i]

  R[2*i - 1]             <- rho[des[2*i - 1]]*s[2*i - 1]*(z[2*i - 1] - u[2*i - 1])

  R[2*i]                 <- rho[des[2*i]]*s[2*i]*(z[2*i] - u[2*i]) + (c3[2*i] - rho[des[2*i - 1]]*s[2*i -
1]*rho[des[2*i]]*s[2*i])*(y[2*i - 1] - mu[2*i - 1])/(pow(s[2*i - 1], 2)*(1 - pow(rho[des[2*i - 1]], 2)))

  vy[2*i - 1]            <- pow(s[2*i - 1], 2)*(1 - pow(rho[des[2*i - 1]], 2))

  vy[2*i]                <- pow(s[2*i], 2)*(1 - pow(rho[des[2*i]], 2)) - (c3[2*i] - rho[des[2*i - 1]]*s[2*i -
1]*rho[des[2*i]]*s[2*i])*(c3[2*i] - rho[des[2*i - 1]]*s[2*i - 1]*rho[des[2*i]]*s[2*i])/(pow(s[2*i - 1],
2)*(1 - pow(rho[des[2*i - 1]], 2)))

  prob.pub[2*i - 1]      <- phi(u[2*i - 1])

  pub.stud[2*i - 1]      <- 1/prob.pub[2*i - 1]

  prob.pub[2*i]          <- 0

  pub.stud[2*i]          <- 0

 }

 for (i in 1:2) {
```

```
  u[42 + 3*i - 2]        <- alpha[des[42 + 3*i - 2]] + beta[des[42 + 3*i - 2]]/S[6 + i]

  u[42 + 3*i - 1]        <- u[42 + 3*i - 2]

  u[42 + 3*i]            <- u[42 + 3*i - 2]

  z[42 + 3*i - 2]        ~ dnorm(u[42 + 3*i - 2], 1)T(0, )

  z[42 + 3*i - 1]        <- z[42 + 3*i - 2]

  z[42 + 3*i]            <- z[42 + 3*i - 2]

  y[42 + 3*i - 2]        ~ dnorm(my[42 + 3*i - 2], w[42 + 3*i - 2])

  y[42 + 3*i - 1]        ~ dnorm(my[42 + 3*i - 1], w[42 + 3*i - 1])

  y[42 + 3*i]            ~ dnorm(my[42 + 3*i], w[42 + 3*i])

  my[42 + 3*i - 2]       <- mu[42 + 3*i - 2] + R[42 + 3*i - 2]

  my[42 + 3*i - 1]       <- mu[42 + 3*i - 1] + R[42 + 3*i - 1]

  my[42 + 3*i]           <- mu[42 + 3*i] + R[42 + 3*i]

  R1[i]                  <- rho[des[42 + 3*i - 2]]*s[42 + 3*i - 2]*(z[42 + 3*i - 2] - u[42 + 3*i - 2])

  R2[i]                  <- rho[des[42 + 3*i - 1]]*s[42 + 3*i - 1]*(z[42 + 3*i - 1] - u[42 + 3*i - 1])

  R3[i]                  <- rho[des[42 + 3*i]]*s[42 + 3*i]*(z[42 + 3*i] - u[42 + 3*i])

  Sigma11[i]             <- pow(s[42 + 3*i - 2], 2)*(1 - pow(rho[des[42 + 3*i - 2]], 2))

  Sigma22[i]             <- pow(s[42 + 3*i - 1], 2)*(1 - pow(rho[des[42 + 3*i - 1]], 2))

  Sigma33[i]             <- pow(s[42 + 3*i], 2)*(1 - pow(rho[des[42 + 3*i]], 2))

  Sigma12[i]             <- c4[3*i - 2] - rho[des[42 + 3*i - 2]]*s[42 + 3*i - 2]*rho[des[42 + 3*i - 1]]*s[42 +
3*i - 1]
```

```
  Sigma21[i]          <- Sigma12[i]

  Sigma13[i]          <- c4[3*i - 1] - rho[des[42 + 3*i - 2]]*s[42 + 3*i - 2]*rho[des[42 + 3*i]]*s[42 +
3*i]

  Sigma31[i]          <- Sigma13[i]

  Sigma23[i]          <- c4[3*i] - rho[des[42 + 3*i - 1]]*s[42 + 3*i - 1]*rho[des[42 + 3*i]]*s[42 + 3*i]

  Sigma32[i]          <- Sigma23[i]

  mu21[i]             <- R2[i] + Sigma21[i]*(y[42 + 3*i - 2] - mu[42 + 3*i - 2])/Sigma11[i]

  mu22[i]             <- R3[i] + Sigma13[i]*(y[42 + 3*i - 2] - mu[42 + 3*i - 2])/Sigma11[i]

  Omega11[i]          <- Sigma22[i] - Sigma21[i]*Sigma12[i]/Sigma11[i]

  Omega22[i]          <- Sigma33[i] - Sigma31[i]*Sigma13[i]/Sigma11[i]

  Omega12[i]          <- Sigma23[i] - Sigma21[i]*Sigma13[i]/Sigma11[i]

  Omega21[i]          <- Omega12[i]

  R[42 + 3*i - 2]     <- R1[i]

  R[42 + 3*i - 1]     <- mu21[i]

  R[42 + 3*i]         <- mu22[i] + Omega21[i]*(y[42 + 3*i - 1] - mu[42 + 3*i - 1])/Omega11[i]

  vy[42 + 3*i - 2]    <- Sigma11[i]

  vy[42 + 3*i - 1]    <- Omega11[i]

  vy[42 + 3*i]        <- Omega22[i] - Omega12[i]*Omega21[i]/Omega11[i]

  prob.pub[42 + 3*i - 2]  <- phi(u[42 + 3*i - 2])

  pub.stud[42 + 3*i - 2]  <- 1/prob.pub[42 + 3*i - 2]
```

```
  prob.pub[42 + 3*i - 1]   <- 0

  pub.stud[42 + 3*i - 1]   <- 0

  prob.pub[42 + 3*i]       <- 0

  pub.stud[42 + 3*i]       <- 0

}

for (i in 1:32) {

  w[i]                  <- 1/vy[i]

  mu[i]                  ~ dnorm(mean[i], prec)

  mean[i]                <- d[t[i]] - d[b[i]]

}

for (i in 33:42)  {

  u[i]                  <- alpha[des[i]] + beta[des[i]]/s[i]

  z[i]                  ~ dnorm(u[i], 1)T(0,)

  y[i]                  ~ dnorm(my[i], w[i])

  R[i]                  <- rho[des[i]]*s[i]*(z[i] - u[i])

  my[i]                 <- mu[i] + R[i]

  vy[i]                 <- pow(s[i], 2)*(1 - pow(rho[des[i]], 2))

  w[i]                  <- 1/vy[i]

  mu[i]                  ~ dnorm(mean[i], prec)

  mean[i]                <- d[t[i]] - d[b[i]]
```

```
  prob.pub[i]            <- phi(u[i])
  pub.stud[i]            <- 1/prob.pub[i]
}
for (i in 43:N) {
  w[i]                 <- 1/vy[i]
  mu[i]                 ~ dnorm(mean[i], prec)
  mean[i]               <- d[t[i]] - d[b[i]]
}
tot.pub[1]              <- sum(pub.stud[1:30])
tot.pub[2]              <- sum(pub.stud[31:32])
tot.pub[3]              <- sum(pub.stud[33:34])
tot.pub[4]              <- sum(pub.stud[35:38])
tot.pub[5]              <- sum(pub.stud[39:40])
tot.pub[6]              <- sum(pub.stud[41:42])
tot.pub[7]              <- sum(pub.stud[43:45])
tot.pub[8]              <- sum(pub.stud[46:48])
d[1]                  <- 0
for (k in 2:NT) {
  d[k]                  ~ dnorm(0, .0001)
}
```

```
prec                  <- 1/pow(tau, 2)

tau                   ~ dnorm(0, 0.01)I(0, )

for (j in 1:D) {

  rho1[j]              ~ dunif(0, 2)

  rho[j]              <- rho1[j] - 1

}

for (i in 1:N) {

  DD[i]               <- w[i]*(y[i] - my[i])*(y[i] - my[i])

}

D.bar                 <- sum(DD[])

for (k in 1:NT) {

  order[k]             <- rank(d[], k)

  most.effective[k]       <- equals(order[k], 1)

  for (j in 1:NT) {

    effectiveness[k, j]    <- equals(order[k], j)

  }

}

for (k in 1:NT) {

  for (j in 1:NT) {

    cumeffectiveness[k, j] <- sum(effectiveness[k, 1:j])
```

```
   }
  }
  for (k in 1:NT) {
    SUCRA[k]              <- sum(cumeffectiveness[k, 1:(NT - 1)])/(NT - 1)
  }
}"
fileConn                  <- file("copas_output.txt")
writeLines(copas_model, fileConn)
close(fileConn)
file.show("copas_output.txt")
# Example 1, run without adjustment
data                      <-
  list(l1   = c(0.9999, 0.9999, 0.9999, 0.9999, 0.9999, 0.9999, 0.9999, 0.9999),
       l2   = c(0.9999, 0.9999, 0.9999, 0.9999, 0.9999, 0.9999, 0.9999, 0.9999),
       u1   = c(0.9999, 0.9999, 0.9999, 0.9999, 0.9999, 0.9999, 0.9999, 0.9999),
       u2   = c(0.9999, 0.9999, 0.9999, 0.9999, 0.9999, 0.9999, 0.9999, 0.9999),
       N    = 48,
       D    = 8,
       NT   = 9,
       c4   = c(0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025,
```

```
        0.025, 0.025, 0.025),

c3  = c(0.025, 0.025, 0.025, 0.025, 0.0025, 0.0025, 0.025, 0.025, 0.025,

        0.025, 0.025, 0.025, 0.025, 0.025, 0.0025, 0.0025, 0.025, 0.025,

        0.025, 0.025, 0.025, 0.025, 0.0025, 0.0025, 0.0025, 0.0025,

        0.0025, 0.0025, 0.0025, 0.0025, 0.025, 0.025),

y   = c(0.843, 1.093, 0.726, 1.308, 0.75, 0.904, 0.628, 0.754, 0.8087,

        1.5063, 0.7975, 1.1756, 0.809, 1.1282, 0.84, 1.03, 0.342,

        0.6235, 0.408, 0.788, 1.02, 1.08, 0.38, 0.541, 0.671, 1.012,

        0.83, 0.9343, 0.31, 0.56, 0.5277, 0.6956, 1.008, 1.17, 0.3646,

        1.082, 0.611, 0.775, 0.514, 0.26, 0.8566, 0.16, 1.123, 1.3745,

        2.109, 0.66, 0.603, 1),

s   = c(0.38, 0.276, 0.263, 0.25, 0.09, 0.086, 0.2211, 0.214, 0.213,

        0.2218, 0.293, 0.2723, 0.298, 0.2757, 0.06, 0.06, 0.31, 0.292,

        0.201, 0.244, 0.32, 0.32, 0.09, 0.09, 0.1176, 0.161, 0.251,

        0.25, 0.0985, 0.121, 0.219, 0.204, 0.1657, 0.345, 0.186, 0.234,

        0.1761, 0.344, 0.248, 0.131, 0.2, 0.122, 0.4959, 0.557, 0.573,

        0.257, 0.242, 0.254),

b   = c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,

        1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 1, 1, 1, 1, 1, 1, 2, 2, 5, 5,

        1, 1, 1, 1, 1, 1),
```

```
    t    = c(2, 6, 2, 6, 2, 6, 2, 6, 2, 6, 2, 6, 2, 6, 2, 6, 2, 6, 2,
             6, 2, 6, 2, 6, 2, 6, 2, 6, 4, 7, 2, 2, 3, 3, 3, 3, 6, 6, 7, 7,
             2, 4, 7, 2, 8, 9),
    des  = c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
             1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 3, 3, 4, 4, 4, 4, 5, 5, 6, 6,
             7, 7, 7, 8, 8, 8),
    rho1 = c(1.1, 1.1, 1.1, 1.1, 1.1, 1.1, 1.1, 1.1))
no adjustment            <- bugs(data,
                         model.file       = "copas_output.txt",
                         inits            = NULL,
                         parameters.to.save =
                          c("d", "SUCRA", "most.effective",
                            "order"),
                         n.chains         = 2,
                         n.iter           = 100000,
                         n.burnin         = 50000)
print(no_adjustment, digits.summary = 3)
# Example 2, run with adjustment
data$l1                  <- c(0.3, 0.3, 0.3, 0.2, 0.3, 0.2, 0.2, 0.1)
data$l2                  <- c(0.8, 0.8, 0.9, 0.75, 0.8, 0.8, 0.9, 0.75)
```

```
data$u1                 <- c(0.9, 0.9, 0.9, 0.3, 0.9, 0.7, 0.75, 0.15)

data$u2                 <- c(1, 1, 1, 0.8, 1, 0.99, 0.99, 0.8)

with_adjustment             <- bugs(data,

                        model.file        = "copas_output.txt",

                        inits             = NULL,

                        parameters.to.save =

                         c("d", "SUCRA", "most.effective",

                           "order"),

                        n.chains          = 2,

                        n.iter            = 100000,

                        n.burnin          = 50000)

print(with_adjustment, digits.summary = 3)
```

WINBUGS code for Part B in the sixth publication in Appendix 9

```
model                                {

for(i in 1:ns2)

 {

   y[i,2] ~ dnorm(delta[i,2],prec[i,2])

   resdev[i] <- (y[i,2]-delta[i,2])*(y[i,2]-delta[i,2])*prec[i,2]

  }

for(i in (ns2+1):(ns2+ns3))

 {
```

```
   for (k in 1:(na[i]-1))
{
      for (j in 1:(na[i]-1))
{
         Sigma[i,j,k] <- V[i]*(1-equals(j,k)) + var[i,k+1]*equals(j,k)
        }
     }
Omega[i,1:(na[i]-1),1:(na[i]-1)] <- inverse(Sigma[i,,])
y[i,2:na[i]] ~ dmnorm(delta[i,2:na[i]],Omega[i,1:(na[i]-1),1:(na[i]-1)])
for (k in 1:(na[i]-1)){  # multiply vector & matrix
      ydiff[i,k]<- y[i,(k+1)] - delta[i,(k+1)]
      z[i,k]<- inprod(Omega[i,k, 1:(na[i]-1)], ydiff[i,1:(na[i]-1)])
                        }
   resdev[i]<- inprod(ydiff[i,1:(na[i]-1)], z[i,1:(na[i]-1)])
}
for(i in (ns2+ns3+1):(ns2+ns3+ns4))
{
for (k in 1:(na[i]-1)) {
for (j in 1:(na[i]-1)) { Sigma2[i,j,k] <- V[i]*(1-equals(j,k)) + var[i,k+1]*equals(j,k) }
}
Omega2[i,1:(na[i]-1),1:(na[i]-1)] <- inverse(Sigma2[i,,]) #Precision matrix
# multivariate normal likelihood for 4-arm trials
y[i,2:na[i]] ~ dmnorm(delta[i,2:na[i]],Omega2[i,1:(na[i]-1),1:(na[i]-1)])
#Deviance contribution for trial i
   for (k in 1:(na[i]-1)){  # multiply vector & matrix
      ydiff[i,k]<- y[i,(k+1)] - delta[i,(k+1)]
```

```
      z[i,k]<- inprod(Omega2[i,k, 1:(na[i]-1)], ydiff[i,1:(na[i]-1)])
                        }
   resdev[i]<- inprod(ydiff[i,1:(na[i]-1)], z[i,1:(na[i]-1)])
  }
for(i in 1:(ns2+ns3+ns4)){
                #   LOOP THROUGH ALL STUDIES
   w[i,1] <- 0    # adjustment for multi-arm trials is zero for control arm
# deltas are the trial specific shrunken relative effect estimates, relative to treatment 1 in that trial
   delta[i,1] <- 0
   for (k in 2:na[i]) {           #  LOOP THROUGH ARMS
      var[i,k] <- pow(se[i,k],2)   # calculate variances
      prec[i,k] <- 1/var[i,k]      # set precisions
     }
   for (k in 2:na[i]) {           # LOOP THROUGH ARMS
# trial-specific LOR distributions
      delta[i,k] ~ dnorm(md[i,k],taud[i,k])
# mean of random effects distributions, with multi-arm trial correction
      md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k]
# precision of random effects distributions (with multi-arm trial correction)
      taud[i,k] <- tau *2*(k-1)/k
# adjustment, multi-arm RCTs
      w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
# cumulative adjustment for multi-arm trials
      sw[i,k] <- sum(w[i,1:k-1])/(k-1)
     }
  }
```

```
totresdev <- sum(resdev[])          #Total Residual Deviance

d[1]<-0      # treatment effect is zero for reference treatment

# vague priors for treatment effects

for (k in 2:nt){  d[k] ~ dnorm(0,.0001) }

sd ~ dunif(0,5)     # vague prior for between-trial SD

tau <- pow(sd,-2)   # between-trial precision = (1/between-trial variance)

# ranking on relative scale

for (k in 1:nt) {

# rk[k] <- nt+1-rank(d[],k)  # assumes events are “good” so don't need this since death is bad!

rk[k] <- rank(d[],k) # assumes events are “bad”

best[k] <- equals(rk[k],1) #calculate probability that treat k is best

for (j in 1:nt) {

effectiveness[k,j] <- equals(rk[k],j )

}

}

for (k in 1:nt) {

for (j in 1:nt) {

cumeffectiveness[k,j] <- sum(effectiveness[k, 1:j])

}

}

# SUCRAS

for (k in 1:nt) {

SUCRA[k] <- sum(cumeffectiveness[k,1:(nt-1)])/(nt-1)

}

}

                        # *** PROGRAM ENDS
```

Data

 # NOTE place multi-arm trials at end

ns2= number of studies with two arms;

ns3= number of studies with three arms;

ns4= number of studies with three arms;

nt=number of treatments; V=variance of baseline arm

#Random number seed set as 23

list(ns2=10, ns3=16, ns4=2, nt=9)

t[,1]	t[,2]	t[,3]	t[,4]	y[,2]	y[,3]	y[,4]	se[,2]	se[,3]	se[,4]	na[]	V[]
	#study										
1	2	NA	NA	1.008	NA	NA	0.1657	NA	NA	2	NA
	#Bristow 2011										
1	2	NA	NA	1.17	NA	NA	0.345	NA	NA	2	NA
	#Cuylan 2018										
1	3	NA	NA	0.3646	NA	NA	0.186	NA	NA	2	NA
	#Polterauer 2012										
1	3	NA	NA	1.082	NA	NA	0.234	NA	NA	2	NA
	#Hofstetter 2013										
1	3	NA	NA	0.775	NA	NA	0.344	NA	NA	2	NA
	#Luger 2020										
2	6	NA	NA	0.514	NA	NA	0.248	NA	NA	2	NA
	#Chan 2003										
2	6	NA	NA	0.26	NA	NA	0.131	NA	NA	2	NA
	#Winter 2008b										
5	7	NA	NA	0.8566	NA	NA	0.2	NA	NA	2	NA
	#Akahira 2001										

5 7 NA NA 0.16 NA NA 0.122 NA NA 2 NA

#Winter 2008c

1 3 NA NA 0.611 NA NA 0.1761 NA NA 2 NA

#Feng 2016

1 2 6 NA 0.7975 1.1756 NA 0.293 0.2723 NA 3 0.025

#Chang 2012a

1 2 6 NA 0.342 0.6235 NA 0.31 0.292 NA 3 0.025

#Teweri 2016

1 2 6 NA 0.809 1.1282 NA 0.298 0.2757 NA 3 0.025

#Chang 2012b

1 2 6 NA 0.8087 1.5063 NA 0.213 0.2218 NA 3 0.025

#Langstraat 2011

1 2 6 NA 0.75 0.904 NA 0.09 0.086 NA 3 0.0025

#Winter 2007

1 2 6 NA 0.726 1.308 NA 0.263 0.25 NA 3 0.025

#Chi 2006

1 2 6 NA 0.843 1.093 NA 0.38 0.276 NA 3 0.025

#Eisenkop 2003

1 2 6 NA 0.628 0.754 NA 0.2211 0.214 NA 3 0.025

#Wimberger 2010

1 2 6 NA 0.31 0.56 NA 0.0985 0.121 NA 3 0.0025

#Tseng 2018

2 1 6 NA -0.84 0.19 NA 0.06 0.05 NA 3 0.0025

#Klar 2016

2 4 7 NA 0.5277 0.6956 NA 0.219 0.204 NA 3 0.025

#Chi 2001

6 1 2 NA -1.08 -0.06 NA 0.32 0.319 NA 3 0.025

#Melamed 2017a

6 1 2 NA -0.541 -0.161 NA 0.09 0.089 NA 3 0.0025

#Melamed 2017b

1 2 6 NA 0.671 1.012 NA 0.1176 0.161 NA 3 0.0025

#Kahl 2017

1 2 6 NA 0.83 0.9343 NA 0.251 0.25 NA 3 0.0025

#Paik 2018

1 2 6 NA 0.408 .788 NA 0.201 0.244 NA 3 0.025

#Ataseven 2016

1 2 4 7 1.123 1.3745 2.109 0.4959 0.557 0.573 4 0.025

#Aletti 2006

1 2 8 9 0.66 0.603 1.0 0.257 0.242 0.254 4 0.025

#Winter 2008a

END

Initial Values

gen inits in spec tool

Chain 1

list(

d = c(

NA,0.6221196725728642,0.6217862538578188,1.132124042164687,0.3508124329078437,

0.871384198107111,0.8635260370246234,-0.06836681013117955,1.000705385618646),

delta = structure(.Data = c(

NA,0.6572202035556228, NA, NA, NA,

0.9111159558190508, NA, NA, NA,0.2453654603951218,

NA, NA, NA,0.9121447307193166, NA,

NA, NA,0.8664151728256154, NA, NA,

NA,0.3845618162421535, NA, NA, NA,

0.2678924686248531, NA, NA, NA,0.8067949536357871,

NA, NA, NA,0.2417366400784014, NA,

NA, NA,0.737628767617755, NA, NA,

NA,0.6240906410704821,0.9459147850346625, NA, NA,

0.1523032154912328,0.6030525757109958, NA, NA,0.72106528425035,

0.8972878490586746, NA, NA,0.7077587335156812,1.351749222190123,

NA, NA,0.7444983891512498,0.8396749458097386, NA,

NA,0.7463008300559638,0.7699948308659057, NA, NA,

0.7194766203051981,1.040586673341126, NA, NA,0.7637831028573335,

0.6636300681033304, NA, NA,0.1705085388301971,0.5459136879426828,

NA, NA,-0.9287999902253876,0.1545680057900695, NA,

NA,0.8242467615090538,0.6707316709170086, NA, NA,

-0.9261539918760694,-0.5038783916995591, NA, NA,-0.5973681213932421,

-0.07035336781371966, NA, NA,0.5744996715226977,1.024904971097157,

NA, NA,0.5671161751225009,0.8328141804101244, NA,

NA,0.6460649095845794,0.8738248074406967, NA, NA,

0.8123456522504535,1.292058189479752,1.203635273189171, NA,0.4764014363658424,

0.1533608984182706,0.9594703190026386),

.Dim = c(28,4)),

sd = 0.1717450535063422)

Chain 2

list(

d = c(

NA,0.8835612106907204,0.5948172151101957,1.28042959917834,0.4440856138637916,
1.066200324515274,1.248420019306943,0.5032386249995585,0.8189973081099853),
delta = structure(.Data = c(
NA,1.057821718415797, NA, NA, NA,
1.271677841968639, NA, NA, NA,0.3885557809363049,
NA, NA, NA,0.9454627467985199, NA,
NA, NA,0.2231899065175175, NA, NA,
NA,0.1871983942059395, NA, NA, NA,
0.3394483527579222, NA, NA, NA,0.7263844616216729,
NA, NA, NA,0.2707040559635077, NA,
NA, NA,0.6786064143556662, NA, NA,
NA,1.04195166634256,1.330339586060714, NA, NA,
0.7623242013186953,1.073898462707973, NA, NA,0.9222100327458661,
0.7840620937007673, NA, NA,0.87446626571409,1.327023226150742,
NA, NA,0.8275415226021652,1.014137477556607, NA,
NA,1.007241390308154,1.375196830234182, NA, NA,
1.090561452490763,1.098985240086203, NA, NA,0.8196921658102825,
0.9557430166168031, NA, NA,0.3520079555453237,0.5395997813127018,
NA, NA,-0.8037906742852767,0.2274253526132164, NA,
NA,0.7372908760911715,0.633071666367231, NA, NA,
-0.9766512192614173,0.20619185507668, NA, NA,-0.5387707083078694,
-0.2376914634521927, NA, NA,0.6958481745168091,0.9488124149242627,
NA, NA,0.8053695451808027,0.8161425379464685, NA,
NA,0.5176747004257884,0.7948189635126102, NA, NA,
1.200952374948967,1.399522354173569,1.528046753609429, NA,0.9545232159619261,
0.7892853279554364,1.185235980497421),

.Dim = c(28,4)),

sd = 0.2160709386888112)