How do preferences for public health interventions differ? A case study using a weight loss maintenance intervention

A thesis presented for the degree of

Doctor of Philosophy at Newcastle University

(Faculty of Medical Sciences)

David John Mott

BSc, MSc (University of Sheffield)

April 2018

Abstract

Preference information is routinely used in healthcare decision-making, such as in health technology assessment, where preferences for health states are used to estimate quality-adjusted life years. However, there is debate over whose preferences should be considered. This debate extends to other areas in health economics where preferences are elicited to inform decision-making. Common stated preference methodologies include discrete choice experiments (DCEs) and contingent valuation (CV), which produce information that can be used in economic evaluations and benefit-risk assessments. This thesis expands the 'preference debate' beyond health utilities by considering the use of DCE and CV studies in decision-making. It also extends the debate by considering more than 'patients' and the general population; introducing four user groups that differ in their relative experience. The thesis' aim is to explore whether, and how, differences in individuals' experience of a healthcare intervention/health issue might influence preferences. It uses a weight loss maintenance intervention from the NULevel trial (ISRCTN14657176) as a case study.

Chapter One introduces the thesis premise and Chapter Two outlines the economic theory underpinning the methodologies used. Chapter Three discusses the use of economic evaluation, providing a review of the 'preference debate' and recent developments. CV and DCE methodologies are next outlined and the use of information from DCEs in healthcare decision-making reviewed. Chapter Four systematically reviews the DCE literature in health economics to determine whether respondents can be classified as either a patient or general population sample. Building upon previous chapters, Chapter Five introduces an alternative framework for classifying respondent samples in DCE studies in terms of their relative experience; this defines the user groups for the empirical work (conducted via NULevel and an online panel). The research questions are next outlined. Chapter Six summarises the NULevel trial, the recruitment plan and the online survey design (containing both a DCE and CV task) used to address the research questions. Chapters Seven to Nine present results and Chapter Ten provides a discussion of implications, strengths and limitations of the work.

The empirical work suggests that preferences differ according to the relative experience of respondents; raising concerns surrounding the generalisability of DCE studies. This is important given the increasing application of preference studies in decision-making. If this trend continues the implications will be even greater in future.

Acknowledgements

This thesis would not have been possible if it were not for the support of a number of individuals. Firstly, I would like to thank my two supervisors based in the Institute of Health & Society: Luke Vale and Laura Ternent. Their enthusiasm kept me motivated throughout the process and I truly appreciate their continued support over the past four years.

I would like to thank the Health Foundation for providing financial support for this PhD. I would also like to thank several other individuals at Newcastle University that provided support whilst I was undertaking this work. These include, but are not limited to, John Wildman, Darren Flynn, Ruth Bell, Frauke Becker, Falko Sniehotta, and Liz Evans. In addition, I am grateful to numerous individuals from the broader academic community for their feedback on various pieces of work from this thesis. I am particularly grateful to those that took the time to discuss my work at conferences: Anne Spencer, Victoria Serra-Sastre, and Jennifer Whitty.

I am fortunate to have started this process alongside two fellow Health Economics PhD students: Tomos Robinson and Matthew Little. The PhD process would have been far less pleasant without our regular coffee breaks and catch-ups.

Finally, I am thankful for the support of my parents and my brother Alex during what must have appeared to be an endless endeavour. Last, but certainly not least, I am truly grateful for the love, support and patience of my partner Jenny.

Contents

List of Tables

List of Figures

Dissemination

Publications

The following publication is based, in part, on Chapter Three:

> Mott, D.J. & Najafzadeh, M., 2016. Whose preferences should be elicited for use in health-care decision-making? A case study using anticoagulant therapy. *Expert Review of Pharmacoeconomics & Outcomes Research*, 16(1), pp.33–39.

Presentations

A review paper based on Chapter Four was discussed by Professor Anne Spencer:

> Mott, D.J., Ternent, L., Vale, L. Whose preferences matter in discrete choice experiments? European Health Economics Association (EuHEA) PhD & Early Career Researcher Conference, Manchester (September 2014).

The systematic review from Chapter Four was also presented via an invited online webinar:

> Mott, D.J., Ternent, L., Vale, L. Whose preferences matter in discrete choice experiments? Health Economics Research Group (HERG) Brunel Virtual Research Seminar (November 2014).

A results paper based on Chapter Seven was discussed by Dr Victoria Serra-Sastre:

> Mott, D.J., Ternent, L., Vale, L. How do preferences for weight loss maintenance interventions differ? A discrete choice experiment. Health Economists' Study Group (HESG) Meeting, Gran Canaria (June 2016).

A results paper based on Chapter Nine was discussed by Professor Jennifer Whitty:

> Mott, D.J., Ternent, L., Vale, L. How do preferences for public health interventions differ and whose preferences should count? A case study using weight loss maintenance. Health Economists' Study Group (HESG) Meeting, Aberdeen (June 2017).

Part I. Introduction, Theory & Literature Review

Chapter 1. Introduction

The overall aim of this thesis is to explore whether, and how, differences in individuals' experience of a healthcare intervention and/or a health issue might influence their preferences for a related healthcare intervention. The importance of this relates to the use of quantitative preference information in healthcare decision-making. If preferences differ across stakeholder groups, and preferences influence decision-making, the decisions that are made may differ depending on whose preferences are considered. On the other hand, if preferences do not differ significantly across stakeholder groups then it does not matter whose preferences are considered. It is therefore important to explore whether, and how, preferences might differ between different groups of individuals.

In this chapter, section one sets out the broad topic area by outlining how preferences are elicited by health economists for use in healthcare decision-making, and the issue of whose preferences to elicit for this purpose. Section two explains the overall motivation and the importance of the project. Section three sets out the four research questions that are addressed in the empirical work. Section four briefly describes the case study that is used in this thesis. Finally, section five outlines the overall structure of the thesis and describes the purpose of each of the following chapters.

1.1 Background

Individuals' preferences are often quantified by health economists using a set of tools known as stated preference methodologies (Drummond et al., 2015). These methodologies are highly beneficial where preferences are not known (or revealed), which could be a result of non-existent markets, or market failures. Quantitative data from stated preferences studies can be, and routinely are, used in healthcare decision-making where revealed preference data is not available (Brazier et al., 2016).

Perhaps the most common example of stated preference data being used in healthcare decision-making is in health technology assessment (HTA). It is common, particularly in the United Kingdom (UK), for cost-utility analysis (CUA) to be conducted in order to evaluate the cost-effectiveness of a healthcare intervention (Drummond et al., 2015; see section 3.1.1). The most common measure of benefit used in a CUA is the quality-adjusted life year (QALY), which combines estimates of both length of life and quality of life (Tsuchiya & Dolan, 2005). Quality of life is typically measured on a zero to one scale, where zero represents being dead and one represents some notion of 'perfect health'. In order to obtain

these values, referred to as health state utilities, stated preference tasks are conducted (Brazier et al., 1999; see section 3.1.2). This is necessary because one cannot observe individuals' preferences for health states; they must be elicited.

When it comes to conducting CUA for HTA, the issue of whose preferences to elicit is a contentious one (Brazier et al., 2005). Specifically, the question relates to whether preferences of patients or the general population are more appropriate to incorporate into analyses that aim to inform healthcare decision-making. The debate, which can be linked directly to the theory of welfare economics (Gandjour, 2010), is yet to be fully resolved by the academic community (see section 3.1.4) due to its normative nature (Versteegh & Brouwer, 2016).

Health state utilities are not the only example where individuals' preferences are elicited. Contingent valuation (CV) studies have been used in healthcare since the 1980s (Smith & Sach, 2010; see section 3.2) to approximate the economic concepts of compensating variation and equivalent variation (see section 2.2). The estimates derived from these studies, known as willingness to pay (WTP) or willingness to accept (WTA), can be used in cost-benefit analysis (CBA) where the market values of benefits are not known (McIntosh et al., 2010). Estimates of WTP and WTA can also be produced using discrete choice experiment (DCE) methodology and the number of published studies that use a DCE approach has increased significantly since their first reported use in health economics in the 1990s (de Bekker-Grob et al., 2012; see section 3.3). As DCEs offer more than the opportunity to estimate WTP and WTA, there has been an increasing interest in using this methodology to assist healthcare decision-making. However, the use of DCEs in decision-making is in its infancy, with only a small number of examples to date (see section 3.4).

1.2 Study Motivation

If current trends continue (see sections 3.3 and 3.4), one might expect DCEs to play a larger role in healthcare decision-making over time. The inevitable question to ask once this happens is: whose preferences should be elicited? Many of the arguments made in the 'preference debate' from the health state valuation literature also apply when considering the use of DCEs in healthcare decision-making (Mott & Najafzadeh, 2016; see section 3.4). Therefore, there is clear justification for exploring how preferences might differ between patients and the general population in this context.

However, it is important not to ignore the fact that there has been an increasing interest in 'patient-centred' decision-making in recent years (Barry & Edgman-Levitan, 2012). This interest has led to calls for a more 'patient-centric' HTA process (Mühlbacher, 2015; Facey et

al., 2017), where patient preferences can potentially play a larger role. DCEs could be used to achieve this aim. Indeed, DCE methodology was originally put forward as a way to capture patient experiences (Ryan et al., 2001).

It should also be noted that the comparison between patients and the general population has been referred to as a 'false dichotomy' in the context of health state valuation on the grounds that experience can differ significantly between individuals within each group (Dolan, 1999). With all of this in mind, there is clear justification for exploring how preferences might differ between groups of patients with differing levels of experience, rather than focusing solely on comparisons of patients and the general population.

Ultimately, if quantitative preference data from DCEs are to be used in analyses that inform healthcare decision-making, it is important to understand how preferences might differ amongst the different stakeholder groups that could be recruited for these studies. If significant differences in preferences are found between the different groups, it will be important to consider why these differences exist and how they might influence decision-making. It might be the case that experienced individuals have 'better defined' preferences than those with less experience and have significantly different preferences to those with less experience. Individuals with no experience might focus solely on clinical effectiveness and efficiency-related aspects of healthcare (Bryan & Dolan, 2004).

Understanding the differences that could occur in a DCE because of the type of respondents could make it easier to provide guidance on the use of DCEs, which would help to improve the credibility of the methodology, as well as improving decision-making processes. However, it is also important to consider that the choice of methodology may have an impact on the results of preference studies too. Given that WTP can be elicited using both CV and DCE methodologies, utilising both methods could provide a useful insight into this area.

1.3 Research Questions

Chapter Five sets out a framework that describes how respondent samples could differ in relation to their relative experience of a health condition and its associated treatment. The empirical work in this thesis subsequently tests whether preferences differ between these four user groups. In order to recruit samples of each user group, an online survey was designed (see Chapter Six) and delivered to individuals that had participated in a randomised controlled trial (RCT), as well as to individuals in an online panel (see section 6.2 for a detailed overview of the sampling process).

The specific research questions to be addressed in the empirical work (see section 5.4 for a more detailed overview) are:

1. To what extent do preferences for a health service, elicited via a DCE, differ between different user groups, and why might these differences occur?
2. In relation to WTP estimates:
    a. To what extent do the estimates differ between different user groups, and why might these differences occur?
    b. To what extent do the estimates differ if they are elicited indirectly (via DCE) or directly (via CV)?
3. To what extent do certain user groups have better defined preferences than other user groups, and why might these differences occur?
4. To what extent might differences in preferences between the user groups be attributed to the recruitment vehicle (RCT vs. online panel)?

1.4 The Case Study: Weight Loss Maintenance

A recent report suggests that over a quarter of adults in the UK are obese (OECD, 2017). In addition, it is estimated that 40% of adults worldwide attempted to lose weight between 2010 and 2015 (Santos et al., 2017). The risks of being obese are increasingly well understood in the general population and, as a result, achieving a healthy weight is increasingly desirable. The case study used in this thesis is that of a weight loss maintenance (WLM) intervention. WLM is becoming increasingly important to public health researchers. This is because, whilst a vast amount of research has been undertaken to improve the lifestyles of those that are obese, little focus has been given to the maintenance of lifestyle improvements following successful and significant weight loss (Dombrowski et al., 2014). Studies have shown that people often fail to maintain weight loss over time, creating a 'yo-yo effect' where individuals' weight fluctuates regularly (Avenell et al., 2004). Specifically, the interventions used in the stated preference tasks in the empirical work of this thesis are specifically based upon a WLM intervention from the NULevel RCT. The NULevel trial was undertaken by a team of researchers at Newcastle University (Evans et al., 2015) and is outlined in detail in section 6.1.

The work contained in this thesis is independent of the NULevel trial itself. NULevel trial participants that had consented to be contacted for future studies were invited to participate in this study, which went through its own ethical approval process (see section 6.8), upon completion of the trial. However, members of the NULevel trial team were involved at the 'defining attributes and levels' stage of the DCE design (see section 6.4).

1.5 Thesis Structure

This chapter has provided an overview of the setting and the motivation behind the thesis. In addition, the research questions and the case study have been outlined. All of this is developed further and in greater detail in the subsequent chapters of this thesis.

Chapter Two outlines the economic theory that underpins the various methodologies used in this thesis. This includes an overview of consumer choice theory and welfare changes, as well an explanation of different types of value. It finishes by outlining key concepts in welfare economics and the theoretical foundation of CBA.

Chapter Three follows on from Chapter Two by discussing the use of economic evaluation methods such as CBA and CUA in health economics. It also provides a narrative review of the 'preference debate' and outlines recent developments in relation to both areas. Following this, CV and DCE methodologies are introduced and outlined. Finally, the use of quantitative preference information from DCEs in healthcare decision-making is reviewed and the 'preference debate' is discussed in this context.

Chapter Four builds upon Chapter Three by systematically reviewing the published DCE literature in health economics. The aim of the review is to determine whose preferences are typically elicited in DCE studies that elicit preferences for healthcare services. Respondent samples are classified as either a patient sample or a general population sample in order to determine whether DCEs are typically used to elicit patient preferences.

Chapter Five reflects the findings of Chapters Three and Four and introduces an alternative framework for classifying respondent samples in DCE studies in terms of their relative experience. This sets out the user groups to be recruited for the empirical work. In addition, this chapter outlines the four research questions from section 1.3 in greater detail.

Chapter Six is a methodological chapter that outlines the case study from section 1.4 in greater detail. The chapter then details the recruitment plan and the overall design of an online survey that is used to address the research questions. The latter involves a set of screening questions to classify the respondents, the design of the DCE and CV tasks, and additional data collection. The chapter also outlines the process undertaken to obtain ethical approval.

Chapters Seven and Eight are results chapters that provide the results for research questions one to three, as well as some discussion points. Chapter Seven provides results from the NULevel trial sample and Chapter Eight provides results from the online panel sample.

Chapter Nine combines the two samples and provides overall results and discussion points relating to all four research questions.

Chapter Ten provides a summary of the results of the empirical work and discusses the implications to different stakeholders such as researchers and policymakers. The chapter also outlines the strengths and limitations of the empirical work, provides suggestions for further research and, finally, concludes the thesis.

Chapter 2. Economic Theory

The aim of this chapter is to outline the economic theory of which the various concepts addressed in this thesis are based upon (e.g. economic evaluation methodologies and CV) or influenced by (e.g. DCEs).

This chapter begins with a section overviewing the basic concepts and assumptions in consumer choice theory. Section two uses these concepts to illustrate how welfare changes can be measured according to economic theory. Section three discusses the value of a good, the existence of 'use' and 'non-use' values, and explains how an alternate theory of consumer demand addresses value. Section four moves beyond the individual consumer by discussing concepts from welfare economics, which is concerned with social welfare. The theory discussed in this final section provides the theoretical basis for an important economic evaluation method (CBA) and hence provides a starting point for the discussion of economic evaluation in health (see Chapter Three).

2.1 The Basics of Consumer Choice Theory

The theory of consumer choice is concerned with the consumer's approach to decision-making in relation to consumption. This section will firstly present an overview of the choice problem and secondly provide further detail on the assumptions that are made with respect to consumer preferences in consumer choice theory.

2.1.1 The Choice Problem

The typical setup of the choice problem is to consider a 'rational consumer' *i* that wishes to consume goods[1] such that his or her utility is maximised. Utility here is used to denote the subjective sensations that are derived from consumption, such as satisfaction or pleasure (Gravelle & Rees, 2004). The problem faced by the consumer is that they are constrained by a finite budget. Hence, the challenge is for the consumer to consume goods in line with his or her preferences and within their budget such that their utility is maximised.

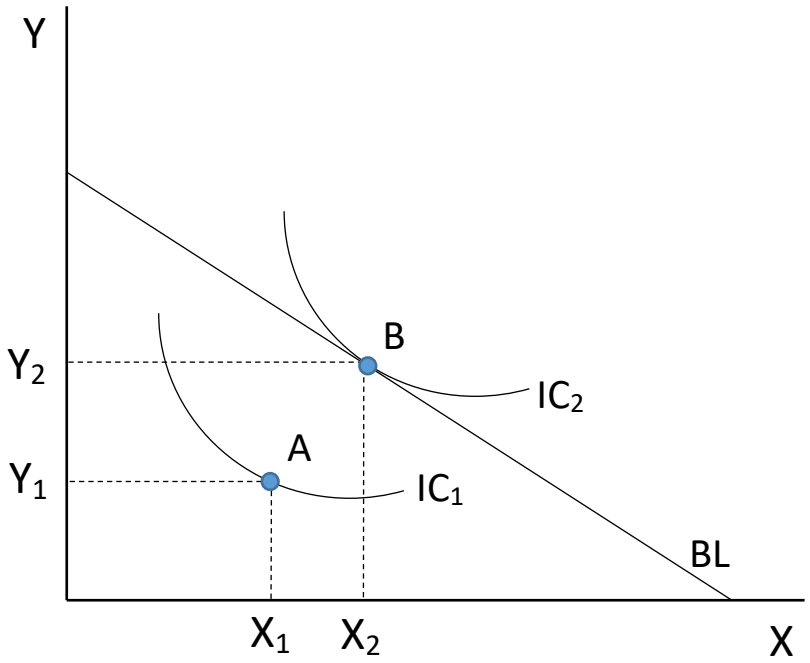
The choice problem is often set out algebraically in a hypothetical world where only two goods exist, typically denoted X and Y. Equation 2.1 sets out the choice problem described above in this world, where U_i is the utility function of consumer *i*, P_x and P_y represent the prices of good X and Y respectively and M represents the consumer's income.

[1] As is convention, the term 'goods' throughout this chapter is meant as an umbrella term for any kind of product or service.

$$\text{Max}\, U_i = f(X, Y)\ subject\ to\ M = P_X X + P_Y Y \qquad (2.1)$$

The choice problem is also often expressed diagrammatically using indifference curves and budget constraints. The former are curved lines that illustrate the different combinations of goods (i.e. bundles of X and Y) that generate a particular level of utility, whereas the latter is a straight line that represents the budget constraint from equation 2.1.

Figure 2.1 Indifference Curve & Budget Constraint Diagram



The budget line (BL in figure 2.1) illustrates the possible combinations of X and Y that can be consumed if the consumer spends all of their income. Hence any combination of X and Y below the budget line is feasible (costing less than the consumer's total income) and any combination of X and Y above the budget line are not feasible (costing more than the consumer's total income).

The two indifference curves in figure 2.1 represent different levels of utility for consumer *i*. A higher indifference curve (away from the origin) represents a higher level of utility, thus the consumer has a higher utility on IC_2 than on IC_1. This is clear when comparing the two bundles A and B. If the consumer consumes bundle A (on IC_1) the consumer receives a combination of X_1 and Y_1, whereas if the consumer consumes bundle B (on IC_2) the consumer receives more of both (X_2 and Y_2). In fact, bundle B is where IC_2 is tangential to the budget line and hence represents a utility maximising point because the consumer cannot

consume a bundle of goods that contains more of one good, but no less of the other, given the funds that they have available.

It should follow that a higher income provides the consumer with the opportunity to reach a utility-maximising point that is on a higher indifference curve as the budget line will shift to the right (they can afford more of X and/or Y). Additionally, as the slope of the budget line is the ratio of prices for X and Y, a favourable change in the ratio (price decreases) will also provide the consumer with the opportunity to reach a higher indifference curve by shifting or tilting the budget line away from the origin. The opposite is also the case; an unfavourable change in the ratio (price increases) will shift or tilt the budget line toward the origin, resulting in the consumer only being able to reach a lower indifference curve.

2.1.2 Preference Axioms

In consumer choice theory a number of assumptions, referred to as axioms by convention, are made about consumer preferences; these are implied in the consumer problem outlined in the previous subsection.

1. Complete Preferences

The consumer has defined preferences for all bundles of goods. That is to say that one of the following will always be true:

i. The consumer prefers bundle A to bundle B ($A \succ B$)
ii. The consumer prefers bundle B to bundle A ($B \succ A$)
iii. The consumer is indifferent between bundles A and B ($A \sim B$)

This implication of this assumption is that any bundle of goods will lie on an indifference curve.

2. Continuous Preferences

There are no ‘jumps’ in the consumer’s preferences. That is to say that if bundle A is preferred to bundle B, other bundles of goods that are very close to bundle A will also be preferred to bundle B.

3. Transitive Preferences

Given three bundles of goods (A, B and C), the consumer’s preferences are such that:

$$If\ A \succ B\ and\ B \succ C\ then\ A \succ C$$

The implication of this assumption, in conjunction with axiom 2, is that indifference curves cannot intersect.

4. Non-satiation

The consumer is not totally satisfied with the amounts of each good that is already obtained and, as such, would always prefer to have more of each good. There is never a situation for the consumer where increased consumption results in a decrease in overall utility (i.e. no good is 'bad'). The implication of this assumption is that any increase in the amount of a good obtained (providing there is no decrease elsewhere) will allow the consumer to reach a higher indifference curve.

5. Diminishing Marginal Rates of Substitution

As more of X is consumed, progressively less of good Y will be sacrificed for each extra unit of X if utility is to be unchanged. This is a consequence of diminishing marginal utility for each good. The implication of this assumption is that indifference curves are convex to the origin (the slope diminishes as we move from left to right along its length).

These assumptions enable consumer choice theory to be explained and illustrated in the manner shown in the previous subsection. Additionally, these assumptions can be tested empirically to determine whether the behaviour of an individual corresponds with that of a rational consumer as defined by economic theory.

2.2 Welfare Changes

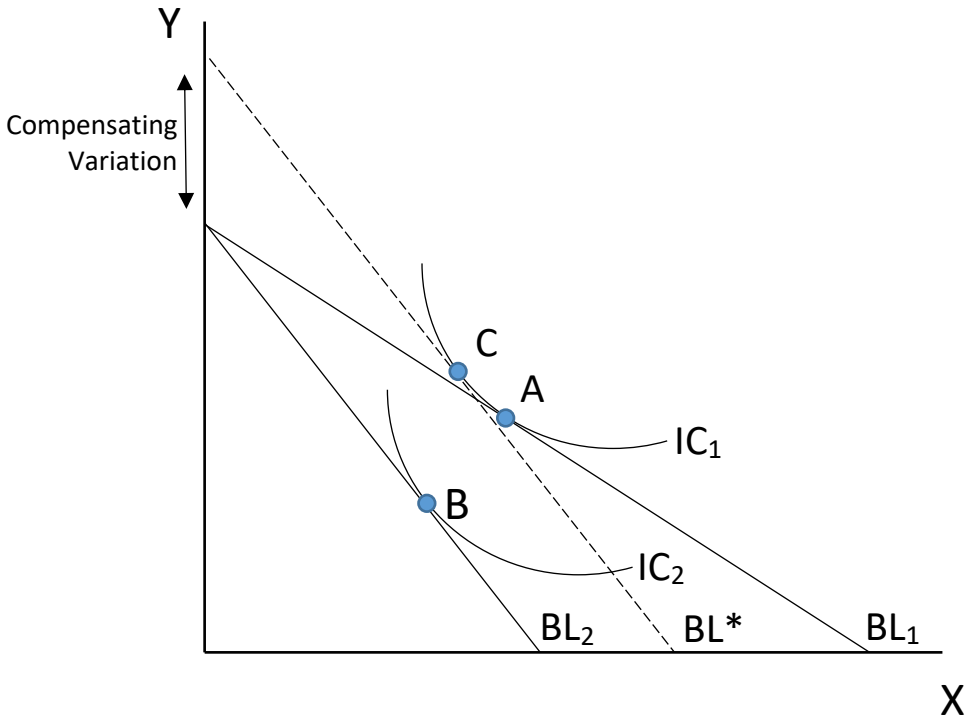
It is often desirable to measure (or estimate) the change in welfare that occurs after an economic change in order to inform future policy decisions. As shown in section 2.1, price changes (just one example of an economic change) directly impact the utility of the consumer because they alter the range of indifference curves that the consumer can reach. The problem that economists face when trying to measure or estimate welfare changes is that changes in utility are unobservable i.e. it is not possible to measure the difference in utility obtained between IC_1 and IC_2 in figure 2.1. As a result, it is necessary for analyses to estimate welfare changes using a measure that is observable. As lump-sum transfers of money can offset economic changes, the common interpretation is that the size of such transfers are proportional to the magnitude of the associated welfare change. Hence, economists use measures based on this principle in order to examine welfare changes. This section will outline the two most common measures used to estimate welfare changes in economics; compensating variation and equivalent variation.

2.2.1 Compensating Variation

The compensating variation is the adjustment in income that returns the consumer to their *original* utility following an economic change. This can be made clearer using an example, such as a price change as previously discussed in this chapter. Consider figure 2.2, set up in a similar fashion to figure 2.1, where the original budget line BL_1 leads to a utility-maximising point A where the indifference curve IC_1 is tangential to BL_1. If the price of good X increases and the price of good Y remains constant the consumer is faced with a new budget line BL_2. On BL_2 the utility-maximising point B is on a lower indifference curve IC_2.

This process can be broken down into two different components: the price rise making the consumer worse-off; and the change in the relative prices of the two goods (Y is now *relatively* cheaper than it was before). The former is the 'income effect' which shifts the budget line and the latter is the 'substitution effect' which rotates it. An intermediate budget line, BL*, that is tangential to the original indifference curve IC_1 at point C can be drawn to illustrate this. In this case, BL* and point C illustrate how the consumer would maximise their utility if the price ratio had changed but their income had not been affected as a result.

Figure 2.2 Compensating Variation



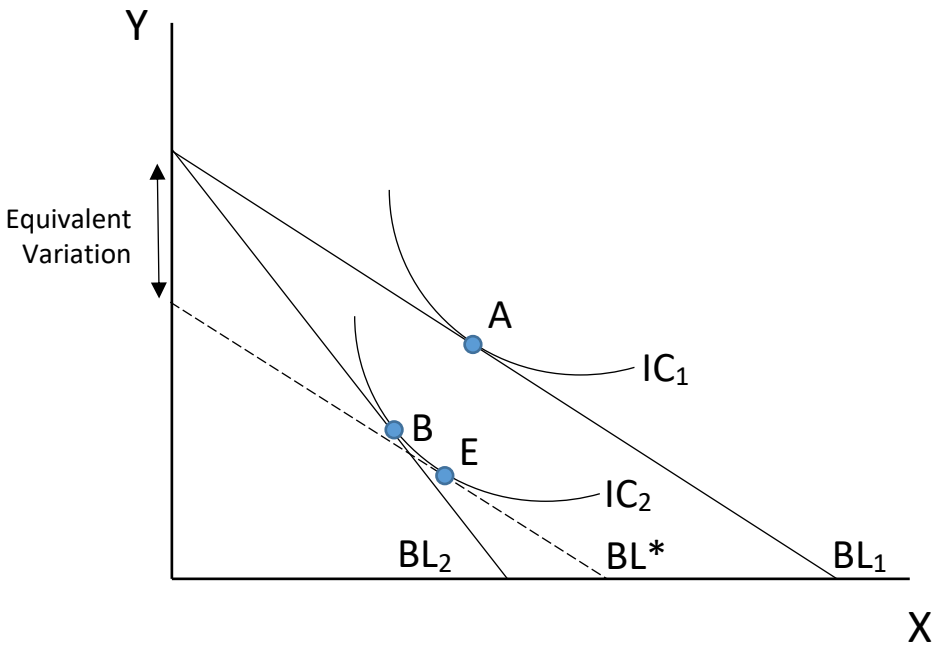
The compensating variation is the magnitude of the shift from the final position of the budget line, BL_2, to the intermediate budget line BL*. It is the minimum amount of extra income that

the consumer would need to be compensated in order to return to the original indifference curve and hence become indifferent to the price change.

2.2.2 Equivalent Variation

The equivalent variation is the adjustment in income that changes the consumer's utility to the level that would occur if the event had happened. Consider again a price increase scenario exactly the same as the one used previously to explain compensating variation. The key difference between the two measures can be seen when comparing figures 2.2 and 2.3. The intermediate budget line BL* in figure 2.3 is instead based on the old price ratio and moved such that it is tangential at point E on the lower (new) indifference curve IC_2. BL* and point E illustrate how the consumer would maximise their utility had their income changed but the price ratio remained the same.

Figure 2.3 Equivalent Variation



The equivalent variation is therefore the magnitude of the shift from the intermediate budget line BL* to the original budget line BL_1. It is the amount of income that would need to be taken away to lower the consumer's utility to the level that would have been experienced had the change happened. Had the example been a price *decrease* in good X (with the price of Y remaining constant), the equivalent variation would instead be interpreted as the increase in income that would give the consumer the same additional utility that they would have received had the change happened.

It is often the case that economists wish to identify the value of a good i.e. the amount consumers are willing to pay for the good. This is not the same as the market price for a good, as at a market equilibrium there will always be consumers that were willing to pay more for the good than the market price[2]. Nonetheless, goods that are available in the market are easier to assess and incorporate into analyses simply because they have a market price and consumers actively reveal their preferences when making purchasing decisions for the good.

There are often situations where goods are not available in the market either due to the fact that they are public goods[3] or because they are simply not provided through the free market e.g. services provided by the National Health Service (NHS) in the UK. In such scenarios, stated preference methods such as CV and DCEs can be used to estimate WTP (i.e. compensating variation) and WTA (i.e. equivalent variation). These methods will be discussed further in the next chapter.

2.3 The Value of a Good

The most common practical application of compensating variation and equivalent variation is when WTP and WTA are estimated. In other words, it is often the case that researchers wish to value goods in a hypothetical setting due to reasons such as those outlined in the previous section. However, it is not necessarily the case that all hypothetical settings give rise to the same types of value. This section will first differentiate between use and non-use values, and describe how Lancaster's theory of consumer demand reflects on value, which is slightly different to consumer choice theory.

2.3.1 Use and Non-Use Values

Eliciting an individual's willingness to pay for a good through stated preference methods provides an estimate of the value of that good to the individual. However, it need not be the case that a good is valued based on the individual's current, or even their likely future, use of the good. That is, individuals may value a good for reasons that relate to more than simply their own expected consumption of the good. In this sense, a distinction can be made between the types of value that can be derived in tasks that aim to elicit willingness to pay i.e. use and non-use values.

One of the most commonly cited examples of a non-use value is the 'option' value (Smith, 2007). This is often seen in environmental economics where individuals express a willingness to pay for the preservation of a natural resource such as a national park. The value expressed

[2] This leads to the consumer surplus, an alternate welfare measure that will not be covered in this chapter.
[3] A public good is defined as a good that is both non-excludable and non-rivalrous, such as street lighting.

is not a use value in the typical sense. The individual is not necessarily stating how much they would pay for access to the national park. Rather, the value provided is considered an option value because they are instead stating that they value having the option to use and benefit from the national park.

Another commonly cited example of a non-use value is the 'caring externality' (Smith, 2007). For example, an individual may express a willingness to pay for the provision of a healthcare service that is not at all relevant to them as an individual (e.g. a gender-specific cancer service). Despite the fact that they will not ever use the service, the individual could certainly value its availability to others and hence the stated WTP value is altruistic and not considered to be a use value.

In practice, the type of value elicited in a hypothetical stated preference task can be determined by the researcher through the wording of the questions. For example, a male participant could, although unlikely in practice, be asked to imagine that they need a health service that is in reality only consumed by females and to provide a WTP value (in this case a use value). It is useful therefore to be aware of the different types of WTP value as they can all potentially play a role in the economic analysis of a good or service.

2.3.2 Lancaster's Theory of Consumer Demand

In contrast to traditional consumer theory outlined earlier in this chapter, in 1966 Lancaster published an alternative theory of consumer demand (Lancaster, 1966). One of the major differences was the notion that individuals derive utility from the characteristics of a good rather than from the good in itself. Hence, while a good could be valued in a stated preference task as a whole (a typical CV task; see section 3.2), it may be more beneficial to value the attributes of a good individually to derive value (possible in a typical DCE task; see section 3.3).

2.4 Beyond the Individual Consumer: Social Welfare

In order to provide economic analyses to support macro-level decision-making, it is necessary to look beyond the utility of the individual consumer and instead focus on the welfare of society as a whole. The field of welfare economics is concerned with the maximisation of social welfare and provides a theoretical basis for such analyses.

This section will briefly outline the key theoretical concepts such as Paretian value judgements, Pareto improvements and Pareto efficiency. Following this, the Kaldor and Hicks criteria will be outlined to provide a theoretical background for future discussion of CBA and related economic evaluation methodologies (Chapter Three).

2.4.1 Key Concepts in Welfare Economics

When moving beyond the individual consumer the problem becomes that of how best to maximise the welfare of society, subject to the constraint of scarce resources. Therefore, in comparison to the previous section on consumer theory, instead of focusing on an individual consumption bundle, the focus becomes that of the complete allocation of resources across society. Hence, the latter not only incorporates the former (i.e. consumption) but also incorporates production.

Maximising social welfare can be split into two components: the identification of a (set of) Pareto efficient allocation(s) of resources and; the identification of a Pareto efficient allocation of resources that optimises societal welfare.

Before defining Pareto improvements and Pareto efficiency, it is important to note that these concepts are all based on a set of value judgements. These Paretian value judgements (sometimes referred to as Pareto postulates) are somewhat controversial and are as follows (Gravelle & Rees, 2004):

1. Process Independence

It is assumed that the way in which an allocation of resources is reached is irrelevant to the maximisation of societal welfare. In other words, it is only the final allocation of resources that matters, not the process.

2. Individualism

It is assumed that societal welfare is simply the sum of every individual's utility. In other words, society is simply considered to consist of a group of individuals.

3. Non-Paternalism

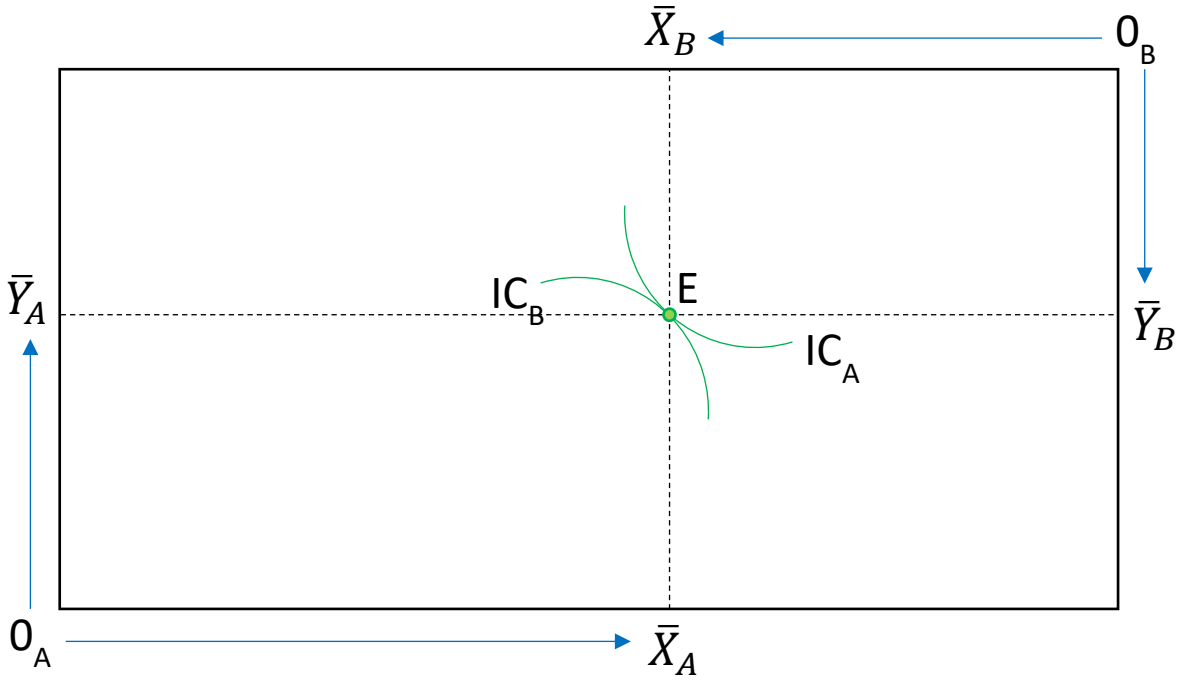
It is assumed that individuals are their own best judge of their welfare. In other words, so long as the individual is maximising their utility, there is no concern as to whether their consumption could be detrimental.

4. Benevolence

It is assumed that an increase in the utility of an individual, with all else remaining equal, will have a positive effect on social welfare. In other words, any potential inequalities between the beneficiary and other individuals are irrelevant to the determination of societal welfare.

Given a specific allocation of resources across all individuals within a society, a Pareto improvement can be defined as the movement to a new allocation of resources where at least one individual is better-off and no individuals are made worse-off. Therefore, it follows that a Pareto efficient allocation of resources is an allocation where it is not possible to improve the utility of an individual without decreasing the utility of another (Varian, 2014).

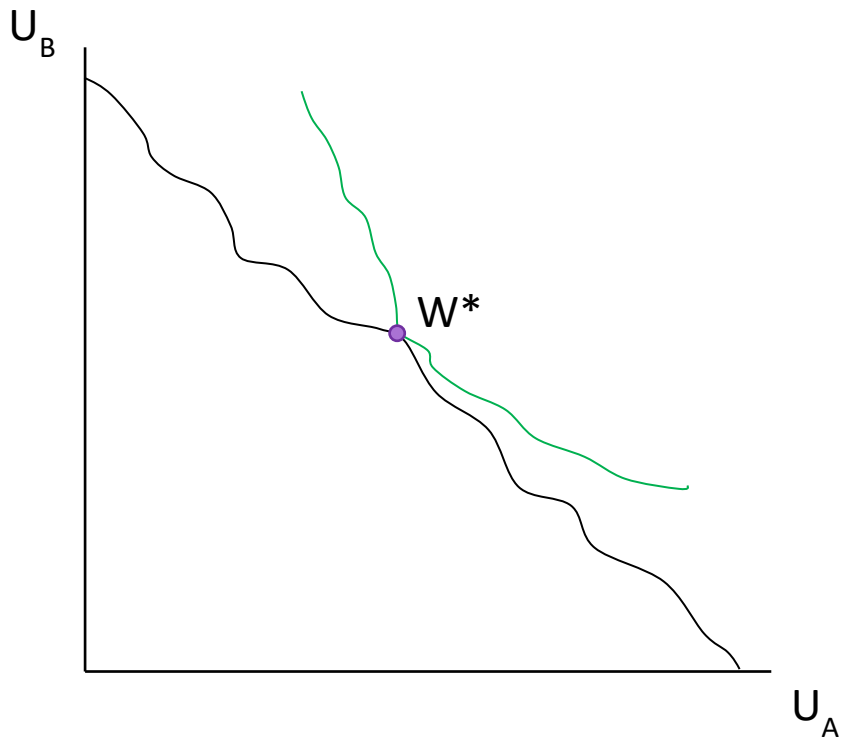
Figure 2.4 Edgeworth Box Diagram



The Edgeworth box in figure 2.4 illustrates the concept of Pareto efficiency. The box depicts a pure trade economy with two individuals (A and B) and two goods (X and Y). The width of the box represents the total amount of good X in the economy and the height of the box represents the total amount of good Y in the economy. At the origins 0_A and 0_B, neither individual has any units of either good. However, this is not the allocation of resources in this economy, as indicated by the indifference curves. IC_A is an indifference curve of individual A, which is convex to A's origin at 0_A. Likewise, IC_B is an indifference curve of individual B, which is convex to B's origin at 0_B. At point E where the two indifference curves are tangential, individual A receives $\bar{X}_A$ units of good X and $\bar{Y}_A$ units of good Y. Similarly, individual B receives $\bar{X}_B$ units of good X and $\bar{Y}_B$ units of good Y. At this point, the total endowment of the two goods is allocated between the two individuals. Now imagine a movement from point E to the north-east; at this point, individual A can reach a higher indifference curve (i.e. this is further from A's origin) but individual B can only reach a lower indifference curve (i.e. this is closer to B's origin). This would not be a Pareto improvement because individual B would be made worse off. In fact, no movement from point E would

result in a Pareto improvement. Hence, point E is a Pareto efficient allocation of resources; one of many possible such allocations within the Edgeworth box[4].

Figure 2.5 Utility Possibility Frontier & Welfare Optimum



It is not the case that a Pareto efficient allocation of resources (a Pareto optimum) is necessarily a welfare optimum (although the reverse is true). That is, there are numerous possible allocations of resources that could be Pareto efficient but only one point that maximises welfare. This is illustrated in figure 2.5; W* indicates a single welfare optimum on the utility possibility frontier (Estrin et al., 2012). The welfare optimum is dependent on the welfare system adopted by the society in question. For example, a society may take an egalitarian view where social welfare is improved only if individual utilities are made more equal. In this case, the welfare optimum would be found at the Pareto optimal point that produces the most equal distribution of utility across individuals.

2.4.2 The Theoretical Foundation of Cost-Benefit Analysis

It should be clear from the previous section that a movement from an existing allocation of resources to a new allocation of resources where one or more individuals are made worse-off

[4] Although Pareto improvements might not be possible from points such as point E in figure 2.4, there are many possible locations in an Edgeworth box where an indifference curve for individual A could be tangential to an indifference curve for individual B. Hence, there are many possible Pareto efficient points. The locus of these tangential points, known as the contract curve, is sometimes drawn to illustrate this point.

is not a Pareto improvement. In practice, achieving a Pareto improvement is likely to be highly difficult. However, the Kaldor and Hicks criteria (sometimes jointly referred to as the compensation test or principle) can provide a justification for the reallocation of resources in circumstances where some individuals are made worse-off.

The Kaldor criterion states that if it is possible at the new allocation for the 'winners' (i.e. those that are made better-off) to jointly compensate the 'losers' (i.e. those that are made worse-off) and that the winners are still better off than they previously were, then the movement to the new allocation should be made. In a similar vein, the Hicks criterion states that if, before the change, the 'losers-to-be' could jointly bribe the 'winners-to-be' so that without the change at least one person is better off and nobody is made worse off, the movement to the new allocation should not be made (Brazier et al., 2016). One of the major issues with the Kaldor and Hicks criteria is that they do not require the compensation to actually be paid, inevitably leading to some individuals being made worse-off and no Pareto improvement being achieved.

Nonetheless, the operationalisation of the Kaldor criterion where the gains to the winners are referred to as 'benefits' and the losses to the losers are referred to as 'costs' is CBA, which is a commonly utilised approach to economic evaluation. In a CBA, all of the relevant societal costs and benefits of an intervention are measured in commensurate units (often money).

2.5 Conclusion

This chapter has outlined the key economic theory that forms the basis of the concepts and techniques that will be referred to, and put into practice, within this thesis. While the discussion of economic evaluation techniques is central to the work within this thesis, the application of such techniques to health-related issues has involved development that goes beyond the remit of traditional welfare economics. Such developments (i.e. the move to an 'extra-welfarist' approach) sit more comfortably within the sub discipline of health economics rather than economics itself and hence will be addressed separately in the next chapter.

Chapter 3. The Preference Debate and the Use of Quantitative Preference Information in Health Economics

The aim of this chapter is to illustrate how the preference debate in health state valuation can be readily applied when considering the use, and potential uses, of quantitative preference information in healthcare decision-making.

In order to achieve this aim the first section in this chapter begins by providing a brief overview of economic evaluation in the context of health and explains the 'preference question' that arises through the use of QALYs. It then explains how economic evaluation guidelines, subsequent published literature and recent economic evaluation developments have shaped the preference debate. The second section provides an overview of the CV method and describes how CV studies have been utilised in health over time. The third section provides an overview of DCEs and outlines how DCE studies have increasingly been published in the health economics literature, in particular those that elicit preferences for healthcare services. The fourth section then describes the numerous potential ways in which the results of a DCE study have, and could be, used to influence healthcare decision-making at both the macro and micro-level. Alongside this, it is argued that these potential applications of DCEs give rise to a debate that is comparable in many ways to the preference debate in health state valuation: whose preferences should be elicited in each application?

3.1 Economic Evaluation & the Preference Debate

3.1.1 Economic Evaluation in Health

In order to maximise output, economic analyses are conducted to examine how best to allocate scarce resources. The Kaldor criterion outlined in section 2.4.2 provides an underlying rationale for CBA; a form of economic evaluation which measures all of the relevant societal costs and benefits of an intervention in commensurate units. Typically conducting a CBA involves pricing the costs and valuing the benefits using existing market prices as a guide. Where markets do not exist, which is often the case with the benefits in a health setting, stated preference methods may be used to provide values[5]. Stated preference methods typically involve surveys where respondents state, directly or indirectly, their WTP for the benefit in question or their WTA compensation for a negative situation. If the benefits of an intervention outweigh its costs (i.e. using the Kaldor criterion 'winners' could

[5] In the past, the human capital approach has also been used for this purpose – see Johannesson (1996).

potentially compensate the ‘losers’; net benefit>0) then resources should be allocated such that the intervention is provided.

The strong theoretical basis for CBA and its ability to directly address the question of whether resources *should* be allocated to a given purpose provides a solid justification for its use. In the United Kingdom (UK), Her Majesty’s Treasury recommend its use for appraisals of policies, programmes and projects in The Green Book (2011). CBA is grounded in welfare economic theory and relies on the assumption that individuals are their own best judge of their welfare, it is therefore often referred to as a ‘welfarist’ approach to economic evaluation.

In health, the welfarist approach is rarely applied (Buchanan & Wandsworth, 2015). One likely explanation for this is the resistance of the medical community and general public towards attaching a monetary value to life. Additionally, CBA requires individual-level assessment of the benefits of a healthcare intervention; this means that the monetary value associated with the benefits of a healthcare intervention must be considered for each individual separately, which is likely to be affected by the individual’s income level. While the potential inequalities caused by the latter could be dealt with by using some form of ‘distributional weights’, the concern around attaching a monetary value to life is unavoidable when using a welfarist CBA to evaluate a healthcare intervention (Brazier et al., 2016).

Cost-effectiveness analysis (CEA) is a form of economic evaluation that is more commonly applied in health and, to an extent, avoids some of the issues relating to the use of CBA. CEA compares different approaches to achieving a particular objective, with the aim of identifying the most cost-effective approach. While CEAs can use a wide range of different measures of benefit, a specific variant of CEA dominates the area of health: CUA. CUAs compare the costs of healthcare interventions with their benefits, where the costs are represented in monetary terms and the benefits are represented using a measure such as the QALY. The QALY is a composite measure of health that combines both quantity and quality of life, where one QALY corresponds to a year in perfect health. QALY gains are calculated by multiplying the change in length of life by the change in quality of life associated with a healthcare intervention. Thus, CUA uses ‘incremental cost per QALY’ as its comparator and hence aims to maximise health rather than individual utility. This approach to economic evaluation in health is often referred to as extra-welfarist for this reason.

3.1.2 Health State Valuation & the Preference Question

For the calculation of QALYs an individual's quality of life is described on a 0 to 1 scale, where 0 represents being dead and 1 represents perfect health[6]. These values are referred to as quality of life weights, or health state utility values (HSUVs), and are generated by eliciting the preferences of individuals such as the general population, patients, health professionals or policymakers.

Various different methods have been used to elicit the preferences of individuals in order to value health states. Early studies typically used rating scales; the most popular, the visual analogue scale (VAS), requires respondents to rate the health state between 0 and 100 (where 0 represents being dead and 100 represents perfect health). It was later argued that choice-based methods could be a better alternative to rating scales because they introduce the notion of opportunity cost (Green et al., 2000). Time trade-off (TTO) and standard gamble (SG) are the two most commonly used choice-based preference elicitation methods for health state valuation.

TTO requires respondents to choose between a given amount of time in the health state that is being valued (followed by death) and a shorter amount of time in full health (also followed by death). The length of time in full health is varied until the respondent expresses indifference between the two choices, and the HSUV is calculated from this indifference point (Torrance, 1986). SG requires respondents to choose between a certain option (a given amount of time in the health state being valued) and a risky option (a gamble between living in full health for the given amount of time, or instant death). The probability of living in full health that renders the respondent indifferent between the gamble and the certain option provides the HSUV (Torrance, 1986).

As the methods for health state valuation developed, various studies in the literature began to explore how preferences for health states might differ according to respondents' characteristics and experiences (Rosser & Kind, 1978; Llewellyn-Thomas et al., 1984; Froberg & Kane, 1989). By the early 1990s studies began to raise an important question regarding the generation and use of HSUVs (Boyd et al., 1990; Hadorn, 1991): whose preferences should be elicited? Should it be individuals that are experiencing the health state (patients) or members of the public that are not, necessarily, experiencing the health state

[6] It is possible, however, that some health states are considered worse than dead and result in negative values (Tilling et al., 2010).

(general population)? This is a particularly important question because, due to the use of HSUVs in CUA, the preferences elicited can directly influence resource allocation decisions.

3.1.3 The Washington Panel & Economic Evaluation Guidelines

As the use of economic evaluation methods in health grew, guidance on how best to conduct CUAs became increasingly desirable. In 1996, a group of academics convened for an in-depth discussion regarding the use of CUA for the economic evaluation of healthcare interventions. The group, now widely referred to as the 'Washington Panel', provided a range of guidance for the conduct of CUAs (Gold et al., 1996). As CUAs use QALYs as the measure of benefit, the Washington Panel discussed a number of issues regarding their calculation, including the question of whose preferences to elicit when generating HSUVs.

The Washington Panel recommended that preferences from the general population should be elicited in order to generate HSUVs. The main argument put forward for this viewpoint was that, due to the influence of CUAs on resource allocation, such evaluations should adopt a societal perspective and incorporate societal preferences. It was also suggested that general population preferences may theoretically be suitable based on the 'veil of ignorance' thought experiment discussed by the likes of Rawls and Harsanyi (Gandjour, 2010). It is theorised that, when acting under a veil of ignorance, a rational public would decide the best (morally acceptable) course of action, as individuals would be blind to their own self-interest. The veil of ignorance is relevant here because with health conditions it is often the case that individuals in the general population have the potential to be affected by them but do not know if this will be the case in the future. In contrast, patients by definition *are* affected by the health condition and hence have a vested interest in decision-making relating to it; they are not likely to be neutral judges.

Arguments for using patient preferences were also considered by the Washington Panel. A key argument is that patients have actually experienced the health state in question, unlike the average member of the general population. In addition, the general population may have biased preferences due to stereotypes or a fear associated with certain conditions, such as cancer (Kahneman, 2006). However, the fact that patients are experiencing the health state has been used as a criticism due to the issue of 'adaptation'. Patients may adapt to their health state and revise their expectations regarding quality of life once they are faced with a chronic condition, resulting in health state valuations that are higher than what might be expected. This has the potential knock-on effect of healthcare interventions appearing less cost-effective (as the benefits could appear less substantial), potentially reducing the likelihood of resources being allocated for the intervention. As a result, it has been suggested that using patient

preferences may actually work against patients' interests, although this is another normative debate in itself (i.e. should adaptation be taken into account or not?).

The advice of the Washington Panel is thought to have been highly influential in the generation of guidelines by regulatory authorities, such as those produced by the National Institute of Health and Care Excellence (NICE) in England and Wales and the Canadian Agency for Drugs and Technologies in Health (CADTH) in Canada (Brazier et al., 2005). Both NICE and CADTH recommend the use of HSUVs generated from general population preferences in their guidelines to this day (NICE, 2013; CADTH, 2017). However, it should be noted that this is not the case in every country, such as in Sweden where utilities from those experiencing the health state are preferred (Pharmaceutical Benefits Board, 2003).

3.1.4 Further Preference Debate Developments

The Washington Panel's recommendations, as well as other guidelines, for the economic evaluation of healthcare interventions provided a practical answer to the question of whose preferences to use when generating HSUVs for use in a CUA. In fact, the second edition of the influential text produced by the Washington Panel has since reiterated their previous recommendation (Neumann et al., 2016). However, this debate has nonetheless remained a regular feature in the health economics literature due to its normative nature (Brazier et al., 2005; Stamuli, 2011; Versteegh & Brouwer, 2016).

A number of studies continued to explore the extent to which patient and general population HSUVs differ (de Wit et al., 2000; Ubel et al., 2003; Oldridge et al., 2008; Little et al., 2014; Papageorgiou et al., 2015) and a relatively recent meta-analysis has verified that disparities typically do exist between the preferences of the two populations (Peeters & Stiggelbout, 2010). More specifically, it is generally found that patients provide larger HSUVs than members of the general population for a given health state. That is, the loss of utility as a result of a deviation from full health is smaller for those that have actually experienced the (suboptimal) health state. These findings confirm the importance of the debate, given that it would make no difference whose preferences are incorporated if they were identical.

Other studies have put forward frameworks to describe how and why these disparities may occur (Dolan, 1999; Stiggelbout & de Vogel-Voogt, 2008). Stiggelbout & de Vogel-Voogt (2008) created a framework that aims to describe the decision-making process in health state valuation based on theories of information processing from social cognition theory. Their framework provides a range of ideas as to why disparities exist such as the lack of scope in

health state descriptions and a range of potential biases that might affect interpretation, judgement and the final response to a health state valuation task.

Dolan (1999), on the other hand, focuses largely on how personal experience, both direct and indirect, may affect an individual's preferences. The typical presentation of the preference debate (patient vs. general population) is criticised, with Dolan arguing that this is a false dichotomy as "the distinction between those with and without experience of illness is very blurred" (p.483). This argument is developed slightly further and applied to another area of preference elicitation in Chapter Five.

Gandjour (2010) examined the theoretical basis for using general population values, by considering the ethical theories cited by the Washington Panel in their justification for the use of general population values (e.g. theories that use the veil of ignorance). In addition to this, Gandjour (2010) considered Sen's capability approach (Sen, 1985) and traditional welfare economics as theoretical justification for the use of general population values. His conclusion was that there is no compelling theoretical basis for their use in this context.

An alternative approach to this debate was taken by Dolan & Kahneman (2008). They suggest that the standard debate considers only types of 'decision utility' (the preference or desire for an outcome), whereas it may be more appropriate to consider 'experienced utility' (the hedonic experience of an outcome). On the one hand, this may sound like support for patient preferences (only patients experience the health state), however this ultimately implies that HSUVs may not be appropriate at all because they are only a measure of decision utility regardless of the source (Dolan, 2008). If, as suggested by Dolan (2008), QALYs are too narrow whereas broader measures such as subjective well-being (SWB) may be more appropriate. A further implication is that a typical CUA may simply be too narrow for assessing the cost-effectiveness of healthcare interventions.

The commonly cited points that are raised in the preference debate have recently been developed even further. Versteegh & Brouwer (2016) revisited the debate and, by examining the most commonly cited arguments for using general population preferences, showed that the common justification for the use of general population preferences is incomplete at best. For example, the authors argue that a general population sample does not exclude individuals that are unhealthy and hence may expect to experience a particular health state, which means that the veil of ignorance argument is imperfect. Additionally, the authors argue that under a veil of ignorance individuals, given that they may struggle to imagine health states, may actually prefer that individuals with experience value them. The authors also question whether

adaptation is a good argument for using general population preferences, pointing out that patients might not always benefit from the lower HSUVs typically reported by the general population. Furthermore, the fact that adaptation is also likely to occur when patients describe their health state is highlighted to show that adaptation cannot be entirely avoided by focusing on general population preferences alone. As a result, the authors conclude that it may be best for economic evaluations to use two sets of HSUVs, utilising values from both the general population and patients, in order to provide decision makers with all of the necessary evidence.

3.1.5 Further Economic Evaluation Developments

While CUA is still the recommended method for economic evaluation method in health in many countries (Rowen et al., 2017) there are several concerns about its narrowness, largely due to the use of QALYs as the measure of benefit. QALYs only consider the health of an individual in terms of quality and quantity of life, whereas it may be the case that other outcomes should be considered too (Coast et al., 2008; Mitchell et al., 2017).

A growing literature is emerging that attempts to operationalise Sen's capability approach (Sen, 1985) for the purpose of economic evaluation in health due to it being considered a broader outcome measure. For example, the "ICEpop CAPability measure for Adults" (ICECAP-A) measure has been developed with an accompanying tariff such that it could be used as an alternative to QALYs (Al-Janabi et al., 2012; Flynn et al., 2015). While using the ICECAP-A might go some way in addressing the narrowness of CUA, it is not free from the issues raised in the preference debate (whose preferences should be elicited in the valuation task?) and does not measure experienced utility as advocated by Dolan (2008).

Another issue with CUA arises when evaluating the cost-effectiveness of public health interventions. Public health interventions aim to promote health within a community or population and hence are broader by nature than clinical interventions that are concerned with individual health. As a result, the benefits of a public health intervention are likely to be far broader than those of a clinical intervention. The use of QALYs in a CUA is perhaps better suited to assessing the benefits of the latter than the former. This has been taken into consideration by NICE in England and Wales; they acknowledge that CBA may sometimes be the most appropriate method to evaluate the cost-effectiveness of public health interventions (NICE, 2014). Specifically, NICE state that in some circumstances:

> *"The tools used for economic evaluation must reflect a wider remit than health and allow greater local variation. The nature of the evidence and that of the outcomes being measured may place more emphasis on cost–consequences analysis and cost–benefit analysis for interventions in these areas."* (NICE, 2014, p.135)

Given that, as discussed earlier, CBA is a welfarist approach and CUA is an extra-welfarist approach to economic evaluation, the advice from NICE to an extent creates an ideological debate when conducting economic evaluations in England and Wales. A public health intervention evaluated using a CBA considers the preferences of the individuals affected by the intervention, whereas a clinical intervention evaluated using a CUA considers the preferences of the general population. Which is the more appropriate approach when trying to influence decision-making surrounding the allocation of resources? Further to this, does this advice suggest that a welfarist CUA (using health state utilities elicited from patients) or an extra-welfarist CBA (eliciting WTP from the general population), both of which are possible (Brazier et al. 2016), would be equally acceptable where the methodology is appropriate?

Economic evaluation methodologies and guidance are constantly evolving to ensure that healthcare interventions are assessed appropriately. This is important in ensuring an efficient allocation of resources, regardless of the desideratum (be it maximising population health or social welfare). A major focus in the health economics literature appears to be the narrowness of CUA which, in addition to the points discussed above, may also be influenced by the increasing pressure to achieve patient-centred healthcare (given that extra-welfarist CUA does not incorporate patient preferences) (Barry & Edgman-Levitan, 2012).

3.1.6 Conclusion

Over time, the preference debate and different approaches to economic evaluation have developed considerably. It would seem that quantitative preference information including, but not limited to, HSUVs have a role to play in informing healthcare decision-making. At this stage, there is nothing to suggest that the importance of such information will decline over time. In fact, there is evidence to suggest that the number of publications utilising stated preference methods such as CV and DCE is increasing (Smith & Sach, 2010; de Bekker-Grob et al., 2012; Clark et al., 2014; Vass, Rigby et al., 2017). This suggests that it may be increasingly important for individuals' preferences to be elicited and, potentially, incorporated into healthcare decision-making. What is less clear is how the preference debate outlined in this section may apply to the use of CV and DCE methodologies. The following two sections examine the application of these methodologies in further detail.

3.2 Contingent Valuation: An Overview

CV is a hypothetical and direct survey-based method used to estimate the monetary value of a good or service (Diener, 1998; Klose, 1999). Monetary values elicited using CV can be used in CBAs when a market for the good or service of interest does not exist. In a CV survey, individuals are asked the maximum that they are willing to pay or the minimum amount of compensation that they are willing to accept for a particular good or service or, sometimes, a broader scenario. WTP is an approximation of the compensating variation and WTA is an approximation of the equivalent variation (recall sections 2.2 and 3.1.1). In the context of health, CV is typically used to elicit the monetary value of a healthcare intervention (Olsen & Smith, 2001; Smith & Sach, 2010). CV is a particularly useful tool in this context given that, in many countries, the healthcare system is publicly funded and thus there is no true market price for healthcare interventions. Furthermore, even in countries where there is a market, market failure may mean that the market price is not a reflection of the true value.

The simplest format of a CV survey is open-ended, where an individual is asked to provide a figure relating to their maximum WTP, or minimum WTA, themselves (Johannesson et al., 1991). Another format is the bidding-game where respondents are asked if they are willing to pay at a given price and, based on their response, are subsequently asked if they would pay at a different price; the prices provided are based on a pre-determined algorithm designed by the researcher (McNamee et al., 2010). Other formats include the payment card where respondents are given a list of potential values and asked to select the maximum that they would be willing to pay (Stewart et al., 2002), and dichotomous choice approaches where respondents are typically asked up to two 'yes or no' questions (Ryan et al., 2004). The various formats have been contrasted and critiqued in the literature over time (Frew et al., 2004; Ryan et al., 2004; Whynes et al., 2004). However, despite an attempt dating as far back as 1993 to agree upon a set of guidelines (Arrow et al., 1993), no clear consensus exists on the best format to use e.g. the least biased and the most precise.

Systematic reviews of the use of the CV method in health have explored how the method has been used over time (Olsen & Smith, 2001; Smith & Sach, 2010). Figure 3.1 illustrates how the number of CV studies has increased over time. The use of the method in health dates back to at least 1985, with notable increases from the early 1990s, although the peak in 2005 corresponds to fewer than 40 published studies. Smith & Sach (2010) note that the method has been used to value drugs, surgery, screening and various other healthcare interventions.

Smith & Sach (2010) also found that the majority of CV studies elicited use values (recall section 2.3.1) from users that, in this context, are typically patients. In fact, only 24% of the

published studies included in their review elicited preferences from the general population. This is perhaps unsurprising given that WTP and WTA values are likely to be used in CBA, a traditionally welfarist method of economic evaluation. It also follows that the output from CV studies arguably has limited use in healthcare decision-making given that CBA is rarely utilised in practice. Indeed, Smith & Sach (2010) note that CV has had a limited influence in decision-making in healthcare. Whilst their review only goes as far as 2005, a recent study has provided evidence to suggest that DCEs are becoming more popular than CV in health as well as in other areas of applied microeconomics (Mahieu et al., 2014). It may be the case that the criticisms of the methodology and its limited potential for use in decision-making in contrast with DCEs are key factors in this finding.

Figure 3.1 Published CV Studies in Health Economics Between 1985 and 2005

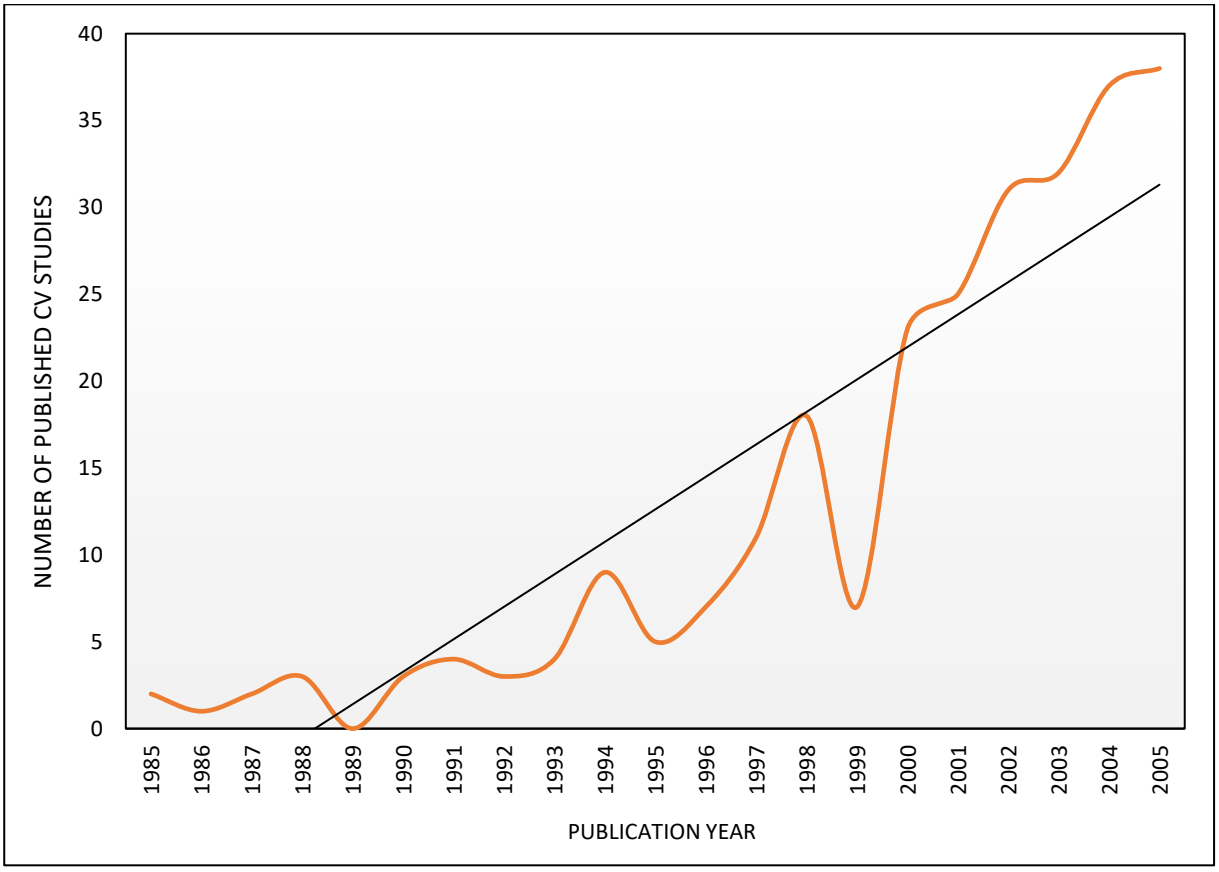


Figure produced using data from Smith & Sach (2010)

3.3 Discrete Choice Experiments: An Overview

DCEs have emerged as an increasingly popular stated preference methodology that is regularly applied in health. A DCE survey requires respondents to choose between two or more alternatives, such as different types of healthcare interventions that are described using a static set of attributes, with alternating levels for each. After a sample of respondents have completed all of the scenarios, the data can be analysed to determine the relative importance

of the different attributes and levels, as well as to estimate marginal rates of substitution (Louviere & Lancsar, 2009).

Although DCEs had been used earlier, they became better known in the health economics literature in the early 2000s after being put forward as a methodology that could potentially be used to incorporate patient preferences into the decision-making process (Ryan, 2004b; Ryan & Gerard, 2003). There are currently three published systematic reviews of DCE studies in health economics covering the period 1990 to 2012 (Ryan & Gerard, 2003; de Bekker-Grob et al., 2012; Clark et al., 2014). In addition, a recent systematic review regarding the use of qualitative research to inform DCE studies has been published that provides overall data on the number of DCE studies beyond this point, up to 2015 (Vass, Rigby et al., 2017). Over this 26 year period, the use of DCE methodology has substantially increased as indicated by the number of publications; figure 3.2 illustrates this trend and the findings in Chapter Four suggest that there is no sign of this trend abating (see section 4.2.1).

Figure 3.2 Published DCE Studies in Health Economics Between 1990 and 2015

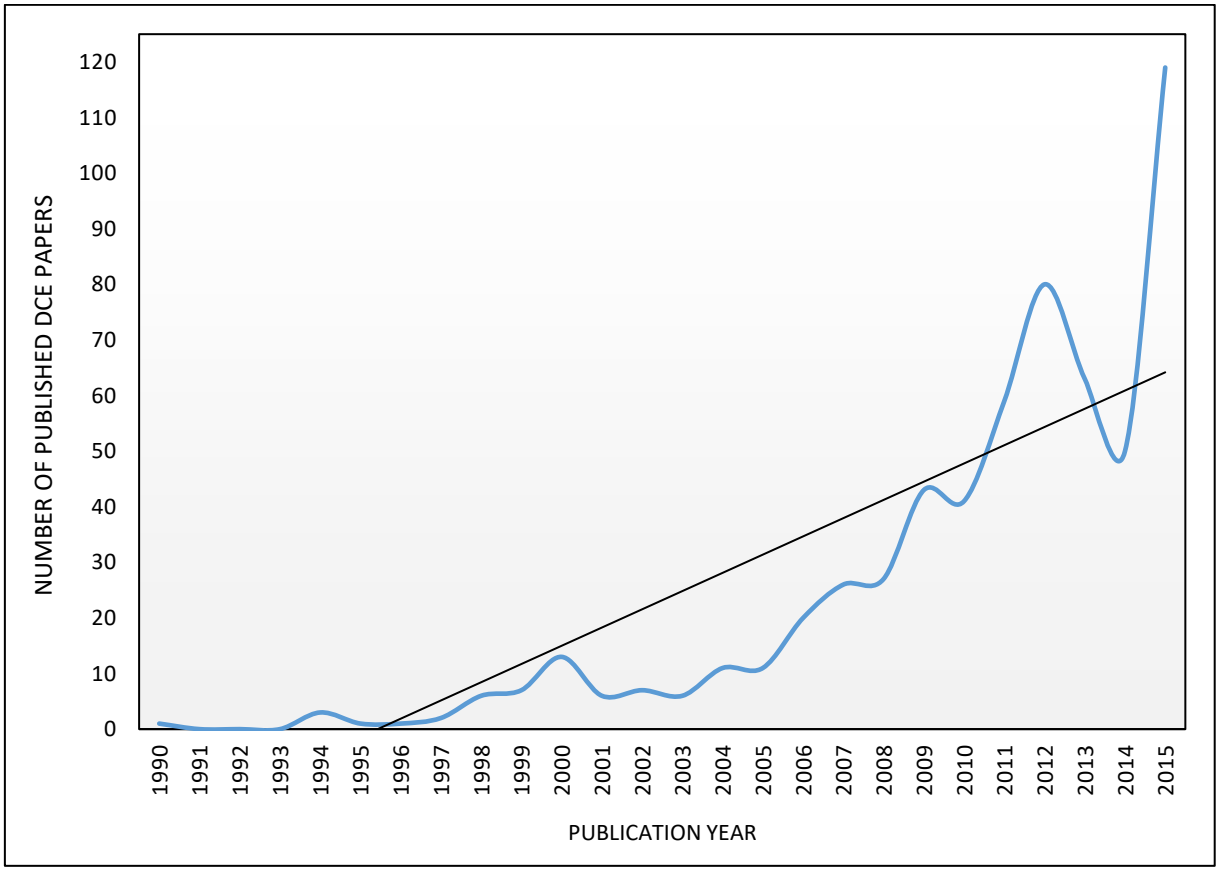


Figure produced using data from Ryan & Gerard (2003), de Bekker-Grob et al. (2012) and Vass, Rigby et al. (2017)[7]

[7] 2009-2012 data was grouped in Clark et al. (2014), hence Vass, Rigby et al. (2017) was used for this period, as well as 2013-2015.

The average number of published DCE studies has continuously increased over time, with 119 studies published in 2015 (Vass, Rigby et al., 2017). The volume of studies may, in part, be due to the flexibility of the methodology. DCEs can be used to elicit preferences in any number of health-related contexts, such as: job choices made by healthcare professionals (Mangham & Hanson, 2008); priority setting in health (Shah et al., 2015); health insurance decisions (Obse et al., 2016); health state valuation (Devlin et al., 2017) and; comparisons of patient and physician preferences (Okumura et al., 2015).

The literature reviews by de Bekker-Grob et al. (2012) and Clark et al. (2014) consider this flexibility by organising the published DCE studies by their main objective using eight categories:

1. Valuing experience factors;
2. Valuing health outcomes;
3. Trade-offs between experience factors and health outcomes;
4. Utility weights within a QALY framework;
5. Job choices;
6. Developing priority-setting frameworks;
7. Health professional's preferences; and
8. Other

The vast majority of published studies fit within the first three categories; 234 of the 327 studies (72%) across the time period covered by de Bekker-Grob et al. (2012) and Clark et al. (2014). Whilst the distinction between these categories is important, in reality the DCEs will have been framed in a similar manner in order to be able to achieve these objectives. That is, respondents will typically have been asked to make a choice between hypothetical alternatives where the alternatives are healthcare services[8]. As such, a logical conclusion based on this information is that the majority of DCEs elicit preferences for healthcare services.

Preference data generated from DCE studies in category four has a clear application in economic evaluation (i.e. to estimate utility tariffs in order to generate QALYs). There is little to say about whose preferences should be elicited that hasn't already been said in the literature outlined in section 3.1. In addition, results from DCE studies in categories five to eight have little clear application to individual economic evaluations.

[8] Here, 'healthcare services' is used as an umbrella term for any treatment, drug, programme or other intervention aimed at improving or maintaining the health of the service user.

In contrast, DCEs that elicit preferences for healthcare services provide preference information that could potentially play a role within or alongside an economic evaluation of a healthcare service (or even a benefit-risk assessment). However, it is not entirely clear how the results might be applied in practice and the question of whose preferences should be elicited has rarely been discussed. In an early editorial regarding the use of DCEs in health economics, Bryan and Dolan (2004) state that:

> *"Whilst patients ex post might place a relatively high value on 'non-health' attributes (as seems to be the case from many, although not all, of the studies), tax-payers ex ante might value a more limited (and possibly more health-focused) set of attributes." (p.200)*

However, there is little evidence to suggest that this has been taken into consideration. In the existing literature, DCE studies that elicit preferences for healthcare services have used both patient and general population samples (Ghijben et al., 2014; Quaife et al., 2017) but the relative merits of each population have not been clearly compared.

This thesis therefore argues that this particular area within the DCE literature is where the preference debate might be most applicable and unique. Hence, the focus will be on DCE studies that elicit preferences for healthcare studies (which, from hereon, will simply be referred to as DCE studies).

3.4 A Further Application of the Preference Debate: Discrete Choice Experiments

In order to discuss whose preferences might be more appropriate to elicit in DCE studies it is important to first identify how the results of DCEs might actually be used in analyses that aim to support healthcare decision-making. This section will discuss how DCEs might be used in macro-level decision-making (section 3.4.1) and whose preferences might be more appropriate to elicit in each case. Following this, for completion, a short note will discuss the proposed use of DCEs in micro-level decision-making (section 3.4.2).

3.4.1 DCEs & Macro-Level Decision-Making

Economic Evaluation

DCEs have rarely been *formally incorporated* into economic evaluations, however those that have been incorporated have done so in several different ways (see table 3.1), each of which will be outlined with examples in this section. There are also other potential ways in which DCEs might be incorporated into economic evaluations that are yet to be put into practice (or,

rather, published); some of these will also be outlined however it is not possible to be exhaustive given the hypothetical nature.

In addition to the generation of utility tariffs, DCEs can also be formally incorporated into a CUA by generating HSUVs that directly relate to the outcomes in a specific economic evaluation. One such example exists where HSUVs were generated for a condition-specific measure of health with the aim of incorporating these values directly into a model-based CUA (Burr, Kilonzo et al., 2007; Burr, Mowatt et al., 2007). The DCE required respondents, a sample of glaucoma patients, to choose the worst description of health relating to living with glaucoma out of two alternatives. The results were then used to generate utility weights for any combination of the attribute levels, such that they could be incorporated into the economic evaluation. As this was one of the first studies that used a DCE for this purpose, the study suffered from the difficulties associated with anchoring health states on a full health-dead scale (i.e. 0 did not represent being dead) and the choice of respondents did not match the NICE reference case guidance of using general population preferences (albeit the measure was condition-specific rather than generic). Nonetheless, this is the only known example of a DCE being formally incorporated into a CUA, where using utilities from a generic tariff generated from a DCE is not considered 'formal incorporation' (e.g. the use of EQ-5D utilities from a health state valuation study that used a DCE).

Another study that formally incorporated a DCE into an economic evaluation used a CEA rather than the more common CUA (Benning et al., 2012). Instead of using QALYs as the measure of benefit, the study used utility more generally by using individual-level preferences elicited from a DCE about different breast cancer follow-up strategies. The DCE required respondents, a sample of patients that had completed breast cancer treatment, to choose between alternative follow-up strategies. The information on individual-level preferences were then combined with cost data in order to examine the cost-effectiveness of various follow-up strategies. This is a unique economic evaluation, not only because it formally incorporated a DCE into a CEA, but because it only considers 'experience factors' on the benefits side (there are no health outcomes included in the DCE) and the preferences incorporated into the analysis were individual-level rather than aggregated.

Table 3.1 Studies that Formally Incorporate a DCE into an Economic Evaluation

Reference(s)	Economic Evaluation	Patient or General Population Preferences?	Purpose of the DCE
Vale (2004) McCormack et al. (2005)	CBA	Patients	To obtain WTP estimates for different types of hernia surgery
Burr, Kilonzo et al. (2007) Burr, Mowatt et al. (2007)	CUA	Patients	To obtain utility weights for a condition-specific preference-based measure for glaucoma
Regier (2008) Regier et al. (2009)	CBA	Patients (by proxy)	To obtain WTP estimates for various genetic tests
Burr et al. (2012)	CBA	General Population	To obtain WTP estimates for interventions relating to ocular hypertension
Benning et al. (2012)	CEA	Patients	To predict individual-level utility for each program (benefit expressed in terms of utility, not QALYs)
Buchanan (2015) Buchanan et al. (2016)	CBA	Patients	To obtain WTP estimates for various genomic tests
Tinelli et al. (2016)	CBA	Patients	To obtain WTP estimates for a collaborative medicine review service between community pharmacists and general practitioners
Boyers et al. (2016) Ramsay et al. (Forthcoming) Boyers (Forthcoming)	CBA	General Population	To obtain WTP estimates for different interventions to prevent periodontal disease
Quaife (2017) Quaife et al. (2017)	CUA	General Population	To predict uptake for HIV prevention products (uptake parameters incorporated into decision model)
Watson et al. (2017)	CBA	Patients and General Population	To obtain WTP estimates for various haemorrhoidal treatments

Other studies that have formally incorporated DCEs into economic evaluations did so by using WTP estimates from DCEs in CBAs, a process that has been outlined in some detail in the literature (McIntosh, 2006; McIntosh, 2010). The first known published example of this involved estimating WTP from the preferences of past hernia patients elicited from a DCE about different types of hernia surgery (McCormack et al., 2005) and was used as part of the evidence presented to NICE. A later example involved using the preferences of the general population for different ocular hypertension monitoring services in order to generate WTP estimates for a CBA (Burr et al., 2012). Three examples of DCEs used to generate WTP estimates for use in CBA exist have been identified that formed components of PhD projects but are currently unpublished. The context of one of these studies was genetic testing for idiopathic developmental disability (Regier, 2008). Another study focused on genomic testing for patients with chronic lymphocytic leukaemia (Buchanan, 2015). The context of the third study is preventative interventions for periodontal disease (Boyers, forthcoming; Ramsay et al., forthcoming). Although the CBAs from these projects have, as yet, not been published in peer-reviewed journals, the DCEs from two of them have (Regier et al., 2009; Buchanan et al., 2016) and the DCE from the other study has been presented at a European conference (Boyers et al., 2016).

A recent publication presented another example of using WTP estimates generated using a DCE in a CBA, but also explored how the estimates differed between different individuals (Tinelli et al., 2016). The context was a collaborative medicine review service between community pharmacists and general practitioners, and the data came from a RCT. The authors found that, for all trial participants, WTP for the intervention was -£26.48. However, once the sample was split into the two trial arms, they found that the treatment arm were willing to pay £3.52 relative to -£56.47 for the control arm. The difference was statistically significant and impacted the results of the CBA. The authors also explored how participants that chose to receive the service beyond the standard follow-up time compared with those that ceased to use the service, however this comparison is affected by selection bias. Finally, a recent *Health Technology Assessment* publication used a DCE to generate WTP estimates for various aspects of haemorrhoidal treatments (Watson et al., 2017). Interestingly, the authors elicit preferences from a patient sample *and* a general population sample. The authors explain that it is unclear whose preferences should be elicited for use in healthcare decision-making and cite an article that is based on work from this thesis (Mott & Najafzadeh, 2016).

Given the existing guidance (e.g. NICE guidelines) and theory (i.e. welfarism vs. extra-welfarism) relating to the three common economic evaluation methods there is little new to

say regarding the choice of preferences in the aforementioned studies. One would expect DCE studies that are used in CUAs and CEAs to typically elicit general population preferences, whereas patient preferences might be elicited for use in CBAs. Such trends are not observed among the studies that formally incorporate a DCE into an economic evaluation. However, given the normative nature of the preference debate and the fact that many of these studies may have exploratory in nature it would be unreasonable to critique their choice of sample for their DCE studies. It is also perhaps unlikely that the preference debate was considered when these studies were devised and conducted. Nonetheless, this thesis argues that the choice of sample *should* be an important consideration in any DCE study that aims to provide information for use in an economic evaluation.

In addition, as DCEs can be used to predict uptake for hypothetical health care interventions (van der Pol et al., 2010; Ghijben et al., 2014), it has been suggested that DCE results could be used to provide uptake parameters for use in decision models for economic evaluations. A recent *Health Economics* letter outlines an approach to do this using uptake for human immunodeficiency virus (HIV) prevention technologies as an example (Terris-Prestholt et al., 2016). The study found that uptake is likely to vary significantly according to user characteristics and the efficacy of the technology; as a result, the authors conclude that this approach is likely to be superior to current practice (where assumptions of uniform uptake that are based on expert opinion are common). A PhD project used this approach to incorporate uptake estimates derived from a DCE into a decision model for a CUA of HIV prevention products in South Africa (Quaife, 2017). Thus far, the protocol and the results of the DCE have been published (Quaife et al., 2016; Quaife et al., 2017) but the CUA has not. With regard to the preference question, it would only seem logical to elicit preferences from those that are likely to undertake, or are currently undertaking, a related intervention to those included in the DCE. As a result, it should be patient preferences that are elicited in this case.

It has also been suggested that DCEs could also be used to estimate minimally important differences (MIDs) in trial outcomes. This could be done by designing a DCE that covers the key attributes of interest, estimating the MRS between those attributes and calculating the MID using this information (Hughes, Charles et al., 2016). The example given by Hughes et al. (2016) is the trade-off between remission and survival, generating the MID in the primary outcome measure (probability of survival) between choosing the trial intervention or not. MIDs are often estimated from the perspective of patients or clinicians and hence it would be expected that the DCE would be administered to a sample of either population; hence there is little scope here to argue for the use of a general population.

Finally, DCEs may also be used to provide preference information prior to an economic evaluation. For example, a DCE could be used in conjunction with a RCT of a healthcare intervention before the economic evaluation stage. During the development of the trial intervention, a DCE could be used to inform the design; the trial team could aim to provide the (feasible) intervention that is expected to maximise uptake and/or utility. Alternatively, a DCE could be delivered to trial participants at the end of the trial in order to obtain preference information based on their experience and to use this to refine the intervention further (as well as to generate uptake estimates for an economic evaluation as described in the previous paragraph). Refinement may be particularly useful when the results of the RCT were equivocal. In either case it would seem logical to elicit the preferences of the trial population, which may be a fairly representative sample of the patient population. However, it is impossible to escape the notion of opportunity cost. If healthcare interventions are designed based on patient preferences, it is possible that a costlier intervention is created than is necessary to achieve the desired health outcome. However, it should be acknowledged that an intervention designed based on patient preferences could lead to better adherence. Regardless, given that the cost is borne by the general population in a public funded healthcare system, an argument remains that the general population's preferences should be favoured.

Although relatively few DCEs have been formally incorporated into published economic evaluations and several of the applications listed above are currently hypothetical, there is clear interest from HTA agencies in utilising DCE data. In Germany, the Institute for Quality and Efficiency in Health Care (IQWiG) conducted a pilot study where a DCE was administered to patients to elicit their preferences for antiviral therapy for chronic hepatitis C (Mühlbacher et al., 2017). The purpose of the study was to examine the relative importance of multiple 'patient-relevant outcomes' according to patient preferences. It is argued that focusing largely on a single primary endpoint, as is standard in a RCT, is too narrow and not reflective of the wide range of factors that affect patients. The authors concluded that DCEs could potentially provide useful information to support decision-making, although no final conclusions were drawn.

Similarly, there is an ongoing project at NICE exploring how preference elicitation methods could be used in a HTA setting. This exploratory project will involve providing a survey to, and conducting a focus group with, myeloma patients in order to determine their needs and subsequently the most appropriate method(s) to use, which may include DCE (L Cowie, Personal Communication, 11 May 2017). This work, in collaboration with Myeloma UK, will continue through to 2018. The focus on patient preferences is interesting when contrasted

with the NICE reference case that recommends general population preferences for HSUVs. One might argue that this guidance is slightly outdated if the intention is to achieve a more patient-centric HTA process (Mühlbacher, 2015). Indeed, a focus on patient (or experienced) HSUVs is supported by various recent preference debate publications outlined in section 3.1. Nonetheless, a normative argument can be made that patients should not be the only ones to have their preferences considered in the healthcare decision-making process.

Regulatory Decisions

It should also be noted that health-related DCEs may also be used to provide preference information for other types of macro-level decision-making. Regulatory bodies such as the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) evaluate medicinal products with regards to their quality, safety and efficacy. Using England & Wales as an example, the EMA would determine whether a medicine can be provided (e.g. whether it is safe enough to justify its use) whereas NICE determine whether it should be made available on the NHS (e.g. whether it is cost-effective).

In a move that is likely to have been motivated by the push towards patient-centred healthcare the Center for Devices and Radiological Health (CDRH), part of the FDA, has recently incorporated preference information from a DCE in a benefit-risk assessment (BRA) for a weight loss device[9]. The DCE contained several attributes describing risks and benefits relating to the device and hence benefit-risk trade-offs could be examined, which is key for BRA (Ho et al., 2015). Subsequent guidance has been published by the CDRH, along with another FDA centre (Center for Biologics Evaluation and Research; CBER), which reaffirms the importance of patient preferences and states that DCEs are a suitable method for eliciting patient preferences for this purpose (FDA, 2016). The passing of legislation such as the FDA Reauthorisation Act and the 21st Century Cures Act has reinforced the focus on patients, and has led to the FDA producing a "plan for issuance of patient-focused drug development guidance" (FDA 2017), which will lead to a series of new guidance documentation being finalised between 2019 and 2021. It would seem likely that this guidance will lead to a rise in the collection and incorporation of patient preference data in regulatory decision-making in the US.

In addition to the use of DCEs in the US, a recent project by the EMA under the Innovative Medicines Initiative (IMI), called the Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium (PROTECT) project, concluded the DCEs were the

[9] For a comprehensive overview of the developments that led to this CDRH study, see Irony et al. (2016)

preferred “utility survey technique” of those that it explored (Hughes et al., 2013; Hughes, Waddingham et al., 2016). Another, ongoing, IMI project called PREFER is looking into this area by taking a broader perspective. The project aims to “develop a systematic approach for considering the use of patient preferences across the medical treatment life cycle”, with final recommendations due in autumn 2021 (de Bekker-Grob et al., 2017).

With regard to the question of whose preferences to elicit in the above examples, it would seem logical to elicit preferences from existing or likely future patients. The guidance from the CDRH only refers to the elicitation of patient preferences, whereas the PROTECT report is less direct but does state that patient preferences are considered important by many stakeholders (Hughes et al., 2013). In addition, the PREFER project is clearly focused on patient preferences. While the preference debate may not have traditionally been considered in the context of BRA directly, it was certainly relevant given that many BRAs use QALYs as the measure of benefit. Assuming that utility tariffs are commonly applied in such cases, it would appear that BRAs have often been incorporating general population preferences given that most tariffs are based on general population preferences (Brazier et al., 2016). It is hard to argue against the suggestion that patient preferences regarding benefit-risk trade-offs are perhaps more relevant, but many issues raised in the preference debate in the context of health state valuation cannot be avoided even when DCEs are used to elicit preferences. For example, patients have a vested self-interest in decisions relating to them and any implication that a study may affect a regulatory decision could lead to biased results. A DCE study found that patients expressed a preference for new anticoagulant therapies over existing therapies regardless of their related benefits and risks (Najafzadeh et al., 2014). As a result, this thesis argues that it is just as important to consider the choice of sample (and, perhaps more importantly, the wording/design of the survey due to the previous example) for a DCE that is intended to contribute to regulatory decision-making.

3.4.2 DCEs & Micro-Level Decision-Making

As DCE studies typically provide aggregate-level information about the preferences of a particular sample, it is somewhat questionable that DCEs can be used to aid micro-level decision-making. Attempting to reach conclusions regarding an individual’s preferences from aggregated information is fundamentally flawed; this is the ecological fallacy (King, 2013). A recent debate arose in the *European Journal for Person Centered Care* on this very issue. Kaltoft et al. (2015) argue that DCE studies should not claim to offer any clinically relevant information (e.g. to assist person-centered care/shared decision-making) for this very reason and use several examples from the DCE literature to illustrate the ‘false’ claims that might be

used (Kaltoft et al., 2015). However, a response by Pedersen et al. (2015) offers a more optimistic view on the matter. They state that while DCE results could not be used in isolation to assist clinical decision-making (between a doctor and patient) they can be used as a starting point and provide a better insight into what the patient's preference might be based on their characteristics (as subgroup analysis is common in DCE studies and indeed in many clinical studies) (Pedersen et al., 2015). In this sense, DCEs contribute to breaking down information barriers between the doctor and the patient. Regardless of the extent to which DCEs may influence micro-level decision-making, it would be hard to argue against the use of patient preferences when the aim is to reduce the asymmetric information between doctors and patients.

3.5 Conclusion

This chapter provided an overview of economic evaluation in health and the preference debate that arose from the use of QALYs in CUA. Following this, CV and DCE methodologies were outlined. Given the abundant potential of the method, the use of DCEs in healthcare decision-making was considered and suggestions were made surrounding the choice of whose preferences to elicit in each context. It should be clear that the vast majority of arguments made in the preference debate could also be made in the context of DCEs when the intention is for the results to be incorporated into analyses that will influence decision-making. Unlike the systematic review of CV studies by Smith & Sach (2010), none of the major systematic reviews of DCE studies in healthcare (Ryan & Gerard, 2003; de Bekker-Grob et al., 2012; Clark et al., 2014) have explored *whose* preferences are being elicited. To fill this gap, the next chapter will review the published DCE literature in order to examine whose preferences are typically elicited.

Chapter 4. Whose Preferences Are Elicited in DCEs? A Systematic Review

The aim of this chapter is to examine the type of respondent sample that has been used in published DCE studies that elicit preferences for health services, and to examine any existing studies that have compared patient and general public preferences for a health service.

In order to achieve this aim the first section in this chapter will outline the review process that was used to identify DCE studies that elicit preferences for health services and the definitions used to classify respondent samples. Section two outlines the main results from the review and section three discusses the implications of the findings and discusses the potential differences in preferences between different types of respondent sample. Finally, section four concludes the chapter.

4.1 Outline of the Review

4.1.1 Review Objectives

The previous chapter outlined how DCE studies that elicit preferences for healthcare services could be used to inform decision-making in various settings. Given that DCEs provide preference data, this gives rise to a normative debate surrounding whose preferences should be elicited. As the direct use of DCEs in healthcare decision-making is relatively novel, there is little insight to be gained by examining the choice of respondent sample in the existing examples. Hence, the next best option in lieu of substantial applications of DCE data in healthcare decision-making would be to search the published DCE literature and examine the choice of respondent samples. It may be the case that relatively few of the published DCE studies were planned with the intention of the results being used to support a specific decision. Regardless, all research intends to provide useful information that addresses a gap in the collective knowledgebase. Therefore, all published DCE studies have the *potential* to influence decision-making, if only indirectly.

Consistent with the preference debate in health state valuation, this review will only compare the number of DCE studies that elicit the preferences of patient or general public samples. It is assumed that DCE studies eliciting the preferences of physicians or policymakers are far less likely to be used in the various decision-making processes outlined in Chapter Three and hence these are not of interest. This is also consistent with the preference debate in health state valuation; physicians and policymakers could also value health states but are relatively rarely ever considered in this area of literature. For simplicity, it is also assumed that DCE studies that make comparisons between different respondent samples primarily aim to provide

information on the potential discordance between groups and thus are less likely to be relevant. For example, it is quite common for DCEs to be used to examine the principal-agent relationship between patients and physicians (Mühlbacher & Juhnke, 2013). While the preference data from the patient samples could also be used elsewhere in isolation, including such studies under the patient category may unfairly inflate the results and therefore such studies will not be considered in this review. However, any studies comparing the preferences of patient and general public samples are clearly highly relevant and will also be reviewed, separately, within this chapter.

Hence, the objectives of this systematic review are:

1. To examine, and attempt to classify, the samples of respondents that have been used in published DCE studies in the health economics literature.
2. To identify any studies that compare the preferences of both a patient and general public sample using the same DCE.

4.1.2 Respondent Sample Definitions

Classifying a respondent sample as 'patients' or 'general public' in the context of DCEs is somewhat more complex than in the case of health state valuation[10] because the task involves both a health issue and some form of health service. The key difference between the two user groups in the preference debate outlined in Chapter Three is the experience of a health state. To remain aligned with this debate, the definitions outlined in table 4.1 use experience as the main differentiator in this context and are the definitions that will be applied in this review.

Table 4.1 Respondent Sample Definitions

User Group	**Definition**
Patient	Individuals that have been recruited due to their experience of the health issue that the DCE addresses (current or past) and/or experience of existing services similar to those described in the DCE scenarios.
General Public	Individuals that have not been recruited due to any specific experience of the health issue that the DCE addresses or of existing services similar to those described in the DCE scenarios.

[10] It has been argued recently this terminology is also complex in health state valuation. Patient samples typically value more than their own health state, hence a patient sample does not exclusively provide 'experienced' utility values. See Brazier et al (2017) for a detailed discussion.

The definitions are intentionally worded carefully as it can be difficult to classify a respondent sample as patients or the general public in certain scenarios. It should be clear from the definition in table 4.1 that there is no requirement for a general public sample to be representative of a country's general population. This allows for a sample of respondents of, for example, a particular gender and age to still be classified as a general public sample if they do not have experience of a particular health issue or service.

Defining the two types of sample based on the recruitment strategy of the study is an attempt at reducing the subjectivity of the classifications. For example, it is common for DCEs to elicit preferences for generic health services such as general practitioner (GP) appointments (Hole, 2008; Gerard et al., 2008) or vaccinations (Brown et al., 2014; Marshall et al., 2016). One might assume that most individuals have experience of both GP appointments and vaccinations and should be classified as patients accordingly; however, the definitions in table 4.1 avoid this assumption. For example, one might recruit a sample for the former from a GP's waiting room (a patient sample according to table 4.1) and a sample for the latter from an online panel with no vaccination-related screening questions (a general public sample according to table 4.1). In contrast, one might recruit a sample for the former from an online panel with no GP-related screening questions (general public) and a sample for the latter from a vaccination clinic (patients). While this classification process may be somewhat convoluted, it means that the classification is not solely reliant on the terminology used within the study to refer to the sample (which can be inconsistent across studies) nor is it solely reliant on the opinion of the reviewer (which is subjective and may not be replicable).

Finally, it should be noted that studies eliciting preferences by proxy are classified in the same manner. That is, if a study recruits caregivers of individuals with a specific health issue and elicits their preferences for a health service *on behalf of* the individuals that they care for, this is classified as a patient sample.

4.1.3 Search Strategy (2011 to 2013)

A literature search was undertaken using PubMed in January 2014 looking for papers published between 2011 and 2013 which contained a DCE in a health economics context. Titles and abstracts were searched using the same keywords that were used by Ryan and Gerard (2003) and de Bekker-Grob et al. (2012); these are listed in box 4.1. The algorithm used to search PubMed was checked by applying it to the range of years (2001 to 2008) used by de Bekker-Grob et al. (2012) and all of the papers included in their review were successfully identified.

Box 4.1 Keywords used in the first literature search

discrete choice experiment	conjoint studies
discrete choice experiments	conjoint choice experiment
stated preference	conjoint choice experiments
part-worth utilities	discrete choice modelling
functional measurement	discrete choice modeling
paired comparisons	
pairwise choices	
conjoint analysis	
conjoint measurement	

4.1.4 Search Strategy (2014 to March 2017)

Two additional searches were conducted to keep up to date with the literature and to prepare a database for a subsequent update to the review. The search was conducted in two parts: the first was in May 2016 which covered 2014 to 25th May 2016; the second was in March 2017 and covered May 2016 to 27th March 2017. The results from both searches were merged into one EndNote library with duplicates removed. These searches varied slightly from the one outlined in the previous section as a subsequent review of DCE studies in health economics had been published to follow on from de Bekker-Grob et al. (2012) that included some additional search terms (Clark et al., 2014). The full selection of search terms used in these additional searches can be found in box 4.2 with the new terms in bold.

Box 4.2 Keywords used in subsequent literature searches

discrete choice experiment	conjoint studies
discrete choice experiments	conjoint choice experiment
stated preference	conjoint choice experiments
part-worth utilities	discrete choice modelling
functional measurement	discrete choice modeling
paired comparisons	**DCE**
pairwise choices	**conjoint**
conjoint analysis	**discrete choice conjoint experiment**
conjoint measurement	**discrete choice conjoint experiments**

4.1.5 Inclusion & Exclusion Criteria

For clarity, the inclusion and exclusion criteria are outlined below.

Studies were eligible for the review if they:

- Contained a DCE that elicited preferences from a general public or patient sample (proxies allowed) for a health service or;
- Contained a comparison of preferences between a general public and a patient sample.

Studies were considered ineligible for the review if they:

- Had been published previously (e.g. methodological studies, further publications of the same study).
- Did not elicit preferences for a health service (other common applications include health states, health insurance, health-related jobs, priority setting).
- Did not report any results (e.g. conference abstracts, protocols).
- Contained a comparison of preferences between any other types of respondent sample (e.g. patient and physician samples).

4.1.6 Data Extraction

Although the objectives of this review result in a relatively simple data extraction process (i.e. classification of the respondent sample), additional data was extracted in the initial review (2011 to 2013) to get a feel for the published literature and explore possible trends. This included the study location and sample size, the type of attributes used in the study, the payment vehicle (if a cost attribute was used) and the type of results presented. However, this supplementary information was not deemed particularly relevant for consideration in this review following feedback from a session at a European Health Economics Association (EuHEA) conference. Hence, this information is not presented within this chapter but the conference paper can be found in Appendix A-1. The full data extraction table for the 2011 to 2013 review can be found in Appendix A-2. The 2014 to March 2017 data extraction process was subsequently simplified to focus only on the classification of the respondent sample, which included the extraction of a direct quote (about the sample) from the study but little else. This decision was made because the scope of the review had been determined at this stage, and it was known that information on results and other aspects of the studies would not be fully utilised. The full data extraction table for the 2014 to March 2017 review can be found in Appendix A-3. Full reference information for all of the included studies can be found in Appendix A-4.

4.2 Results

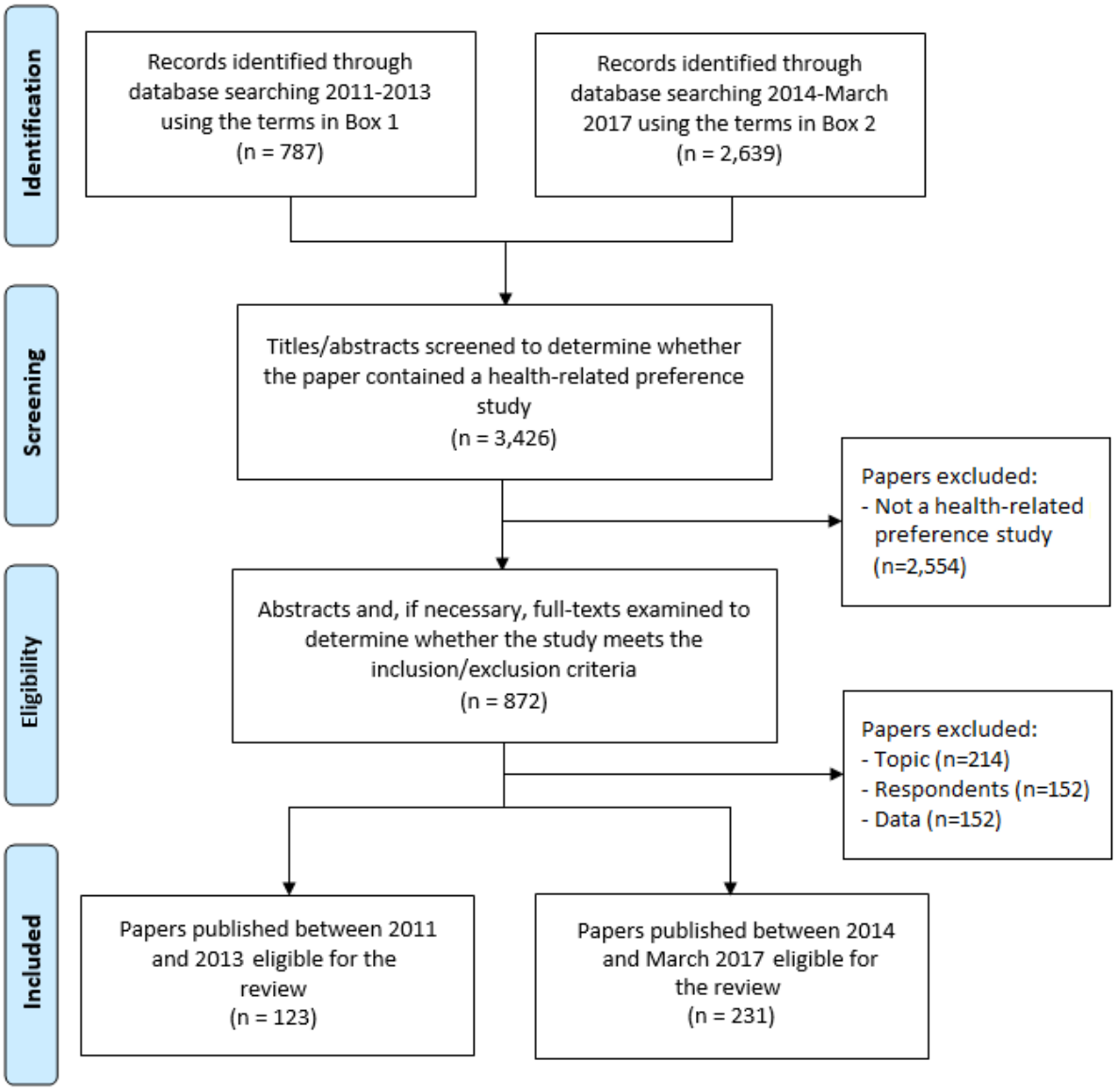
4.2.1 Literature Search

The 2011 to 2013 search identified a total of 787 records and the 2014 to March 2017 search identified a total of 2,639 records. While an increase in published DCE studies in health economics would be expected over this period given the trends in previous reviews, the large number of records in the more recent search was largely due to the additional keywords. For instance, a large number of irrelevant records were identified as a result of the inclusion of 'DCE' and 'conjoint' as keywords. The former resulted in numerous records about dynamic-contrast enhanced magnetic resonance imaging (DCE MRI) and the latter is a common term across medical studies. The title and abstract screening process resulted in the exclusion of 2,554 records across the 2011 to March 2017 time period. The primary objective at this stage was to remove ineligible records, hence any preference studies were included for more detailed examination in the next stage. In general, it was easy to determine whether the record was irrelevant based on the information provided within these fields.

The final screening stage involved a closer examination of abstracts and, where necessary, the examination of full-texts, of 872 records in order to determine whether the DCE elicited preferences from a general public or patient sample for a health service. Full-text examination was required when clarity was needed regarding the preference elicitation method used in the study as DCEs are often incorrectly labelled as conjoint analyses and, occasionally, the opposite is true[11]. In total, 518 studies were removed at this stage: 214 due to an irrelevant topic (i.e. not a health service); 152 due to an irrelevant study sample such as physicians (or irrelevant comparisons); and 152 due to data concerns (e.g. not a DCE or no primary data). This resulted in a total of 123 studies from 2011 to 2013 being included in the review and 231 studies from 2014 to March 2017 being eligible for review. Figure 4.1 presents a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram that describes the aforementioned search process (Moher et al., 2009).

[11] Conjoint analyses are often ranking tasks and are not based on random utility theory (outlined in 6.4.7). The importance of the latter is explained in detail in Louviere et al. (2010).

Figure 4.1 PRISMA Flow Diagram



4.2.2 Respondent Sample Classification

All of the 123 eligible studies between 2011 and 2013 that were identified were included in the review. Table 4.2 contains the key extracted information: 68% of the included studies elicited preferences from a respondent sample that, based on the definitions in table 4.1, was classified as a patient sample.

Table 4.2 Classification of Respondent Samples (2011 to 2013)

Classification of Respondent Sample	**Number of Studies**
Patient Sample	82 (68%)
General Public Sample	39 (32%)

For the subsequent review period (2014 to March 2017), the decision was made to review only 10% of the included studies due to the high number of eligible records (n=231). It was believed that 10% would be a sufficient sample to establish a trend, given that it was not expected that there would be a major change between review periods. The main reason that no change was expected in the 2014 to March 2017 review period was the continued, and seemingly increasing, focus on patient preferences in health economics as identified from informal evidence. This included unsystematic literature reviewing (i.e. keeping up with DCE developments in health economics), attending conferences and engaging with researchers conducting health preference research (where there was an increasing focus on patient preferences) as well as keeping up to date with guidance released by, and case studies conducted by, various HTA bodies and regulatory agencies. In the latter case, it became clear that *patient* preferences were a key focus, especially in the case of the FDA's CDRH (see Chapter Three). In fact, a number of the studies described in Chapter Three were recent developments (i.e. published following the initial search period in this review – January 2014).

Hence, data was extracted from only 23 studies to examine whether the trend identified between 2011 and 2013 had continued. The studies were selected at random using a random number generation function in Microsoft Excel. Table 4.3 contains the key extracted information: 74% of the included studies elicited preferences from a respondent sample that, based on the definitions in table 4.1, was classified as a patient sample (relative to 68% in the previous period).

Table 4.3 Classification of Respondent Samples (2014 to March 2017)

Classification of Respondent Sample	**Number of Studies**
Patient Sample	17 (74%)
General Public Sample	6 (26%)

This is only 10% of the identified studies, hence it cannot be concluded with certainty that the trend identified in the previous period had indeed strengthened. A 95% confidence interval was calculated around this estimated proportion using equation 4.1.

$$74\% \pm 1.96 \times \sqrt{\frac{74\% \ (1 - 74\%)}{N = 23}} \tag{4.1}$$

This produces a 95% confidence interval with a lower bound of 56% and an upper bound at 92% (74% ± 18%). Hence, with considerable confidence it can be concluded that the majority of studies in the latter review period are likely to have elicited preferences from a patient sample (as defined by table 4.1).

4.2.3 Patient & General Public Comparisons

The search for published studies that compared the preferences of a patient and a general public sample for a health service using a DCE was conducted as part of the inclusion/exclusion criteria screening stage. Hence, all 872 records across the full 2011 to March 2017 time period were checked. However, only three published studies were identified (Najafzadeh et al., 2013; Finkelstein et al., 2015; Landfeldt et al., 2016).

The first study, based in Canada, elicited preferences for hypothetical genomic testing for guiding cancer treatment for a cancer described as aggressive but curable (Najafzadeh et al., 2013). The patient sample was made up of current and former cancer patients (n=38) and a selection of the general public sample were matched to the patient sample using propensity scoring (n=533 to n=83). The authors found that the two samples expressed different preferences regarding the value of the various aspects of genomic testing, with patients caring more about the sensitivity of the tests and being more willing to opt-out of a test if it was not sensitive enough relative to the matched public sample. The differences in coefficients lead to large differences in WTP estimates, both for individual attribute levels as well as genomic tests overall:

> *"…the public was willing to pay as high as $6050 for having a genetic test while patients' WTP for genetic testing was only $919."* (p.8)

The second study, based in Singapore, elicited preferences for end of life (EOL) treatment (Finkelstein et al., 2015). The patient sample was made up of individuals with advanced (stage four) cancer (n=320) whereas the general public sample was made up of individuals aged over 50 described in the study as "community dwelling older adults" (n=522). The authors found that the two samples expressed different preferences, such as with regard to the source of payment for treatment and the prolongation of life. The differences in coefficients lead to large differences in WTP estimates, with patients expressing a higher WTP for nearly all attributes compared to the general public sample. For example, the patient sample were willing to pay S$11,043 (95% CI: S$3,061, S$16,426) for an additional life year compared with the general public sample that were willing to pay S$1,587 (95% CI: -S$1,299, S$4,379). Despite the wide and slightly overlapping confidence intervals, the difference is statistically significant at the 5% level.

The final study, based in Germany and Sweden, elicited preferences for hypothetical treatments for relapsed/refractory (r/r) chronic lymphocytic leukaemia (CLL) (Landfeldt et al., 2016). The patient sample was made up of individuals with CLL (n=43), and the general public sample contained respondents that were over 18 and did not know anybody with a malignant hematologic disease (n=196). Physicians that had treated patients with CLL (n=72) were also included as a separate sample. The only statistically significant differences (at the 5% level) between the patient and the general public sample were in relation to the relative importance of progression free survival (9% and 11.1%, respectively) and treatment administration (18% and 13.2%, respectively). No welfare estimates were provided within this study.

4.3 Discussion

4.3.1 The Emphasis on Patient Preferences

It is somewhat unsurprising that more studies elicit the preferences of patients. Early introductory articles about the use of DCEs discussed their potential use for incorporating the *patient* voice (Ryan, 2004b) and others identified a similar trend to the one identified in this review when the literature base was in its infancy (Bryan & Dolan, 2004). To this day it very much appears to be the case that DCEs are seen as a tool to elicit *patient preferences* and little has been written about the justification and implications of this patient focus since the commentary by Bryan & Dolan (2004).

That is not to say that there are not a substantial number of DCE studies eliciting the preferences of an inexperienced, general public sample. However, many of these studies elicit preferences for fairly generic services e.g. those that are relevant for a wide range of potential users (Boonen et al., 2011; McAteer et al., 2015). In addition, many of these studies elicit preferences from a subset of the general public whom may be more appropriate than a completely random or representative sample. For example, studies often elicit preferences from a sample of respondents that fit within a certain age range that is appropriate for a particular health issue (Ghijben et al., 2014; Kistler et al., 2015).

One might attribute this lack of random or representative general public samples to the guidance provided by experienced choice modellers. It is often recommended that a DCE is only provided to individuals that will, or may, have to make a similar decision in real life (S Hess, Personal Communication, 3 April 2014). This advice is provided with data quality in mind, but combined with the way in which DCEs were introduced into the health economics literature, it provides a strong basis for rejecting inexperienced respondent samples in DCE studies. Additionally, many of the examples provided in the previous chapter surrounding the

incorporation of DCE data into analyses that will be used to inform decision-making have considered patient preferences in the first instance.

However, at least one study that elicited preferences from a general public sample justified their choice on a normative basis as outlined in this thesis. The study elicited preferences for treatments for Fabry disease (Lloyd et al., 2017). The authors stated that:

> *"The general public (rather than people with Fabry disease) were recruited because this study was done from a societal perspective. Decision makers such as the National Institute of Health and Care Excellence (NICE) in the UK state that decisions should reflect the preferences of the general public because of the central role of taxation in funding health care. Outcome measures like the widely used EQ-5D are based on societal preferences (rather than weights derived from patients)."* (p.25)

It may be the case that more studies in the future will consider this approach if the use of preferences elicited from DCE data begins to influence decision-making to a greater extent. However, a sceptic may question the true rationale of the sample used in Lloyd et al. (2017), given that Fabry disease may only have an incidence rate of 1 in 117,000 (Germain, 2010).

4.3.2 Patient & General Public Comparisons

There is a clear lack of evidence regarding the potential differences in preferences between a patient and a general public sample that might occur in a DCE study that elicits preferences for a health service. The existing published studies provide some evidence to suggest that preferences vary to a significant extent, which in turn creates large differences in welfare estimates (Najafzadeh et al., 2013; Finkelstein et al., 2015). However, it could be the case that differences are not always as extreme, as appears to be the case in Landfeldt et al. (2016). It is interesting to note, however, that Najafzadeh et al. (2013) did try to match respondents using propensity scoring and Finkelstein et al. (2015) restricted their general public sample to over 50s. Hence, the existing evidence, while mixed, indicates a significant probability that preferences for health services may differ between general public and patient samples.

Landfeldt et al. (2016) did not provide any rationale for their examination of both patient and general public preferences (and physicians). However, the other two studies did provide a rationale for examining preferences of both samples, and these varied slightly. Najafzadeh et al. (2013) expressed the need for public preferences in order to assist reimbursement decisions, as discussed by Lloyd et al. (2017) and in the previous chapter:

> *"In general, approval and use of genomic tests varies widely across different jurisdictions and for different populations. Publicly (or privately) funded health care benefit providers are often interested in learning about tax payers' (or privately insured populations') opinion about the value of these genomic tests. Knowledge*

> *about these preferences will enable health benefit providers to select genomic tests with the highest perceived value when making funding decisions."* (p.2)

Finkelstein et al. (2015) justified their decision based on the fact that the most relevant group of individuals affected by coverage decisions may vary depending on the decision-maker's definition and preferences may differ between groups:

> *"In making coverage decisions, many governments consider preferences of its relevant constituents [7,8]. In Singapore, for EOL care, the largest constituents consist of older adults and those with life limiting illnesses as they are the ones most immediately affected by any health care reform concerning coverage for EOL treatments [9]. Yet, it is likely that preferences and willingness to pay for life extending treatments among these two groups differ."* (p.1483)

It should also be noted that other DCE comparison studies exist but were excluded based on the exclusion criteria. For example, Marshall et al. (2016) elicited preferences from both adults and adolescents for meningococcal B vaccines. While the vaccine may be more relevant for adolescents, the sampling strategy meant that both samples were considered to be from the general public based on the definitions in table 4.1. Additionally, Sossong et al. (2016) elicited preferences of patients and non-patients for health states related to rheumatoid arthritis. While one of the attributes related to treatment duration, the alternatives were described/framed as health states and hence this study was excluded. Finally, the Tinelli et al. (2016) paper described in section 3.4.1 was identified during this review was of high relevance to this thesis but does not compare patient and general public preferences and hence was excluded.

4.3.3 The Definition of a Patient Sample

Throughout the review process it became clear that the definition of a patient sample often varied across studies. Hence, two or more DCE studies that aim to elicit the preferences of a sample of the same general patient population may not be particularly comparable. Logically, this also means that the results of a single DCE study may not be particularly generalisable either without further clarification of the sample used.

Some differences in patient sample characteristics were fairly minor. Two studies with almost identical titles aimed to elicit patient preferences for disease-modifying therapies for treatment of multiple sclerosis (Utz et al., 2014; Garcia-Dominguez et al., 2016). However, one study required that respondents had a diagnosis of *relapsing-remitting* multiple sclerosis (Utz et al., 2014) whereas the other only required a diagnosis of multiple sclerosis (Garcia-Dominguez et al., 2016). As a result, only 72% of respondents in the latter had relapsing-remitting multiple sclerosis; it may be the case that preferences differ for those individuals relative to others.

Differences were more substantial between other studies. Two studies aimed to elicit patient preferences for anticoagulation therapy in atrial fibrillation (Ghijben et al., 2014; Böttger et al., 2015). However, only one of these studies actually had a diagnosis of atrial fibrillation as an inclusion criterion (Böttger et al., 2015). The other study instead elicited preferences from members of the general public aged over 40 (a general public sample by the definitions in table 4.1); only 6 respondents had a history of atrial fibrillation. Hence in this example, the type of patients differed not only between the two studies but also within the Ghijben et al. (2014) study.

Differences in patient characteristics across similar studies could occur as a result of the source of the patient sample. For example, some studies may use online panels (Giles et al., 2016) whereas others may use clinical trials (Tinelli et al., 2016) to recruit their sample. Clinical trials have their own inclusion/exclusion criteria which researchers will be restricted by. Additionally, online panels may prevent certain inclusion/exclusion criteria being used because researchers might be unable to identify and/or verify certain patient characteristics.

4.3.4 Limitations of this Review

This review is not without its limitations. First, the definitions used within this review to classify a sample as patients or the general public would not necessarily be consistent with those used if others conducted a similar review. However, the definitions are explicit and hence can be questioned and replicated if necessary. Furthermore, attempts were made to keep the definitions in line with the preference debate in health state valuation and to try and base the classification on the inclusion/exclusion criteria used within each study. Second, the search strategies may have been suboptimal due to the use of different keywords in the second and third searches as well as the use of only one database (PubMed). However, published systematic reviews of DCE studies from the literature were closely followed when determining the search strategy in this review. As a result, improvements by Clark et al. (2014) to the search strategy used by de Bekker-Grob et al. (2012) were therefore implemented into the updates in this review. In addition, it is quite unlikely other databases would have contained a significant number of relevant articles that could feasibly affect the conclusion of this review; for example, there is no reason to believe that articles contained within Embase and not PubMed are any less likely to elicit patient preferences. Finally, the results from this review cannot be used to conclude that most decision-making using DCE data *has* used patient preferences. However, the clear emphasis on patient preferences does appear to suggest that DCEs are generally considered to be a tool for eliciting *patient*

preferences and hence future guidance from HTA or regulatory bodies that suggest the use of data from DCEs may well be a recommendation for using *patient* preference information.

4.4 Conclusion

This review has identified that the vast majority of published DCE studies from 2011 to 2013 that elicit preferences for health services do so from a sample of patients. It also appears (based on a 10% random sample) that this trend continued between 2014 and March 2017. Only one example (Lloyd et al., 2017) was identified where a general public sample was used on the normative grounds that were outlined in the previous chapter (e.g. that resource allocation decisions should be based on societal preferences in tax-funded systems). However, in reality this sample may be better characterised as a convenience sample given the rarity of the disease that the study is based on.

This review also identified a very small number of studies that compared the preferences of a patient and a general public sample. In two of the three examples, there were large and statistically significant differences that translated to very large differences in WTP estimates. However, little can be concluded from three studies and further research would be highly beneficial in this area. In particular, it may be useful to see if such differences exist when the subject of the DCE is easier for the general public to understand e.g. a public health intervention rather than, for example, genomic testing. Finally, this review also identified that there is a range of heterogeneity in the characteristics of patient samples that are used in DCE studies, which may translate to considerable preference heterogeneity within and across studies. It may also mean that DCE studies are not, in general, particularly generalisable i.e. it may be inaccurate to conclude that patients prefer a particular treatment attribute based on only one study with a specific patient sample.

Chapter 5. A Framework for Classifying Respondent Samples in DCE Studies & Related Research Questions

The aim of this chapter is to put forward an alternative framework for classifying respondent samples in DCE studies that elicit preferences for health services.

The first section of this chapter follows on from the previous chapter by discussing the usefulness of comparing patients with the general public in the context of DCEs where there may be considerable differences *within* each type of respondent sample when they are classified in this manner. The subsequent section puts forward a framework that could be used to classify respondent samples in relation to their level of experience, which could be used to compare preferences across different types of user group. Section three summarises the overall research aims of the thesis and the final section outlines the four specific research questions that will be examined within the empirical work in this thesis.

5.1 Patients vs. General Public: A False Dichotomy?

In the context of health state valuation, it is fairly simple to categorise a survey sample as patients or the general public. The key distinction is simply whether or not the sample of respondents are experiencing the health state(s) in question. As DCE studies that elicit preferences for healthcare services are so broad, this distinction is not as clear.

One possible approach, used in the previous chapter, is to define a sample of patients as individuals that have experienced the health issue, and/or have used an equivalent service in the past. In contrast, a general public sample would be defined as individuals that have not *necessarily* experienced the health issue or equivalent services. This broad approach keeps experience as the primary differential between the two populations which, whilst enabling most of the issues raised in the preference debate to be easily applied to DCE studies, may be over simplistic when discussing the choice of whose preferences to elicit for different purposes. For example, a sample of individuals over the age of 65 that has been generated at random for a DCE about bowel cancer treatments would be classified as a general public sample under this definition, despite this being a highly relevant sample given the topic of research. It is also quite plausible that, in an example like this, that respondents in a certain age range will be aware of the relevance of the health issue to them. This could mean that the respondents have vastly different preferences compared to younger individuals. Hence, the sample in this case may be better off classified in a different way.

Similarly, there may be some merit in distinguishing between respondent samples that have been generated from a RCT, which indicates that the respondents have experience with some form of intervention for the health issue, and those with the health issue identified from an online panel or patient registry. This is because preferences for health services may be highly influenced by the respondents' prior experience of related treatments.

It is not the aim of this chapter, nor this thesis, to provide a comprehensive framework of all of the different types of sample that could be used to elicit preferences in a DCE study, while this may be possible and comparable to the frameworks of *perspectives* that have been introduced and extended (Dolan et al., 2003; Tsuchiya & Watson, 2017). However, this thesis argues that whilst the preference debate is appropriate and should be considered in the context of DCE studies, limiting the debate to patients vs. general public preferences is a false dichotomy because it is hypothesised that preferences will differ across narrower user groups. This argument is in line with the seminal paper by Dolan (1999) that raises similar concerns in the context of health state valuation. Hence, an alternative framework to classify respondent samples beyond simply patients or the general public will be put forward in the next section.

5.2 An Alternative Preference Framework for DCE Studies

A simple alternative approach to classifying samples of respondents in DCE studies involves more careful consideration of the recruitment process and the extent of the respondents' experience of the healthcare issue and/or service in question (i.e. related to the topic of the DCE). Note that this refers to the recruitment of samples, hence while individuals within the sample may fit into several classifications, the sample itself should have one overall classification based on the recruitment process. Table 5.1 outlines a simplistic framework that could be used to classify different samples in DCE studies into four different categories.

Using this framework would enable samples of respondents to be categorised more closely in line with their true characteristics, which may help to reduce preference heterogeneity in DCE studies and improve the generalisability of their results. This framework allows for at-risk samples to be classified separately from more general samples, as well as opening up the definition of a patient by considering respondent's experience of related healthcare services, which is a likely source of preference heterogeneity. It should be clear that placing a group of individuals into one of these categories will be a somewhat subjective process, however these categories should be distinct considering that they are determined by the recruitment process itself.

Table 5.1 User Groups and Their Definitions

Group	Definition
Service User (SU)	A group of individuals that have (or have had) an observable need for the healthcare service in question and have used a related healthcare service in the past.
Potential Service User (PSU)	A group of individuals that have an observable need for the healthcare service in question but have not necessarily used a related healthcare service in the past.
Potential Beneficiary (PB)	A group of individuals that can reasonably be considered as being at-risk of developing a future need for the healthcare service.
Non-User (NU)	A group of individuals that have no specific need for the healthcare service, or prior experience of related healthcare services.

Considering the different ways in which preference information from DCEs could be used in macro-level decision-making, this framework could be used to help justify the choice of sample in such studies and potentially improve the relevance of the results that are used in decision-making. For example, if the aim of the DCE is to provide information that will help to design or refine a healthcare service that is in development (e.g. a trial intervention) it would make sense to elicit the preferences of 'service users' because these individuals will have experienced how related healthcare services work (or the trial intervention itself if the RCT participants are used as the sample). If the aim of the DCE is to provide information regarding expected user uptake it may be more useful to elicit the preferences of 'potential service users' where possible, or alternatively 'potential beneficiaries' could be used as a proxy where it is not possible to recruit a sample of the former. DCEs designed for incorporation into economic evaluations should, as always, follow the relevant guidance but typically may elicit the preferences of 'non-users' (e.g. in extra-welfarist evaluations) or (potential) service users (e.g. in welfarist evaluations). In any case, this thesis argues that this framework would be an improvement over the typical use of classifying a sample of respondents as patients or the general public.

It could be the case that service users express preferences that are largely in line with those of potential service users as both groups suffer from some kind of health issue and the only difference is that the former have experienced some treatment. On a similar note, potential service users may not express vastly different preferences to potential beneficiaries if the former do not personally consider themselves to be in significant need of treatment.

5.3 Research Aims

The empirical work in this thesis aims to examine the differences in preferences between all four groups in this framework in order to provide some evidence regarding the appropriateness of these classifications. In a typical health setting it may be difficult to identify potential service users because it may be unrealistic that individuals suffering from a health issue could be identified and recruited to a DCE study prior to receiving any treatment. However, in the case of public health interventions this may be more plausible, given that such interventions tend to focus on dealing with broad health issues that are not likely to cause morbidity in the short-term. Furthermore, public health interventions are a good example where a sample of respondents that has not been recruited based on any specific characteristics is likely to contain individuals that fit into several of the different classifications. Additionally, public health interventions are typically easy to understand in that they will rarely involve clinical jargon and be associated with risks, such as harmful side effects. This has the added benefit that differences in preferences in this context are perhaps more meaningful in that they cannot be readily discarded as differences in respondent understanding of the context.

For these reasons, the empirical work in this thesis will conduct a DCE that elicits preferences for different compositions of a public health intervention that aims to promote weight loss maintenance. Respondent samples that fit within the described framework will be recruited. This allows for a comparison of patients vs. the general public (where the former could be any of the first three user groups and the latter the non-user group) but also allows for the framework to be tested by making comparisons across all four groups. The next section formally outlines the specific research questions.

5.4 Research Questions

5.4.1 Research Question One

RQ1: To what extent do preferences for a health service, elicited via a discrete choice experiment, differ between different user groups, and why might these differences occur?

This research question comes directly out of the discussion of the preference debate (Chapter Three), the lack of evidence in the literature (Chapter Four) and the limitations of the patient and general public dichotomy, hence the potential need for an alternative classification framework. The different user groups to be examined will be those within the framework in table 5.1: service users; potential service users; potential beneficiaries and; non-users. As discussed in previous chapters, if preferences between these groups differ significantly then the choice of whose preferences to elicit in DCE studies that aim to influence decision-making becomes a more difficult question. If statistically significant differences in preferences are found, it is hoped that additional information could shed some light on the reasons behind the differences.

5.4.2 Research Question Two

RQ2a: To what extent do WTP estimates differ between different user groups, and why might these differences occur?

RQ2b: How might WTP estimates differ if they are elicited indirectly or directly?

Part (a) of this research question links directly with RQ1: if preferences differ between the four user groups then how does this translate into differences in WTP estimates? As WTP estimates could be used in the likes of CBA the use of one set of WTP estimates could lead to different resource allocation decisions being made when compared to those based on another set of WTP estimates. Hence it would be of interest to examine how WTP estimates differ between the four user groups.

In addition to this, part (b) of this research question aims to explore a methodological question: the difference between WTP estimates when they are elicited directly (via CV) or indirectly (via DCE). Existing studies are limited and use various different types of CV tasks, however the general finding in a health context is that WTP estimates from DCEs are typically larger than those from CV questions (van der Pol et al., 2008; Ryan and Watson, 2009; Danyliv et al., 2012), although in some cases they may not differ at all (Ryan, 2004a). Part (b) of this research question aims to provide additional insight into this sparsely explored area of the literature. The design and implementation of a CV task also provides the

opportunity for additional exploration of how WTP estimates differ across user groups i.e. part (a) of this research question.

5.4.3 Research Question Three

RQ3: To what extent do certain user groups have better defined preferences than other user groups, and why might these differences occur?

It may be the case that individuals in certain user groups have better defined preferences and are able to express these clearly when responding to a DCE, whereas individuals from other user groups may not have such clearly defined preferences. This could be seen in the quality of the preference data, but also in the success rates for embedded rationality tests within the DCE that test the preference axioms outlined in Chapter Two, such as non-satiation and transitivity tests (Miguel et al., 2005). Lower fail rates in certain groups might indicate that individuals within these groups have better defined preferences, which could provide support for their use over another group's preferences. Another interpretation might be that people that pass the rationality tests are better able to cope with the cognitive demands from a DCE.

There may be other factors, beyond the classification of the individual (into one of the four user groups), that affect the probability of a respondent failing a rationality test or providing poor quality preference data, which can also be examined by collecting additional data such as demographic characteristics and attitudinal information. This could provide some insight into some of the questions raised by Ryan et al. (2009) surrounding the decision over whether it is right to exclude responses when rationality tests are failed.

5.4.4 Research Question Four

RQ4: To what extent might differences in preferences between the user groups be attributed to the recruitment vehicle?

A potential criticism when it comes to examining differences between the four user groups is that the groups might not be recruited from the same place. In the empirical work in this thesis this is the case; a clinical trial is used to recruit potential service users and an online panel is used to recruit the other three groups (see section 6.2). Although this may be unavoidable, it is important to consider that the recruitment vehicle may explain some of the differences in preferences that might occur between user groups. Therefore, research question four aims to explore this. In relation to research questions one and two, this can involve pooling the data for the groups recruited via the online panel and comparing this with the trial sample. For research question three, rationality test results can also be pooled across the online panel

groups, and a variable relating to the recruitment vehicle can be used in regression analyses (see section 6.4.3).

5.5 Conclusion

This chapter has critically assessed the idea that respondents can be described as either patients or the general public and put forward a generic framework for classifying respondent samples in DCE studies that elicit preferences for health services. Following this, the overall aims of the empirical work were outlined (i.e. to make comparisons between the user groups) and the specific research questions that will be addressed were summarised in turn.

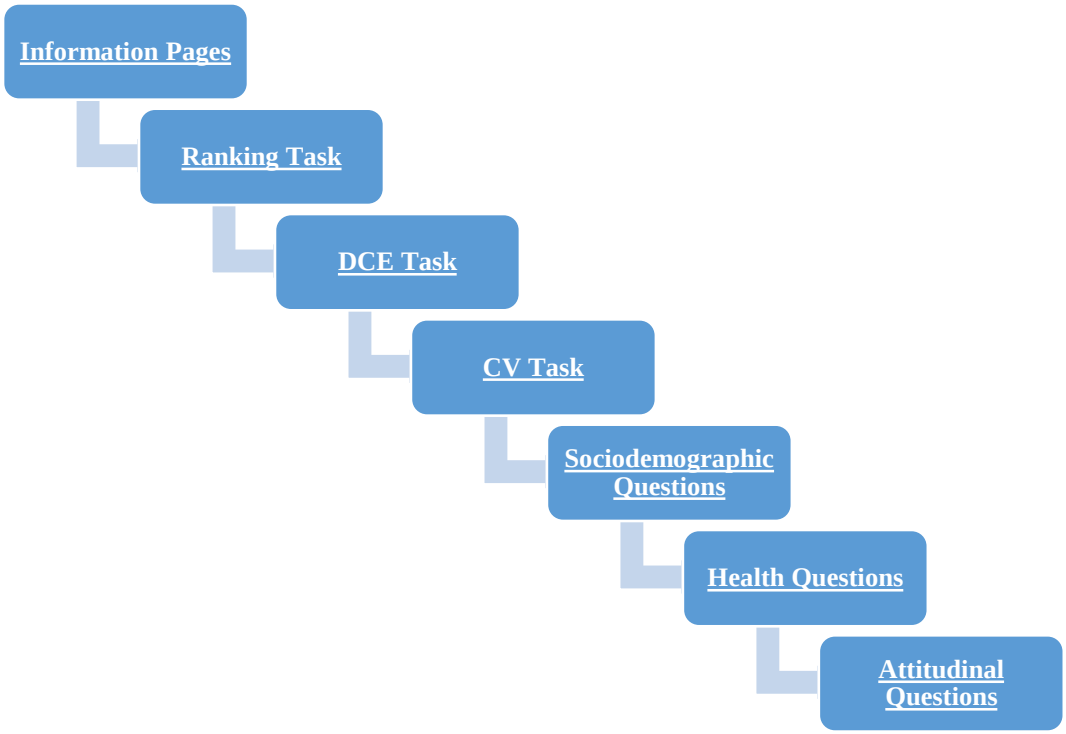
Part II. Methodology

Chapter 6. Designing a Survey to Test the Research Questions

The aim of this chapter is to outline the design process of the online survey used to collect data in order to examine the research questions outlined in the previous chapter.

Section one describes the case study that was used in this project. Section two provides an overview of the sampling process that was followed in order to obtain samples of the four user groups of interest. Section three explains how screening questions were devised in order to classify the respondent samples that were recruited from the online panel. Section four outlines the design of the DCE in detail, covering: the selection of attributes and levels; the generation of an experimental design; the design of rationality tests; the piloting process; the determination of sample sizes; the online programming and; the proposed analyses. Section five outlines the design of the CV tasks in detail, covering; the perspective of the tasks; the object being valued and; the use of different health outcomes. Section six describes the sociodemographic and health-related questions that were included in the survey. Section seven describes the attitudinal questions that were included in the survey. Section eight details the process undertaken in order to obtain ethical approval for the empirical work. Finally, section nine concludes the chapter. For reference, the full survey can be found in Appendix B-1, but to provide a clear overview the overall structure of the survey is illustrated in figure 6.1.

Figure 6.1 Structure of the Online Survey



6.1 The Case Study: Weight Loss Maintenance & The NULevel Trial

While a vast amount of research has been undertaken within public health to improve the lifestyles of those that are obese, little focus has been given to the maintenance of lifestyle improvements following successful and significant weight loss (Dombrowski et al., 2014). Studies have shown that people typically fail to maintain weight loss over time, leading to what is sometimes referred to as the yo-yo effect where weight fluctuates regularly (Avenell et al., 2004). As a result, there is an increasing interest in researching how weight loss can be maintained over time.

The NULevel RCT that was undertaken by a team of researchers at Newcastle University examined the effectiveness of an intervention aimed at weight loss maintenance (Evans et al., 2015). The trial involved the application of a novel technology-based intervention that was provided to a treatment group of 144 participants. The key component of the trial intervention required participants to weigh themselves daily on a set of scales provided to them as part of the trial. The scales sent weight information to the trial team via a cellular network, enabling the trial team to access the information and for the information to be displayed graphically online for the participants. In the first two weeks, participants were reminded to weigh themselves (via text message) if they failed to do so within a 24-hour period; this was extended to a 72-hour period thereafter. In a single face-to-face consultation at baseline, the participants agreed weight-related targets with a facilitator of the trial intervention such that three different 'zones' can be identified: green, yellow and red. The different zones indicated how close a participant was to achieving their agreed target, with green indicating that they were on target, yellow indicating that their weight is increasing and red indicating that weight re-gain is substantial. Whilst the intervention was being delivered, participants also received a variety of automated and tailored remote feedback via text message (or phone call if requested by the participant). The amount of feedback received by the participant related to the zone that they were in, with a 'lighter touch' provided to those in the green zone. The control group consisted of 144 participants that received a set of scales but were not advised on how to use them. In addition, rather than receiving any feedback they were provided with quarterly text messages that contained links to evidence-based, weight management guidance.

The area of WLM, and the NULevel trial intervention in particular, is used as a case study for the purpose of examining the research questions in this empirical work. The research questions could be readily applied to a preference elicitation study about almost any healthcare service, but this is an ideal case study for two reasons in particular:

(1) The question of whose preferences should be elicited for use in healthcare decision-making is particularly relevant in the context of public health interventions, as it is more likely that CBA may be used to evaluate them, relative to the economic evaluation of a drug. As DCEs that elicit preferences for a health service can provide WTP estimates for use in a CBA, using a public health intervention in this project may be particularly relevant.

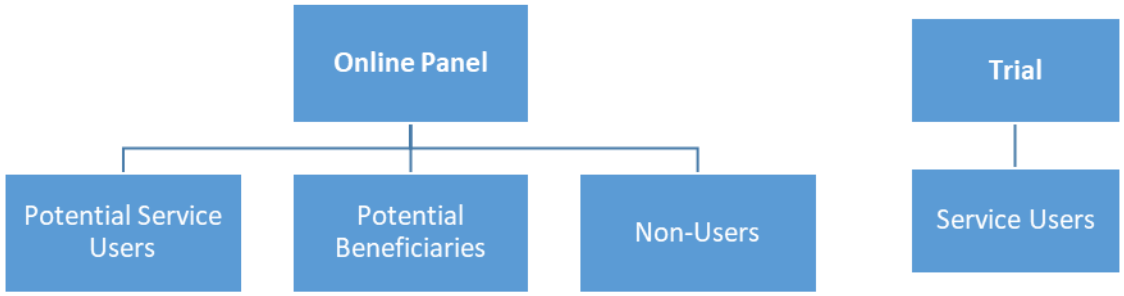
(2) It is hypothesised that the vast majority of individuals can relate to weight loss, weight gain and hence WLM, with minimal difficulty. As a result, preference information from individuals that would not necessarily need a WLM intervention should be of a higher quality than that expected from, for example, a survey about a niche, clinical intervention. Should it be the case that substantial differences in preferences are found for this type of intervention, one might hypothesise that preferences will differ even more in a niche, clinical setting.

In the context of WLM research, information about individual's preferences for WLM programmes is currently not readily available; it is expected that this project will be the first to apply DCE and CV methodology in this area of public health. Furthermore, the results from the project could be particularly useful for future development of the NULevel trial intervention, given that the research will be framed around this particular WLM intervention.

6.2 Sampling Overview

In order to examine the research questions in this thesis, samples of the four different user groups from the framework in Chapter Five need to be recruited. The NULevel trial provided individuals with experience of both the health issue and related treatment (i.e. service users). However, recruiting individuals for the other three user groups is somewhat more challenging. In this case, given that this is an easily relatable health issue, the most effective option (in terms of both recruitment speed and cost) is likely to be delivering the survey to online panel with detailed screening questions used to categorise respondents into the other three user groups. Figure 6.2 illustrates this recruitment process.

Figure 6.2 Recruitment Vehicles for the Four User Groups



As the framework in Chapter Five is so broad, such that it can be readily applied to a range of different research areas, it is necessary to define the four user groups specifically for this context. Table 6.1 defines the four user groups in the context of WLM.

Table 6.1 User Groups in a WLM Context

User Group	**Definition (in a WLM context)**
Service Users (SU)	Obese individuals that have lost weight, attempted to maintain the weight loss and have experienced a WLM intervention (NULevel) to help them achieve this.
Potential Service Users (PSU)	Obese individuals that have lost a clinically significant amount of weight recently and could benefit from WLM intervention today, but have not *necessarily* experienced a WLM intervention.
Potential Beneficiaries (PB)	Individuals that are currently overweight or obese that have not lost a clinically significant amount of weight recently, but could potentially benefit from a WLM intervention in the near future if they do so.
Non-Users (NU)	Individuals that are currently a normal weight with no clear need to lose weight at present or in the foreseeable future. They cannot feasibly be considered potential beneficiaries of a WLM intervention at present.

6.3 Screening Questions (for the Online Panel Sample)

The NULevel participants, the SU group in this project, were subject to a wide range of inclusion and exclusion criteria before their acceptance into the trial. In order to reduce the subjectivity of the PSU group definition, and because the DCE is based on the NULevel intervention specifically, the classification of individuals to this group is also based on these criteria. The full inclusion and exclusion criteria for the NULevel trial are listed below (Evans et al., 2015).

Individuals are eligible to participate if they meet the following eligibility criteria:

1. A body mass index (BMI) of ≥30 kg/m^2 any time in the 24 calendar months preceding trial entry (i.e. the date of consent); the BMI threshold is ≥28 kg/m^2 for individuals of South Asian descent (WHO Expert Consultation, 2004);
2. A weight loss of ≥5 % in the 12 calendar months preceding trial entry. Written verification of this weight loss should be provided by a physician, weight loss counsellor or friend; if no such evidence from a third party is available, participants may self-certify their weight loss;
3. Ordinarily living or working in the North East of England;
4. Access to and willingness to use an internet-enabled mobile telephone in order to receive messages from the research team, containing embedded links to relevant online content;
5. Ability to use a standing scale for weight measurements.

Individuals are excluded from participation on the following grounds:

1. Participation in prior development studies of the intervention;
2. Weight loss due to illness or surgical procedures, including bariatric surgery;
3. Pregnancy or plans to become pregnant in the next year;
4. Breastfeeding an infant under 6 months of age;
5. Current involvement in other weight intervention research studies;
6. Inability to understand written material or telephone conversations in English;
7. A diagnosis of anorexia nervosa, bulimia nervosa or purging disorder, or of any condition which may preclude increasing mild to moderate physical activities such as walking;
8. Baseline weight of >175 kg (due to the measurement range of the provided scales);
9. Plans to leave the area or to undertake long-term travel in the forthcoming 12 months.

The inclusion and exclusion criteria are comprehensive and pose a challenge for recruiting a significant number of similar individuals for the PSU group, however not all are relevant for

this project and hence some can be reasonably ignored. It should also be noted that, for pragmatic reasons, minimising the number of screening questions was seen as desirable. With regards to the inclusion criteria it is argued here that there is no need for this project to replicate the third criterion; the specific inclusion of individuals living in the North East of England is practical for the trial but not necessary for this project. Otherwise, the other inclusion criteria are highly relevant for the PSU group. With regards to the exclusion criteria it is argued that several of the criteria are unnecessary for use in generating the PSU group. Specifically, exclusion criteria one, two, five and nine are not considered as part of the screening questions for the recruitment of the PSU group due to concerns over their relevance and practicality. All other exclusion criteria are considered, to a varying extent, within the screening questions.

After careful consideration, the following 12 screening questions were considered to be the minimum number necessary in order to classify respondents into one of the three user groups:

1. What is your age?
2. What is your gender?
3. Are you currently pregnant or breastfeeding?
4. What are your preferred units of measurement for height & weight?
5. What is your height?
6. What is your current weight?
7. What is your ethnic group?
8. Have you attempted to lose weight in past 12 months?
9. Have you ever attempted to lose weight?
10. What was your highest weight in the past 12 months?
11. Are you able to use a set of scales to weigh yourself? (It doesn't matter if you don't own a set of scales)
12. Do you have access to a mobile phone that can connect to the Internet?

Questions one to three are used to remove respondents that are under 18 or currently pregnant or breastfeeding (covering exclusion criteria three and four for the most part). Questions four to seven enable the calculation of body mass index (BMI) and hence classification into a BMI category (covering inclusion criterion one, exclusion criterion eight and partially exclusion criterion seven by excluding underweight respondents). BMI was calculated using the equation 6.1 and the classification system from the NULevel trial, illustrated in table 6.2, was used.

$$BMI = \frac{weight\ in\ KG}{(height\ in\ metres)^2}\ or\ 703\ \times \frac{weight\ in\ lbs}{(height\ in\ inches)^2} \quad (6.1)$$

Questions eight to 10 provide an insight into the respondent's weight management history (covering inclusion criterion two and, to an extent, exclusion criterion two by use of the word 'attempted'). Questions 11 and 12 are used to determine whether an obese person that has lost ≥5% of their weight in the past 12 months, according to their responses in questions four to 10, meet the other relevant criteria from the NULevel trial (covering inclusion criteria four and five). Figure 6.3 illustrates the way these screening questions were used to classify respondents into one of the three user groups.

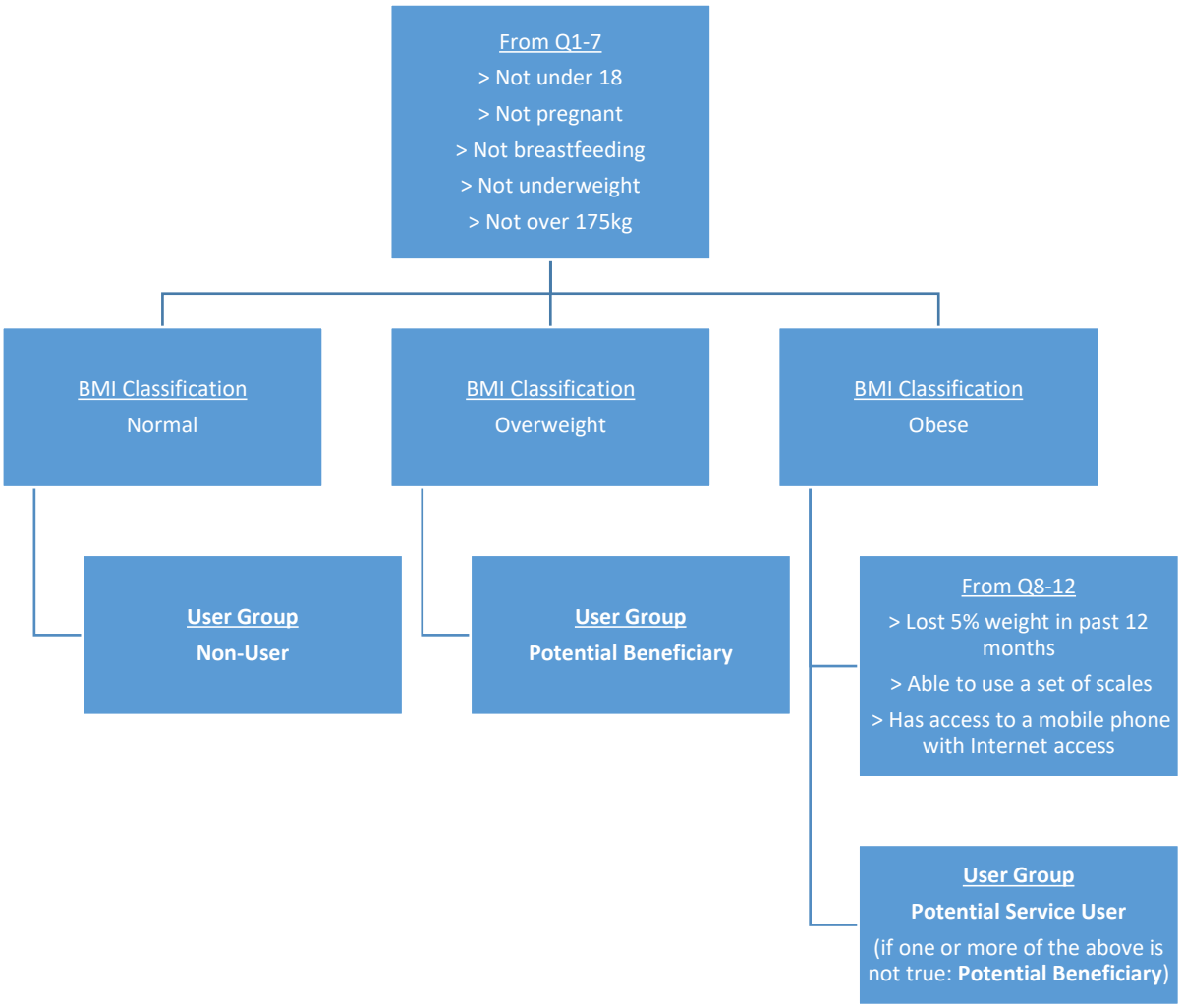
Table 6.2 BMI Classifications

Classification	**BMI (South Asian)[12]**	**BMI (All Other Ethnicities)**
Underweight	<18.5	<18.5
Normal	18.5-22.9	18.5-24.9
Overweight	23-27.4	25-29.9
Obese	27.5+	30+

Defined in full, the PSU group is a group of individuals that: have been obese in the past 12 months and are still obese; have lost ≥5% of their weight in the past 12 months; are able to use a mobile phone to access the Internet; are able to use a set of standing scales; are not currently pregnant or breastfeeding; and do not weigh over 175kg. This group is therefore subtly different to the SU group, because at the baseline of the NULevel trial it was not necessary that the participants were still obese (e.g. an individual that *was* obese and lost ≥5% of their weight, moving them into the overweight classification, are excluded here whereas they wouldn't have been in the trial). This is a consequence of arranging the screening questions to make them as efficient as possible, as well as to reduce the programming requirements of the survey. For completion, the PB group are defined as group of individuals that are overweight, or are obese but do not meet the NULevel trial criteria and the NU group are defined as a group of individuals that are a normal weight.

[12] BMI classifications are different for individuals of South Asian descent, see WHO Expert Consultation (2004).

Figure 6.3 Screening Questions for the Online Panel Sample



The online panel sample was recruited using Research Now (http://www.researchnow.com), a global market research company that owns large nationally representative panels of individuals that can be invited to respond to online surveys. Participants are provided with rewards upon completion of surveys with a value in the region of £1-2. It was requested that each user group contained respondents with a split of sociodemographic characteristics that is comparable to the UK general population and that the split was comparable across the three user groups.

6.4 The Discrete Choice Experiment

The key component of the survey is the DCE that will be used to elicit the preferences of the respondents for hypothetical WLM interventions, hence significant efforts were made to follow best practice guidelines throughout (Lancsar and Louviere, 2008; Bridges et al., 2011; Reed Johnson et al., 2013). This subsection will outline the design of the DCE in seven stages: defining attributes & levels; experimental design; embedded rationality tests; piloting; determining sample sizes; online programming and; analysis.

6.4.1 Defining Attributes & Levels

In this empirical work, the primary aim of the DCE is to examine the differences in preferences between different user groups. As a result, once the context of the DCE was identified as WLM, and the NULevel trial intervention in particular, its specific application was determined during discussions with key members of the NULevel trial team and supplemented by reviewing relevant literature. The latter included a highly relevant systematic review conducted by members of the trial team and others (Dombrowski et al., 2014). As the ideal application of the DCE from the perspective of the trial team was to provide additional useful information for refining the trial intervention once the trial was complete, no additional qualitative research was conducted at this stage as it was not deemed necessary. For example, it was felt that a focus group with members of the general public to discuss WLM interventions may well have provided interesting information but would not have provided superior guidance on how best to frame the NULevel trial intervention in a DCE compared to guidance from the trial team. Another concern was that involving NULevel trial participants in qualitative work (or piloting) could potentially interfere with the trial and possibly result in fewer eligible participants for the SU group (i.e. if participation in qualitative work ruled them out of completing the survey at a later date).

During the discussions with the trial team, several research questions were identified that could be investigated using the DCE. Three important questions were:

- *Would reminders delivered by a different medium be preferred?*

Reminders for individuals (in the treatment arm) to weigh themselves were sent via text message. As the efficacy of these reminders is so important to the intervention, the trial team were interested in determining whether it would be better to deliver them via an alternative delivery mode.

- *How important is face-to-face feedback?*

As one aim of the NULevel trial is to produce a *scalable* intervention, the trial team believed that it may be beneficial if the intervention could be delivered without the need for any face-to-face feedback. The trial team wished to know whether this would be acceptable to participants and hence it was considered important to examine the importance of face-to-face feedback relative to other delivery modes.

- *Would individuals be willing to pay for a WLM intervention?*

A recent DCE study regarding lifestyle interventions found that individuals would need to be incentivised/compensated to take part in the interventions (Ryan et al., 2015). As a result, the

initial expectation was that the DCE in this project would consider financial incentives rather than personal payments. However, following discussions with the NULevel trial team it was decided that it would be unrealistic for incentives to be offered in practice and that one avenue for a future roll-out of the trial intervention might be through a weight loss company (rather than the NHS). As a result, an examination of willingness to pay was considered extremely useful by the trial team.

Considering the research questions raised during discussions with the NULevel trial team, as well as the design of the trial intervention itself, the following six attributes were generated: length of the programme; delivery of reminders to weigh yourself; delivery of feedback from programme staff; availability of online tool(s) to track your progress; weight re-gain; personal cost (per month). Table 6.3 contains full definitions of these attributes as well as the levels assigned to each.

The levels for 'delivery of reminders to weigh yourself', 'delivery of feedback from programme staff' and 'availability of online tools to track your progress' were based on the existing design of the NULevel trial intervention and feasible alternatives as determined by the trial team. The use of four attribute levels was sufficient for these three attributes, hence it was decided from this point to aim for four attribute levels (or multiples of four) for the other attributes as this would be beneficial in the experimental design stage with respect to level balance (Reed Johnson et al., 2013).

For 'personal cost (per month)' the prices of commercial weight loss services such as Slimming World and Weight Watchers were examined in order to provide a feasible price range. On 20th April 2015, Slimming World were charging approximately £26 a month for online membership and approximately £21 a month for group membership (approximations due to membership bundles). On the same date, Weight Watchers were charging £12.95 a month for their online service, and £21.45 a month for their "monthly pass". While these are not strictly WLM interventions and hence not necessarily competing services, the price range of £0-30 per month was deemed suitable for use in the DCE based on this information as individuals regularly pay for these services.

The levels for 'length of the programme' were based on the typical programme lengths for existing WLM interventions based on a recent systematic review (Dombrowski et al. 2014). The review found that the vast majority of trials lasted between 12 and 18 months, with trial interventions rarely being delivered beyond 24 months.

Table 6.3 Attributes and Levels for the DCE

Attribute	Definition	Possible Options	Continuous/Categorical	Variable Name(s)
Length of the Programme	How long the programme will last in total (in months).	6 months; 12 months; 18 months; 24 months	Continuous	*Length*
Delivery of Reminders to Weigh Yourself	How you are reminded if you forget to weigh yourself for over 48 hours (if you are reminded at all).	No reminders; via phone call; via text message; via the online tool(s)	Categorical	*RENone*, REText, REPhone, REOnline*
Delivery of Feedback from Programme Staff	How you will receive feedback on your progress from programme staff.	Via phone call; via text message; via the online tool(s); face to face	Categorical	*FBFace*, FBPhone, FBOnline, FBText*
Availability of Online Tools to Track Your Progress	The type of online tool(s) provided, if any, so that you can track your progress independently.	No online tool; website only; mobile phone application only; website & mobile phone application	Categorical	*OTNone*, OTApp, OTWeb, OTBoth*
Weight Re-gain	The amount of weight that you re-gain, as a percentage of the amount that you lost originally.	0%; 10%; 20%; 40%; 60%; 80%; 90%; 100%	Continuous	*Outcome*
Personal Cost (per month)	The cost, to you, of the programme each month.	£0; £10; £20; £30	Continuous	*Cost*

*Chosen as base level; see section 6.4.7

Finally, it was decided that the levels for "weight re-gain" should be more comprehensive than simply using four levels as the extreme ends of the scale (0% and 100%) are both perfectly plausible in this context. To choose only an additional two levels was therefore a challenge. Hence the decision was made to include eight attribute levels for the 0-100% weight re-gain range in order to provide a wider range of possible outcomes. The aim was to increase the sensitivity at the extreme ends, hence the 20 percentage point gaps in the middle of the range and 10 percentage point gaps at either end of the scale.

One could argue that not conducting additional qualitative work with a group, or groups, of potential survey participants in order to identify attributes and levels is potentially problematic. Indeed, a recently published review has highlighted the importance of qualitative work to inform DCE studies (Vass, Rigby et al., 2017). However, Clark et al. (2014) state the following in relation to an observed reduction in the use of qualitative methods in DCE studies between 2009 and 2012:

> *"It would be of little concern, however, if the recent reduction in the use of qualitative methods to inform attribute selection was triggered by the wider use of DCEs in contexts in which the decision framework is already known (for example, if DCEs are conducted alongside clinical trials)."* (p.892)

Hence in this case it is believed that the work undertaken to generate the attributes and levels as described in this section is sufficient because the 'decision framework' was already known. Additional attributes of apparent importance may have been identified had additional qualitative work been conducted, but these may not have been particularly relevant to the NULevel trial. In addition, the use of four different samples in this study means that qualitative work aimed at potential beneficiaries or non-users (for example) may have produced irrelevant additional attributes for service users or potential service users. It is also possible that the findings of qualitative studies are influenced by the prior experience of the participants, hence using potential service users to identify additional attributes might have produced attributes that were not important to actual service users (those that had experienced 'treatment'). In this case, it was not plausible to conduct qualitative work with NULevel trial participants (i.e. service users) as the trial was on-going, as well as the fact that this might have reduced the potential number of survey respondents from the trial (i.e. if they were excluded from participating as a result of being involved at an earlier stage).

6.4.2 Experimental Design

It was decided that the DCE should include two alternatives as well as an opt-out option (defined as 'no programme'). The opt-out option adds an element of realism as individuals

would not be forced to partake in a WLM intervention in real life. It is generally accepted that this also allows for more realistic uptake rates to be predicted (Clark et al., 2014), which may be a useful route for analysis outside of the research questions addressed in this thesis. However, it is unclear whether the presence of an opt-out option might have an influence on responses. Veldwijk et al. (2014) explored this by providing individuals with both a forced-choice task (no opt-out) and an unforced choice task (with opt-out). Their results suggest that including an opt-out option may influence choice behaviour and they conclude that its inclusion may lead to an unnecessary loss of effectiveness in relation to the efficiency of the DCE, as many individuals may choose this option.

In order to minimise any effect of the opt-out option on choice behaviour and to minimise the risk associated with too many respondents choosing the opt-out in the first instance (i.e. providing unusable data), a second (forced) choice was required for each scenario. The second choice asked respondents "of the remaining two options, which would you choose?" This format is sometimes referred to as a 'best-best' DCE (Ghijben et al., 2014; Lancsar et al., 2017). The plan at an early stage was to treat the second choice data as a backup. That is, the data would only be used in the 'worst-case scenario' where too many respondents chose the opt-out option in the first (unforced) choice. Therefore, the planned data analyses (see subsection 6.4.7) for this empirical work does not include the analysis of the data from the second (forced) choice.

As with any DCE, combining all of the attributes and levels results in a large number of potential alternatives. If all of the attributes and levels in table 6.3 were to be used in the DCE, this would result in a total of 8,192 possible alternatives ($4^5.8^1$). Clearly this number of alternatives is too large to be considered by one individual, hence the DCE has to be designed with a far smaller subset of these. Several different software packages can be used to generate a range of different types of designs; for this DCE, Ngene (Rose & Bliemer, 2012) was used to generate an efficient design. Such designs use the D-optimality criterion in order to minimise the joint confidence sphere around the complete set of estimated model parameters by maximising the determinant of the inverse of the variance-covariance matrix in maximum-likelihood estimation (Reed Johnson et al., 2013). Efficient designs are the most commonly applied in the health economics DCE literature (Clark et al., 2014). Efficient designs allow for attribute interactions to be considered at the design stage. Given the attributes and levels in table 6.3, it was thought that interactions between: 'personal cost' and 'length of the programme' (total cost); 'length of the programme' and 'weight re-gain' (rate of re-gain); as well as 'personal cost' and 'weight re-gain' (cost per % weight re-gain) might be relevant to

include in the analysis. Hence, these interactions were incorporated into the experimental design. Efficient designs also allow for constraints to be included; in the case of this DCE constraints were required because if no online tool is provided, it would not make sense for reminders or feedback to be provided via the online tool.

It is generally considered beneficial to include as many choice scenarios as possible in order to generate more data for the choice models. As the intention was to include a transitivity test that would require three scenarios (see section 6.4.3), the final design contained forty choice scenarios split into four blocks such that respondents would see 10 choice scenarios from the DCE design and 13 in total (excluding a practice scenario). This was considered a feasible number of scenarios based on existing evidence within the literature (Clark et al., 2014). See Appendix B-2 for the full Ngene code used to generate the design, consisting of the number of rows and blocks, the constraints and the specifications of the utility functions. Appendix B-3 contains the full experimental design (i.e. 40 rows) generated from the Ngene code.

6.4.3 Embedded Rationality Tests

Two rationality tests were included as part of the DCE component of the survey in order to examine research question three: a dominance (or non-satiation) test and a transitivity test.

The dominance test was included as part of a practice scenario prior to the 13 choice scenarios where respondents were given a practice choice between two WLM programmes (and an opt-out). In this particular scenario, the attribute levels assigned to each of the WLM programmes were based on rankings directly-elicited from the respondent earlier in the survey. Programme A contained the attribute levels that the respondent ranked as "most preferred" for the following attributes: length of the programme; delivery of reminders to weigh yourself; delivery of feedback from programme staff; and availability of online tool(s). Programme B contained the attribute levels that the respondent ranked as "least preferred" for the same four attributes. The attribute levels for 'weight re-gain' and 'personal cost (per month)' were fixed at 0% and £0 for programme A respectively while for programme B these were fixed at 100% and £30 respectively as respondents were not asked to rank the levels of these attributes. Respondents pass this test if they select programme A over programme B as the former is expected to be their most preferred WLM intervention. It should be noted however that as this test is part of a practice scenario, and because errors could be made during the ranking task, the margin for error is likely to be higher than in dominance tests used in other DCEs.

The transitivity test, a test of the axiom of transitive preferences (see section 2.1.2), is slightly more complex. This test requires a series of choices to be made such that a full-ranking can be

determined between three hypothetical WLM programmes (labelled here as X, Y and Z). Hence in this two-alternative DCE (ignoring the opt-out) this test requires three choice scenarios. As the scenarios are not part of the experimental design of the DCE, the attribute levels for X, Y and Z had to be chosen manually. There is no clear guidance in the literature regarding the best approach for this decision, but to reduce potential confusion it was decided that the combination of attribute levels for X, Y and Z should be unique. In other words, the alternatives should not be identical to any alternatives in the experimental design and there should be no level overlap across the three scenarios. Table 6.4 defines programmes X, Y and Z by their respective attribute levels.

Table 6.4 The Three Alternatives Used in the DCE Transitivity Test

Attribute	**Alternative X**	**Alternative Y**	**Alternative Z**
Length of the Programme	6 months	18 months	12 months
Delivery of Feedback from Programme Staff	Via phone call	Via the online tool	Via text message
Delivery of Reminders to Weigh Yourself	No reminders	Via the online tool	Via phone call
Availability of Online Tools to Track Your Progress	Mobile application only	Website only	Mobile application and website
Weight Re-Gain	60%	20%	40%
Personal Cost	£20 a month	£10 a month	£0 a month

The key to passing the transitivity test is the following condition:

If $X \succ Y$ and $Y \succ Z$ then it must be the case that $X \succ Z$

If the opt-out option is ignored, there are a total of 8 (2^3) rank-order combinations across the three choice scenarios where 6 would pass the test (75%). However, if the opt-out is included in the specification of the test it becomes more complicated. For example, where O denotes the opt-out option:

If X ≻ Y ≻ O and Y ≻ O ≻ Z then it must be the case that X ≻ O ≻ Z

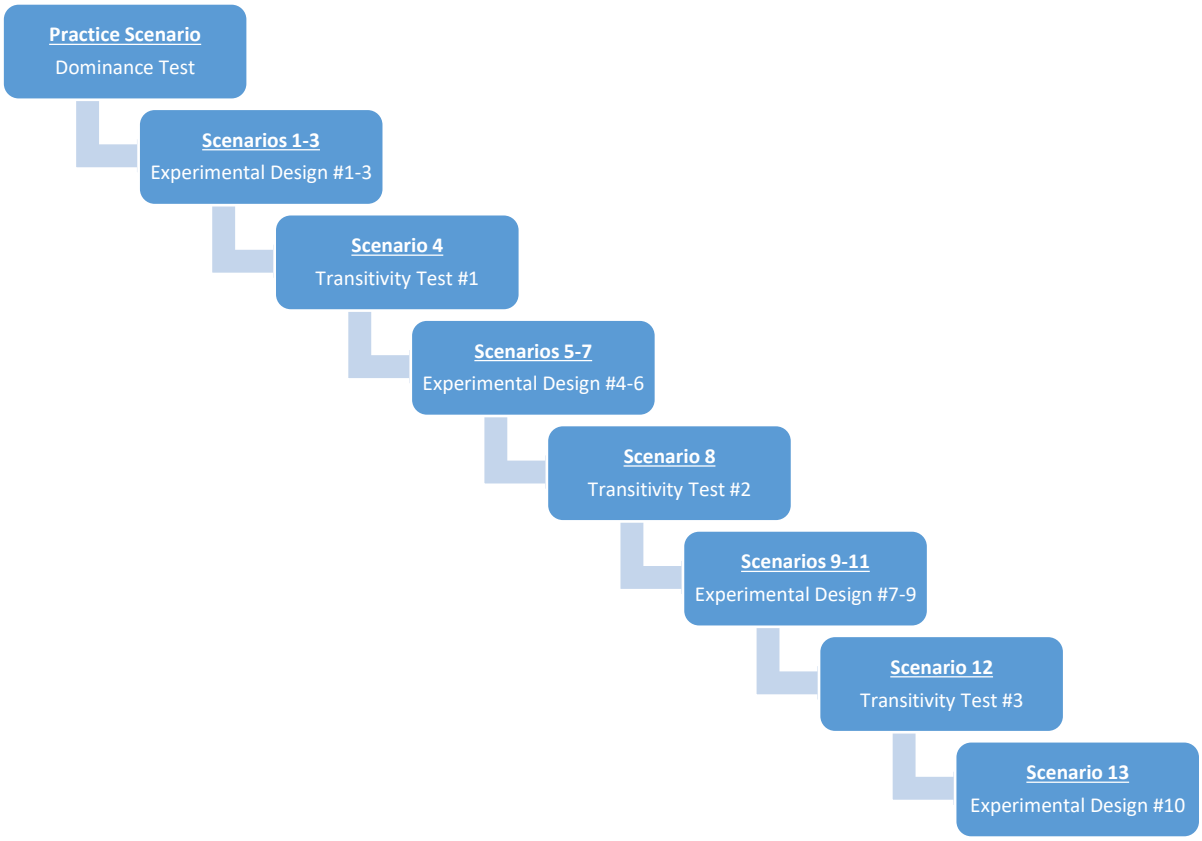
When the opt-out option is considered, there are a total of 216 (3^3) rank-order combinations across the three choice scenarios where 24 would pass the test (11%).

It was decided that the scenarios relating to the transitivity test should be spread out as evenly as possible across the choice task, hence the test was structured in the following way:

- Choice Scenario #4: Programme A = X, Programme B = Y
- Choice Scenario #8: Programme A = Y, Programme B = Z
- Choice Scenario #12: Programme A = X, Programme B = Z

The positioning and configuration of these choice scenarios remained the same for all respondents, regardless of the block (of the experimental design) assigned to them. For clarity, figure 6.4 illustrates how the tests fit within the overall choice task seen by respondents.

Figure 6.4 Structure of the Choice Tasks



To examine research question three, comparisons across user groups of the pass rates for each rationality test will be made. A simplistic percentage comparison can be made in the first instance, but a superior approach would be to estimate a series of regression models with a range of independent variables such that other factors can be controlled for whilst exploring

whether there are differences across user groups. For example, equation 6.2 could be estimated:

$$y_i = \boldsymbol{\delta}_i + \boldsymbol{\gamma}_i + \boldsymbol{\theta}_i + \beta_1 SU_i + \beta_2 PSU_i + \beta_3 PSU_i \quad (6.2)$$

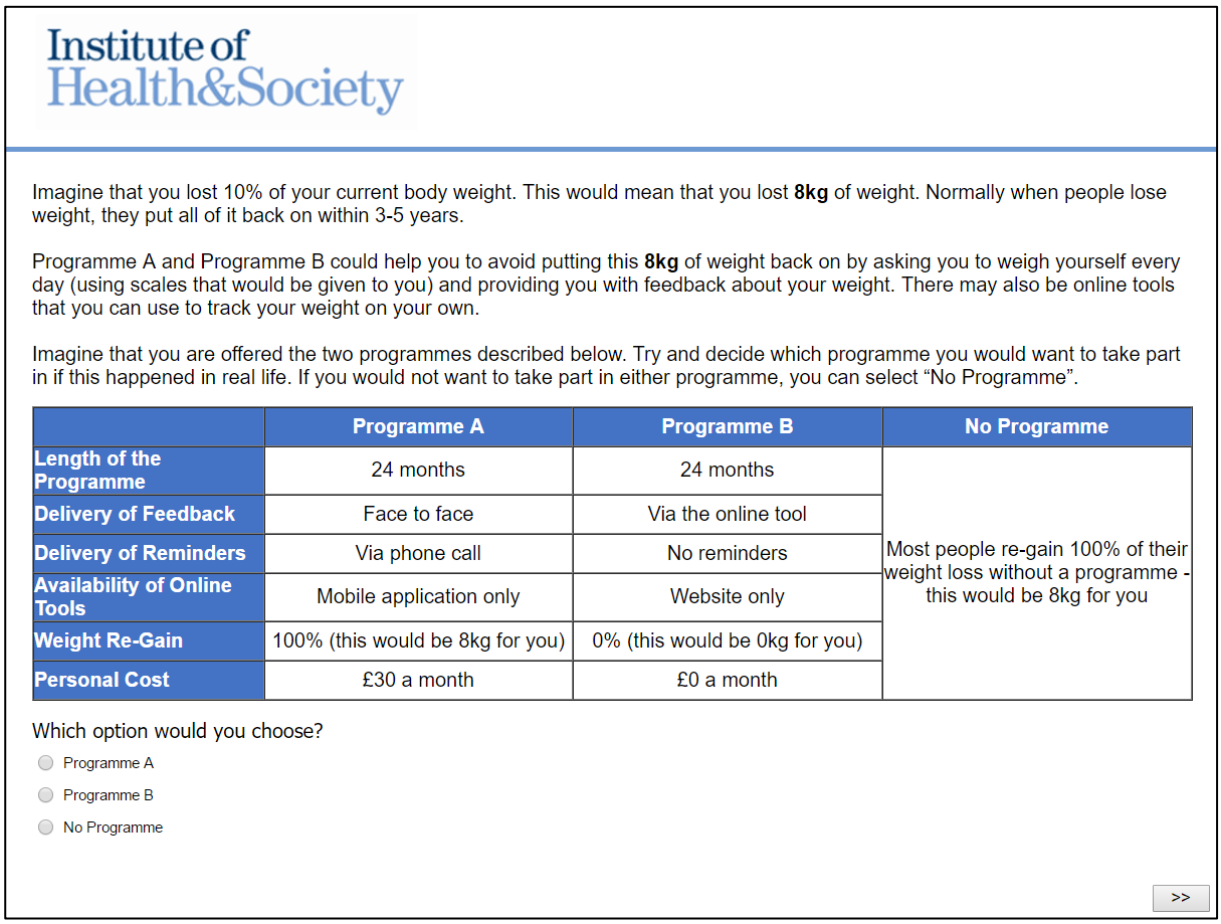
Where y_i is a dummy dependent variable that is equal to one if the rationality test was passed by individual i, $\boldsymbol{\delta}_i$ is a vector of sociodemographic variables, $\boldsymbol{\gamma}_i$ is a vector of DCE-related variables (e.g. the block, difficulty rating) and $\boldsymbol{\theta}_i$ is a vector of attitudinal variables (see section 6.7). The user group variables (SU_i, PSU_i and PB_i) are dummy variables indicating membership of the user group and the coefficients (β's) will indicate whether pass rates are likely to differ across user groups. Given the dummy dependent variable, the use of ordinary least squares (OLS) may be considered suboptimal as the linear probability model (LPM) suffers from well-known limitations (Gujarati & Porter, 2009). Therefore, it would be more appropriate to use logit models and estimate marginal effects in order to explore how pass rates may differ across user groups.

6.4.4 Piloting

Once the experimental design and rationality tests had been decided, a test version of the DCE task was created for pilot testing. This used only one block of the experimental design and respondents were asked to assume that they weigh 80kg and lost 10% of their weight. The plan in the final survey was to customise these figures according to respondents' self-reported weight. However, as this requires highly complex programming it was not put into practice for the pilot tests. Figure 6.5 shows an example choice task in the pilot.

Pilots were conducted as 'think aloud' pilots; these are conducted face-to-face and require the participant to explain their thought processes to the interviewer as they complete the task, and are thought to provide more useful information than online pilots (Ryan et al., 2009). Five think aloud pilots were conducted in total, with adjustments made to the wording and display of the DCE task in between pilots. The participants consisted of two PhD students from the Health Economics Group within the Institute of Health & Society at Newcastle University, as well as three individuals that had no link to the project nor any economics education.

Figure 6.5 An Example Choice Task Used in the Pilots



Institute of Health&Society

Imagine that you lost 10% of your current body weight. This would mean that you lost **8kg** of weight. Normally when people lose weight, they put all of it back on within 3-5 years.

Programme A and Programme B could help you to avoid putting this **8kg** of weight back on by asking you to weigh yourself every day (using scales that would be given to you) and providing you with feedback about your weight. There may also be online tools that you can use to track your weight on your own.

Imagine that you are offered the two programmes described below. Try and decide which programme you would want to take part in if this happened in real life. If you would not want to take part in either programme, you can select "No Programme".

	Programme A	Programme B	No Programme
Length of the Programme	24 months	24 months	Most people re-gain 100% of their weight loss without a programme - this would be 8kg for you
Delivery of Feedback	Face to face	Via the online tool	
Delivery of Reminders	Via phone call	No reminders	
Availability of Online Tools	Mobile application only	Website only	
Weight Re-Gain	100% (this would be 8kg for you)	0% (this would be 0kg for you)	
Personal Cost	£30 a month	£0 a month	

Which option would you choose?

- Programme A
- Programme B
- No Programme

>>

It was clear from the pilots that the combination of the 'weight re-gain' and 'personal cost (per month)' attributes were often driving the choices being made, however only one respondent focused exclusively on one attribute (weight re-gain). The pilots also reinforced the expectations of interaction effects between the 'length of programme' attribute and others such as 'personal cost (per month)' (i.e. relating to the total cost of the programme) and 'weight re-gain' (i.e. relating to the rate of weight re-gain).

Changes made to the survey as a result of these pilots primarily related to clarifications about the way the programmes worked as well as general wording. An example of the former is that, prior to these pilots, the frequency of feedback was not stated during the DCE task however this was subsequently added to one of the information pages before the scenario questions. An example of the latter is that, prior to these pilots, the fact that most individuals re-gain 100% of their weight loss between 3-5 years (Avenell et al., 2004) was part of the wording above the table containing the different alternatives (see figure 6.5). It was felt that this was confusing when contrasted with the timeframes used in the "length of programme" attribute and hence this fact was moved to the information pages prior to the scenarios.

In addition to the think aloud pilots, Research Now conducted two pilots of the (near-finalised) survey on 2nd February (n=177) and 10th February 2016 (n=113) in order to ensure that the data collection process was running as expected. No issues were identified with the DCE data in either pilot, although no choice modelling was undertaken with the pilot data. In health economics, where prior knowledge of DCE parameter estimates is rarely available, pilots are sometimes used to generate priors which can then be used to create a more efficient experimental design (Reed Johnson et al., 2013). In this case, it did not seem appropriate to take this approach. This is because the intention was to compare preferences across four different user groups; using one set of priors would have contradicted the hypothesis for RQ1, whereas using four sets of priors would have resulted in four experimental designs and reduce the comparability of the end results. Additionally, with such a small sample size (n=288) in the NULevel trial and the possibility of recruiting an insufficient number of respondents to run the choice models, it would have been too risky to use NULevel trial participants in a pilot (as this would render them ineligible for the final survey).

6.4.5 Determining Sample Sizes

The ideal method for determining sample sizes for a DCE is a point of contention in the literature. A recent paper states that 70% of the healthcare-related DCE studies published in 2012 did not clearly report whether and what kind of sample size method was used (de Bekker-Grob et al., 2015). Of those that did explain their method, there were three general approaches: rules of thumb; referring to existing literature; and parametric approaches. For this DCE, the former two approaches were utilised.

Specifically, the rule of thumb by Johnson and Orme (2003) was used once the attributes and levels were determined. This involves the use of equation 6.3:

$$N > \frac{500c}{t \times a} \tag{6.3}$$

When considering main effects: c is equal to the highest number of levels used for an attribute, t is the number of choice tasks and *a* is the number of alternatives. For this DCE this translates to:

$$N > \frac{500 \times 8}{10 \times 3} = N > 133.33 \tag{6.4}$$

This means that in order to estimate the main effects for the DCE the sample size should be greater than 133. As three different user groups will be generated from the online panel sample, this minimum sample size of 134 is applicable to each user group. This number also

satisfies the rules of thumb by Pearmain et al. (1991) and Lancsar and Louviere (2008) as these suggest significantly lower numbers would be sufficient. In addition, from the literature review in Chapter Four it was clear that this figure is in line with many existing published DCE studies. However, in an attempt to ensure that sample sizes would not cause any future issues due to analyses involving more than just main effects (for example) this number was inflated to a target of 200 respondents per user group (hence a target of n=600 for the online panel sample as a whole). The same sample size calculation is relevant for the trial sample, however due to the limited numbers of trial participants (n=288) the planned approach was simply to maximise the response rate.

While it has been shown to be commonplace to use rules of thumb it should be noted that this is a suboptimal approach. Even the more advanced parametric approaches have been shown to be unsuitable (de Bekker-Grob et al., 2015). de Bekker-Grob et al. (2015) suggest a superior alternative, however this requires prior knowledge about the significance level, statistical power level, the statistical model that will be used in the analysis, initial beliefs about the parameter values (i.e. priors) and the DCE design itself. With this information, the variance-covariance matrix is estimated and the minimum sample size is determined using the Equation 6.5:

$$N > \left(\left(z_{1-\beta} + z_{1-\alpha} \right) \sqrt{\sum_{\gamma k} / \delta} \right)^2 \qquad (6.5)$$

Where $\sum_{\gamma}$ is the variance-covariance matrix, δ are the effect sizes, $1\text{-}\beta$ is the power level, α is the confidence level and N is the required sample size. As no priors were estimated at the design stage for the DCE, it was not possible to use this method.

6.4.6 Online Programming

With the experimental design finalised and the rationality tests embedded, the next stage was to have the DCE exercise set up as part of the online survey. A market research company, Research Now (http://www.researchnow.com), was employed to undertake the online programming of the survey (in addition to their role of recruiting the online panel sample).

A simplistic ranking exercise was included in the survey design prior to the DCE scenarios. This involved ranking the attributes in order of importance, as well as the attribute levels for each attribute. The attribute levels for the weight re-gain and cost attributes did not have to be ranked as these have an obvious rank-order and it was not felt that testing respondents in this

exercise would be necessary. Responses to the ranking questions were incorporated into the rationality tests as well as the CV task that followed the DCE (see section 6.5).

In principle, the DCE task should be relatively straightforward for respondents however, due to concerns about respondent comprehension of the 'weight re-gain' attribute, it was decided that absolute values should be provided in addition to the *relative* attribute level (i.e. the percentages). This approach was taken in another DCE study and no issues were reported (see the supplementary material of Ho et al., 2015). It was implemented by taking the respondent's self-reported current weight from the screening questions, calculating a 10% weight loss, and using this to provide values for the level of the "weight re-gain" attribute. For example, for a respondent weighing 80kg the DCE scenario would start with:

> *"Imagine that you lost 10% of your current body weight. This would mean that you lost 8kg of weight."*

Hence, in the table containing the DCE attributes and levels instead of stating "50%" (for example) in the "weight re-gain" cell it would say "50% (this means 4kg for you)". It was hoped that using absolute weights would improve respondent understanding of the concept of WLM, aid respondents' understanding of the task as percentages can be easily misunderstood and add an element of reality to the task for those that had not previously considered losing weight. The decision was made to choose an initial 10% weight loss (rather than the 5% used in the NULevel trial inclusion criteria and PSU group classification process) because for individuals with a normal BMI classification (i.e. the non-user group) a 5% weight loss may be too minute to relate to, and/or too minute to be able to feasibly distinguish between different attribute levels for 'weight re-gain'. By way of example, consider a respondent that weighs 50kg being told to imagine that they lost 2.5kg (5% weight loss) and being asked to compare an intervention where they would re-gain 0.25kg (10%) with an intervention where they would re-gain 0.5kg (20%). Neither value is particularly large, nor is the difference between them, hence this may be difficult for the respondent to consider. Figure 6.6 shows the online presentation of a DCE scenario for a respondent that provided a self-reported weight of 14 stone (preferred units of measurement: stones and pounds).

6.4.7 Analysis

As outlined in Chapter Two, DCEs draw upon Lancaster's economic theory of value, which suggests that it is the characteristics of a good that determine the amount of utility it provides, rather than the good as a whole (Lancaster, 1966). DCEs are typically modelled upon a random utility framework where individuals have some construct of utility for choice alternatives and have perfect discrimination capability (McFadden, 1973). Researchers cannot

observe all of the factors affecting the preferences of an individual *n*, hence the latent utility of an alternative *i* is considered to be made up of two components. The first is an explainable component (*V*), known as an indirect utility function, that is specified as a function of the attributes and the second is a random component ($\mathcal{E}_{in}$) representing unmeasured variation in preferences; this is expressed in equation 6.6:

$$U_{in} = V(X_{in}, \beta) + \varepsilon_{in} \tag{6.6}$$

When faced with the choice between a number of options in a DCE task it is believed that a rational individual would choose the option that they prefer, which is assumed to be the option that generates the highest utility. Hence alternative *i* is chosen if and only if that alternative maximises their utility amongst all *J* alternatives included in the choice set C_n. This is shown mathematically in equation 6.7:

$$(V_{in} + \varepsilon_{in}) > (V_{jn} + \varepsilon_{jn}) \ \forall j \neq i \in C_n \tag{6.7}$$

Rearranging this gives equation 6.8:

$$(V_{in} - V_{jn}) > (\varepsilon_{jn} - \varepsilon_{in}) \ \forall j \neq i \in C_n \tag{6.8}$$

As ($\mathcal{E}_{jn} - \mathcal{E}_{in}$) cannot be observed, we can only consider the probabilities of choice outcomes, as in Equation 6.9:

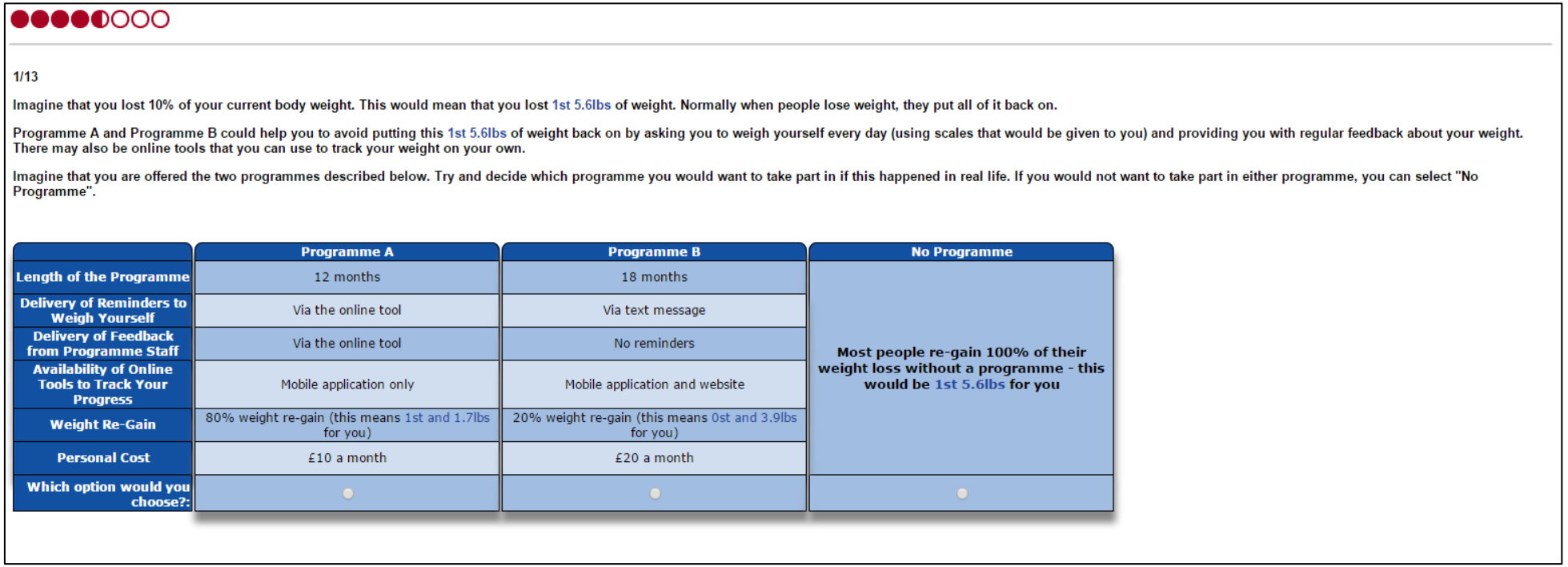
$$P_{in} = Prob(V_{in} - V_{jn} > \varepsilon_{jn} - \varepsilon_{in}) \ \forall j \neq i \in C_n \tag{6.9}$$

The difficulty here is that the actual distribution of ($\mathcal{E}_{jn} - \mathcal{E}_{in}$) across the population is not known and hence assumptions must be made about which distribution or density function it follows in order to estimate choice probabilities. The various different ways in which the distribution of ($\mathcal{E}_{jn} - \mathcal{E}_{in}$) can be defined gives rise to a variety of probabilistic discrete choice models. The specification of the (indirect) utility function to be estimated is another issue when attempting to model choice data. It is commonly assumed, for practical reasons, that it will take the form of a linear-in-parameters function (Ryan et al., 2007).

$$V_{in} = ASC_i + \beta_1 X_{i1} + \cdots + \beta_K X_{iK} \tag{6.10}$$

Equation 6.10 is a linear-in-parameters indirect utility function where there are K attributes and generic coefficients (β_k) across all alternatives. ASC_i is an alternative specific constant which captures the mean effect of the unobserved factors in the error terms for each of the alternatives.

Figure 6.6 Screenshot of a Typical DCE Task in the Online Survey



1/13

Imagine that you lost 10% of your current body weight. This would mean that you lost 1st 5.6lbs of weight. Normally when people lose weight, they put all of it back on.

Programme A and Programme B could help you to avoid putting this 1st 5.6lbs of weight back on by asking you to weigh yourself every day (using scales that would be given to you) and providing you with regular feedback about your weight. There may also be online tools that you can use to track your weight on your own.

Imagine that you are offered the two programmes described below. Try and decide which programme you would want to take part in if this happened in real life. If you would not want to take part in either programme, you can select "No Programme".

	Programme A	Programme B	No Programme
Length of the Programme	12 months	18 months	Most people re-gain 100% of their weight loss without a programme - this would be 1st 5.6lbs for you
Delivery of Reminders to Weigh Yourself	Via the online tool	Via text message	
Delivery of Feedback from Programme Staff	Via the online tool	No reminders	
Availability of Online Tools to Track Your Progress	Mobile application only	Mobile application and website	
Weight Re-Gain	80% weight re-gain (this means 1st and 1.7lbs for you)	20% weight re-gain (this means 0st and 3.9lbs for you)	
Personal Cost	£10 a month	£20 a month	
Which option would you choose?:	○	○	○

The multinomial logit (MNL) model is the typical first choice when analysing choice data where there were two or more alternatives to choose from (McFadden, 1973). The underlying assumption of the MNL model is that the disturbances, ε_{in}, are independent and identically distributed (IID) extreme value type 1 (Gumbel) with mode zero and variance $\frac{\mu^2\pi^2}{6}$ where μ is a positive scale parameter. This gives rise to equation 6.11:

$$P_{in} = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}} \tag{6.11}$$

In a single sample where a linear-in-parameters utility function is assumed, it is not possible to separate the scale parameter (μ) from that of tastes (β's); we can only identify the product of the two. For this reason it is typically assumed that the scale parameter is equivalent to one; this assumption can be relaxed with alternative models.

The MNL model is estimated by finding the values of the β's that maximise the following log-likelihood function (equation 6.12):

$$LnL = \sum_{n=1}^{N} \sum_{i \in C_n} y_{in}(ln(P_{in})) = \sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \left\{ \beta' X_{in} - ln \sum_{j \in C_n} e^{\beta' X_{jn}} \right\} \tag{6.12}$$

The assumption of independent errors leads to the independence of irrelevant alternatives (IIA) property (equation 6.13):

$$\frac{P_{in}}{P_{kn}} = \frac{e^{V_{in}} / \sum_j e^{V_{jn}}}{e^{V_{kn}} / \sum_j e^{V_{jn}}} = \frac{e^{V_{in}}}{e^{V_{kn}}} = e^{V_{in} - V_{kn}} \tag{6.13}$$

This assumption suggests that adding or removing an alternative will not affect the choice probabilities of the other alternatives in the same proportion. The IIA assumption might not always hold and a number of alternative models exist that relax it (Lancsar & Louviere, 2008).

The type of choice behaviour that a researcher expects to observe from a sample will typically influence the type of choice model used. Researchers are encouraged to consider the possible existence of preference heterogeneity in their data (Hauber et al., 2016). This is the assumption that preferences differ across individuals, even after any observable differences in their characteristics are taken into account. The MNL model does not allow for preferences to differ across individuals, hence it assumes preference homogeneity. Given the inability of the MNL model to take preference heterogeneity into account, as well as the problematic IIA assumption, it is rarely considered appropriate to use this model in isolation.

In anticipation of preference heterogeneity, the primary model of interest in this project is the mixed logit (MXL) model, also known as the random parameters logit model (Hess & Train, 2017). The mixed logit choice probability is given by equation 6.14:

$$P_{in} = \int \frac{e^{\boldsymbol{x'}_{in}\boldsymbol{\beta}}}{\sum_{j \in C_n} e^{\boldsymbol{x'}_{jn}\boldsymbol{\beta}}} f(\boldsymbol{\beta}|\boldsymbol{\theta}) d\boldsymbol{\beta} \tag{6.14}$$

Where $f(\boldsymbol{\beta}|\boldsymbol{\theta})$ is the density function of $\boldsymbol{\beta}$. Given that respondents will make several choices in the DCE, the probability of a particular sequence of choices is given by equation 6.15:

$$S_n = \int \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\frac{e^{\boldsymbol{x'}_{int}\boldsymbol{\beta}}}{\sum_{j \in C_n} e^{\boldsymbol{x'}_{jnt}\boldsymbol{\beta}}} \right]^{y_{jnt}} f(\boldsymbol{\beta}|\boldsymbol{\theta}) d\boldsymbol{\beta} \tag{6.15}$$

Where $y_{jnt} = 1$ if the individual chose alternative j in choice situation t and 0 otherwise. The θ parameters can be estimated by maximising the simulated log-likelihood function (equation 6.16):

$$SLL = \sum_{n=1}^{N} ln \left\{ \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\frac{e^{\boldsymbol{x'}_{int}\boldsymbol{\beta}_n^{[r]}}}{\sum_{j \in C_n} e^{\boldsymbol{x'}_{jnt}\boldsymbol{\beta}_n^{[r]}}} \right]^{y_{jnt}} \right\} \tag{6.16}$$

Where $\boldsymbol{\beta}_n^{[r]}$ is the r-th draw for individual n from the distribution of $\boldsymbol{\beta}$. This approach can be implemented in Stata using the user-written *mixlogit* command (Hole, 2007a).

In lieu of any evidence that there are interactions that should be taken into account, a main effects MXL model is the *a priori* preferred specification for analysing the choice data. All variables will be modelled as random and normally distributed. The indirect utility function will be linear and additive and specified as in equation 6.17:

$$\begin{aligned} V_j = \alpha ASC &+ \beta_1 Length + \beta_2 FBPhone + \beta_3 FBOnline + \beta_4 FBText \\ &+ \beta_5 REText + \beta_6 REPhone + \beta_7 REOnline + \beta_8 OTApp \\ &+ \beta_9 OTWeb + \beta_{10} OTBoth + \beta_{11} Outcome + \beta_{12} Cost \end{aligned} \tag{6.17}$$

Where ASC is an alternative-specific constant that is equal to one in alternatives one and two of each choice task (indicating a WLM programme). *Length*, *outcome* and *cost* are all continuous variables. The other variables are categorical and dummy coded.

The sign on the model coefficients indicates whether a change in the attribute level has a positive or negative effect on utility, however the magnitude of the coefficients in their own right provide little meaningful information. Additionally, given the fact that scale will differ across models for each user group, the coefficients cannot be meaningfully compared across

models (Vass, Wright et al., 2017). This poses a challenge when it comes to determining whether differences in the results are due to differences in preferences or differences in scale.

Given that this analysis will compare preferences across four user groups (as opposed to two) and intends to use the MXL model, the traditional method of identifying and adjusting for differences in scale using MNL models i.e. the Swait and Louviere test (Swait & Louviere, 1993) cannot easily be applied. Adjustments using this approach are only suitable for MNL models. As preference heterogeneity is likely to exist (and therefore MNL models are not desirable in this context), the preferred approach is to focus on the issue of preference heterogeneity rather than scale heterogeneity.

As coefficients will not be adjusted as a result of any possible scale heterogeneity between user groups, two approaches are taken to interpret the data which are not affected by differences in scale. One approach is to estimate the relative importance for each attribute as a percentage for each user group using the method used by Jiang & Fraenkel (2017). The first stage of this process is to determine the utility range for each attribute. For example, imagine a categorical variable where the smallest coefficient is -0.5 and the largest coefficient is 0.5; here, the range is equal to one. For a continuous variable that has levels spanning 0 to 30 (e.g. the cost coefficient in this DCE), the range is 30 multiplied by the coefficient. The relative importance is then calculated by dividing the range for each individual attribute by the total range (all attributes). This information can be compared across models as is necessary to examine RQ1 in this project because the effect of the scale parameter is removed.

Another approach to compare the results from each model is to estimate marginal rates of substitution (MRS). A MRS that is of particular interest for RQ2 is willingness to pay (WTP). This can be calculated using equation 6.18:

$$WTP = -\frac{\beta_X}{\beta_{Cost}} \tag{6.18}$$

Where β_X is the coefficient for the attribute or attribute level of interest. It is not necessarily the case that MRS examined should be WTP, however this provides the easiest information to interpret. The user-written command *wtp* in STATA allows for WTP estimates to be calculated post-estimation with accompanying confidence intervals (Hole, 2007b).

It should also be noted that scale heterogeneity can exist between individuals within each user group (Vass, Wright et al., 2017). One approach to deal with this when using MXL models is to estimate a full covariance matrix, as differences in scale within a model is a form of correlation across coefficients (Hess & Train, 2017). Therefore, MXL models will be

estimated with a full covariance matrix in the first instance. However, this requires that a significant number of additional parameters are estimated and, as a result, may be a challenge to run.

6.5 The Contingent Valuation Tasks

In order to provide additional data for research question two, CV questions were included as part of the survey. It was decided that the format of the CV questions would be open-ended to reduce the likely time commitment required from the respondents. In addition, as the concept of WLM and WLM interventions specifically are defined within the DCE task, and because the DCE task is more cognitively demanding, it was decided that the CV questions should be presented afterwards. The CV questions used in the survey are unique in three ways: the individual respondent determines the perspective of the question; the WLM programme being valued is determined by the respondent's own direct rankings of the attribute levels used in the DCE; and WTP is elicited multiple times for the same programme, but with various different outcomes (i.e. levels of weight re-gain). These are addressed in turn in more detail below.

6.5.1 Perspective

The type of value elicited from a CV question varies depending on the wording of the question (Smith, 2007). For example, an individual that has a normal BMI classification could be asked how much they are willing to pay for a WLM intervention to use now (i.e. a use value); in this example, the respondent must imagine that they do in fact need a WLM intervention. As explained in Chapter Two, other common perspectives give rise to option values or an estimate of a caring externality.

The main focus of this project is to examine the differences in preferences between different samples, rather than the effect of alternate perspectives on preferences. Despite this it was decided that respondents should only see a relevant perspective for them (rather than just one), as it was hypothesised that this pragmatic approach would generate more realistic data. The hypothesis is that forcing an individual that would happily state outright that they have absolutely no interest in a WLM programme to provide a WTP value for a WLM programme is not likely to provide valuable or reliable responses, even if they are asked to *pretend* that they do need a WLM programme. The selection of a relevant perspective was done by using the following survey question to split the respondents into groups (hereby referred to as the 'self-selection question'):

"Would you like a weight loss maintenance programme, similar to those described in the choice scenarios, to be made available? (Please choose one option)"

The possible responses to this question and the CV task that would follow each response are described in table 6.5.

Table 6.5 Responses to the CV Self-Selection Question & Associated CV Task

Responses to the Self-Selection Question	**Associated CV Task**
1. Yes – because I would like to use a programme like this	Use Value
2. Yes – because I would like to use a programme like this & I want it to be available for others to use too	Use Value
3. Yes – because I might want to use a programme like this in future	Use Value
4. Yes – because I might want to use a programme like this in future & I want it to be available for others to use too	Use Value
5. Yes – I would not use it myself, but I want it to be available for others to use	Caring Externality
6. No – I do not think a programme like this should be made available	None

If a respondent selected options one to four and were faced with the use value CV task, they would initially see the following question (the programme described varies across respondents – see section 6.5.2):

"*Would you be willing to pay to take part in this weight loss maintenance programme if it was offered to you?"*

If the respondent answered "no" to the above question the task would end. However, if they answered "yes", the respondent would then be asked:

"What is the maximum amount you would be willing to pay (per month, for 12 months) to take part in this weight loss maintenance programme?"

Respondents were given the option to state that they are not sure how much they are willing to pay. After responding to this question, respondents would see the previous question again, with a marginally different programme defined in terms of weight re-gain (see section 6.5.3) and the process would continue until they stated that they are not willing to pay for a programme (or until they reached the 100% weight re-gain outcome; see section 6.5.3).

Respondents choosing option five in the self-selection question would face the caring externality question, where the WLM programme is not specifically defined:

"Would you be willing to pay (as a one-off payment) to enable a weight loss maintenance programme to be provided to those that need it?"

If the respondent answered "no" to the above question the task would end. However, if they answered "yes", the respondent would then be asked:

> *"What is the maximum amount that you would be willing to pay (as a one-off payment) in order to enable a weight loss maintenance programme to be provided to one individual for 12 months?"*

Respondents were given the option to state that they are not sure how much they are willing to pay. Once an answer is given the task ends for these respondents.

Following the CV tasks, all respondents are directed to an open-ended comment box. For those that selected option six in the self-selection question, they would be directed to this comment box straight away (i.e. they would not complete a CV task). The assumption is that these respondents would not be willing to pay for a WLM programme and might not provide meaningful data if they were asked to pretend that they needed a WLM programme, given that they do not think that WLM programmes (as described in the survey) should be made available.

6.5.2 The Object Being Valued

With any CV question, it is important to be clear as to what the respondent is being asked to value. In 1993, the National Oceanic and Atmospheric Administration (NOAA) stated the following in their report:

> *"If CV surveys are to elicit useful information about willingness to pay, respondents must understand exactly what it is they are being asked to value (or vote upon) and must accept the scenario in formulating their responses. Frequently, CV surveys have provided only sketchy details about the project(s) being valued and this calls into question the estimates derived therefrom."* (Arrow et al., 1993, p.13)

As the questions were delivered after the DCE task it was decided to utilise the attributes & levels from the DCE to describe the WLM programmes to the respondent in the use value CV task. However, as no clear *a priori* expectations existed regarding how individuals might value some of the attributes in the DCE task, it would have been somewhat arbitrary to specify a single WLM programme. As a result it was decided that the respondent's directly-elicited rankings (from the questions prior to the DCE task) would be used and the combination of the respondents "most preferred" attribute levels would be shown for the following three attributes: delivery of feedback about your weight; delivery of reminders to weigh yourself; and availability of online tool(s). The intention was to try and ensure that the respondents saw a programme that, at the very least, had the potential to be of value to them. In theory, this would improve the response rate. Whilst some individuals may have a strong

preference regarding the length of the programme, this attribute was not varied between respondents because it was decided to elicit WTP per month for a WLM programme with an overall length of 12 months (to enable clearer comparisons with the DCE WTP estimates). The outcome attribute was changed iteratively (see section 6.5.3) but in the first WTP question this was set to 0% weight re-gain. This meant that, if the respondents' directly-elicited rankings were accurate, or at least stable, across the survey, it could be concluded that if the respondent was *not* willing to pay for the WTP programme shown in the first iteration then the respondent would be willing to pay for *any* WLM programme. In other words, the assumption is that if the respondent would not pay for their most-preferred programme with the best possible outcome, they would not be willing to pay for any programme.

In the caring externality CV task, the WLM programme was not specifically defined. It was decided that defining the WLM programme was not necessary considering that it would be consumed by someone else and that defining a specific programme could lead to biased responses. For example, defining the programme using the respondents' preferred attribute levels from the ranking exercise could artificially inflate the WTP estimate (and vice versa).

6.5.3 Iterative Outcomes

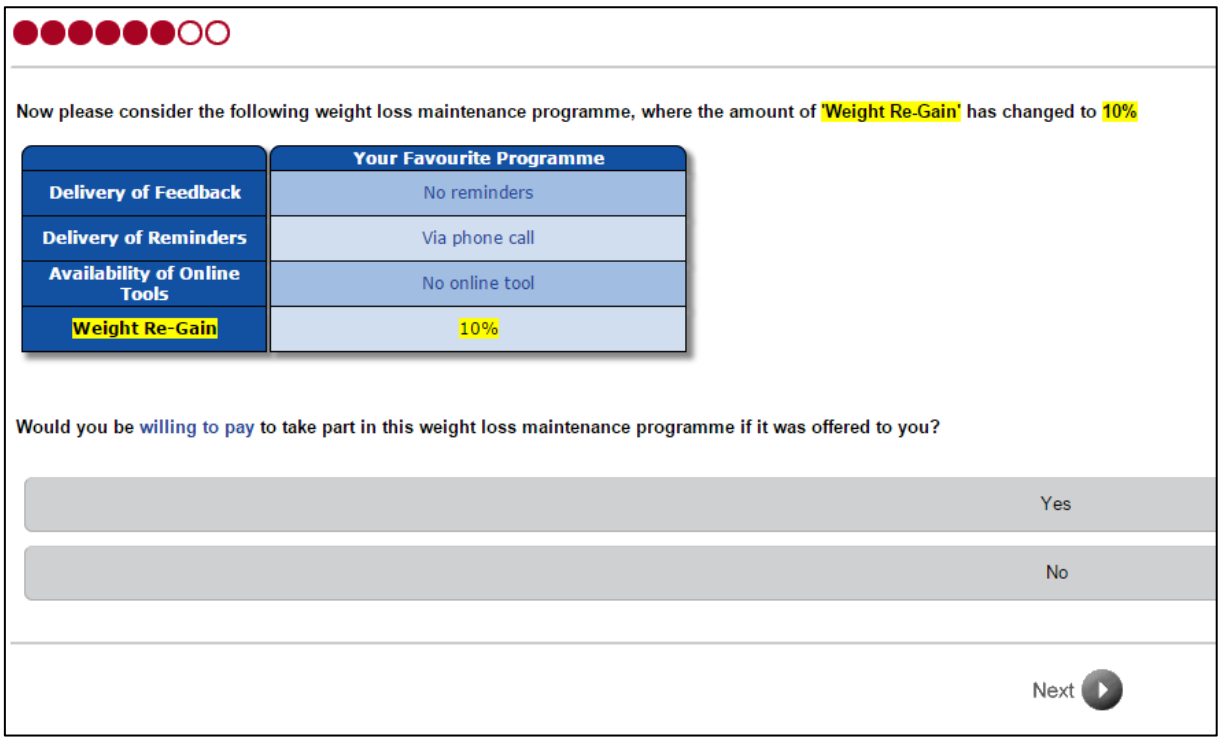
This is only relevant for respondents that self-selected options one to four (use value CV task).

As stated previously, the first CV question used the 0% weight re-gain outcome. In order to examine the importance of the outcome on respondents' willingness to pay, this was changed iteratively. If a respondent stated that they were willing to pay for a WLM programme that results in 0% weight re-gain, then the respondent would be asked if they would be willing to pay for a programme that results in 10% weight re-gain (where all else would remain equal). This continued up to 8 times (i.e. all of the attribute levels for the 'weight re-gain' attribute), with the process ceasing if the respondent expressed that they would not be willing to pay for a particular WLM programme. The idea was that this could provide more useful information for the NULevel trial, given that the trial outcomes had not been determined at this stage.

After an examination of the pilot data from the Research Now pilot on 2nd February 2016, it became clear that some respondents did not notice the change in the outcome attribute between the different iterations due to unexpected results and some comments in the open-ended comment box that followed the CV task. This was addressed by highlighting the changes to the weight re-gain attribute on the page in yellow. In the subsequent pilot on 10th February 2016 it appeared that confusion regarding this process had been significantly

reduced. Figure 6.7 shows how a typical question in a use value CV task was displayed to respondents if they made it to the second iteration (10% weight re-gain).

Figure 6.7 Screenshot of a Typical Use Value CV question in the Online Survey



6.6 Sociodemographic & Health Questions

Sociodemographic information is highly useful when comparing the characteristics of different samples. It is common to use such information as covariates in different analyses (e.g. the logit models for the rationality tests, see section 6.4.3). The sociodemographic questions used in the survey included: marital status; highest educational qualification; number of dependent children; age of youngest dependent child; number of individuals that the individual lives with; employment status and; household income (banded). The wording of these questions was taken directly from the wording used in the NULevel trial in the baseline questionnaire. This was due to the fact that the initial plan was to not include these questions in the trial sample and to instead link trial participants' data from the trial with the data obtained in this project. This plan was later changed (see section 6.8) but the wording of the questions remained the same on the basis that they were appropriate for use in this project.

Generic health information is also useful when comparing the characteristics of different samples. Whilst information about respondents' weight is collected during the screening questions, generic health information can provide a broader insight into the health status of respondents. Due to this, the EuroQol Group's EQ-VAS and EQ-5D-5L generic measures of health status were included following the sociodemographic questions (Herdman et al., 2011).

The EQ-VAS is a tool that asks respondents to indicate how good or bad their own health is on that day on a scale of 0 (worst imaginable health state) to 100 (best imaginable health state). The EQ-5D-5L describes health on five dimensions: pain/discomfort; (ability to take part in) usual activities; anxiety/depression; mobility; self-care. Within each dimension there are five options (levels) ranging from extreme problems or no problems at all. The descriptive data is informative in its own right but is commonly also converted into a utility value on a scale where 0 is equivalent to being dead and 1 is equivalent to perfect health, with negative values (worse than dead) possible. The values for each health state vary depending on the tariff used but given the currently limited options available for this particular instrument, the tariff for England developed by Devlin et al. (2017) will be used in this project.

In addition to the EQ-VAS and EQ-5D-5L, both samples faced an additional question asking respondents about the BMI classification that they believed that they fit into. It was thought that this question might provide useful information when combined with the BMI classification derived from the height and weight information in the screening questions. In other words, a variable indicating whether respondents got their BMI classification incorrect (and in which direction) could potentially be used as a covariate in some analyses.

Finally, two additional questions were provided to the online panel sample only. The first was to explore whether individuals that were classified as overweight or obese according to BMI may be due to bodybuilding or other athletic activity rather than, for example, a sedentary lifestyle (Jonnalagadda et al., 2004). This is because there may be substantial differences in the needs of such individuals in relation to WLM programmes. The second was a question asking whether the individuals had taken part in a weight loss maintenance programme (of any kind) before. This was included to explore the assumption that most individuals in the online panel sample would not have experienced WLM programme.

6.7 Attitudinal Questions

Whilst not critical for addressing the research questions in this thesis, attitudinal questions were also included in the survey in the final section. Of particular interest was the possible effect on preferences of negative attitudes towards obese individuals, given that respondents in the non-user group may never have had any individual experience of being overweight or obese. To explore this the Belief About Obese Persons (BAOP) scale was included (Allison et al., 1991). This scale consists of eight statements where respondents are asked to indicate whether they agree or disagree (slightly, moderately or strongly). The scoring system can be found in Appendix B-4; higher scores indicate a stronger belief that obesity is *not* under the obese person's control. An alternate scale, the Universal Measure of Bias: Fat version (UMB-

FAT) was also considered (Latner et al., 2008) but later rejected in favour of the BAOP scale due to concerns about the term 'fat' from the NULevel trial team and a member of the Newcastle University Ethics Committee (see section 6.8 and Appendix B-6).

A short-form version of the Marlowe-Crowne scale was also included (Reynolds, 1982). This is a well-known scale from clinical psychology that examines whether respondents are likely to be susceptible to socially desirable responding (i.e. answering how they *think* they should answer). This was included in part because issues around weight can be quite sensitive to some individuals but also as a potentially useful measure when examining various aspects of the survey such as stated WTP in the caring externality CV task and self-reported BMI. The original scale consists of 33 items (Crowne & Marlowe, 1960), however the version used in this survey contains only 13 items (Short Form C). A shorter questionnaire was preferable due to the length of the survey. The scoring system can be found in Appendix B-5; higher scores potentially indicate that individuals answer in ways that exaggerate their good qualities and minimise their bad ones.

The final question of the survey asks to what extent respondents agree with the following statement:

> *"Answers to this survey will affect the organisation of/delivery of weight loss maintenance programmes."*

The purpose of this question was to explore whether any differences in preferences across the user groups could be explained by the fact that, in some user groups, more people agree/disagree with this statement. For example, it could be the case that respondents that believe their responses to the survey could be influential may have stated higher WTP values in the CV task.

6.8 Ethical Approval

Ethical approval for the delivery of the survey to the online panel sample was sought from Newcastle University Faculty of Medical Sciences Ethics Committee, with the original application sent on 26th March 2015. Responses from two members of the Committee were received on 19th May 2015 which raised some minor issues with the information page (shown to users prior to starting the survey; see the full survey in Appendix B-1 for the final version) and an attitudinal scale that had previously been considered for inclusion in the survey (UMB-Fat). These responses can be found in Appendix B-6. Amendments to the survey were made and responses were sent to the Ethics Committee later that day and ethical approval was granted on 18th June 2015 (Code 00874/2015).

As alluded to earlier, the original plan was to make an ethics amendment to the NULevel trial's NHS ethics application in order to provide the survey to the NULevel trial participants. This would have allowed data linkage between this project and the NULevel trial. However, this plan later changed due to time constraints and ethical approval for the delivery of the survey to NULevel participants was sought from Newcastle University Faculty of Medical Sciences Ethics Committee. This was obtained via an ethics amendment to the existing (approved) ethics application for the online sample, which also included a range of updates to the survey following piloting. The amendment was requested on 26th October 2015 and granted on 6th January 2016 (Code 00874_1/2016). The final approval letter can be found in Appendix B-7.

6.9 Conclusion

This chapter outlined the design process for a survey aimed at providing sufficient data in order to test the three research questions outlined in Chapter Five. This included the context for the survey (case study), the intended sampling process, a section-by-section breakdown of the design process for the survey including an overview of intended data analyses, and the ethical approval process. The following three chapters will provide the results from the data collection and analysis for the online panel sample, the trial sample and the two combined, respectively.

Part III. Results

Chapter 7. Results from the NULevel Trial Sample

The aim of this chapter is to present and summarise the results from the NULevel trial sample in relation to the three research questions presented in Chapter Five. The first research question relates to whether preferences for a health service differ significantly between different user groups (e.g. trial arms), as shown by the results of a DCE. The second, split into two parts, first examines whether WTP estimates are different between user groups and secondly whether estimates differ between methodologies (DCE and CV). The final research question examines whether certain user groups have better defined preferences using the results of embedded rationality tests in the survey.

The chapter begins with a detailed description of the recruitment of the NULevel trial participants (section 7.1) and their characteristics (section 7.2). Each research question is then addressed in turn with both a description of the results as well as discussion (sections 7.3-7.6). Finally, a summary of the results is provided to end the chapter (section 7.7).

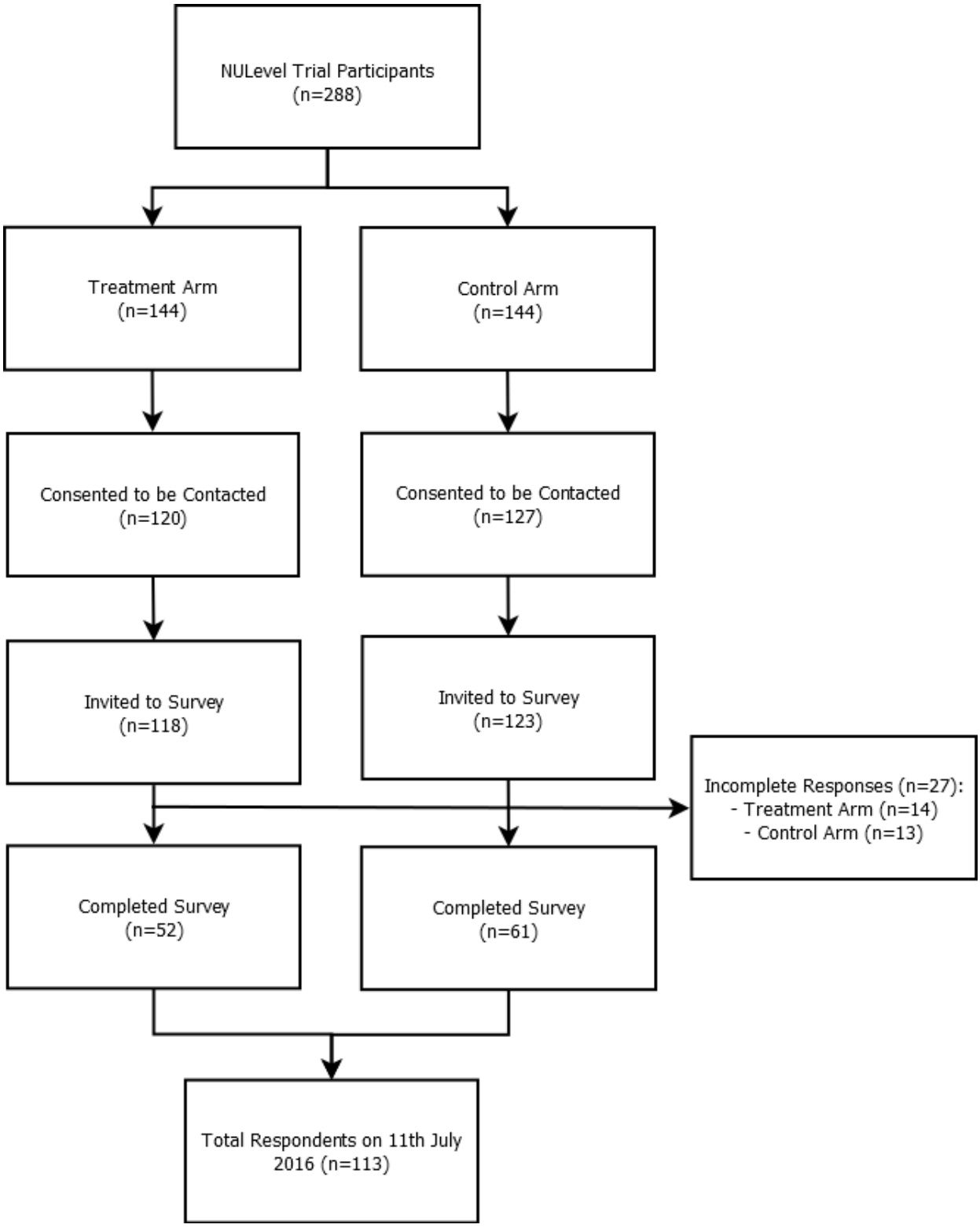
7.1 Response Rates

The survey was hosted online by a market research company (Research Now; http://www.researchnow.com) and NULevel trial participants were invited to participate via email. The URL that linked participants to the survey contained a unique code for each trial participant in order to keep track of the number of unique responses and to avoid contacting individuals that had already completed the survey.

NULevel trial participants were only invited to participate in the survey if they had previously consented to being contacted about future related studies at Newcastle University. This was the case for 247 of the 288 trial participants (86%). The details of participants that had consented were passed on iteratively by a member of the NULevel trial team because participants could not be contacted until after their 12-month assessment in the trial. As a result, email invitations to participate in the survey were sent out on three different occasions in 2016: 25th February (n=150), 21st April (n=34) and 27th June (n=63). Participants that had not responded within four weeks were sent an email reminder, with the exception of the 63 participants that were invited on 27th June. This is because data collection for the PhD project formally ended two weeks later on 11th July in order to allow enough time for data analysis to be completed and the results to be written up in time to meet the thesis submission target. Additionally, while 247 invitations were sent, six invitations were not successfully delivered. Of the 241 successfully invited participants the total number of complete responses was 113,

providing an effective 47% response rate (although in practice recruitment continued beyond 11th July 2016). A flowchart of the recruitment process can be found in figure 7.1.

Figure 7.1 Flowchart of the Trial Sample Recruitment Process



Analysis of the codes that were embedded in the invitation URLs to identify respondents indicated that 27 unique trial participants accessed the survey but failed to complete it (up

until 11th July 2016). These respondents were evenly split across the two arms of the trial; 14 were in the treatment arm and 13 were in the control arm. It was not possible to contact participants that accessed the survey but failed to complete it to determine their reasons for this.

However, there were some technical issues with the delivery of the survey that may explain the incomplete responses (as well as the response rate on the whole). First, when the survey was initially rolled out, the market research company's usual security settings were enabled. These settings are in place to prevent individuals in their online panels from filling out surveys multiple times. Due to the settings, some NULevel trial participants could not access the survey and others could not return to the survey if they left in error. Secondly, in the NULevel trial most correspondence occurred over mobile phone. As a result, a significant percentage of respondents attempted to access the survey using a mobile phone. Once this was realised by the market research company, they indicated that the survey was unstable when accessed using a mobile phone and future invitation emails were adjusted accordingly. However, this is likely to have had an effect as the first wave of invitations covered more than half of the total participants that were available to be contacted.

7.2 Respondent Characteristics

The 113 respondents that completed the survey were spread relatively evenly across trial arms: 52 were in the treatment arm and 61 were in the control arm. Table 7.1 summarises the characteristics of respondents across the whole trial, as well as split by trial arm.

On the whole, respondents are closer to retirement age than working age, very well educated and much more likely to be female than male. Most respondents are employed (full-time or part-time) and those that are not are typically retired. Household income is slightly more varied across the sample, with a significant number of respondents in each of the categories. With regards to health-related characteristics, the majority of the sample are overweight or obese (in equal proportions) however the general health measures suggest that health is good on the whole; the average VAS value is 79 and the average EQ-5D-5L utility value is 0.887.

There is very little evidence to suggest that characteristics differ significantly between respondents from each of the trial arms, as indicated by the p-values in table 7.1. The only characteristic where there is some (moderate) evidence to suggest that the two trial arms differ is the BMI classifications. The control arm has far more obese individuals and fewer overweight individuals relative to the treatment arm. This is a logical difference, given that

the control arm had not recently received a weight loss maintenance intervention for a 12-month period.

Table 7.1 Respondent Characteristics (Trial Sample)

	All	Treatment	Control	*p-values*[1]
n	113	52	61	
Age (mean)	49	48	50	*0.493*
18-24	0 (0%)	0 (0%)	0 (0%)	
25-34	12 (11%)	4 (8%)	8 (13%)	
35-44	34 (30%)	14 (27%)	20 (33%)	
45-54	32 (28%)	17 (33%)	15 (25%)	
55+	35 (31%)	17 (33%)	18 (30%)	
Gender				*0.251*
Male	29 (26%)	16 (31%)	13 (21%)	
Female	84 (74%)	36 (69%)	48 (79%)	
BMI Classification				*0.041*
Normal	7 (6%)	2 (4%)	5 (8%)	
Overweight	53 (47%)	31 (60%)	22 (36%)	
Obese	53 (47%)	19 (37%)	34 (56%)	
Education				*0.498*
Higher (University)	65 (58%)	34 (65%)	31 (51%)	
Further (College)	23 (20%)	10 (19%)	13 (21%)	
Secondary (School)	20 (18%)	6 (12%)	14 (23%)	
None	2 (2%)	1 (2%)	1 (2%)	
Other	3 (3%)	1 (2%)	2 (3%)	
Employment				*0.406*
Employed (full-time)	64 (57%)	29 (56%)	35 (57%)	
Employed (part-time)	13 (12%)	5 (10%)	8 (13%)	
Student	1 (1%)	0 (0%)	1 (2%)	
Retired	20 (18%)	13 (25%)	7 (11%)	
Homemaker/Caregiver	4 (4%)	2 (4%)	2 (3%)	
Unemployed	2 (4%)	1 (2%)	1 (2%)	
Other	9 (2%)	2 (4%)	7 (11%)	
Household Income				*0.439*
£20,000 or less	18 (16%)	7 (13%)	11 (18%)	
£20,001 - £40,000	39 (35%)	22 (42%)	17 (28%)	
£40,001 - £60,000	33 (29%)	13 (25%)	20 (33%)	
£60,001 and above	23 (20%)	10 (19%)	13 (21%)	
Health (Visual Analogue Scale; 0=Dead, 100=Perfect Health)				*0.888*
Mean	79.0	79.2	78.8	
(Standard Deviation)	(17.11)	(17.52)	(16.89)	
Health (EQ-5D-5L Utilities[2]; 0=Dead, 1=Perfect Health)				*0.999*
Mean	0.887	0.888	0.887	
(Standard Deviation)	(0.128)	(0.145)	(0.112)	

[1]p-values based on chi2 tests (categorical) and one-way ANOVA (continuous)

[2]English population utilities from Devlin et al. (2017)

7.3 Research Question One

RQ1: To what extent do preferences for a health service, elicited via a discrete choice experiment, differ between different user groups, and why might these differences occur?

7.3.1 Results

Table 7.2 presents the output from mixed logit models with all parameters set as random (and normally distributed) using 1,000 Halton draws. This means that preference heterogeneity is controlled for and the coefficients are the mean values for each parameter, hence the standard deviations are also reported. Models were also estimated with a full covariance matrix in order to try and address (within-sample) scale heterogeneity; however, the results were deemed unreliable. The results in table 7.2 are generally consistent with the results from conditional logit models that were also estimated (see Appendix C-1), indicating that the results are relatively robust. The statistical significance of numerous standard deviations across the three models in table 7.2 indicates the presence of preference heterogeneity between respondents, suggesting that mixed logit models are more suitable than conditional logit models in this case.

The results from model one in table 7.2 suggest that respondents from the NULevel trial did not typically make choices based on the length of the programme (as indicated by the lack of statistical significance for the length coefficient). This model suggests that respondents strongly prefer to receive feedback via text message relative to face to face (1.006, $p<0.01$). There is also some mild evidence to suggest that respondents prefer to receive feedback via phone call relative to face to face (0.500, $p<0.1$). With respect to receiving feedback via online tool(s) relative to face to face it appears that respondents are indifferent, however there is evidence of preference heterogeneity here indicated by the statistically significant standard deviation. Regardless of the delivery mode, respondents prefer to receive reminders relative to not receiving any. There is a clear hierarchy here, with reminders via the online tool(s) the most preferred and reminders via text message the least preferred. It is also the case that, while indifferent to each online tool individually, respondents appear to prefer the combination of a mobile phone application and a website over not having any form of online tools (0.543, $p<0.05$). Finally, respondents prefer lower weight re-gain (-0.071, $p<0.01$) and programmes that cost less (-0.232, $p<0.01$), although there is clear preference heterogeneity associated with these attributes.

Models two and three in table 7.2 represent the treatment and control arms of the trial, respectively. There are some noticeable differences in the magnitudes and statistical significance of coefficients between the two models. For example, it appears that the control

arm has a stronger preference for the delivery of feedback via phone call as well as via text message relative to the treatment arm. In addition, it appears that the strong preference for all forms of reminders in model one may largely be driven by the control arm, as the coefficients are all larger and more consistent in their statistical significance (all significant at the 1% level). It also appears that the overall preference for the combination of a mobile phone application and website in model one is driven by the control arm, as the coefficient is statistically insignificant for the treatment arm in model two. Finally, the outcome coefficient is nearly 50% larger and the cost coefficient is around 38% larger for the control arm in model three compared with the treatment arm in model two, however the overall scale of the coefficients across model three appears to be larger relative to model two.

Table 7.2 Results from the Mixed Logit Models (Trial Sample)

	1. All Participants		**2. Treatment Arm**		**3. Control Arm**	
	Coefficient	*Std. Dev.*	*Coefficient*	*Std. Dev.*	*Coefficient*	*Std. Dev.*
Alt-Specific Constant	-0.039	3.492 ***	0.591	2.978 ***	-0.115	3.937 ***
	(0.493)	(0.538)	(0.733)	(0.633)	(0.928)	(0.832)
Length (months)	0.016	0.066 ***	0.030	0.087 **	-0.009	-0.030
	(0.015)	(0.020)	(0.023)	(0.039)	(0.023)	(0.027)
Delivery of Feedback[1]						
via Phone Call	0.500 *	-0.079	0.285	-0.426	1.049 **	0.819
	(0.266)	(0.384)	(0.375)	(0.622)	(0.515)	(0.652)
via the Online Tool(s)	-0.324	-1.555 ***	-0.694 *	-0.139	-0.150	2.815 ***
	(0.308)	(0.465)	(0.388)	(0.679)	(0.557)	(0.748)
via the Text Message	1.006 ***	0.752 **	0.81 **	0.578	1.693 ***	-0.538
	(0.251)	(0.302)	(0.356)	(0.466)	(0.496)	(0.468)
Reminders[2]						
via Text Message	0.665 ***	0.553	0.570 *	-0.075	1.182 ***	1.268 ***
	(0.238)	(0.409)	(0.321)	(0.446)	(0.448)	(0.447)
via Phone Call	0.879 ***	0.723 **	0.668 *	-0.224	1.557 ***	0.619
	(0.294)	(0.320)	(0.398)	(0.585)	(0.571)	(0.457)
via the Online Tool(s)	1.414 ***	0.018	1.392 ***	0.018	1.462 ***	-1.172 *
	(0.317)	(0.404)	(0.447)	(0.437)	(0.552)	(0.606)
Online Tool[3]						
App Only	0.329	0.238	0.200	-0.579	0.786 *	1.046 *
	(0.244)	(0.635)	(0.339)	(0.641)	(0.449)	(0.545)
Website Only	-0.003	-0.382	-0.277	-0.616	0.621	0.520
	(0.274)	(0.344)	(0.378)	(0.655)	(0.489)	(0.404)
App & Website	0.543 **	1.293 ***	-0.166	1.065 **	1.442 ***	2.251 ***
	(0.260)	(0.405)	(0.364)	(0.467)	(0.516)	(0.721)
Weight Re-gain (%)	-0.071 ***	0.054 ***	-0.059 ***	-0.037 ***	-0.089 ***	0.085 ***
	(0.009)	(0.007)	(0.010)	(0.007)	(0.018)	(0.019)
Personal Cost (£/month)	-0.232 ***	0.181 ***	-0.201 ***	0.195 ***	-0.278 ***	0.27 ***
	(0.032)	(0.023)	(0.039)	(0.038)	(0.052)	(0.055)
Observations (Sample Size)	3,390		1,560		1,830	
Log Likelihood	-701		-333		-358	
AIC	1,426		692		742	
BIC	1,506		762		814	

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. [1] Base level "face to face" feedback; [2] Base level "no reminders"; [3] Base level "no online tool"

Additional analysis must be conducted in order to meaningfully compare the results across the three models, given the potential differences in scale between samples. Figure 7.2 illustrates the relative importance of each attribute for each user group, based on the method outlined in section 6.4.7. This involves comparing the range of each attribute relative to the total range for all attributes. The ranges for each attribute were calculated using the coefficients from table 7.2 i.e. the main effects mixed logit models. When calculating the ranges for each attribute, coefficients were treated as equal to zero if they were not statistically significant at the 10% level. Hence, for some attributes relative importance is equal to zero.

Figure 7.2 Relative Importance of the Attributes (Trial Sample)

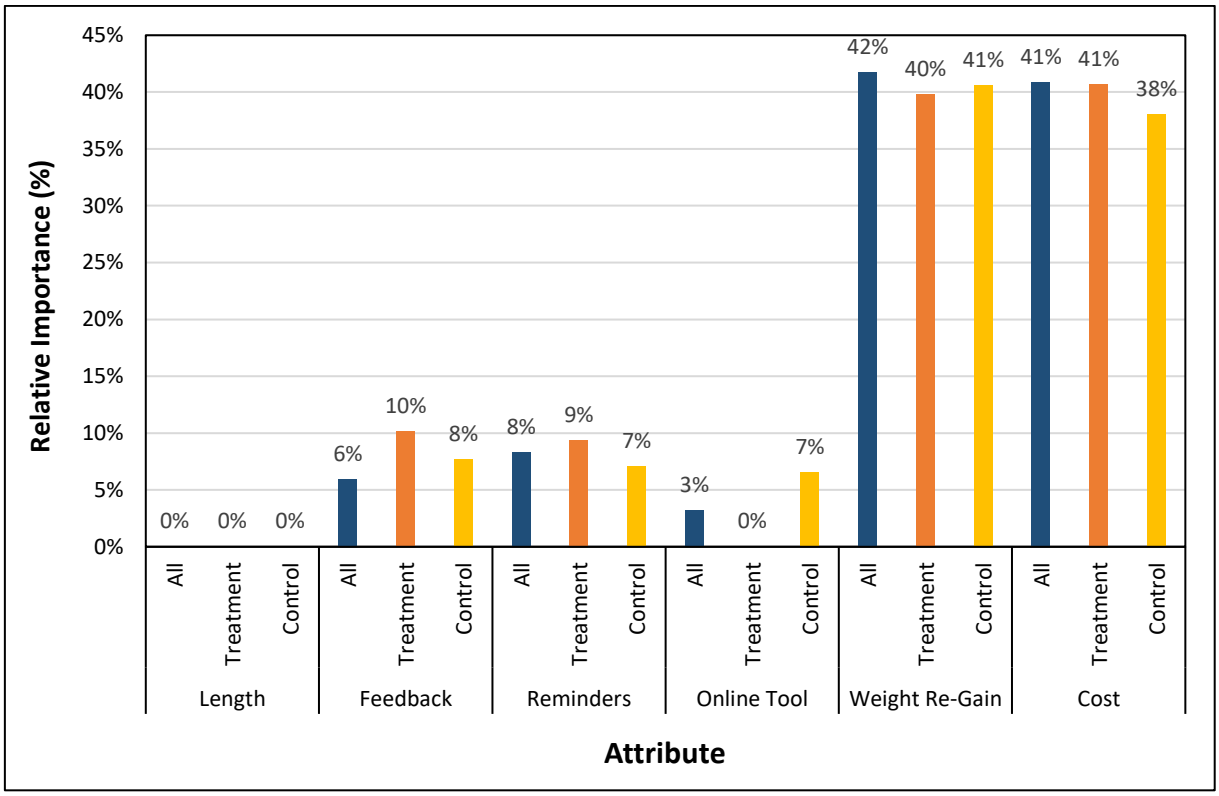


Figure 7.2 provides some insight into how the preferences of respondents from each trial arm differed. The greatest difference is for the relative importance of the online tool attribute; this is because, with no statistically significant coefficients in model two of table 7.2, the relative importance of this attribute is calculated as 0% for the treatment arm. Other differences in the relative importance of the attributes between the two trial arms are minimal.

In addition to relative importance calculations, an additional area of analysis that can provide insight relates to the opt-out option. In total, only one individual in the trial sample chose the opt-out option (no programme) in every scenario; this individual was part of the treatment arm. On average, respondents in the treatment arm opted out 2.1 times throughout the DCE task compared to an average of 1.7 in the control arm (the median value was one for both

arms). On the whole, this suggests that respondents from the trial sample were quite unlikely to select the opt-out option.

Additional choice models were also estimated in order to test the significance of attribute interactions and to identify whether differences in preferences may be observed based on observable characteristics of the respondents (i.e. those that were not varied in the user group definitions).

Results from conditional logit models that contained three interaction terms based on *a priori* expectations (length & cost; outcome & cost; length & outcome) can be found in Appendix C-2. In the full sample model there was moderate evidence ($p<0.05$) to suggest that there is an interaction between length and outcome, as well as length and cost. The coefficients on both terms were positive but of extremely small magnitude. Likelihood ratio tests between the conditional logit models with and without interactions provided weak evidence ($p<0.1$) that the models with interactions provided a better model fit for the full sample but no evidence for the models split by trial arm. As a result, it was not felt that it would be beneficial to incorporate the interaction terms into the more flexible mixed logit models as this would reduce their likelihood of converging, with little expected benefit to model fit.

Whilst the respondent characteristics in table 7.1 indicate very few observable differences between the two trial arms, additional models were estimated containing interactions between the alternative-specific constant and various demographic variables. The variables considered related to age groups, gender, highest educational qualification, employment status and income band. Appendix C-3 provides the results from mixed logit models that included these interactions terms, which were not modelled as random parameters. Some of the interaction coefficients were statistically significant across the three models, however the overall conclusion regarding the preferences of the trial sample were not drastically different. Larger differences were observed between the trial arms, including differences in WTP estimates (not reported), with the vast majority of the attribute coefficients in the control arm model being statistically significant. However, the large alternative specific constant (21.393, $p<0.01$) raised concerns about the reliability of these results. In addition, likelihood ratio tests indicated that model fit had improved for the individual arm models but had not improved for the pooled, whole trial sample. This raised additional concerns, therefore it was determined that these models were not reliable. Given the lack of statistically significant differences in the characteristics of the two trial arms, it was felt that the main effects models are likely to be reliable on their own.

7.3.2 Discussion

The DCE results appear to be relatively robust across different specifications. It is interesting that the length attribute does not appear to have influenced the choices made by the respondents. In the context of WLM this is a sensible finding, given that longer programmes are likely to provide better outcomes but are also associated with higher costs. Respondents appear to value feedback via text message but there is little evidence to suggest that other delivery modes are acceptable. This could be explained by the fact that text messages were used in the NULevel trial for communication purposes. Respondents also appear to prefer reminders over having no reminders, which ultimately is an important part of the intervention. However, it is interesting that respondents do not have a stronger preference in the DCE results for reminders via text message, given that this is the delivery mode used in the NULevel trial. It also appears that respondents do not have a strong preference for the online tools, with only the coefficient on the combination of an app and a website being positive and statistically significant (at the 5% level). This may be a reflection of the fact that the trial itself provided a website which might not have been enjoyed relative to the numerous free, commercial resources that exist in the market. However, the prospect of an app in combination with this may have been tempting, or at least indicated to respondents that a programme that provides both is better value. Finally, it is unsurprising that the outcome and cost attributes were highly dominant given that these have strong implications for service users. The negative coefficients and the magnitude of the relative importance of these two attributes are important when it comes to considering the face validity of these results.

The comparisons of choice models and relative importance between the trial arms provides little evidence to suggest there are any significant differences between the two arms, although the control arm may have a stronger preference for online tools. This would be consistent with the theory that the online tools in the trial were suboptimal and therefore not considered particularly important to the treatment arm.

There was no evidence to suggest that the included attribute interactions were statistically significant, nor was there any convincing evidence that demographic interactions were statistically significant. While this may be seen as positive, given that it provides a relatively strong justification to focus solely on the main effects models, it may be the case that there is a lack of statistical power in the models. In other words, due to the small sample size the models with interaction terms may not have enough statistical power to identify possible effects.

7.3.3 Conclusion

The results from the DCE suggest that trial participants have preferences for various different components of WLM programmes including "process" attributes (those that affect the delivery of the service) as well as the outcome and cost of the programmes. That being said, the outcome and cost attributes appear to be highly dominant when it comes to the choice made by respondents. There is little clear evidence to suggest that preferences differ across the two arms of the trials. Additionally, there is next to no evidence that any attribute or demographic interactions are statistically significant although this could be due to issues with statistical power.

7.4 Research Question Two Part A

RQ2a: To what extent do WTP estimates differ between different user groups, and why might these differences occur?

7.4.1 Results (CV)

Table 7.3 contains the results for each group to the self-selection question; the question that determined whether respondents would be provided with a CV exercise or not. A chi-square test suggests that the responses to this question do not significantly vary across the trial arms ($p>0.1$). However, if similar responses are collapsed (i.e. one and two; three and four - see table 6.5 in Chapter Six), a chi2 test provides moderate evidence ($p<0.05$) to suggest that the responses differ across the trial arms. The apparent differences between the two trial arms are that far more people in the control arm selected the option "I would not use X but I want it to be available for others" and a higher proportion of respondents in the treatment arm indicated that they would "like to use" a WLM programme. Regardless, the vast majority of respondents were forwarded to the use value CV task; 94% of respondents from the treatment arm and 89% of respondents from the control arm.

As outlined in Chapter Six, the use value CV questions included in the survey provide WTP estimates for each individual respondent's preferred set-up of a WLM intervention, excluding the length (as this was held static at 12 months) and weight re-gain (as this changed in each question). Table 7.4 describes the most commonly valued WLM programme from the use value CV task, which did not differ between the two trial arms. However, whilst this was the most frequent, it was by no means valued by the majority of respondents. In fact, a very large range of different programmes were valued across the two arms of the sample; 11 in the treatment arm and 19 in the control arm.

Table 7.3 Responses to the CV Self-Selection Question (Trial Sample)

Response to the "Interest Question"	**All**		**Treatment**		**Control**	
	n	**%**	**n**	**%**	**n**	**%**
1. I would like to use x	17	15%	8	15%	9	15%
2. I would like to use x & want it to be available for others	42	37%	25	48%	17	28%
3. I might want to use x	13	12%	4	8%	9	15%
4. I might want to use x & want it to be available for others	31	27%	12	23%	19	31%
5. I would not use x but I want it to be available for others	9	8%	2	4%	7	11%
6. I do not think x should be made available	1	1%	1	2%	0	0%
Total	**113**	**100%**	**52**	**100%**	**61**	**100%**

x = a programme like this (referring to the hypothetical programmes from the DCE task)

Responses 1-4 – forwarded to the "use value" WTP exercise;

Response 5 – forwarded to the "externality" WTP exercise;

Response 6 – forwarded to open-ended comment box.

Table 7.4 Most Commonly Valued Programmes in the Use Value CV Task (Trial Sample)

Attribute	Control Arm	Treatment Arm
Delivery of Feedback	via Text Message	via Text Message
Delivery of Reminders	via Text Message	via Text Message
Availability of Online Tool(s)	Website & App	Website & App
	N=12/54 (22%)	N=22/49 (45%)

The questions began by asking respondents if they would be willing to pay for a programme with a weight re-gain outcome of 0% and subsequently asked for a maximum price that respondents would be willing to pay (per month for 12 months). The outcome was then adjusted to the next weight re-re-gain level (i.e. 10%) and the task was repeated. This happened until the respondent stated that they were not willing to pay for the specified programme. Table 7.5 presents the full results from the use value CV exercise. Three respondents were removed from the analysis for indicating that they would be willing to pay more at a higher level of weight re-gain (referred to as the 'logic rule' hereafter). It was assumed that such instances represented erroneous responses or a failure of the respondent to fully understand the task. Note that these respondents are still incorporated into the 'stated' column because they indicated a willingness to pay regardless.

Of the 103 respondents that faced the task, 81 (79%) stated that they would be willing to pay for the WLM programme at 0% weight re-gain. Of those, 67 provided WTP values with a median value of £10.00 and a mean value of £17.81. The results split by trial arm indicate that much of the variation in WTP came from respondents in the control arm; the maximum value in the treatment arm was £30 whereas the maximum value in the control arm was £250. In terms of median WTP, the values did not vary at 10% or 20% weight re-gain (i.e. it remained at £10.00) and median WTP was identical across trial arms for the 0%, 10% and 20% outcome levels. By 60% weight re-gain only 13 respondents indicated that they were willing to pay for the programme and only 9 respondents provided a WTP value. When comparing the two arms of the trial, it appears that similar numbers of respondents were willing to pay (and provided a value) for a programme at each outcome level. One-way ANOVA suggests that the average values for each trial arm are not statistically significantly different from one another.

Table 7.5 WTP Estimates from the Use Value CV Task (Trial Sample)

All Respondents (n=113)								
Weight Re-Gain	n	Stated[1] n	Stated[1] %	Provided[2] n	Provided[2] %	Willingness to Pay (WTP) Median	Mean	SD
0%	103	81	79%	67	65%	£10.00	£17.81	32.19
10%	81	63	78%	50	62%	£10.00	£13.66	16.36
20%	63	43	68%	30	48%	£10.00	£10.07	5.79
40%	43	24	56%	14	33%	£8.00	£8.29	5.61
60%	24	13	54%	9	38%	£5.00	£7.11	6.13
80%	13	5	38%	3	23%	£5.00	£8.67	10.02
90%	5	4	80%	2	40%	£3.00	£3.00	2.83
100%	4	4	100%	2	50%	£3.00	£3.00	2.83

	Treatment Arm (n=52)								Control Arm (n=61)							
Weight Re-Gain	n	Stated[1] n	Stated[1] %	Provided[2] n	Provided[2] %	Willingness to Pay (WTP) Median	Mean	SD	n	Stated[1] n	Stated[1] %	Provided[2] n	Provided[2] %	Willingness to Pay (WTP) Median	Mean	SD
0%	49	39	80%	32	65%	£10.00	£12.91	6.39	54	42	77%	35	65%	£10.00	£22.29	43.95
10%	39	35	71%	26	53%	£10.00	£11.27	6.17	42	28	52%	24	44%	£10.00	£16.25	22.70
20%	35	22	45%	13	27%	£10.00	£10.62	6.85	28	21	39%	17	31%	£10.00	£9.65	5.01
40%	22	12	24%	7	14%	£5.00	£7.57	6.50	21	12	22%	7	13%	£10.00	£9.00	4.97
60%	12	5	10%	3	6%	£2.00	£7.67	10.69	12	8	15%	6	11%	£7.50	£6.83	3.76
80%	5	4	8%	2	4%	£10.50	£10.50	13.44	8	1	2%	1	2%	£5.00	£5.00	-
90%	4	3	6%	1	2%	£1.00	£1.00	-	1	1	2%	1	2%	£5.00	£5.00	-
100%	3	3	6%	1	2%	£1.00	£1.00	-	1	1	2%	1	2%	£5.00	£5.00	-

Percentages are based on the number of individuals that were forwarded to the exercise from each user group in the first instance (underlined).

[1]Those that selected "yes" when asked whether they would be willing to pay (those that select "no" exit the exercise at that stage).

[2]Those that provided a value when asked for the maximum amount that they would be willing to pay *and* did not increase their value at any point during the task ('logic rule')

7.4.2 Discussion (CV)

The results suggest that the vast majority (n=103; 91%) of respondents from the trial sample would, or might, want to use a WLM programme like the ones from the DCE task. Of these, 81 (79%) would be willing to pay for a WLM programme with their preferred set-up (in terms of delivery of reminders, delivery of feedback and availability of online tools) and the best possible outcome (0% weight re-gain). There is little difference in the proportion of respondents that were willing to pay at this outcome level in each treatment arm; 80% of respondents in the treatment arm and 77% of respondents in the control arm. Hence, while these are positive results from the perspective of the NULevel trial, there is no evidence to suggest that those with experience of the full intervention are more willing to pay than those that only received the control intervention.

A difference between the trial arms can be seen when looking at how many respondents drop out of the CV task when weight re-gain changes from 0% to 10%. In the treatment arm, 4 respondents drop out (10%) whereas in the control arm 14 respondents drop out (33%). It may be the case that the treatment arm have a greater tolerance for minor weight re-gain, which could be explained by their active participation in the trial. That is, these individuals might have a better understanding of the difficulties of achieving 0% weight re-gain through their experience of monitoring their weight on a regular basis. Therefore, these individuals may be more likely to view a small amount of weight re-gain as a positive outcome, which is how it would be regarded by healthcare professionals, considering the outcomes observed in trials of WLM interventions (Dombrowski et al., 2014).

On the whole the data performs as expected, with fewer people being willing to pay as the efficacy of the programme decreases. Median WTP decreases steadily as efficacy decreases and mean WTP decreases at every outcome level with one exception between the 60% and 80% outcome levels. Due to the use of the 'logic rule', this increase is explained by an individual with a low WTP value leaving the task after the 60% outcome level, rather than individuals increasing their values at 80%. As a result, it seems fair to conclude that this atypical CV task was relatively successful in practice.

7.4.3 Results (DCE)

WTP estimates can also be generated from the DCE by calculating the marginal rate of substitution (MRS) between an attribute or attribute level and the cost attribute as explained in section 6.4.7. Only three of the attributes and attribute levels are statistically significant at the 1% significance level in the models for each trial arm in table 7.2: reminders via the online tool(s); weight re-gain; and cost. Table 7.6 contains the WTP estimates for reminders via the

online tool(s) for each trial arm using the models from table 7.2.

Table 7.6 WTP Estimates from the DCE for the Online Tool(s) Attribute (Split by Trial Arm)

Group	WTP for Reminders via the Online Tool(s)	Confidence Interval [Lower, Upper]
Treatment	£6.94	[£2.77, £11.11]
Control	£5.27	[£1.48, £9.05]

The WTP estimates, considering that they relate to monthly payments, are substantive. When comparing the magnitude of the estimates between the two trial arms, there is a sizeable difference of £1.67. However, the confidence intervals are particularly wide for both estimates, with considerable overlap. As a result, there is no evidence that the estimates differ between the trial arms. Table 7.7 contains the WTP estimates for weight re-gain for each trial arm using the models from table 7.2.

Table 7.7 WTP Estimates from the DCE for the Weight Re-Gain Attribute (Split by Trial Arm)

Group	WTP for % weight re-gain	Confidence Interval [Lower, Upper]
Treatment	-£0.30	[-£0.40, -£0.20]
Control	-£0.32	[-£0.39, -£0.25]

The results are interpreted as such: the treatment arm is willing to pay £0.30 on average to avoid a percentage point increase in weight re-gain, relative to £0.32 in the control arm. It should be clear given the overlap in the confidence intervals that these estimates are *not* statistically significantly different from one another. Figure 7.3 illustrates the WTP estimates for weight re-gain for both the pooled sample and the individual trial arms.

Given this lack of difference across the trial arms, table 7.8 lists the WTP estimates for every attribute or attribute level that is statistically significant at the 10%, 5% or 1% level using the full sample results in model 1 from table 7.2. There are some substantive WTP estimates for various delivery modes of feedback and reminders as well as for the combination level of the online tool attribute. For example, respondents would be willing to pay an extra £2.87 per month on average to receive reminders via text message. While there is only one instance

where the confidence interval covers both a positive and a negative value (for feedback via phone call), however the confidence intervals are wide for most estimates. This means that, although it is generally clear whether the estimates are positive or negative, there is little confidence with respect to whether the estimates are close to the true values. In contrast, the confidence interval for the WTP estimate for weight re-gain is particularly narrow providing some evidence to suggest that the estimate is reasonably reliable.

Figure 7.3 WTP Estimates from the DCE for Weight Re-Gain (Trial Sample)

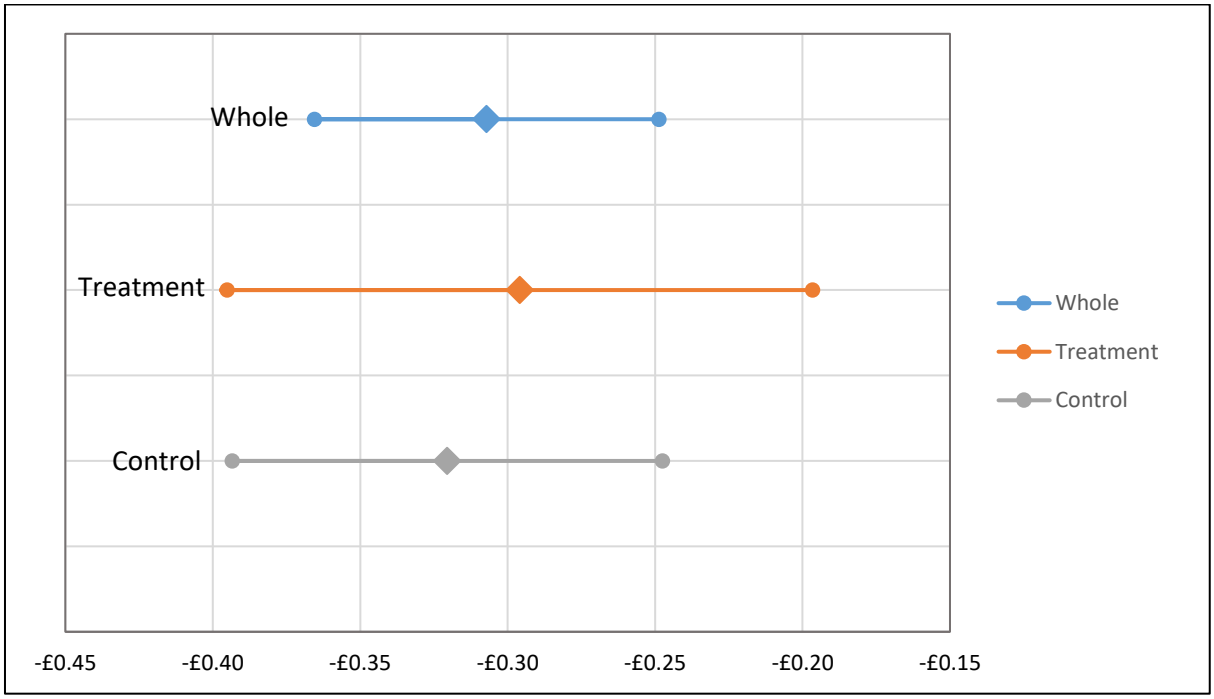


Table 7.8 WTP Estimates from the DCE (All Trial Participants)

Attribute/Attribute Level	WTP Estimate	Confidence Interval [Lower, Upper]
Feedback: Phone	£2.16	[-£0.08, £4.40]
Feedback: Text	£4.34	[£2.15, £6.53]
Reminders: Text	£2.87	[£0.80, £4.94]
Reminders: Phone	£3.79	[£1.34, £6.25]
Reminders: Online	£6.10	[£3.51, £8.69]
Online: Website & App	£2.34	[£0.08, £4.60]
Weight Re-Gain (%)	-£0.31	[-£0.37, -£0.25]

7.4.4 Discussion (DCE)

There is no evidence to suggest that WTP estimates differ substantially between trial arms, which is not surprising given the results of models two and three in table 7.2. However, a substantial number of WTP estimates can be estimated based on model one from table 7.2 for all respondents. It appears that respondents are willing to pay significant amounts per month to receive feedback via phone call or via text message. Respondents were also willing to pay substantial amounts per month to receive reminders (in particular via the online tool(s)), as well as to have access to a combination of both a website and an app. Of all of the WTP estimates, the one with the narrowest confidence interval is unsurprisingly the WTP estimate for weight re-gain (given the statistical significance of the corresponding coefficients in table 7.2).

7.4.5 Conclusion

With respect to RQ2a, there appears to be little evidence to suggest that WTP estimates differ between respondents in each arm of the NULevel trial. The inability to identify a difference may be due to low sample sizes in both tasks, given the complexity of the choice models and the lack of observations in the CV task. Nonetheless, the preferred specification of the choice model provides evidence that respondents are willing to pay for various components of WLM programmes and the CV results suggest that a large proportion of respondents are willing to pay for an efficacious programme.

7.5 Research Question Two Part B

RQ2b: How might WTP estimates differ if they are elicited indirectly or directly?

7.5.1 Results

Comparisons between the WTP estimates from the CV task (direct) and the DCE (indirect) could be made in a number of different ways. One approach would be to calculate an average WTP estimate from the CV questions for the most preferred programme (as determined by the ranking questions at the start of the survey) and compare this with an estimate from the DCE results based on the same programme; this is similar to what has been done elsewhere (van der Pol et al., 2008). However, given the statistical insignificance of numerous attribute levels in the DCE analysis this approach is unlikely to provide a convincing comparison.

An alternative approach, given the results of the two individual tasks, is to use the iterative nature of the CV questions to identify a WTP estimate for a percentage point change in weight re-gain. As these have already been identified in the DCE (WTP_{DCE} hereafter), the estimates from the CV task can be directly compared (WTP_{CV} hereafter). To generate the WTP_{CV} estimates, a variable was generated whereby the WTP value at 10% weight re-gain

was subtracted from the WTP value at 0% weight re-gain and divided by 10. Table 7.9 contains the mean WTP_{CV} as well as the WTP_{DCE} estimate.

Table 7.9 DCE and CV Comparison in WTP for 1% Increase in Weight Re-Gain (Trial Sample)

Group	WTP_{DCE}	WTP_{CV}	Difference	p-value
All Respondents	-£0.31	-£0.62	£0.31	0.4967

The average WTP_{CV} values are not statistically significantly different across the trial arms (p-values>0.1; not reported). Additionally, as indicated by the p-value in table 7.9, there does not appear to be a statistically significant difference between the WTP_{DCE} and the average WTP_{CV} value for all respondents. Due to the exceptionally large WTP estimates from the CV task, there is a high variance associated with the WTP_{CV} value. This, coupled with the small sample sizes, has resulted in the large p-value.

7.5.2 Discussion

There appears to be no statistically significant difference between the average direct and indirect WTP estimates for the trial participants. As alluded to in the previous section, the WTP_{CV} estimate appears to be inflated due to some large values provided in task at the 0% outcome level. Even after removing any WTP values that exceed £100 per month, which decreases the WTP_{CV} estimate to -£0.17, there is no evidence of a difference between the WTP_{DCE} and WTP_{CV} estimates. Previous studies have provided mixed evidence as to whether WTP estimates differ between these two methodologies, with at least one study indicating no difference between estimates (Ryan, 2004a). However, the comparison being made here is perhaps atypical given the isolation of a single component of the WLM programme that was valued in the CV task.

7.5.3 Conclusion

On the whole there is no evidence to suggest that the WTP estimates from the two preference elicitation tasks differ. However, it could be the case that there is not enough data to identify a difference in this case. Regardless, the approach taken here is somewhat atypical and it may be unwise to compare these results with those from the existing literature.

7.6 Research Question Three

RQ3: To what extent do certain groups have better defined preferences than other user groups, and why might these differences occur?

7.6.1 Results

Two rationality tests were embedded within the DCE in order examine how well-defined respondents' preferences are, as described in Chapter Six. Table 7.10 presents the pass rates for the transitivity test (using two definitions), the dominance test and a combination of the three for all respondents as well as split across trial arms.

Table 7.10 Embedded DCE Rationality Test Results (Trial Sample)

Group	Dominance	Weak Transitivity	Strong Transitivity	All 3
All	110/113 (97%)	111/113 (98%)	78/113 (69%)	75/113 (66%)
Treatment	51/52 (98%)	52/52 (100%)	39/52 (75%)	38/52 (73%)
Control	59/61 (97%)	59/61 (97%)	39/61 (64%)	37/61 (61%)

The results in table 7.10 are extremely positive; almost all respondents passed the dominance test and the weak transitivity test. Furthermore, the pass rate for the strong transitivity test is particularly high at 69% of respondents, given that the probability of passing the test at random is 11%. The pass rates for all three tests combined is largely reflective of the strong transitivity test pass rates due to the exceptionally high pass rates for the other two tests. When comparing the two trial arms, the differences in proportions of respondents passing the tests are not statistically significant at even the 10% significance level.

As mentioned in section 6.4.3, logit models could be estimated in order to formally examine the factors that may influence the probability of an individual passing one or all of the rationality tests. However, the lack of variation in the dependent variables (dummy variables indicating whether the test was passed) caused by the high pass rates for the weak transitivity and the dominance tests meant that these models would not run due to collinearity. Logit models for the strong transitivity test and the combination of all three did not produce logical, or statistically significant results either. This is revisited in section 9.5 where data from the trial and the online panel are pooled.

7.6.2 Discussion

The rationality test results suggest that respondents from the trial participant sample largely behave in a 'rational' manner. Only three respondents in the sample failed the dominance test and only two failed the weak transitivity test. The latter is particularly interesting. As explained in Chapter Six, passing the weak transitivity test means that the respondent has not, at any stage within the three test scenarios, contradicted their rank-order of the three WLM programmes. This is a necessary but insufficient condition for passing the strong transitivity test. A respondent that passes the weak transitivity test but fails the strong transitivity test does so because their rank-order is contradicted in relation to their rank-ordering of the opt-out option (no programme). Even if the rank-ordering of the opt-out option is considered important, the majority of respondents also passed the strong transitivity test too, suggesting that respondents in this sample were highly consistent in their responses. Whether this constitutes a sample of individuals with 'well-defined' preferences is up for debate, however it certainly appears that respondents in this sample rarely contradicted themselves when expressing their preferences in the DCE task.

One might speculate that the impressive pass rates are due to the high education levels of the respondents. Ultimately it is not possible to determine whether this is the case using this sample alone; this hypothesis is revisited in Chapter Nine.

7.6.3 Conclusion

The results of the rationality tests are particularly strong for the trial participant sample. The difference in pass rates between the two arms of the trial are minimal and not statistically significant. The factors that influence pass rates could not be explored with this small, and relatively homogenous, sample.

7.7 Overall Conclusion

On the whole, the DCE results suggest that respondents from the NULevel trial have relatively strong preferences for a wide range of characteristics of WLM programmes. These preferences go beyond the basic characteristics such as cost and the health outcome and include characteristics that relate to the process of care. The DCE results also allow the calculation of several WTP estimates, given that many attributes and attribute levels are statistically significant. The vast majority of these WTP estimates were sizeable but confidence intervals were very wide. In addition, the CV results suggest that the majority of respondents from the NULevel trial would be willing to pay for a WLM programme providing that it is efficacious. The average WTP values at 0% and 10% weight re-gain are substantial, suggesting that respondents value such programmes if they are clinically

effective. Finally, it also appears that the respondents have relatively well-defined preferences in that the majority passed the embedded rationality tests in the DCE. Throughout, there was little evidence to suggest that preferences differed between participants in the two arms of the trial, although sample sizes were relatively low for both arms.

Chapter 8. Results from the Online Panel Sample

The aim of this chapter is to present and summarise the results from the online panel sample in relation to the three research questions presented in Chapter Five. The first research question relates to whether preferences for a health service differ significantly between different user groups, as shown by the results of a DCE. The second, split into two parts, first examines whether WTP estimates are different between user groups and secondly whether estimates differ between methodologies (DCE and CV). The final research question examines whether certain user groups have better defined preferences using the results of embedded rationality tests in the survey.

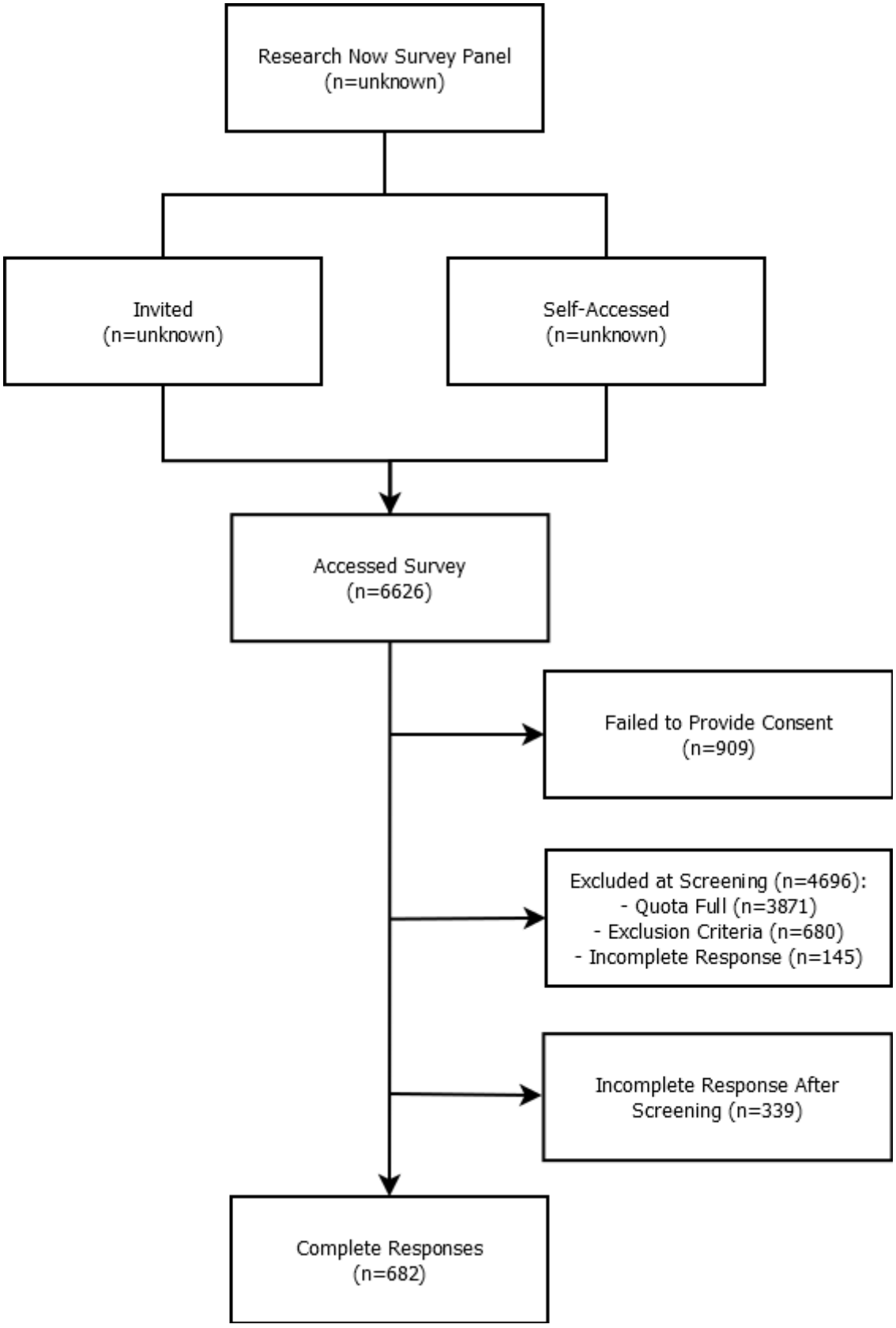
The chapter begins with a detailed description of the recruitment of the online panel (section 8.1) and the characteristics of each user group (section 8.2). Each research question is then addressed in turn with both a description of the results as well as discussion (sections 8.3-8.6). Finally, a summary of the results is provided to end the chapter (section 8.7).

8.1 Response Rates

Recruitment for the online panel sample began on 10th February and concluded on 18th February 2016. A total of 682 individuals completed the online survey; a flowchart of the recruitment process described below can be found in figure 8.1. The survey was hosted by a market research company (Research Now; http://www.researchnow.com), who used their extensive panel of respondents to generate the sample. Some individuals were directly invited to participate in the survey, whereas others were not invited but could access the survey through hyperlinks when logged into the website using their personal user accounts. Due to this, it is not possible to generate a true response rate for the online panel sample as the number of individuals that were invited will be exceeded by the number of individuals that actually accessed the survey.

However, some information is known about the individuals that accessed the survey and failed to complete it. A total of 909 individuals accessed the survey but made it no further than the information pages (i.e. they exited before the consent page). This may have been due to concerns about the content or the time commitment of the survey.

Figure 8.1 Flowchart of the Online Panel Recruitment Process



An additional 4,696 individuals provided consent but did not make it any further than the screening questions. The main purpose of the screening questions was to classify the respondents into either: the 'potential service user' (PSU) group (those that met the NULevel

trial criteria); the 'potential beneficiary' (PB) group (those that did not meet the criteria but are obese or overweight); or the 'non-user' (NU) group (those with a normal BMI). Of these 4,696 individuals, 3,871 could not continue beyond the screening questions due to full quotas. This means that, for example, if an individual that was eligible for the PB group filled out the screening questions once that particular group had received enough responses, they would be unable to progress any further. Contact with Research Now during recruitment confirmed that the PSU group was the reason for the high number of rejected respondents. This was expected due to the challenging inclusion criteria described in Chapter Six. A further 680 individuals were excluded from the survey based on their responses to the screening questions (irrespective of the user group quotas). These individuals may have been underweight or pregnant, for example. Finally, 145 respondents failed to complete the screening questions once they had begun the survey. It may have been the case that these individuals would have been excluded, or removed due to full quotas, had they completed the screening questions.

In total, 339 respondents successfully made it past the screening questions but failed to go on to complete the survey. Of these individuals, only 79 made it to the DCE task (note that only the screening and attribute/attribute level ranking questions preceded the DCE task). As a result, it would seem fair to conclude that relatively little survey drop-out occurred during the critical components of the survey (i.e. the DCE or the CV task).

8.2 Respondent Characteristics

Due to the quotas in place that required at least 210 respondents for each user group, the 682 respondents that completed the survey were spread relatively evenly across the three user groups: 210 potential service users; 243 potential beneficiaries and; 229 non-users. Table 8.1 summarises the characteristics of respondents from the whole sample and across each of the three user groups. On the whole, the respondents are closer to retirement age than working age, well-educated and slightly more likely to be female compared to male. With regards to employment, those that are employed (full-time or part-time) or retired make up the vast majority of the sample. General health is good on average, as determined by the average visual analogue scale (VAS) scores and EQ-5D-5L utility values.

As a result of the recruitment process, there are significant differences between the three user groups with regards to weight-related characteristics. All respondents in the PSU group are obese, 63% of respondents in the PB group are overweight and 37% obese, while all respondents in the NU group have a normal BMI. All respondents in the PSU group had attempted to lose weight in the previous 12 months, compared with 68% and 55% of the PB and NU groups, respectively.

The fact that over half of the NU group attempted to lose weight in the last 12 months could be considered surprising given that these individuals have a normal BMI classification. It could be the case that some of these respondents were previously overweight, however it is not possible to determine whether this was the case. Regardless, this is arguably positive considering the topic of the survey as it is likely to aid respondent understanding of WLM and potentially reduce hypothetical bias. The general health measures suggest that there are differences between the user groups with regards to health that go beyond the BMI classification differences. Both the VAS and EQ-5D-5L average values are significantly different across the three user groups ($p<0.01$) according to one-way ANOVA. It appears that the NU group has significantly better health according to these measures relative to the other two groups.

Other differences exist between the user groups that were not expected. Average age, the gender-split, and employment statuses across the three groups are significantly different from one another according to chi2 tests ($p<0.01$). It appears that this can be explained by the PB group in particular. The PB group contains older respondents on average, more male respondents than any other group, as well as more retired respondents. This may have had an effect on the education variable, given that younger individuals are more likely to have attended university (Blanden & Machin, 2004), although the evidence to suggest that highest educational qualifications differ significantly across the three groups is not particularly strong ($p=0.05$).

It is not ideal that the characteristics of the groups differ (beyond differences in weight-related characteristics that occurred due to the user group definitions) because results will be compared between the user groups. If these characteristics were similar between the groups it would be easier to attribute any differences in preferences to the user group classifications. However, it should be noted that household income does not appear to differ across each of the groups ($p=0.55$). This is important because ability to pay has been shown to affect WTP (Klose, 1999), and this means that there is no significant evidence to suggest that ability to pay will differ substantially between respondents in each of the three user groups.

Table 8.1 Respondent Characteristics (Online Panel Sample)

	Potential Service User (PSU)	Potential Beneficiary (PB)	Non-User (NU)	*p-value*[1]
n	210	243	229	
Age (mean)	44	50	45	*0.000*
18-24	9 (4%)	7 (3%)	14 (6%)	
25-34	48 (23%)	41 (17%)	54 (24%)	
35-44	55 (26%)	38 (16%)	54 (24%)	
45-54	50 (24%)	52 (21%)	35 (15%)	
55+	48 (23%)	105 (43%)	72 (31%)	
Gender				*0.000*
Male	87 (41%)	139 (57%)	89 (39%)	
Female	123 (59%)	104 (43%)	140 (61%)	
BMI Classification				*0.000*
Normal	0 (0%)	0 (0%)	229 (100%)	
Overweight	0 (0%)	153 (63%)	0 (0%)	
Obese	210 (100%)	90 (37%)	0 (0%)	
Attempted to lose weight in the 12 months prior to the survey?				*0.000*
Yes	210 (100%)	166 (68%)	125 (55%)	
No	0 (0%)	77 (32%)	104 (45%)	
Highest Level Qualification				*0.054*
Higher (University)	92 (44%)	97 (40%)	109 (48%)	
Further (College)	52 (25%)	78 (32%)	67 (29%)	
Secondary (School)	60 (29%)	56 (23%)	46 (20%)	
None	6 (3%)	8 (3%)	4 (2%)	
Other	0 (0%)	4 (2%)	3 (1%)	
Employment Status				*0.004*
Employed (full-time)	107 (51%)	113 (47%)	94 (41%)	
Employed (part-time)	39 (19%)	41 (17%)	57 (25%)	
Student	2 (1%)	3 (1%)	8 (3%)	
Retired	25 (12%)	60 (25%)	37 (16%)	
Homemaker/Caregiver	19 (9%)	13 (5%)	15 (7%)	
Unemployed	16 (8%)	12 (5%)	10 (4%)	
Other	2 (1%)	1 (<1%)	8 (3%)	
Household Income				*0.545*
£20,000 or less	54 (26%)	68 (28%)	48 (21%)	
£20,001 - £40,000	80 (38%)	72 (30%)	84 (37%)	
£40,001 - £60,000	41 (20%)	58 (24%)	57 (25%)	
£60,001 and above	35 (17%)	45 (19%)	40 (17%)	
Health (Visual Analogue Scale; 0=Dead, 100=Perfect Health)				*0.001*
Mean	73.5	74.9	80.2	
(Standard Deviation)	(20.09)	(18.39)	(17.62)	
Health (EQ-5D-5L Utilities[2]; 0=Dead, 1=Perfect Health)				*0.001*
Mean	0.870	0.881	0.927	
(Standard Deviation)	(0.187)	(0.188)	(0.116)	

[1]p-values based on chi2 tests (categorical) and one-way ANOVA (continuous);

[2]English population utilities from Devlin et al. (2017)

8.3 Research Question One

RQ1: To what extent do preferences for a health service, elicited via a discrete choice experiment, differ between different user groups, and why might these differences occur?

8.3.1 Results

Table 8.2 presents the results using mixed logit models with all parameters set as random (and normally distributed) using 1,000 Halton draws. Hence, the coefficients are the mean values for each parameter. Models were also estimated with a full covariance matrix in order to try and address (within-sample) scale heterogeneity; however, the results were deemed unreliable. The results in table 8.2 are generally consistent with results from the conditional logit models that were also estimated, although slightly more coefficients are statistically significant in the conditional logit models. The conditional logit models can be found in Appendix D-1.

The results in table 8.2 suggest that respondents in every user group prefer shorter programmes that are more efficacious and less costly. Across each of the three models the coefficients for these variables are statistically significant at the 1% level and have identical signs throughout. The finding that all respondents prefer to spend less and prefer better health outcomes is important with respect to the face validity of the choice models.

In contrast to the consistent statistical significance of the other attributes, the coefficients for the levels of the feedback, reminders and online tool(s) attributes vary in their significance across the three models. Model one suggests that respondents in the PSU group are indifferent to the majority of the levels associated with these three attributes. There is weak evidence to suggest that respondents in this group have a preference for receiving an app (0.242, $p<0.1$) as well as both an app and website (0.240, $p<0.1$). Model two suggests that respondents in the PB group prefer to receive feedback via text message relative to face to face (0.337, $p<0.01$), and there is weak evidence to suggest that feedback via the online tool(s) is also preferred to face to face feedback (0.219, $p<0.1$). This model also suggests that respondents prefer to receive reminders via the online tool(s) relative to receiving no reminders (0.343, $p<0.01$) and there is weak evidence of a preference for reminders via text message (0.216, $p<0.1$). Model three suggests that respondents in the NU group prefer feedback via both online tool(s) and via text message relative to face to face feedback (0.566, $p<0.01$ and 0.615, $p<0.01$, respectively). Furthermore, this model also suggests that the respondents in this group prefer to receive reminders via text messages and via the online tool(s) relative to not receiving any (0.425, $p<0.01$ and 0.352, $p<0.05$, respectively).

Table 8.2 Results from the Mixed Logit Models (Online Panel Sample)

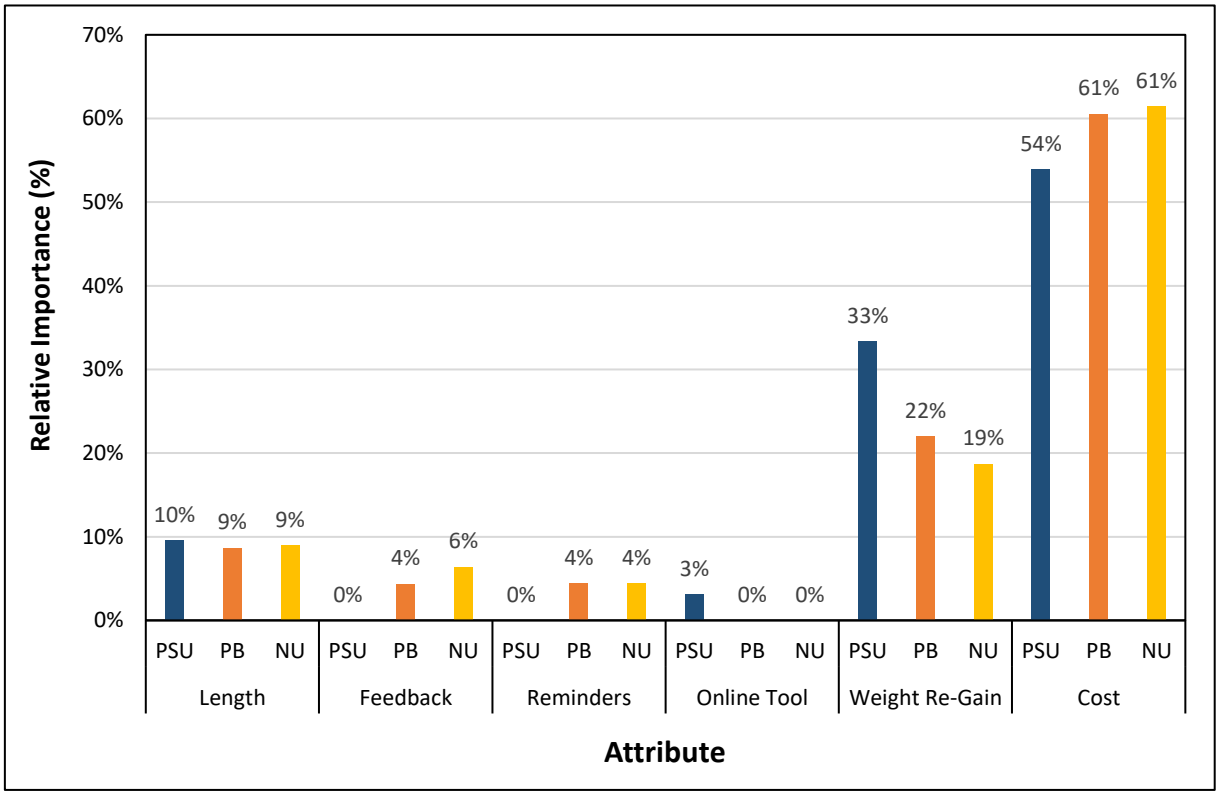
	1. Potential Service User		2. Potential Beneficiary		3. Non-User	
	Coefficient	*Std. Dev.*	*Coefficient*	*Std. Dev.*	*Coefficient*	*Std. Dev.*
Alt-Specific Constant	2.380 ***	2.357 ***	3.012 ***	2.675 ***	2.897 ***	3.569 ***
	(0.301)	(0.256)	(0.315)	(0.294)	(0.379)	(0.527)
Length (months)	-0.031 ***	0.029 *	-0.028 ***	0.040 ***	-0.036 ***	-0.070 ***
	(0.008)	(0.015)	(0.008)	(0.015)	(0.009)	(0.013)
Delivery of Feedback[1]						
via Phone Call	0.179	0.422	0.218	0.914 ***	0.129	-0.089
	(0.144)	(0.343)	(0.146)	(0.257)	(0.149)	(0.315)
via the Online Tool(s)	0.199	0.670 ***	0.219 *	-0.073	0.566 ***	0.088
	(0.148)	(0.256)	(0.130)	(0.526)	(0.150)	(0.329)
via the Text Message	0.181	-0.148	0.337 ***	-0.138	0.615 ***	-0.355
	(0.131)	(0.290)	(0.123)	(0.316)	(0.142)	(0.246)
Reminders[2]						
via Text Message	0.144	-0.255	0.216 *	0.701 ***	0.425 ***	0.352
	(0.124)	(0.376)	(0.125)	(0.192)	(0.135)	(0.264)
via Phone Call	-0.040	-0.617 ***	0.142	0.279	0.103	-0.231
	(0.150)	(0.223)	(0.134)	(0.301)	(0.156)	(0.476)
via the Online Tool(s)	0.020	-0.238	0.343 ***	0.226	0.352 **	-0.382 *
	(0.143)	(0.336)	(0.132)	(0.303)	(0.153)	(0.231)
Online Tool[3]						
App Only	0.242 *	-0.003	0.129	0.102	0.224	0.138
	(0.134)	(0.253)	(0.127)	(0.234)	(0.140)	(0.294)
Website Only	0.211	0.446	-0.054	0.499 *	0.197	-0.831 ***
	(0.146)	(0.289)	(0.143)	(0.269)	(0.159)	(0.223)
App & Website	0.240 *	0.144	0.137	0.575 ***	-0.044	-0.623 ***
	(0.130)	(0.359)	(0.127)	(0.221)	(0.140)	(0.234)
Weight Re-gain (%)	-0.026 ***	0.029 ***	-0.017 ***	0.023 ***	-0.018 ***	0.026 ***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.004)
Personal Cost (£/month)	-0.140 ***	0.138 ***	-0.156 ***	0.128 ***	-0.197 ***	0.158 ***
	(0.014)	(0.013)	(0.013)	(0.012)	(0.018)	(0.016)
Observations (Sample Size)	6,300		7,290		6,870	
Log Likelihood	-1,505		-1,752		-1,587	
AIC	3,036		3,530		3,200	
BIC	3,124		3,620		3,289	

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

[1] Base level "face to face" feedback; [2] Base level "no reminders"; [3] Base level "no online tool"

As outlined in Chapter Six, additional analysis must be conducted in order to compare the results across the three models. Figure 8.2 illustrates the relative importance of each attribute for each user group, based on the method outlined in section 6.4.7 of Chapter Six. The ranges for each attribute were calculated using the results from table 8.2 i.e. the main effects mixed logit models. When calculating the ranges for each attribute, coefficients were treated as equal to zero if they were not statistically significant at the 10% level. Hence, for some attributes relative importance is equal to zero.

Figure 8.2 Relative Importance of the Attributes (Online Panel Sample)



It is clear from figure 8.2 that the outcome and cost attributes were highly dominant regardless of the user group. The relative importance of the outcome attribute is greater for the PSU group (33%) relative to the PB group (22%) and greater for the PB group relative to the NU group (19%). The difference in the relative importance of the outcome attribute for the PSU group compared to the NU group is particularly large (14 percentage points). The opposite trend is apparent for the cost attribute, although the relative importance of the cost attribute for the PB and NU groups is identical (61%).

In addition to relative importance calculations, an additional area of analysis that can provide insight relates to the opt-out option. Four participants in the PSU group chose the opt-out option (no programme) in every scenario relative to three in the PB group and nine in the NU group. Furthermore, the average number of opt-outs per respondent were 2.2 for the PSU and PB groups compared with 2.9 for the NU group.

Additional choice models were also estimated in order to test the significance of attribute interactions and to identify whether differences in preferences may be observed based on observable characteristics of the respondents (i.e. those that were not varied in the user group definitions).

Results from conditional logit models that contained three interaction terms based on *a priori* expectations (length & cost; outcome & cost; length & outcome) can be found in Appendix

D-2. These suggest that there is weak evidence that the interaction between length and outcome is significant for the PSU group (-0.0002, $p<0.1$) and moderate evidence that the interaction between length and cost is significant for the PB group (-0.0009, $p<0.05$). No other interactions were found to be statistically significant. Likelihood ratio tests between the conditional logit models with and without interactions provided weak evidence ($p<0.1$) that the models with interactions provided a better model fit for the PSU and PB groups but no evidence for the NU group. Given this weak evidence overall of the importance of the interaction terms, it was not felt that it would be beneficial to incorporate them into the more flexible mixed logit models as this would reduce their likelihood of converging, with little expected benefit to model fit.

The respondent characteristics in table 8.1 indicate substantial differences between the three user groups. Therefore, additional models were estimated containing interactions between the alternative-specific constant and various demographic variables. The variables considered related to age groups, gender, highest educational qualification, employment status and income band. Appendix D-3 provides the results from mixed logit models that included these interactions terms, which were not modelled as random parameters. Across the three models, very few interactions were statistically significant. In addition, the effect on the coefficients of the attribute/attribute level variables was minimal and WTP estimates (not reported) did not differ substantially from those generated using the main effects models (see 8.4.3). Likelihood ratio tests confirmed that the addition of the demographic interactions did not lead to a statistically significant improvement in log-likelihood and therefore model fit. As a result, the main effects models presented in table 8.2 remain as the preferred specification.

8.3.2 Discussion

The differences between models one to three in table 8.2, and illustrated in figure 8.2, would appear to suggest that individuals in the PSU group are more sensitive to negative outcomes (higher weight re-gain) relative to the other two groups, whereas individuals in the NU group are more price sensitive relative to the other two groups. These results are intuitive as it would be expected that respondents in the PSU group care more about the outcome of the service relative to others, and that they would be less price sensitive relative to the NU group, as respondents within the PSU group are more likely to be interested in using the service. On the other hand, as the outcome attribute was also presented as an absolute value, it could be the case that the driver of the coefficient differences was the fact that the absolute differences between attribute levels were larger for those in the PSU group relative to others (due to the higher average weight).

It is perhaps surprising that the attribute interactions were not statistically significant across the models estimated for each user group (Appendix D-2). It would seem logical to consider: the total cost of the programme (cost & length); the rate of weight re-gain (outcome & length); or even the cost per weight re-gain avoided (cost & outcome). However, DCE tasks are cognitively demanding and it may simply be too much for the average respondent to consistently assess these combinations of attributes, alongside all the other attributes.

There is no evidence to suggest that demographic characteristics have influenced the results, despite the differences in observable characteristics between the three user groups. It can therefore be concluded with relative confidence that the differences in preferences observed is likely to be due to the way in which the groups were defined i.e. their relative experience of the health issue and current weight status.

8.3.3 Conclusion

The results suggest that preferences for WLM programmes do differ across the three user groups, however the differences are somewhat limited. For example, there are no instances where a statistically significant coefficient is positive in one model but negative in another. Nonetheless, there are clear differences in the significance of attributes and attribute levels, as well as in the relative importance of the attributes.

8.4 Research Question Two Part A

RQ2a: To what extent do WTP estimates differ between different user groups, and why might these differences occur?

8.4.1 Results (CV)

Table 8.3 contains the results for each group to the self-selection question; the question that determined whether respondents would be faced with a CV exercise or not. Chi2 tests provide convincing evidence ($p<0.01$) that the responses to this question significantly vary across the three user groups, even if similar responses are collapsed (e.g. one and two; three and four). Combining responses one and two, it is clear that a higher proportion of the PSU group wish to use a WLM programme like the ones described in the DCE task relative to the other two groups (the specific order is PSU>PB>NU). The proportion of respondents stating that they *might* want to use a WLM programme (responses three and four) is similar between the PSU and PB groups, with a lower proportion in the NU group. Combining all four responses shows that a significantly different proportion of each user group went on to face the use value CV exercise; 84% of the PSU group, 74% of the PB group and 54% of the NU group.

Table 8.3 Responses to the CV Self-Selection Question (Online Panel Sample)

Response to the "Interest Question"	**PSU**		**PB**		**NU**		**Total**	
	n	**%**	**n**	**%**	**n**	**%**	**n**	**%**
1. I would like to use x	37	18%	26	11%	24	10%	**87**	**13%**
2. I would like to use x & want it to be available for others	50	24%	45	19%	20	9%	**115**	**17%**
3. I might want to use x	44	21%	52	21%	37	16%	**133**	**20%**
4. I might want to use x & want it to be available for others	45	21%	57	23%	44	19%	**146**	**21%**
5. I would not use x but I want it to be available for others	21	10%	34	14%	77	34%	**132**	**19%**
6. I do not think x should be made available	13	6%	29	12%	27	12%	**69**	**10%**
Total	**210**	**100%**	**243**	**100%**	**229**	**100%**	**682**	**100%**

x = a programme like this (referring to the hypothetical programmes from the DCE task)

Responses 1-4 – forwarded to the "use value" WTP exercise;

Response 5 – forwarded to the "externality" WTP exercise;

Response 6 – forwarded to open-ended comment box.

As outlined in Chapter Six, the use value CV questions included in the survey provide WTP estimates for each individual respondent's preferred set-up of a WLM intervention, excluding the length (as this was held static at 12 months) and weight re-gain (as this changed in each question). The questions began by asking respondents if they would be willing to pay for a programme with a weight re-gain outcome of 0% and subsequently asked for a maximum price that respondents would be willing to pay. The outcome was then adjusted to the next level (i.e. 10%) and the task was repeated. This happened until the respondent stated that they were not willing to pay for the specified programme.

Table 8.4 summarises the most common set-up of the WLM programme valued in the use value task, which did not differ across the three user groups. However, whilst this was the most common programme, very few participants overall actually valued this particular set-up. In total, participants in the PSU group valued 31 different variants relative to 41 in the PB group and 34 in the NU group. Thus, there was a wide range of variety across the respondents.

Table 8.4 Most Commonly Valued Programmes in the Use Value CV Task (Online Panel Sample)

Attribute	**Potential Service User**	**Potential Beneficiary**	**Non-User**
Delivery of Feedback	via Text Message	via Text Message	via Text Message
Delivery of Reminders	via Text Message	via Text Message	via Text Message
Availability of Online Tool(s)	Website & App	Website & App	Website & App
	N=40/176 (23%)	N=28/180 (16%)	n=21/125 (17%)

Table 8.5 presents the full results from the use value CV exercise. In total, 31 respondents were removed from the analysis for indicating that they would be willing to pay more at a higher level of weight re-gain (the 'logic rule'). It was assumed that such instances represented erroneous responses or a failure of the respondent to fully understand the task. Note that these 31 respondents are still incorporated into the 'stated' column because they indicated a willingness to pay regardless. The number of individuals that provided a WTP value sharply decreased as weight re-gain increased and the majority of respondents were not willing to pay beyond 40% weight re-gain. The proportion of individuals within each group that provided a WTP value when weight re-gain was set at 0% follows the expected pattern (PSU > PB > NU), but the differences in proportions diminishes as the level of weight re-gain increases.

Table 8.5 WTP Estimates from the Use Value CV Task (Online Panel Sample)

Potential Service User (PSU)								
Weight Re-Gain	**n**	**Stated[1] n**	**%**	**Provided[2] n**	**%**	**Willingness to Pay (WTP) Median**	**Mean**	**SD**
0%	176	118	67%	93	53%	£10.00	£14.28	13.50
10%	118	91	52%	60	34%	£10.00	£11.93	8.52
20%	91	59	34%	29	16%	£10.00	£11.90	6.20
40%	59	32	18%	11	6%	£5.00	£9.00	6.32
60%	32	17	10%	4	2%	£12.50	£11.00	5.23
80%	17	12	7%	2	1%	£10.00	£10.00	0.00
90%	12	11	6%	2	1%	£10.00	£10.00	0.00
100%	11	11	6%	2	1%	£10.00	£10.00	0.00

	Potential Beneficiary (PB)								Non-User (NU)							
Weight Re-Gain	**n**	**Stated[1] n**	**%**	**Provided[2] n**	**%**	**Willingness to Pay (WTP) Median**	**Mean**	**SD**	**n**	**Stated[1] n**	**%**	**Provided[2] n**	**%**	**Willingness to Pay (WTP) Median**	**Mean**	**SD**
0%	180	115	64%	81	45%	£10.00	£13.84	15.89	125	76	61%	58	46%	£10.00	£13.91	16.66
10%	115	95	53%	62	34%	£10.00	£11.97	10.58	76	62	50%	45	36%	£10.00	£13.13	18.64
20%	95	68	38%	46	26%	£10.00	£10.48	8.39	62	47	38%	30	24%	£10.00	£12.57	20.97
40%	68	32	18%	20	11%	£8.50	£8.30	4.05	47	22	18%	11	9%	£5.00	£16.27	31.49
60%	32	16	9%	9	5%	£10.00	£10.33	4.97	22	18	14%	7	6%	£10.00	£20.43	35.43
80%	16	13	7%	6	3%	£10.00	£11.00	6.32	18	13	10%	4	3%	£12.50	£32.50	41.73
90%	13	12	7%	4	2%	£12.50	£13.75	4.79	13	9	7%	3	2%	£10.00	£11.67	2.89
100%	12	7	4%	3	2%	£15.00	£15.00	5.00	9	7	6%	2	2%	£12.50	£12.50	3.54

Percentages are based on the number of individuals that were forwarded to the exercise from each user group in the first instance (underlined).

[1]Those that selected "yes" when asked whether they would be willing to pay (those that select "no" exit the exercise at that stage).

[2]Those that provided a value when asked for the maximum amount that they would be willing to pay *and* did not increase their value at any point during the task ('logic rule').

This effect happens almost instantly; when weight re-gain increases from 0% to 10%, 33 (35%) individuals in the PSU group were no longer willing to pay compared to 19 (24%) and 13 (22%) of the PB and NU groups, respectively.

In the PSU and PB groups average WTP increases after 40% weight re-gain (after 20% for the NU group). A closer look at the data shows that while most individuals decrease their values as weight re-gain increases, there is a small percentage of individuals that state a constant WTP value throughout the task. This latter group of individuals are far more likely to remain in the task until the later levels of weight re-gain, making them a bigger proportion of the sample at these stages, increasing the average WTP values. One-way ANOVA suggests that the mean values at each weight re-gain level are not statistically significantly different across the three user groups. The similarity between the values provided by each group are clearest when looking at the median values, which are identical for every group for the first three weight re-gain levels (£10 at 0%, 10% and 20%).

8.4.2 Discussion (CV)

The findings from the CV task are generally sensible in that fewer respondents were willing to pay for a WLM programme as the efficacy decreased. Additionally, it is logical that fewer respondents in the NU group would be willing to pay for a WLM programme given that they seemingly have less need for such a programme.

Perhaps the biggest concern with the CV results is the increasing average WTP values as weight re-gain increases in table 8.5. As explained, this is largely due to individuals stating that they would be willing to pay the same amount regardless of the outcome. While this may seem like irrational behaviour, it may be unjustified to remove these individuals from the analysis for this reason alone. Constant WTP values may be an expression of the individuals' strong preference for the service to be made available. Additionally, with an intervention such as this the health outcome is largely in control of the service user. It may be the case that some of these individuals consider the stated outcome level to be irrelevant for this reason and choose to ignore it. However, while this issue may appear significant, it is important to note that very few respondents remain in the analysis when the average WTP values start to increase (no more than 11 in each user group).

A particularly interesting finding from the CV task is that a greater proportion of individuals in the PSU group exit the task when weight re-gain increases to 10% from 0%. This is in line with the findings from the DCE results in section 8.3 that suggest that individuals in the PSU group are more negatively affected by an increase in weight re-gain than the other groups.

It was hypothesised that individuals in the PSU group would be willing to pay more than those in the PB group and that the latter would be willing to pay more than those in the NU group. It may be the case that individuals are providing values that represent the amount that they would *expect to pay* if the service were to be made available, rather than the maximum amount that they would be *willing to pay*. The former may be more consistent across the whole sample as many individuals are aware of the costs of lifestyle services such as gyms and weight loss programmes. Alternatively, it could be the case that the levels in the DCE task (provided before the CV task for all respondents) have framed the responses to some extent; the mean WTP at 0% weight re-gain was around £14, which is roughly in the middle of the range from the DCE (£0-30).

8.4.3 Results (DCE)

As explained in Chapter Six, WTP estimates can also be generated from the DCE by calculating the marginal rate of substitution (MRS) between an attribute or attribute level and the cost attribute. Only a small number of attributes and attribute levels were statistically significant in all three models in table 8.2: cost, length & outcome. As the idea of paying for a change in programme length makes little sense in this context, comparisons will be made across the three groups by looking at WTP estimates for a change in outcome.

Table 8.6 contains WTP estimates for a unit change in the outcome attribute, split by each group. These are calculated by taking the ratio of the coefficients for weight re-gain and cost, as explained in Chapter Six. The estimates represent willingness to pay for a single percentage point increase in weight re-gain. As this is an undesirable change, the WTP estimates are negative.

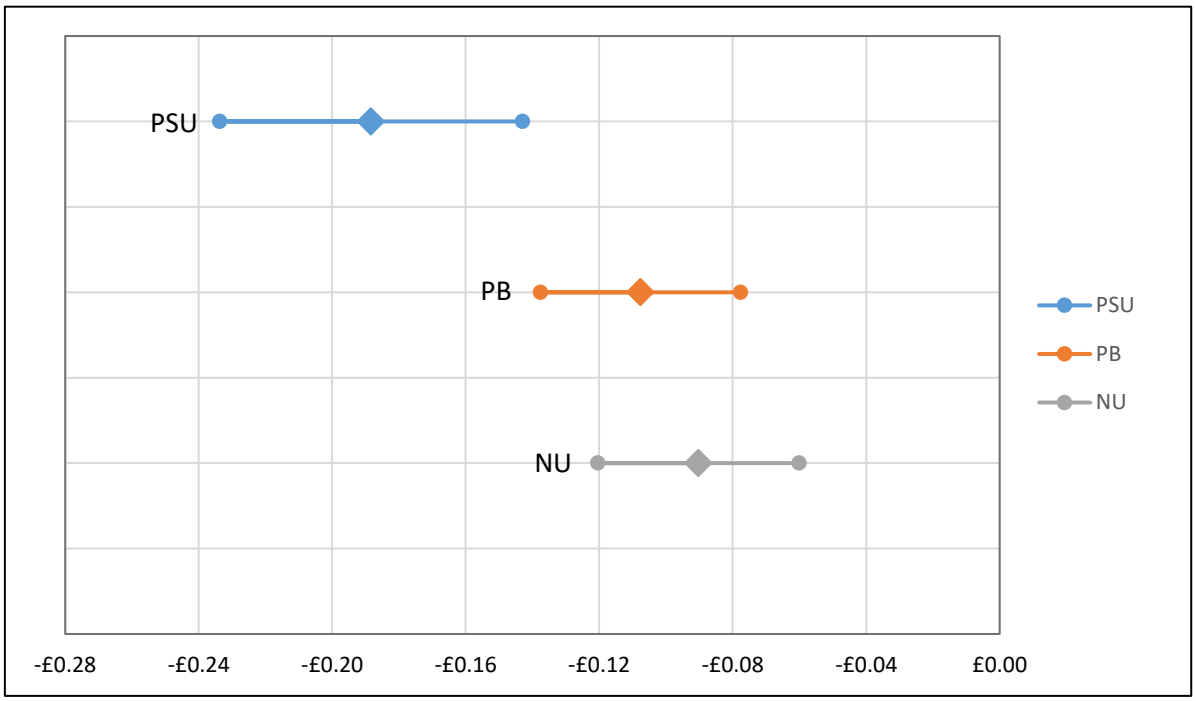
Table 8.6 WTP Estimates from the DCE (Online Panel Sample)

Group	**WTP for % weight re-gain**	**Confidence Interval [Lower, Upper]**
Potential Service User (PSU)	-£0.19	[-£0.23, -£0.14]
Potential Beneficiary (PB)	-£0.11	[-£0.14, -£0.08]
Non-User (NU)	-£0.09	[-£0.12, -£0.06]

It appears that the WTP estimate for the PSU group differs from the PB and NU groups, with the latter two groups providing relatively similar estimates. This can be seen in figure 8.3

where there is no overlap between in the confidence interval for the PSU group with the other two groups. The significant overlap between the confidence intervals for the estimates from the PB and NU groups suggest that the difference between the WTP estimates is unlikely be statistically significant.

Figure 8.3 WTP Estimates from the DCE for Weight Re-Gain (Online Panel Sample)



To understand the interpretation of these results, consider the following comparison between a programme with 40% weight re-gain and a programme with 20% weight re-gain. The PSU group would pay £3.80 more per month for the latter programme than the former, compared to £2.20 for the PB group and £1.80 for the NU group.

8.4.4 Discussion (DCE)

The differences in WTP estimates are not surprising given the DCE results illustrated in figure 8.2. From this figure, it was clear that as you move across the groups (i.e. from PSU to PB to NU) the relative importance of weight re-gain decreased and the relative importance of cost increased. This has created differences in WTP estimates, given that these are calculated by taking the ratio of the two coefficients. The results fit *a priori* expectations that more 'experienced' respondents would be willing to pay more for a service. However, as is the case for any DCE results of this nature, the WTP estimates should be taken with caution as they are derived in an indirect manner. That is, individuals are not stating their willingness to pay but implying it.

8.4.5 Conclusion

It appears that WTP estimates for the CV task do not differ significantly between the three user groups. However, the WTP estimates derived from the DCE task do appear to differ between PSU group and the other two groups. Hence, these results are somewhat inconclusive with regard to overall differences in willingness to pay.

8.5 Research Question Two Part B

RQ2b: How might WTP estimates differ if they are elicited indirectly or directly?

8.5.1 Results

The approach taken to compare WTP estimates from the DCE and CV task from the previous chapter will be used here. This involves using the iterative nature of the CV questions to identify a WTP estimate for a percentage point change in weight re-gain. As these have already been identified in the DCE (WTP_{DCE} hereafter), the estimates from the contingent valuation task can be directly compared (WTP_{CV} hereafter). To generate the WTP_{CV} estimates, a variable was generated whereby the WTP value at 10% weight re-gain was subtracted from the WTP value at 0% weight re-gain and divided by 10. Table 8.7 contains the mean WTP_{CV} as well as the WTP_{DCE} estimates from table 8.6, split by group.

Table 8.7 DCE and CV Comparison in WTP for 1% Increase in Weight Re-gain (Online Panel Sample)

Group	WTP_{DCE}	WTP_{CV}	Difference	p-value
PSU	-£0.19	-£0.34	£0.15	0.178
PB	-£0.11	-£0.22	£0.11	0.087
NU	-£0.09	-£0.19	£0.10	0.045

The average WTP_{CV} values are not statistically significantly different across the three user groups (p-values>0.1, not reported). Additionally, as indicated by the p-values in table 8.7, there does not appear to be a statistically significant difference between the WTP_{DCE} and the average WTP_{CV} value for the PSU groups. A closer look at the data suggests that this is due to the large variance in WTP_{CV} for this group that results from respondents that provided values suggesting that they would pay multiple pounds less per month for a single percentage point increase in weight re-gain. On the other hand, the difference between WTP_{DCE} and WTP_{CV} for the PB group is significant at the 10% level and the difference for the NU group is statistically significant at the 5% significance level. Whilst these groups also suffered from large values in

the CV task, it appears that there may be a significant difference between estimates generated from the two methodologies. In this case, the estimates from the CV task are approximately twice the size of those from the DCE task.

8.5.2 Discussion

One possible reason for the lack of a statistically significant difference between the two estimates for the PSU group, despite the seemingly large difference, is due to the range of values expressed within the CV task and the fact that a mean estimate was used. As indicated in the previous subsection, some of the differences between WTP at 0% and 10% in the CV task were infeasibly large and inflated the mean value. However, a median could not be used for the WTP_{CV} estimates instead because for every group the median value was zero, indicating that most individuals did not adjust their WTP estimate between 0% and 10% weight re-gain in the CV task. As a result it is hard to conclude whether there is a significant difference between WTP values when elicited directly or indirectly. This is clearly a weakness with the approach taken to generate WTP estimates for a percentage point increase in weight re-gain in the CV task, which has arisen due to the nature of the task itself.

There is weak evidence ($p<0.1$) to suggest that the estimates for the PB group are significantly different and moderate evidence ($p<0.05$) for the NU group, with the magnitudes of the WTP_{CV} estimates larger than the WTP_{DCE} estimates. This finding would be in contrast with the findings of other studies that make similar comparisons (van der Pol et al., 2008; Ryan & Watson, 2009; Danyliv et al., 2012). However, it could be the case that the comparison being made in this study is too different to compare with prior findings in the literature. This is because this comparison looks at WTP estimates for 'parts' of a WLM programme rather than comparing 'whole' WLM programmes. Hence, the comparison approach employed here may have reduced the potential differences in WTP estimates that previous studies identified, specifically those that could be attributed to 'part-whole bias' (Cookson, 2003).

When comparing two methodologies it is also important to consider framing effects. While the CV and DCE task in this survey were similar in that programmes were described using the same attributes, there were some inevitable differences. Firstly, the length of the programme in the CV task was set at 12 months in order to ensure that WTP estimates were comparable across individuals. Had this been set at a different level, the WTP estimates may have differed. For example, the DCE results suggested that respondents preferred shorter programmes therefore WTP_{CV} estimates may have been higher (per month) if the level had been set at six months. On the other hand, there is no specific reason to believe that this would have resulted in differences in the estimates when compared across user groups.

It should be noted that some details from the DCE task were not included in the CV task due to concerns over respondent fatigue. For example, it was not reiterated in the CV task that individuals are expected to re-gain 100% of their lost weight without a programme. As well as this, corresponding absolute weight re-gain values were not listed alongside the proportional levels in the CV task as they were in the DCE. This latter point may be important if respondents in the DCE were strongly influenced by the absolute values. As discussed in section 8.3, it could be that respondents in the PSU group were more strongly influenced by the weight re-gain attribute as a result of their higher absolute values relative to the other two groups. This could be the reason for the larger WTP_{DCE} estimate (in absolute terms) for the PSU group relative to the other two groups, and hence this may explain why no such differences between groups were found in the CV task.

8.5.3 Conclusion

On the whole, the results are inconclusive regarding the potential differences between indirect (DCE) and direct WTP estimates (CV). There is moderate evidence that WTP_{CV} estimates are larger than WTP_{DCE} estimates for one of the user groups, however the comparison made in this study is atypical and it is therefore difficult to contrast this finding with the existing literature.

8.6 Research Question Three

RQ3: To what extent do certain groups have better defined preferences than other user groups, and why might these differences occur?

8.6.1 Results

It was hypothesised in Chapter Six that, as the most experienced individuals in the online panel sample based on observable characteristics, the PSU group may have 'better defined' preferences towards WLM programmes than the other two user groups. To test this hypothesis, two rationality tests were embedded into the DCE task. Table 8.8 presents the pass rates for the transitivity test (using both definitions), the dominance test and a combination of all three.

The results in table 8.8 show that the majority of respondents pass the weak transitivity and dominance tests. If respondents randomly responded to the DCE questions, there would be a 75% probability that the weak transitivity test would be satisfied. As such the results appear favourable for this test across all user groups with at least 93% of respondents passing in each user group. In addition, strictly speaking, the probability that a respondent would pass the dominance test at random is 50% and therefore the results for this test are positive in this

respect with at least 90% of respondents passing in each user group. Pass rates are substantially lower for the strong transitivity test, under 57% for each group but never less than half. However, the probability of passing this test at random is far lower than the other tests (11%). On the whole, it appears that the results of the rationality tests are favourable for all tests and across all user groups. There is little variation in the pass rates between the three groups, suggesting that their classification into these groups has little to no influence on the probability of passing.

Table 8.8 Embedded DCE Rationality Test Results (Online Panel Sample)

Group	Dominance	Weak Transitivity	Strong Transitivity	All 3
PSU	193/210 (92%)	201/210 (96%)	117/210 (56%)	110/210 (52%)
PB	223/243 (92%)	226/243 (93%)	125/243 (51%)	114/243 (47%)
NU	207/229 (90%)	221/229 (97%)	118/229 (52%)	106/229 (46%)

In order to formally examine the factors that may influence the probability of an individual passing one or all of the rationality tests, a series of models were estimated. In each model in table 8.9, a dummy dependent variable was used to indicate whether individuals passed the dominance test (model one), the weak transitivity test (model two), the strong transitivity test (model three) or all of the tests (model four). The models were estimated using logistic regression and as such the coefficients have little practical meaning, therefore marginal effects at the means were calculated and are reported alongside the coefficients in table 8.9.

Model one suggests that being heavier and being interested in a WLM programme (as determined by the self-section question) are associated with a decreased probability of passing the dominance test. The marginal effect for the former is small (0.2% for every kilogram) and significant at the 1% level, whereas the marginal effect for the latter is larger (3.5%) but is only significant at the 10% level. There is also weak evidence (10% level) to suggest that older age is associated with an increased probability of passing the dominance test. Being in the PSU group (relative to the NU group) is associated with a 7.1% increase in the probability of passing the dominance test. Similarly, being in the PB group (relative to the NU group) is associated with a 4.7% increase in the probability of passing the dominance test.

Table 8.9 Rationality Test Logit Models (Online Panel Sample)

	1. Dominance		2. Weak Transitivity		3. Strong Transitivity		4. All Tests	
	Coeff.	*Mfx.*	*Coeff.*	*Mfx.*	*Coeff.*	*Mfx.*	*Coeff.*	*Mfx.*
Age	0.135 *	0.009 *	0.104	0.004	0.018	0.004	0.040	0.010
	(0.073)	(0.005)	(0.098)	(0.003)	(0.041)	(0.010)	(0.041)	(0.010)
Age2	-0.001	-0.000	-0.001	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Higher Education	-0.196	-0.013	0.086	0.003	0.488	0.122	0.425	0.106
	(1.072)	(0.069)	(1.092)	(0.038)	(0.500)	(0.125)	(0.499)	(0.125)
Further Education	-0.497	-0.032	0.517	0.018	0.397	0.099	0.154	0.038
	(1.075)	(0.070)	(1.117)	(0.038)	(0.504)	(0.126)	(0.504)	(0.126)
Secondary Education	0.297	0.019	1.161	0.040	0.425	0.106	0.357	0.089
	(1.114)	(0.072)	(1.180)	(0.040)	(0.511)	(0.127)	(0.510)	(0.127)
Other Education	-0.918	-0.059	-0.241	-0.008	0.356	0.089	-0.247	-0.062
	(1.554)	(0.100)	(1.602)	(0.055)	(0.935)	(0.233)	(0.991)	(0.247)
Employed (FT/PT)	-0.526	-0.034	-0.445	-0.015	-0.096	-0.024	-0.153	-0.038
	(0.402)	(0.026)	(0.507)	(0.017)	(0.199)	(0.049)	(0.198)	(0.049)
Married/Cohabiting	0.089	0.006	-0.499	-0.017	0.096	0.024	0.093	0.023
	(0.312)	(0.020)	(0.455)	(0.015)	(0.180)	(0.045)	(0.180)	(0.045)
DCE Block 1	0.119	0.008	1.013 **	0.035 **	0.305	0.076	0.281	0.070
	(0.395)	(0.026)	(0.489)	(0.017)	(0.221)	(0.055)	(0.223)	(0.056)
DCE Block 2	0.087	0.006	0.725	0.025	0.542 **	0.135 **	0.572 **	0.143 **
	(0.403)	(0.026)	(0.471)	(0.016)	(0.226)	(0.056)	(0.227)	(0.057)
DCE Block 3	-0.037	-0.002	1.393 **	0.048 **	0.804 ***	0.200 ***	0.735 ***	0.184 ***
	(0.404)	(0.026)	(0.589)	(0.019)	(0.229)	(0.057)	(0.228)	(0.057)
Found DCE Difficult	0.335	0.022	-0.190	-0.007	-0.449 ***	-0.112 ***	-0.356 **	-0.089 **
	(0.301)	(0.019)	(0.369)	(0.013)	(0.162)	(0.040)	(0.162)	(0.040)
Weight (kg)	-0.025 ***	-0.002 ***	0.011	0.000	0.005	0.001	-0.002	-0.000
	(0.009)	(0.001)	(0.013)	(0.000)	(0.006)	(0.001)	(0.005)	(0.001)
Interested in Using a WLM Programme	-0.546 *	-0.035 *	-0.298	-0.010	0.251	0.062	0.060	0.015
	(0.299)	(0.019)	(0.387)	(0.013)	(0.179)	(0.045)	(0.178)	(0.045)
Potential Service User	1.099 **	0.071 **	-0.640	-0.022	-0.033	-0.008	0.279	0.070
	(0.489)	(0.031)	(0.675)	(0.023)	(0.271)	(0.067)	(0.269)	(0.067)
Potential Beneficiary	0.731 *	0.047 *	-1.124 **	-0.039 **	-0.165	-0.041	0.019	0.005
	(0.419)	(0.027)	(0.534)	(0.018)	(0.228)	(0.057)	(0.227)	(0.057)

	1. Dominance		2. Weak Transitivity		3. Strong Transitivity		4. All Tests	
	Coeff.	*Mfx.*	*Coeff.*	*Mfx.*	*Coeff.*	*Mfx.*	*Coeff.*	*Mfx.*
Constant	1.241		-0.072		-1.449		-1.591	
	(1.817)		(2.314)		(1.047)		(1.049)	
Observations	682		682		682		682	
Pseudo R2	0.0751		0.0908		0.0289		0.0272	

Model two suggests that being in the PB group (relative to the NU group) is associated with a 3.9% decrease in the probability of passing the weak transitivity test (significant at the 5% level). The block of the DCE also appears to have some influence on the probability of passing the treat. Model three suggests that finding the DCE task difficult to complete is associated with an 11.2% decrease in the probability of passing the strong transitivity test (significant at the 1% level). Similar to model two for the weak transitivity test, this model also suggests that the block of the DCE that the respondents faced influenced their probability of passing. However, the coefficients on the user group variables are not statistically significant. Finally, model four suggests that finding the DCE task difficult to complete is associated with an 8.9% decrease in the probability of passing all the tests (although this is only significant at the 5% level). Similar to models two and three, this model also suggests that the block of the DCE that the respondents faced influenced their probability of passing all of the tests.

8.6.2 Discussion

The results indicate that very few of the included independent variables were statistically significant. There were some differences between the user groups in relation to the likelihood of passing the rationality tests, although there were relatively minor and none of the coefficients on the user group variables were highly statistically significant (i.e. at the 1% level). Table 8.8 indicated that there was a relatively large difference between the combined (all tests) pass rate for the PSU group relative to the other two groups. Therefore, it is surprising that no user group dummies were found to be significant in model four in table 8.9.

With regards to the other variables that were found to be significant, the most logical is the finding that difficulty in completing the DCE task was associated with a reduced likelihood of passing the strong transitivity test (as well as the combination of all three). It is easy to imagine how a respondent that is struggling to make decisions in the DCE may be less likely to pass the test; their preferences may not have been well defined and hence been inconsistent throughout the task. In contrast, there are several unexpected statistically significant variables. It is unclear why weight may influence the pass rate of the dominance test, although this effect was minimal. It is also not clear why the block of the DCE that respondents were assigned to influences the likelihood of passing the transitivity tests. It may be the case that the scenarios that preceded or succeeded the transitivity test scenarios influenced the difficulty of the test for respondents. It could be the case that this effect would have been minimised had the order of the (regular i.e. non-test) scenarios been randomised.

8.6.3 Conclusion

While the pass rates for the rationality tests are positive, there is little evidence to suggest that pass rates vary across the three user groups. Several factors have been identified that may influence the likelihood of respondents passing the rationality tests that may warrant further exploration in future research.

8.7 Overall Conclusion

The results presented in this chapter provide mixed evidence relating to research questions one to three. With regards to research question one, it appears that preferences do differ across the three user groups when examining the results of the DCE. Those in the PSU group seem to be more strongly affected by worse weight re-gain outcomes than those in the other two groups, as hypothesised. Additionally, those in the PB and NU groups appear to be more price sensitive than the PSU group. This is in line with the *a priori* expectations that non-users would be less interested in the service (as reflected in the fact that they are less willing to pay for a programme). Research question two involved the examination of differences in WTP estimates across the three user groups. While the DCE provided evidence that WTP estimates are larger (in absolute terms) for the PSU group as hypothesised, the CV task found no differences between the three user groups. There was also only mild evidence that the WTP estimates from the DCE and the CV tasks differed, with the CV estimates being larger than those from the DCE. Finally, analysis relating to research question three indicated that there are next to no differences in the pass rates of the rationality tests across the three user groups. As such, it would appear that there is little evidence to suggest that any of the groups have 'better defined' preferences over another group in the online panel sample. On the whole, it appears that only the basic DCE results provide convincing evidence that preferences differ across the three user groups.

Chapter 9. Comparison of Results from the Trial and the Online Panel

The aim of this chapter is to collate and compare the results from the two samples in Chapters Seven and Eight in relation to the three research questions presented in Chapter Six. The first research question relates to whether preferences for a health service differ significantly between different user groups, as shown by the results of a DCE. The second, split into two parts, first examines whether WTP estimates are different between user groups and secondly whether estimates differ between methodologies. The third research question examines whether certain user groups have better defined preferences using the results of embedded rationality tests in the survey. The fourth and final research question examines whether differences in preferences could be explained by the recruitment vehicle.

The chapter begins with a comparison of the characteristics of each user group (section 9.1). In this chapter, the trial sample will not be split due to the lack of evidence to suggest that characteristics and preferences significantly differ between the two trial arms, as shown in the previous chapter. Following the comparison of characteristics, each research question is addressed in turn with a comparison of the results and a discussion (sections 9.2-9.6). Finally, a summary of the results is provided to end the chapter (section 9.7).

9.1 Respondent Characteristics

9.1.1 Results

Table 9.1 combines tables 7.1 and 8.1 from the previous two chapters in order to examine how respondent characteristics may differ across the four key user groups that will be examined in this chapter. From this point, the trial sample will be considered as a single user group, referred to as ‘service users’ (SU) due to their experience in the trial.

One-way ANOVA suggests that the average ages of each user group differ ($p<0.01$). Post-hoc analysis using Tukey-Kramer pairwise comparisons suggest differences occur between the PSU and PB groups, the PB and NU groups, as well as the SU and PSU groups. Additionally, chi2 tests suggest that the spread of ages across the five categories displayed in table 9.1 differ across the four groups ($p<0.01$).

The gender split also differs across the four groups ($p<0.01$), with far more females present in the SU group relative to any other. Additionally, only the PB group has a higher proportion of males relative to females. The spread of BMI classifications is also variable across the four groups; however, this is due to the screening/recruitment process.

There also appears to be a slightly different spread of educational qualifications across the four groups ($p<0.05$). This is most noticeable when focusing on the proportion of individuals with university-level qualifications, which range from 58% of the SU group to 40% of the PB group. No other qualification level has such a wide range across the four groups.

Employment status also differed across the four groups ($p<0.01$) and this can be seen clearly in a number of areas. For example, 57% of the SU group are employed full-time compared to 41% in the NU group. The latter has the highest proportion of part-time workers (25%) of the four groups. Additionally, 25% of the PB group are retired compared to 12% of the PSU group. On the other hand, the spread of household incomes did not vary significantly across the four user groups ($p>0.1$). Unsurprisingly, given the screening process, average weight varies across the four user groups ($p<0.01$). Pairwise comparisons suggest that there are statistically significant differences in the average values in all comparisons except for that of the SU and PB group (mean difference of 1.6kg).

One-way ANOVA suggests that the average values from the generic health measures (VAS and EQ-5D-5L) also differ across the four user groups ($p<0.01$). Post-hoc analysis suggests that the differences occur between the larger average values in the NU group relative to the PSU and PB groups. Statistically significant differences were not detected between the NU group and the SU group for either measure, however this may be more reflective of the relatively small sample size for the SU group than a lack of a difference.

9.1.2 Discussion

It is unsurprising that there are differences between the user groups with regards to weight, BMI and generic health measures as all of the groups were defined by a combination of their weight and in many cases their weight loss history. It is, however, a concern that the groups differ widely in their other characteristics given that comparisons will be made across the four groups. On the other hand, it is promising that there are no statistically significant differences between the four groups with respect to the spread of household income, however each category covers a range of £20,000 or more and is not particularly sensitive for this reason. The household income question in the survey had double the number of categories than those presented in table 9.1, however no significant differences are found when looking at the full spread of categories either ($p>0.1$).

Table 9.1 Respondent Characteristics (All Four User Groups)

Sample	Trial	Online Panel		
User Group	**Service User (SU)**	**Potential Service User (PSU)**	**Potential Beneficiary (PB)**	**Non-User (NU)**
n	113	210	243	229
Age (mean)	49	44	50	45
18-24	0 (0%)	9 (4%)	7 (3%)	14 (6%)
25-34	12 (11%)	48 (23%)	41 (17%)	54 (24%)
35-44	34 (30%)	55 (26%)	38 (16%)	54 (24%)
45-54	32 (28%)	50 (24%)	52 (21%)	35 (15%)
55+	35 (31%)	48 (23%)	105 (43%)	72 (31%)
Gender				
Male	29 (26%)	87 (41%)	139 (57%)	89 (39%)
Female	84 (74%)	123 (59%)	104 (43%)	140 (61%)
BMI Classification				
Normal	7 (6%)	0 (0%)	0 (0%)	229 (100%)
Overweight	53 (47%)	0 (0%)	153 (63%)	0 (0%)
Obese	53 (47%)	210 (100%)	90 (37%)	0 (0%)
Respondent Weight (Kilogrammes)				
Mean	86.3	98.0	87.9	64.9
(Standard Deviation)	(17.93)	(16.72)	(16.10)	(10.05)
Highest Level Qualification				
Higher (University)	65 (58%)	92 (44%)	97 (40%)	109 (48%)
Further (College)	23 (20%)	52 (25%)	78 (32%)	67 (29%)
Secondary (School)	20 (18%)	60 (29%)	56 (23%)	46 (20%)
None	2 (2%)	6 (3%)	8 (3%)	4 (2%)
Other	3 (3%)	0 (0%)	4 (2%)	3 (1%)
Employment Status				
Employed (full-time)	64 (57%)	107 (51%)	113 (47%)	94 (41%)
Employed (part-time)	13 (12%)	39 (19%)	41 (17%)	57 (25%)
Student	1 (1%)	2 (1%)	3 (1%)	8 (3%)
Retired	20 (18%)	25 (12%)	60 (25%)	37 (16%)
Homemaker/Caregiver	4 (4%)	19 (9%)	13 (5%)	15 (7%)
Unemployed	2 (2%)	16 (8%)	12 (5%)	10 (4%)
Other	9 (8%)	2 (1%)	1 (<1%)	8 (3%)
Household Income				
£20,000 or less	18 (16%)	54 (26%)	68 (28%)	48 (21%)
£20,001 - £40,000	39 (35%)	80 (38%)	72 (30%)	84 (37%)
£40,001 - £60,000	33 (29%)	41 (20%)	58 (24%)	57 (25%)
£60,001 and above	23 (20%)	35 (17%)	45 (19%)	40 (17%)
Health (Visual Analogue Scale; 0=Dead, 100=Perfect Health)				
Mean	79.0	73.5	74.9	80.2
(Standard Deviation)	(17.11)	(20.09)	(18.39)	(17.62)
Health (EQ-5D-5L Utilities[1]; 0=Dead, 1=Perfect Health)				
Mean	0.887	0.870	0.881	0.927
(Standard Deviation)	(0.128)	(0.187)	(0.188)	(0.116)

[1]English population utilities from Devlin et al. (2017)

9.2 Research Question One

RQ1: To what extent do preferences for a health service, elicited via a discrete choice experiment, differ between different user groups, and why might these differences occur?

9.2.1 Results

Table 9.2 combines the majority of the information from tables 7.2 and 8.2 from the previous two chapters. Therefore, the results presented in table 9.2 are from mixed logit models using 1000 Halton draws where all parameters are modelled as random and normally distributed. Other models that have been estimated using this data in Chapters Seven and Eight and will not be considered in this chapter for the reasons previously described. The differences in preferences between the four user groups are clear to see in table 9.2 with regards to the magnitudes and statistical significance of the coefficients in each model. The remainder of this subsection will describe the results one attribute at a time.

The first clear difference between the four models is the significance of the length attribute. Unlike the other three groups, the statistically insignificant coefficient in model one suggests that the SU group are, on average, indifferent to the length of the WLM programme. There is significant preference heterogeneity for this attribute in all models suggesting that, although many people in the PSU, PB and NU groups prefer shorter programmes, there is some disparity in preferences within these groups.

The results for the delivery of feedback attribute are slightly more mixed across the four models. There is weak evidence to suggest that the SU group prefer to receive feedback via phone call relative to face to face ($p<0.1$), however models two to four suggest that the other three groups are indifferent. There is strong evidence to suggest that the NU group prefer to receive feedback via the online tool(s) relative to face to face ($p<0.01$) and weak evidence of a similar preference in the PB group ($p<0.1$). In contrast, there is evidence for all groups apart from the PSU group (where the coefficient is insignificant) of a preference for receiving feedback via text message relative to face to face ($p<0.1$).

Table 9.2 Results from the Mixed Logit Models (All Four User Groups)

Variables	1. Service User		2. Potential Service User		3. Potential Beneficiary		4. Non-User	
	Coefficients	Std. Devs.	Coefficients	Std. Devs.	Coefficients	Std. Devs.	Coefficients	Std. Devs.
Alternative-Specific Constant	-0.0390	3.492***	2.380***	2.357***	3.012***	2.675***	2.897***	3.569***
	(0.493)	(0.538)	(0.301)	(0.256)	(0.315)	(0.294)	(0.379)	(0.527)
Length (months)	0.0163	0.0657***	-0.0313***	0.0285*	-0.0285***	0.0398***	-0.0356***	-0.0697***
	(0.0154)	(0.0197)	(0.00769)	(0.0154)	(0.00758)	(0.0149)	(0.00913)	(0.0126)
Feedback[1]								
via Phone Call	0.500*	-0.0792	0.179	0.422	0.218	0.914***	0.129	-0.0890
	(0.266)	(0.384)	(0.144)	(0.343)	(0.146)	(0.257)	(0.149)	(0.315)
via the Online Tool(s)	-0.324	-1.555***	0.199	0.670***	0.219*	-0.0731	0.566***	0.0876
	(0.308)	(0.465)	(0.148)	(0.256)	(0.130)	(0.526)	(0.150)	(0.329)
via Text Message	1.006***	0.752**	0.181	-0.148	0.337***	-0.138	0.615***	-0.355
	(0.251)	(0.302)	(0.131)	(0.290)	(0.123)	(0.316)	(0.142)	(0.246)
Reminders[2]								
via Text Message	0.665***	0.553	0.144	-0.255	0.216*	0.701***	0.425***	0.352
	(0.238)	(0.409)	(0.124)	(0.376)	(0.125)	(0.192)	(0.135)	(0.264)
via Phone Call	0.879***	0.723**	-0.0400	-0.617***	0.142	0.279	0.103	-0.231
	(0.294)	(0.320)	(0.150)	(0.223)	(0.134)	(0.301)	(0.156)	(0.476)
via the Online Tool(s)	1.414***	0.0178	0.0200	-0.238	0.343***	0.226	0.352**	-0.382*
	(0.317)	(0.404)	(0.143)	(0.336)	(0.132)	(0.303)	(0.153)	(0.231)
Online Tool[3]								
App Only	0.329	0.238	0.242*	-0.00304	0.129	0.102	0.224	0.138
	(0.244)	(0.635)	(0.134)	(0.253)	(0.127)	(0.234)	(0.140)	(0.294)
Website Only	-0.00326	-0.382	0.211	0.446	-0.0537	0.499*	0.197	-0.831***
	(0.274)	(0.344)	(0.146)	(0.289)	(0.143)	(0.269)	(0.159)	(0.223)
App & Website	0.543**	1.293***	0.240*	0.144	0.137	0.575***	-0.0442	-0.623***
	(0.260)	(0.405)	(0.130)	(0.359)	(0.127)	(0.221)	(0.140)	(0.234)
Outcome (% weight re-gain)	-0.0712***	0.0539***	-0.0264***	0.0289***	-0.0168***	0.0226***	-0.0178***	0.0263***
	(0.00884)	(0.00687)	(0.00300)	(0.00308)	(0.00222)	(0.00271)	(0.00282)	(0.00376)
Cost (£ per month)	-0.232***	0.181***	-0.140***	0.138***	-0.156***	0.128***	-0.197***	0.158***
	(0.0323)	(0.0227)	(0.0145)	(0.0134)	(0.0133)	(0.0116)	(0.0177)	(0.0165)
Sample Size (Observations)	113 (3,390)		210 (6,300)		243 (7,290)		229 (6,870)	
Log Likelihood	-701		-1,505		-1,752		-1,587	
AIC	1,426		3,036		3,530		3,200	
BIC	1,506		3,124		3,620		3,289	

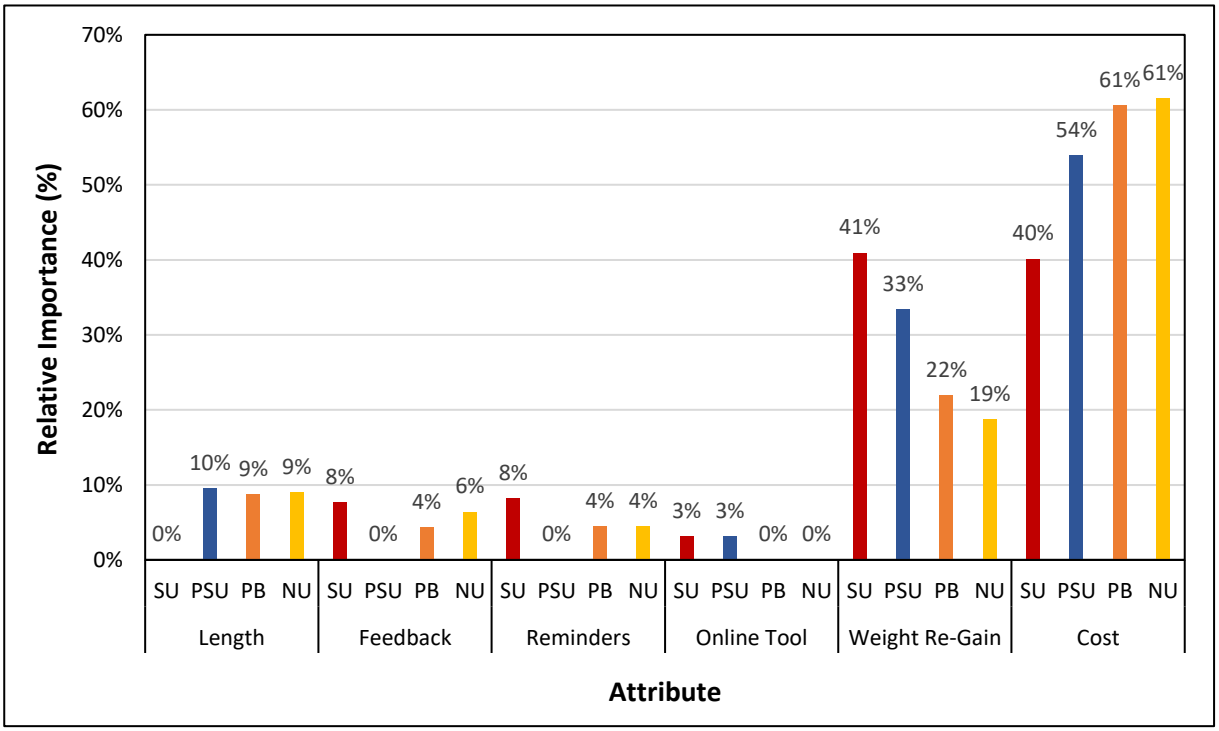
Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; [1] Base level "face to face" feedback; [2] Base level "no reminders"; [3] Base level "no online tool"

The results for the delivery of reminders attribute are also mixed across the models. Every delivery method is preferred to not receiving reminders by the SU group (model one), with a clear hierarchy from text message being the least preferred to online tools being the most preferred. In contrast, no delivery method is clearly preferred (or considered worse) to receiving no reminders in the PSU group (model two), as evidenced by the lack of statistically significant coefficients. There is weak evidence ($p<0.1$) to suggest that the PB group (model three) prefer to receive reminders via text message, and strong evidence to suggest that they prefer to receive reminders via the online tool(s) ($p<0.01$). There is no evidence to suggest that this group prefer receiving reminders via phone call relative to not receiving any. This latter point is also the case for the NU group (model four), where there is strong evidence to suggest a preference for receiving reminders via text message ($p<0.01$) and moderate evidence to suggest a preference for receiving reminders via online tool(s) ($p<0.05$), relative to not receiving any reminders.

When comparing the results of all four groups, the availability of online tool(s) appears to be one of the least important attributes across the board. The strongest evidence of a preference for a level of this attribute is in the SU group (model one) where there is moderate evidence ($p<0.05$) of a preference for having both a mobile application and a website, relative to not having either. There is also weak evidence ($p<0.1$) to suggest that the PSU group (model two) prefer to have a mobile application alone, or a combination of a website and mobile application, relative to not having either. No coefficients are statistically significant for this attribute in the PB and NU groups (models three and four). It should be noted that there is a lot of preference heterogeneity associated with this attribute, particularly for the combination of a mobile application and website where the standard deviations are statistically significant at the 1% level in all models apart from model two.

The weight re-gain and cost attributes were both highly statistically significant ($p<0.01$) in all models and the coefficients had relatively large, negative coefficients. Relative importance calculations provide an insight into how important these attributes were relative to the others and avoids the scale issues when comparing coefficients across models. Figure 9.1 presents the relative importance of each attribute for each user group and illustrates the dominance of these two attributes. However, there are clear differences across the four groups. The weight re-gain attribute is more important for the SU group relative to the PSU, PB and NU groups (with a preference order of SU > PSU > PB > NU). In contrast, the cost attribute is more important for the NU and PB groups relative to the PSU and SU groups (with a preference order of NU = PB > PSU > SU). This is all roughly in line with *a priori* expectations.

Figure 9.1 Relative Importance of the Attributes (All Four User Groups)



9.2.2. Discussion

The models in table 9.2 suggest that preferences differ in many ways across the four user groups. It is interesting that those involved in the trial, the SU group, do not have a clear tendency to select a programme over the opt-out option whereas the other three groups do (based on the coefficient for the alternative-specific constant). It is likely that this group has a better comprehension of what it is like to take part in such a programme, even if they did not partake (control arm), and hence are more willing to select the opt-out option when the characteristics do not suit them.

It is also interesting that those in the SU group do not appear to be affected by the length of the WLM programme. In practice, healthcare professionals would hope that individuals do not base their decision to take part in such a programme on this particular characteristic. It could be the case that this groups' relatively greater experience of the issue (WLM) is being reflected here. In other words, those with a good understanding of the difficult nature of achieving WLM may be more likely to accept that there is no quick-fix and that a longer programme has its merits for this reason.

It was not obvious *a priori* when it came to potential differences that might occur with regards to the 'process' characteristics (i.e. feedback, reminders, online tools). The results across the four models present a varied picture, although there are no dramatic differences between the groups that are statistically significant (e.g. different signs). Feedback delivery preferences

vary slightly, however there is broad agreement that feedback via text message is preferable to face to face feedback (with the exception of the PSU group; model two). Online tools also appear to not be particularly desirable, with little evidence to suggest that people based their choices on this attribute, unless both a mobile application and a website were included. That being said, this attribute level was still insignificant for both the PB and NU groups (models three and four).

However, it is interesting to see that the SU group has such strong preferences for reminders relative to the other three groups. Those involved in the both arms of the trial are likely to have, at times, struggled to remember to weigh themselves each day (note that only the treatment arm received reminders). It appears that the groups from the online panel may have underestimated this element of the WLM programmes presented to them due to their relative lack of experience.

Some of the most striking differences between the four user groups occur in relation to the weight re-gain and cost attributes; figure 9.1 illustrates this clearly. The trends in relative importance of these two attributes align with *a priori* expectations. It was expected that those in the SU group would care more about weight re-gain relative to other groups, and it was hypothesised that this group would care less about cost as a result. This is important in terms of the face validity of the results.

9.2.3 Conclusion

Overall, preferences do appear to differ across the four user groups in several different ways. The clearest differences can be seen when looking at the relative importance of attributes such as weight re-gain and cost, with the most experienced group caring relatively more about weight re-gain and less about cost than any other group.

9.3 Research Question Two Part A

RQ2a: To what extent do WTP estimates differ between different user groups, and why might these differences occur?

9.3.1 Results (CV)

Table 9.3 contains the results for each group to the self-selection; the question that determined whether respondents would be faced with a CV exercise or not. Chi2 tests provide convincing evidence ($p<0.01$) that the responses to this question significantly vary across the four user groups, even if similar responses are collapsed (e.g. one and two; three and four).

Table 9.3 Responses to the CV Self-Selection Question (All Four User Groups)

Response to the "Interest Question"	**SU**		**PSU**		**PB**		**NU**		**Total**	
	n	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
1. I would like to use x	17	15%	37	18%	26	11%	24	10%	**104**	**13%**
2. I would like to use x & want it to be available for others	42	37%	50	24%	45	19%	20	9%	**157**	**20%**
3. I might want to use x	13	12%	44	21%	52	21%	37	16%	**146**	**18%**
4. I might want to use x & want it to be available for others	31	27%	45	21%	57	23%	44	19%	**177**	**22%**
5. I would not use x but I want it to be available for others	9	8%	21	10%	34	14%	77	34%	**141**	**18%**
6. I do not think x should be made available	1	1%	13	6%	29	12%	27	12%	**70**	**9%**
Total	**113**	**100%**	**210**	**100%**	**243**	**100%**	**229**	**100%**	**795**	**100%**

x = a programme like this (referring to the hypothetical programmes from the DCE task)

Responses 1-4 – forwarded to the "use value" WTP exercise;

Response 5 – forwarded to the "externality" WTP exercise;

Response 6 – forwarded to open-ended comment box.

Table 9.4 WTP Estimates from the Use Value CV Task (All Four User Groups)

	Service User (SU)								Potential Service User (PSU)							
		Stated[1]		Provided[2]		Willingness to Pay (WTP)				Stated[1]		Provided[2]		Willingness to Pay (WTP)		
Weight Re-Gain	n	n	%	n	%	Median	Mean	SD	n	n	%	n	%	Median	Mean	SD
0%	103	81	79%	67	65%	£10.00	£17.81	32.19	176	118	67%	93	53%	£10.00	£14.28	13.50
10%	81	63	61%	50	49%	£10.00	£13.66	16.36	118	91	52%	60	34%	£10.00	£11.93	8.52
20%	63	43	42%	30	29%	£10.00	£10.07	5.79	91	59	34%	29	16%	£10.00	£11.90	6.20
40%	43	24	23%	14	14%	£8.00	£8.29	5.61	59	32	18%	11	6%	£5.00	£9.00	6.32
60%	24	13	13%	9	9%	£5.00	£7.11	6.13	32	17	10%	4	2%	£12.50	£11.00	5.23
80%	13	5	5%	3	3%	£5.00	£8.67	10.02	17	12	7%	2	1%	£10.00	£10.00	0.00
90%	5	4	4%	2	2%	£3.00	£3.00	2.83	12	11	6%	2	1%	£10.00	£10.00	0.00
100%	4	4	4%	2	2%	£3.00	£3.00	2.83	11	11	6%	2	1%	£10.00	£10.00	0.00

	Potential Beneficiary (PB)								Non-User (NU)							
		Stated[1]		Provided[2]		Willingness to Pay (WTP)				Stated[1]		Provided[2]		Willingness to Pay (WTP)		
Weight Re-Gain	n	n	%	n	%	Median	Mean	SD	n	n	%	n	%	Median	Mean	SD
0%	180	115	64%	81	45%	£10.00	£13.84	15.89	125	76	61%	58	46%	£10.00	£13.91	16.66
10%	115	95	53%	62	34%	£10.00	£11.97	10.58	76	62	50%	45	36%	£10.00	£13.13	18.64
20%	95	68	38%	46	26%	£10.00	£10.48	8.39	62	47	38%	30	24%	£10.00	£12.57	20.97
40%	68	32	18%	20	11%	£8.50	£8.30	4.05	47	22	18%	11	9%	£5.00	£16.27	31.49
60%	32	16	9%	9	5%	£10.00	£10.33	4.97	22	18	14%	7	6%	£10.00	£20.43	35.43
80%	16	13	7%	6	3%	£10.00	£11.00	6.32	18	13	10%	4	3%	£12.50	£32.50	41.73
90%	13	12	7%	4	2%	£12.50	£13.75	4.79	13	9	7%	3	2%	£10.00	£11.67	2.89
100%	12	7	4%	3	2%	£15.00	£15.00	5.00	9	7	6%	2	2%	£12.50	£12.50	3.54

Percentages are based on the number of individuals that were forwarded to the exercise from each user group in the first instance (underlined)

[1]Those that selected "yes" when asked whether they would be willing to pay (those that select "no" exit the exercise at that particular stage)

[2]Those that provided a value when asked for the maximum amount that they would be willing to pay *and* did not increase their value at any point during the task ('logic rule')

Combining responses one and two, it is clear that a higher proportion of the SU group wish to use a WLM programme like the ones described in the DCE task relative to the other three groups (the specific order is SU > PSU > PB > NU). The proportion of respondents stating that they *might* want to use a WLM programme (responses three and four) are fairly similar across the four user groups. Combining all four responses shows that a significantly different proportion of each user group went on to face the "use value" contingent valuation exercise; from 91% of the SU group to 54% of the NU group (once again following the order of SU > PSU > PB > NU).

Table 9.4 presents the full results from the use value CV exercise. One-way ANOVA suggests that the mean values at each weight re-gain level are not statistically significantly different across the four user groups. The similarity between the values provided by each group are clearest when looking at the median values, which are identical for every group for the first three weight re-gain levels (£10 at 0%, 10% and 20%). As explained in the previous chapters, the average values increase as the sample size decreases (from around 60% weight re-gain onward) due to the way that some individuals responded to the questions (not changing their value as weight re-gain increases).

9.3.2 Discussion (CV)

The differences in interest between the four groups, as determined by the self-selection question, provide clear evidence that these groups differ in more than just their weight loss and weight loss maintenance history. In other words, there is a notable difference in their desire to participate in a WLM programme. The spread of proportions of people responding that they wish to use a WLM programme is largely as expected (SU > PSU > PB > NU). However, it would not have been surprising if the SU group had a slightly lower proportion, given that roughly half of this group had been 'treated'. In fact, there is no statistically significant difference between the responses to the self-selection question between the two trial arms (as shown in Chapter Seven). This could be due to the fact that trial participants found that the intervention was beneficial, and/or that an alternate version of the trial intervention would be helpful to them in future.

Despite the differences in interest, the average WTP values between the four user groups are not statistically significantly different from one another. It could be the case that, although the user groups differ on average, the individuals that believe that a WLM programme is relevant to them have a similar perception of the market value of the service. While a respondent with a normal BMI would have been classified as a 'non-user' it cannot be determined that the respondent has no history of a higher BMI classification or attempted weight loss. In fact,

many individuals that are not considered high-risk (e.g. not obese) are still concerned about their weight and make regular attempts to lose weight (Santos et al., 2017). As such, while there may be fewer individuals in the NU group that feel that they would use a WLM programme on the whole, the individuals that are interested in such a programme may have strong and well-defined preferences that are in line with those with clear experience of WLM programmes (e.g. the SU group).

9.3.3 Results (DCE)

Table 9.5 contains WTP estimates for weight re-gain for each of the four user groups; these are calculated by taking the ratio of the coefficients for weight re-gain and cost, as explained in Chapter Six.

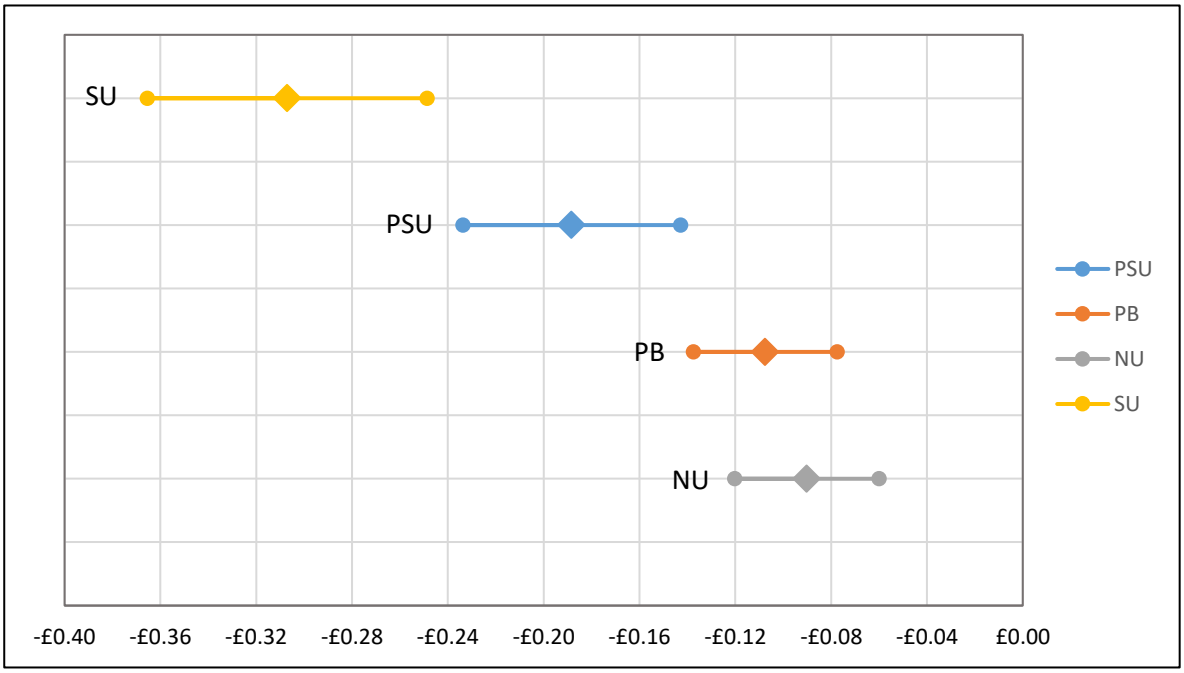
Table 9.5 WTP Estimates from the DCE (All Four User Groups)

Group	WTP for % weight re-gain	95% Confidence Interval [Lower, Upper]
Service User (SU)	-£0.31	[-£0.37, -£0.25]
Potential Service User (PSU)	-£0.19	[-£0.23, -£0.14]
Potential Beneficiary (PB)	-£0.11	[-£0.14, -£0.08]
Non-User (NU)	-£0.09	[-£0.12, -£0.06]

The results are interpreted as such: individuals from the SU group would pay £3.07 less per month for a programme if weight re-gain increased by 10 percentage points. In comparison, the PSU group would pay £1.90 less per month, the PB group £1.10 less per month and the NU group £0.90 less per month.

The lack of overlap in the 95% confidence intervals of the SU and PSU groups relative to all other groups suggest that these estimates are statistically significantly different from each other and the other groups. However, it would appear that the estimates for the PB and NU groups are not significantly different from one another, given the significant overlap in the confidence intervals. This is illustrated in figure 9.2.

Figure 9.2 WTP Estimates from the DCE for Weight Re-Gain (All Four User Groups)



9.3.4 Discussion (DCE)

It is interesting that there are statistically significant differences in WTP between the user groups. The differences are roughly in line with *a priori* expectations: it was expected that individuals from the SU group would be willing to pay more to avoid weight re-gain relative to the other groups. It was also expected that the PSU group would be willing to pay more to avoid weight re-gain relative to the PB and NU groups. However, it is somewhat surprising that there is no statistically significant difference between the PB and NU groups. On the other hand, the distinction between these two groups is relatively minor (overweight vs. normal BMI classification).

Despite the fact that differences exist it is not necessarily the case that they are economically meaningful. For example, if a programme set at a price of £10 per month turned out to be less effective than expected by 10 percentage points (weight re-gain) then individuals in the SU group would only be willing to pay £6.93 per month, on average. In comparison, individuals in the NU group would be willing to be pay £9.10 per month, on average. These are relatively small differences, considering that one user group has an observable need for a WLM programme whereas the other group have no observable need at all. However, it should be noted that these are monthly WTP estimates and hence the absolute difference would be much larger over the course of a typical WLM programme (e.g. 6-24 months).

9.3.5 Conclusion

There appears to be differences between the user groups in relation to the WTP estimates derived from the DCE, with the most experienced groups (SU and PSU) appearing to be willing to pay significantly more than the less experienced (PB and NU) groups. In contrast, no such differences can be identified from the WTP estimates from the CV task.

9.4 Research Question Two Part B

RQ2b: How might WTP estimates differ if they are elicited indirectly or directly?

9.4.1 Results

As explained in previous chapters, there are several methods that have been used in the past to compare WTP estimates from a DCE and a CV task. In this empirical work, the iterative nature of the CV questions was used to generate a WTP estimate for a percentage point change in weight re-gain. As these have already been identified in the DCE (WTP_{DCE} hereafter), the estimates from the CV task can be directly compared (WTP_{CV} hereafter). Table 9.6 contains the mean WTP_{CV} as well as the WTP_{DCE} estimates from table 9.5, split by group.

Table 9.6 DCE and CV Comparison in WTP for 1% Increase in Weight Re-Gain (All Four User Groups)

Group	WTP_{DCE}	WTP_{CV}	Difference	p-value
SU	-£0.31	-£0.62	£0.31	0.497
PSU	-£0.19	-£0.34	£0.15	0.178
PB	-£0.11	-£0.22	£0.11	0.087
NU	-£0.09	-£0.19	£0.10	0.045

The average WTP_{CV} values are not statistically significantly different across the four user groups (p-value>0.1). Additionally, as indicated by the p-values in table 9.6, there are no statistically significant difference at the 1% level between the WTP_{DCE} and the average WTP_{CV} value for any of the four user groups. There is some moderate evidence to suggest that the estimates vary between the two tasks for the NU group, specifically that the WTP_{CV} estimate is larger (in absolute terms). This trend is repeated in the other user groups, however there is no convincing evidence that the estimates are statistically significantly different.

9.4.2 Discussion

There is little convincing evidence that the WTP estimates from the CV and the DCE are statistically significantly different. Part of the issue may be that the CV estimates are based on very few responses; the approach taken here requires individuals to have provided values in the CV task when weight re-gain was set at both 0% and 10% (n=62 for the PB group and n=45 for the NU group). The self-selection question might not have helped either, as this will have guided some respondents away from the use value CV task. That being said, there was a clear justification for taking this approach in that individuals that would not like to use a WLM programme should be less likely to be willing to pay for one.

9.4.3 Conclusion

There appears to be no evidence of a difference between WTP estimates when they are elicited via CV and DCE. It could be the case that data limitations are responsible here, as sample sizes were not large for the WTP_{CV} estimates. As a result, it would be wise to avoid concluding that there *isn't* a difference. Ideally an adequately powered study should be designed; the current study however provides valuable information to help inform the size of that study.

9.5 Research Question Three

RQ3: To what extent do certain groups have better defined preferences than other user groups, and why might these differences occur?

9.5.1 Results

Two rationality tests were embedded within the DCE in order examine how well-defined respondents' preferences are, as described in Chapter Six. Table 9.7 presents the pass rates for the dominance test, the transitivity tests (two definitions), and a combination of the three for all four user groups.

Table 9.7 Embedded DCE Rationality Test Results (All Four User Groups)

Group	Dominance	Weak Transitivity	Strong Transitivity	All 3
SU	110/113 (97%)	111/113 (98%)	78/113 (69%)	75/113 (66%)
PSU	193/210 (92%)	201/210 (96%)	117/210 (56%)	110/210 (52%)
PB	223/243 (92%)	226/243 (93%)	125/243 (51%)	114/243 (47%)
NU	207/229 (90%)	221/229 (97%)	118/229 (52%)	106/229 (46%)

The results in table 9.7 are positive on the whole. The vast majority of respondents in each user group passed the dominance and weak transitivity tests. In addition, at least half of all respondents in each user group passed the strong transitivity test. There are very few differences in the pass rates between the PSU, PB and NU groups, however the SU group has better pass rates for all tests relative to these three groups.

In order to formally examine the factors that may influence the probability of an individual passing one or all of the rationality tests, a series of models were estimated. The models consisted of a dummy dependent variable that indicated whether respondents passed the test and were estimated using logistic regression. As such, the coefficients have little practical meaning, therefore marginal effects at the means were calculated and are reported alongside the coefficients in table 9.8.

Model one suggests that individuals with a higher weight are slightly less likely to pass the dominance test ($p<0.01$); 0.1% less likely for every additional kilogram. Additionally, there is strong evidence ($p<0.01$) to suggest that individuals from the SU group are more likely, by 10.3%, to pass the dominance test and moderate evidence ($p<0.05$) to suggest that individuals from the PSU group are more likely, by 6.1% to pass the dominance test, relative to the NU group.

Model two suggests that there are some differences in the pass rates of the weak transitivity test depending on the block of the DCE that the respondent was assigned to. Specifically, there is moderate evidence ($p<0.05$) to suggest that individuals in blocks one and three of the DCE were more likely, by around 3-4%, to pass the test relative to block four. Additionally, there is moderate evidence ($p<0.05$) to suggest that individuals in the PB group were less likely pass the test relative to the NU group, by around 3%.

Model three also suggests that there are some differences in the pass rates of the strong transitivity test depending on the block of the DCE that the respondent was assigned to. Specifically, respondents assigned to block three appear to be around 19% more likely to pass this test and the effect is significant at the 1% level. It also appears that respondents that found the DCE difficult (according to a difficulty question post-DCE) were 11% less likely to pass the test relative to those that did not find it difficult ($p<0.01$). Additionally, there is moderate evidence ($p<0.05$) to suggest that respondents that stated that they were interested in using a WLM programme (responses one and two to the self-selection question; see table 9.3) were

8% more likely to pass the test. Finally, there is moderate evidence ($p<0.05$) to suggest that individuals in the SU group were 18% more likely to pass this test.

Model four combines the all three rationality tests. It provides moderate-strong evidence that respondents assigned to blocks two and three of the DCE were more likely, by 12% and 18% respectively, to pass all three tests relative to block four. There is also moderate evidence ($p<0.05$) to suggest that respondents that found the DCE difficult were less likely, by 9%, to pass all three tests. Finally, this model provides strong evidence to suggest that respondents in the SU group are 23% more likely to pass the combination of all three rationality tests relative to the other user groups.

9.5.2 Discussion

The finding that heavier respondents are less likely to pass the dominance test may be due to a misunderstanding surrounding the weight re-gain attribute. As the absolute value would be greater for heavier respondents, it may be more tempting for respondents to select the dominated option if they incorrectly believe that this attribute refers to weight loss, despite the fact that the dominated option is also associated with a higher cost.

It is interesting to see that the block of the DCE that respondents were assigned to appears to have an impact on the pass rates of transitivity tests. It may be the case that certain blocks contained alternatives that were slightly more similar to those in the transitivity test scenarios, which impacted the likelihood of a respondent passing the test. Alternatively, this could be a random effect. Regardless, it may have been wise to randomise the order of the scenarios in the DCE such that any potential effect of the scenarios that come before and after the transitivity test scenarios is minimised. An exploration into the reasons behind this finding may be interesting as part of future research.

Despite all of this, the most striking result from the models in table 9.8 is that respondents from the trial (SU group) are 23% more likely to pass the combination of all three rationality tests. While the importance attributed to these tests is debateable, it is undeniable that higher pass rates are preferable. For one group to have such a significantly higher probability of passing the tests relative to the others is therefore hard to ignore. In addition, the fact that it is the SU group that are more likely to pass the tests fits nicely with the hypotheses of this project. It was expected that those with the most relevant experience, of both the health issue and of treatment, would provide ‘better quality’ preference data. The results here suggest that this hypothesis was correct, provided that higher rationality test pass rates can reasonably be considered an indicator of better quality preference data.

Table 9.8 Rationality Test Logit Models (All Four User Groups)

	1. Dominance		2. Weak Transitivity		3. Strong Transitivity		4. All Tests	
	Coeff.	*Mfx.*	*Coeff.*	*Mfx.*	*Coeff.*	*Mfx.*	*Coeff.*	*Mfx.*
Age	0.131 *	0.007 *	0.095	0.003	0.035	0.009	0.053	0.013
	(0.071)	(0.004)	(0.096)	(0.003)	(0.039)	(0.010)	(0.039)	(0.010)
Age^2	-0.001	-0.000	-0.001	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Higher Education	-0.237	-0.013	0.053	0.002	0.289	0.071	0.217	0.054
	(1.069)	(0.060)	(1.093)	(0.034)	(0.477)	(0.118)	(0.476)	(0.119)
Further Education	-0.508	-0.029	0.469	0.015	0.293	0.072	0.067	0.017
	(1.073)	(0.061)	(1.118)	(0.035)	(0.481)	(0.119)	(0.480)	(0.120)
Secondary Education	0.188	0.011	0.857	0.027	0.192	0.047	0.120	0.030
	(1.104)	(0.062)	(1.157)	(0.036)	(0.487)	(0.120)	(0.486)	(0.121)
Other Education	-0.885	-0.050	-0.160	-0.005	-0.282	-0.070	-0.755	-0.189
	(1.533)	(0.086)	(1.588)	(0.050)	(0.828)	(0.205)	(0.865)	(0.216)
Employed (FT/PT)	-0.617	-0.035	-0.239	-0.007	-0.120	-0.030	-0.198	-0.049
	(0.399)	(0.022)	(0.478)	(0.015)	(0.187)	(0.046)	(0.186)	(0.047)
Married/Cohabiting	0.030	0.002	-0.566	-0.018	0.009	0.002	-0.012	-0.003
	(0.306)	(0.017)	(0.449)	(0.014)	(0.169)	(0.042)	(0.168)	(0.042)
DCE Block 1	0.267	0.015	1.038 **	0.032 **	0.160	0.040	0.193	0.048
	(0.381)	(0.021)	(0.480)	(0.015)	(0.203)	(0.050)	(0.204)	(0.051)
DCE Block 2	0.207	0.012	0.745	0.023	0.406 *	0.100 *	0.472 **	0.118 **
	(0.389)	(0.022)	(0.463)	(0.015)	(0.210)	(0.052)	(0.211)	(0.053)
DCE Block 3	0.040	0.002	1.192 **	0.037 **	0.766 ***	0.189 ***	0.729 ***	0.182 ***
	(0.383)	(0.022)	(0.536)	(0.016)	(0.213)	(0.053)	(0.211)	(0.053)
Found DCE Difficult	0.353	0.020	-0.255	-0.008	-0.462 ***	-0.114 ***	-0.365 **	-0.091 **
	(0.292)	(0.016)	(0.360)	(0.011)	(0.153)	(0.038)	(0.153)	(0.038)
Weight (kg)	-0.025 ***	-0.001 ***	0.008	0.000	0.003	0.001	-0.003	-0.001
	(0.009)	(0.000)	(0.012)	(0.000)	(0.005)	(0.001)	(0.005)	(0.001)
Interested in Using a WLM Programme	-0.472	-0.027	-0.118	-0.004	0.337 **	0.083 **	0.185	0.046
	(0.290)	(0.016)	(0.377)	(0.012)	(0.164)	(0.040)	(0.162)	(0.041)
Service User	1.819 ***	0.103 ***	0.648	0.020	0.717 **	0.177 **	0.930 ***	0.233 ***
	(0.694)	(0.037)	(0.857)	(0.026)	(0.279)	(0.069)	(0.277)	(0.069)
Potential Service User	1.079 **	0.061 **	-0.498	-0.016	0.013	0.003	0.301	0.075
	(0.477)	(0.027)	(0.659)	(0.021)	(0.258)	(0.064)	(0.256)	(0.064)

	1. Dominance		2. Weak Transitivity		3. Strong Transitivity		4. All Tests	
	Coeff.	*Mfx.*	*Coeff.*	*Mfx.*	*Coeff.*	*Mfx.*	*Coeff.*	*Mfx.*
Potential Beneficiary	0.718 *	0.041 *	-1.059 **	-0.033 **	-0.125	-0.031	0.043	0.011
	(0.409)	(0.023)	(0.525)	(0.016)	(0.222)	(0.055)	(0.221)	(0.055)
Constant	1.375		0.149		-1.316		-1.427	
	(1.796)		(2.264)		(0.997)		(0.998)	
Observations	795		795		795		795	
Pseudo R2	0.0814		0.0898		0.0400		0.0392	

9.5.3 Conclusion

On the whole, the results to the rationality tests are extremely positive for all user groups, perhaps indicating that the preference data is of a high quality. However, the SU group are 23% more likely pass the combination of all three tests relative to the other user groups. This is a substantial considering that observable characteristics were controlled for. It may be the case that individuals that have experienced the health issue as well as a related healthcare intervention are more likely to be engaged in a preference elicitation task. This would perhaps suggest that samples made of such individuals are preferable from a researcher's perspective.

9.6 Research Question Four

RQ4: To what extent might differences in preferences between the user groups be attributed to the recruitment vehicle?

The focus of the previous sections has been on differences in preferences between the four user groups that were outlined in the framework in Chapter Five. In contrast, the focus of research question four is the overall difference in preferences (and in the data more generally) between the trial and online panel samples. This section will summarise the overall differences between the two samples in relation to research questions one to three, with additional analysis provided as necessary where data from the three user groups recruited via the online panel are pooled.

9.6.1 Comparison of the Trial & Online Panel Samples: Research Question One

Section 9.2 illustrated how the preferences of individuals in the SU group differed from those of the PSU, PB and NU groups. In contrast, when comparing the latter three groups, the differences were not as substantive. Table 9.9 provides the DCE results from the SU group (the trial sample) with results from the PSU, PB and NU groups combined (the online panel sample).

Even though coefficients cannot be meaningfully compared, some differences are clear from observing the statistical significance of the coefficients. For example, those in the trial sample do not appear to care about the length of the programme, whereas the online panel sample prefer shorter programmes. There is also fairly strong evidence that every feedback delivery mode is preferred to face to face feedback for the online panel sample, but the trial sample do not appear to prefer feedback via the online tool(s) over face to face feedback. Reminders appear to be important in both samples, however the online panel do not appear to favour reminders via phone call over not receiving reminders at all. In addition, the online panel sample slightly prefer reminders via text message over reminders via the online tool(s),

whereas the opposite is the case for the trial sample. Neither sample expressed overwhelmingly strong preferences for online tools, with the online panel appearing to prefer having an app over both a website and an app. In contrast, the trial sample only appear to have a preference for having both a website and an app. Unsurprisingly, and a measure of face validity, both samples prefer a lower amount of weight re-gain and lower cost.

Table 9.9 Results from the Mixed Logit Models (Trial vs. Online Panel)

	1. Trial Sample		2. Online Panel Sample	
	Coefficient	*Std. Dev.*	*Coefficient*	*Std. Dev.*
Alternative Specific Constant	-0.039	3.492 ***	2.782 ***	2.82 ***
	(0.493)	(0.538)	(0.186)	(0.164)
Length (months)	0.016	0.066 ***	-0.03 ***	0.044 ***
	(0.015)	(0.020)	(0.005)	(0.008)
Delivery of Feedback[1]				
via Phone Call	0.500 *	-0.079	0.168 **	-0.367
	(0.266)	(0.384)	(0.080)	(0.224)
via the Online Tool(s)	-0.324	-1.555 ***	0.296 ***	-0.149
	(0.308)	(0.465)	(0.078)	(0.363)
via the Text Message	1.006 ***	0.752 **	0.355 ***	-0.110
	(0.251)	(0.302)	(0.073)	(0.187)
Reminders[2]				
via Text Message	0.665 ***	0.553	0.255 ***	-0.537 ***
	(0.238)	(0.409)	(0.072)	(0.123)
via Phone Call	0.879 ***	0.723 **	0.077	-0.455 ***
	(0.294)	(0.320)	(0.082)	(0.164)
via the Online Tool(s)	1.414 ***	0.018	0.239 ***	-0.086
	(0.317)	(0.404)	(0.078)	(0.177)
Online Tool[3]				
App Only	0.329	0.238	0.191 **	0.033
	(0.244)	(0.635)	(0.075)	(0.149)
Website Only	-0.003	-0.382	0.098	0.388 **
	(0.274)	(0.344)	(0.082)	(0.188)
App & Website	0.543 **	1.293 ***	0.122 *	0.425 ***
	(0.260)	(0.405)	(0.073)	(0.148)
Weight Re-gain (%)	-0.071 ***	0.054 ***	-0.02 ***	0.025 ***
	(0.009)	(0.007)	(0.001)	(0.002)
Personal Cost (£ per month)	-0.232 ***	0.181 ***	-0.159 ***	0.138 ***
	(0.032)	(0.023)	(0.008)	(0.007)
Observations (Sample Size)	3,390		20,460	
Log Likelihood	-701		-4,876	
AIC	1,426		9,778	
BIC	1,506		9,881	

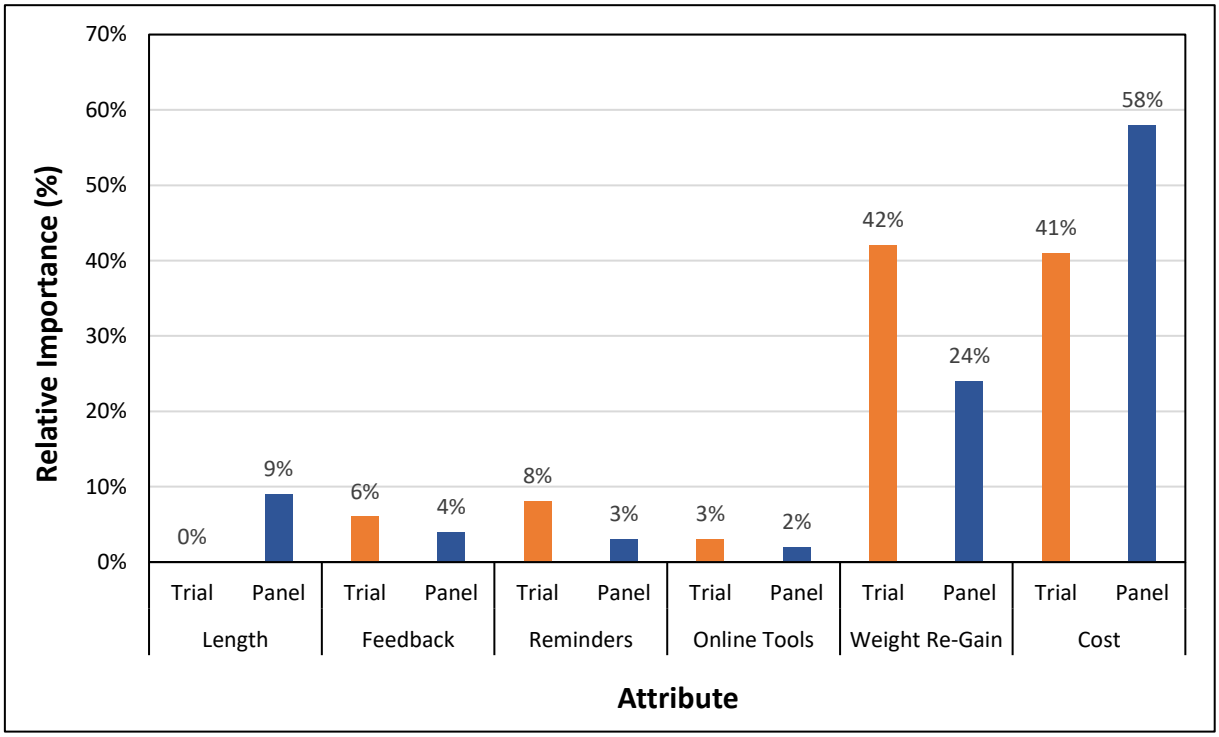
Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1;
[1] Base level "face to face" feedback; [2] Base level "no reminders"; [3] Base level "no online tool"

As has been done throughout the results chapters of this thesis, figure 9.3 provides estimates of the relative importance of each attribute according to the two models in table 9.9. There are

some major differences in the relative importance of the attributes between the two samples. Whilst the relative importance of the weight re-gain attribute is 42% for the trial sample, it is only 24% for the online panel sample. In addition, the relative importance of the cost attribute for the trial sample is 41%, whereas it is 58% for the online panel sample.

On the whole, figure 9.3 suggests that the online panel sample care more about cost, less about weight re-gain and more about the length of the programme relative to the trial sample. One could conclude that this is a result of a lack of focus on the issue at hand (weight management and weight loss maintenance) from those in the online panel relative to those in the trial. Additionally, figure 9.3 is eye-opening in that it highlights that the trial sample also care more about the key process attributes (delivery modes for feedback and reminders, availability of the online tools) relative to the online panel sample. Slight differences in the relative importance of these attributes between the user groups within the online panel made this effect less prominent in figure 9.2. On the whole, it seems that there are significant differences in preferences between the trial sample and the online panel sample.

Figure 9.3 Relative Importance of the Attributes (Trial vs. Online Panel)



9.6.2 Comparison of the Trial & Online Panel Samples: Research Question Two

Section 9.3 illustrated how differences in WTP estimates from the four user groups. Table 9.10 contains the DCE_{WTP} estimates for the two samples separately. The difference between the two estimates is clear, with the estimate almost triple the size in the trial sample relative to the online panel. In addition, the confidence intervals do not overlap at all and hence these

two estimates are likely to be statistically significantly different from one another. This is perhaps unsurprising, given that there were statistically significant differences between the DCE_{WTP} estimates of the SU and the PSU groups.

Table 9.10 WTP Estimates from the DCE (Trial vs. Online Panel)

Group	WTP for % weight re-gain	95% Confidence Interval [Lower, Upper]
Trial Sample	-£0.31	[-£0.37, -£0.25]
Online Panel Sample	-£0.12	[-£0.14, -£0.11]

Table 9.11 highlights how the trial sample differed from the online sample overall in terms of the self-selection question. Over half of the trial sample (52%) stated that they would like to use a WLM intervention compared with 30% of the online panel sample (42% of the PSU group; 30% of the PB group; and 19% of the NU group). When combining these figures with responses three and four (where respondents stated that they *might* want to use a WLM programme), 91% of the trial sample would, or might want, to use a WLM intervention compared with 71% of the online panel sample (84% of the PSU group; 74% of the PB group; and 54% of the NU group). Given the differences identified in table 9.3, a large part of this can be put down to the NU group where almost half of the sample responded with option five or six (46%); which is a measure of face validity.

Table 9.11 Responses to the CV Self-Selection Question (Trial vs. Online Panel)

Response to the "Interest Question"	Trial	Online Panel
1. I would like to use x	17 (15%)	87 (13%)
2. I would like to use x & want it to be available for others	42 (37%)	115 (17%)
3. I might want to use x	13 (12%)	133 (20%)
4. I might want to use x & want it to be available for others	31 (27%)	146 (21%)
5. I would not use x but I want it to be available for others	9 (8%)	132 (19%)
6. I don't think x should be made available	1 (1%)	69 (10%)
	113 (100%)	682 (100%)

x = a programme like this (referring to the hypothetical programmes from the DCE task)
Responses 1-4 – forwarded to the "use value" WTP exercise; Response 5 – forwarded to the "externality" WTP exercise; Response 6 – forwarded to open-ended comment box

Table 9.12 contains the WTP_{CV} estimates after pooling the results from the three user groups that were recruited via the online panel. Despite the differences in WTP_{DCE} estimates in table 9.10 and the differences in responses to the self-selection question shown in table 9.11, the median and mean values in table 9.12 do not vary substantially between the two samples. One-way ANOVA confirms that there are no statistically significant differences between the two samples in relation to the WTP_{CV} estimates at any level of weight re-gain.

Table 9.12 DCE and CV Comparison in WTP for 1% Increase in Weight Re-Gain (Trial vs. Online Panel)

Trial Sample								
Weight Re-Gain		**Stated[1]**		**Provided[2]**		**Willingness to Pay (WTP)**		
	n	**n**	**%**	**n**	**%**	**Median**	**Mean**	**SD**
0%	103	81	79%	67	65%	£10.00	£17.81	32.19
10%	81	63	61%	50	49%	£10.00	£13.66	16.36
20%	63	43	42%	30	29%	£10.00	£10.07	5.79
40%	43	24	23%	14	14%	£8.00	£8.29	5.61
60%	24	13	13%	9	9%	£5.00	£7.11	6.13
80%	13	5	5%	3	3%	£5.00	£8.67	10.02
90%	5	4	4%	2	2%	£3.00	£3.00	2.83
100%	4	4	4%	2	2%	£3.00	£3.00	2.83
Online Panel Sample								
Weight Re-Gain		**Stated[1]**		**Provided[2]**		**Willingness to Pay (WTP)**		
	n	**n**	**%**	**n**	**%**	**Median**	**Mean**	**SD**
0%	481	309	64%	232	48%	£10.00	£14.03	15.12
10%	309	248	52%	167	35%	£10.00	£12.27	12.62
20%	248	174	36%	105	22%	£10.00	£11.47	12.82
40%	174	86	18%	42	9%	£6.00	£10.57	16.47
60%	86	51	11%	20	4%	£10.00	£14.00	20.85
80%	51	38	8%	12	2%	£10.00	£18.00	24.66
90%	38	32	7%	9	2%	£10.00	£12.22	3.63
100%	32	25	5%	7	1%	£10.00	£12.86	3.93

Percentages are based on the number of individuals that were forwarded to the exercise from each user group in the first instance (underlined);
[1]Those that selected "yes" when asked whether they would be willing to pay (those that select "no" exit the exercise at that particular stage);
[2]Those that provided a value when asked for the maximum amount that they would be willing to pay and did not increase their value at any point during the task ("logic rule").

9.6.3 Comparison of the Trial & Online Panel Samples: Research Question Three

The results in table 9.8 highlight that there are very few differences with respect to the rationality test pass rates between the three user groups recruited via the online panel. In model four, the coefficients on the PSU and PB variables are not statistically significant whereas the coefficient on the SU variable is. The interpretation provided in the previous section was that being in the SU group is associated with a 23% increase in the probability of

passing all of the rationality tests. An alternative interpretation is that being in the *trial sample* is associated with a 23% increase in the probability of pass all of the rationality tests relative to being in the online panel sample (as the trial sample is the SU group).

9.6.4 Discussion (Research Question Four)

There are clear differences between the trial sample and the online panel sample in most of the key analyses. However, due to the sampling process, there are also clear differences in the level of experience of individuals within each sample. It is therefore very challenging to identify whether the recruitment source or the user group classification is the primary source of differences in preferences. That being said, the results from previous sections of this chapter show that there are differences between the three user groups recruited via the online panel. Therefore, it would seem logical to conclude that differences in experience are likely to lead to differences in preferences without ruling out that the recruitment vehicle could also play a role.

9.7 Overall Conclusion

The comparisons of the results from each of the four user groups indicate that there are some substantive differences in preferences. The DCE results illustrate that certain aspects of WLM interventions are not important to some user groups but are important to others (on average). The relative importance of the weight re-gain and cost attributes clearly differs across the four user groups in a logical manner, with the most experienced group caring the most about weight re-gain and caring the least about cost. These differences give rise to the differing WTP estimates from the DCE, with the SU group willing to pay three times more to avoid a percentage increase in weight re-gain relative to the NU group. In contrast, no differences were identified between the four user groups in relation to the WTP estimates from the CV task. This could be due to data limitations, framing, or it could be a legitimate finding. If the latter is true, researchers might wish to consider the payment vehicle (i.e. DCE vs CV) if the primary study aim is to estimate WTP.

It also appears that more experienced individuals (i.e. those in the SU group) may provide higher quality data in the sense that they are less likely to fail rationality tests. This particular finding is the most relevant with respect to research question four, as no statistically significant differences were identified between the three user groups recruited via the online panel. Therefore, if any of the differences identified in the empirical work are to be explained by the recruitment vehicle rather than the level of experience, this would appear the most likely. In contrast, differences in preferences and WTP estimates from the DCE do not appear to be fully explained by the recruitment vehicle alone.

Chapter 10. Discussion and Conclusions

The aim of this chapter is to provide a discussion of the findings from this thesis. The first section provides an overview of what the thesis aimed to achieve. The second section outlines the main contributions to knowledge from this thesis in relation to the four research questions. The third section provides the main implications of the thesis to stakeholders such as researchers and policymakers. The fourth section outlines some strengths and broad limitations of the approach taken in the thesis. The fifth section provides a series of suggestions surrounding related future research. The sixth and final section concludes the chapter and the thesis.

10.1 Thesis Outline

This thesis began with an outline of the relevant economic theory (Chapter Two) and two reviews of existing literature relating to the increasing use of preference studies in health economics and healthcare decision-making (Chapters Three and Four). It then presented a theoretical framework describing how respondent samples might differ when recruited for preference studies that focus on healthcare interventions. It then set out a series of research questions to test the framework (Chapter Five). The design of an online survey was then outlined. This survey required two recruitment processes, both with complex screening questions. It contained both a DCE and a CV task along with additional data collection relating to general health, demographic and attitudinal information (Chapter Six). The remaining chapters reported the results for the trial sample (Chapter Seven), the online panel sample (Chapter Eight) and the combined analyses (Chapter Nine) and discussed the implications of the results.

The overall aim of this thesis was to explore whether, and how, differences in individuals' experience of a healthcare intervention and/or a health issue might influence their preferences for a related healthcare intervention. The following four research questions were examined to achieve this aim:

1. To what extent do preferences for a health service, elicited via a DCE, differ between different user groups, and why might these differences occur?
2. In relation to WTP estimates:
    a. To what extent do the estimates differ between different user groups, and why might these differences occur?

b. To what extent do the estimates differ if they are elicited indirectly (via DCE) or directly (via CV)?

3. To what extent do certain user groups have better defined preferences than other user groups, and why might these differences occur?
4. To what extent might differences in preferences between the user groups be attributed to the recruitment vehicle?

The case study used in this thesis was that of a WLM intervention from the NULevel trial (Evans et al., 2015; see section 6.1). Whilst lifestyle interventions are increasingly improving when it comes to helping individuals to lose weight, very few are successful in promoting long-term maintenance of weight loss (Dombrowski et al., 2014). Weight loss and WLM are well understood and widely applicable to the general population; it is estimated that 40% of adults attempted to lose weight between 2010 and 2015 worldwide (Santos et al., 2017). As a result, this case study is particularly useful because observed differences in preferences are perhaps more likely to be attributable to genuine tastes rather than misunderstandings.

10.2 Contributions to Knowledge

10.2.1 Research Question One

Very little evidence exists regarding the differences in results that might occur when samples of respondents with different levels of prior experience complete a DCE. In fact, a recently published paper (August 2017) that set out a research agenda for patient preference studies highlighted this as an important area for further research (Levitan et al., 2017). This thesis has provided evidence to suggest that preferences elicited using a DCE differ across all of the user groups presented in the framework in Chapter Five. Therefore, preferences may indeed differ according to the relative experience of the respondents.

This is particularly interesting in the case of the SU, PSU, and PB groups because all of these groups could be considered as samples of 'patients' and previous chapters have highlighted the increasing interest in patient preference studies (see section 3.4 in particular). The relatively large difference between the SU and PSU groups suggests that results from preference studies may differ even if detailed inclusion and exclusion criteria from clinical trials are replicated when recruiting via an online panel. This difference may be due to relative experience, the recruitment vehicle, or a mixture of the two. Nonetheless, this may have implications for future preference studies with regard to their approach to recruiting patient samples. Furthermore, the results from this thesis suggest that the PB group are not particularly different from non-users in relation to their preferences for WLM interventions. This could be a useful finding for preference studies that focus on areas such as preventative

interventions or vaccinations. In existing studies, the focus may have been on identifying an at-risk group of the general population, whereas the results from this thesis suggest that there might not be a big difference between an at-risk group and a random sample.

However, the case study used here may have had an effect. In the PB group, 63% of individuals were overweight and the remainder were obese, whereas 100% of the NU group had a normal BMI classification. Whilst the make-up of the two groups are significantly different with respect to BMI, it is not necessarily the case that an overweight individual in the PB group believes that they are significantly different to an individual in the NU group with a normal BMI. That is, the individual in the PB group may not be aware that they are at-risk in this context. In another context, these groups might be defined such that at-risk individuals are more likely to be aware that they are at-risk, which could result in greater differences in preferences between the at-risk (PB) group and a group of non-users.

10.2.2 Research Question Two

It follows from research question one that there is currently little evidence regarding the differences in WTP estimates that might occur when samples of respondents with different levels of prior experience complete a DCE. Nonetheless, the limited published evidence suggests that there can be very large disparities in WTP between different groups (Najafzadeh et al, 2013; Finkelstein et al. 2015; Tinelli et al., 2016). The results from this thesis add to this sparse literature by providing evidence that WTP differs between the groups defined in this study. In contrast to the literature (see section 4.2.3), the differences identified in the DCE are not of a large magnitude, perhaps due to the scale of the cost attribute (£0-30). The key WTP estimate that is focused on in this thesis (% change in weight re-gain) ranged from -£0.31 for the SU group to -£0.09 for the NU group. However, these estimates are for a percentage point change in weight re-gain, and the payment frequency is monthly, therefore the differences become substantial when comparing two WLM interventions with vastly different levels of efficacy. For example, imagine an intervention (A) that is entirely effective (0% weight re-gain) and compare it with an intervention (B) that is not entirely effective (say 40% weight re-gain). On average, individuals from the SU group would pay approximately £12.40 less per month for intervention B relative to intervention A. In contrast, individuals from the NU group would pay £3.60 less per month for intervention B relative to intervention A (on average). Two interventions with such drastically different outcomes may be an extreme example but nonetheless the magnitude of, and differences in, WTP are substantial. It should also be stressed that these are monthly price differences, hence over the length of a WLM intervention (e.g. 6-24 months) the estimates, as well as the difference between them, become

even larger in terms of magnitude. The existing evidence suggests that different samples or subgroups can lead to different WTP estimates (Najafzadeh et al, 2013; Finkelstein et al. 2015; Tinelli et al., 2016), and the evidence provided in this thesis adds to this evidence base, as well as providing credibility to the hypothesis that this is due to differences in relative experience. Thus, it is important for future DCE studies to consider the effect of the sample composition on the WTP estimates generated in the study.

The CV task also generated WTP estimates that were compared across the groups. Unlike the DCE task, the WTP estimates across the groups were not statistically significantly different from one another. This is a curious finding, given the differences in the DCE task. However, it may be that the timing of the task (directly following the DCE, which suggested prices for a WLM intervention), the nature of the task (iterative and based on preferred attribute levels) and the manner of the responses (wide variance) has influenced this finding. It would therefore be unwise to conclude on the basis of the evidence in this thesis that CV tasks are any less susceptive to differences in WTP values between groups with differing experience. Future studies that aim make comparisons between different user groups' responses to a CV task would benefit from ensuring that the task does not follow a cognitively demanding task such as a DCE and would also benefit from larger sample sizes than those obtained the empirical work of this thesis in order to improve precision and to explore preference heterogeneity.

In addition to the examination of differences in WTP, the CV task in this thesis provides some insight into the appropriateness of using ranking information and an iterative approach that focuses on the key outcome of the intervention being valued. Whilst the responses were generally logical, a significant number of respondents stated the same value throughout the task, which is a concern as the only change in the programmes between questions was a negative one. Therefore, future studies may be better advised to use a more typical CV approach, using a format such as dichotomous choice or payment card (see section 3.2).

Regarding the second part of research question two, differences in WTP estimates between the DCE and CV, the results from this thesis are inconclusive. The DCE results (i.e. numerous insignificant coefficients) meant that WTP estimates could only compared by looking at WTP for weight re-gain in both tasks. For the CV task, this meant taking the difference between the first two questions to isolate the weight re-gain attribute, which reduced the sample size considerably and is not a typical approach to analysing CV data. Therefore, whilst no convincing evidence of a difference was identified, it would not be wise to conclude that there is no difference between WTP estimates that are elicited directly (CV) or indirectly (DCE).

10.2.3 Research Question Three

The results from this thesis suggest that, even after controlling for observable characteristics such as age, gender, highest educational qualification and employment status, individuals in the SU group were more likely to pass the dominance and the strong transitivity tests that were embedded into the DCE task. In addition, table 9.8 suggests that there could be a hierarchy with respect to the probability of passing the dominance test (SU > PSU > PB > NU), which fits with the hypothesis that those with more experience have 'better defined' preferences and are thus more likely to pass a rationality test. However, such a clear effect is not evident for the transitivity tests, with the least likely group to pass the weak transitivity test being the PB group (as opposed to the NU group). Nonetheless, it does appear that the most experienced group with respect to WLM and WLM interventions (the SU group) are the most likely to pass the combination of all three tests. This could have implications for future research because, if the tests are considered to be meaningful in terms of data quality, one might conclude that more experienced samples will provide better quality data. Whilst one interpretation of this is that individuals in the SU group have 'better defined' preferences relative to those in the other groups, another is simply that these individuals have a more vested interest in engaging in this research (as participants of the NULevel trial).

The dominance test used information from a direct ranking exercise to provide a dominant alternative from the individual respondents' perspective. This was necessary because the best option for delivery of reminders or feedback, and the availability of online tools was not clear from the researchers' perspective. The high pass rates for each group, with an overall combined pass rate of 92% is particularly promising and future research where dominant attribute levels are unclear could follow the approach used in this thesis. The fact that so many different WLM programmes were valued in the CV task (also based on the rankings) provides support for the decision to use the ranking data in the dominance test rather than pre-specifying a 'dominant' alternative.

The pass rates for the transitivity tests were also particularly high. Most DCE studies in healthcare do not contain a transitivity test (de Bekker-Grob et al., 2012; Clark et al., 2014), which could be due to the extra time required at the design stage or because dominance tests are simply better known and generally accepted. However, the pass rates for the transitivity tests were reasonably high considering the probability of passing them at random, providing some evidence to suggest that such tests could be a good alternative to dominance tests. Given that dominance tests have been criticised for being too easy to pass (Miguel et al., 2005), the

findings from this thesis might justify an increased utilisation of transitivity tests in studies that use DCEs.

10.2.4 Research Question Four

Whilst there are instances where preferences differ across the three user groups recruited from the online panel (PSU, PB and NU groups), the largest differences exist between the SU group and these three groups. As such, the exact driver of the differences identified in this thesis is unclear. It could be the case that individuals that are recruited to clinical trials are unique and cannot be easily compared with a group of individuals that appear similar on paper (e.g. the PSU group). Trial participants may be atypical to the general patient population; they are willing to dedicate a considerable amount of time to be involved in research despite not being guaranteed to receive any health benefits. In addition, individuals from the NULevel trial may have had a more vested interest than those in the online panel because the survey was based on an intervention that they understood and potentially experienced. It may also be the case that online panel participants are atypical; they are incentivised to complete surveys and these incentives may attract a specific type of individual. This might help to explain the limited differences between the PSU, PB and NU groups that were recruited from the online panel.

However, it is challenging to disentangle the effect of experience and the effect of the recruitment vehicle. There are observable differences in the preferences of the online panel groups (PSU, PB and NU groups) with respect to the general DCE results and the WTP estimates from the DCE. However, there were no significant differences in the pass rates of the rationality tests between these three groups. Therefore, it would appear that the rationality tests are the main area where differences between the four user groups might be explained by the recruitment vehicle. It could be the case that individuals in the trial had a greater incentive to focus on the survey, given that they might have been interested in providing feedback, and possibly interested in receiving the NULevel trial intervention beyond the clinical trial. However, it is not possible to prove that the recruitment vehicle is the main driver of the differences in rationality test pass rates. If a future study replicated the approach used in this thesis, this effect could be explored in further detail with larger sample sizes. A larger sample size drawn from both arms of the trial could enable clear comparisons between the two arms and, accompanied by the recruitment of a PSU group from an online panel, a comparison could be made between the control arm and the PSU group. This might provide a clearer insight into this research question because the control arm could be more similar to the PSU

group than the treatment arm. However, this would not fix the potential issue of trial participants being atypical to the general patient population.

10.3 Thesis Implications to Different Stakeholders

10.3.1 Researchers

The results from this thesis provide a relatively clear message to researchers: it is important to consider the relative experience of the respondents in your sample. Ultimately, preference heterogeneity will always exist but attempts can (and arguably should) be made to minimise this by keeping the experience of the treatment or condition fairly consistent across respondents. The importance of this will only increase over time, given the increased interest in eliciting patient preferences (see section 10.5).

Researchers should also keep in mind the possible differences that might occur between 'patients' recruited from trials and those from online panels. This will always be a challenging area for research, as individuals from trials are highly experienced by definition and potentially atypical because of their willingness to enter a study that in itself might be considered an onerous undertaking without any direct personal gains in health; their preferences may simply reflect this. However, if respondents that appear similar to the trial participants, but are recruited from online panels, express very different preferences to those from trials then it is important to be aware that this difference could be down to more than the relative experience of the respondents.

Finally, will all of this in mind, it should be increasingly important for researchers to *describe and justify* their choice of respondent sample. Very few studies provide descriptive data regarding the relative experience of their respondents in relation to the health condition and its treatment, which could easily be changed. Furthermore, very few studies provide justification for their initial choice of respondent sample. It may often be the case that the sample recruited is the easiest sample to identify, and perhaps linked to a wider piece of work. This is an understandable justification but it is often missing in subsequent publications.

10.3.2 Policymakers

Policymakers are increasingly under pressure to consider the patient voice in decision-making. This is considered a positive for most individuals involved in health-related research and policymaking. However, policymakers need to be aware of the existence of preference heterogeneity and its potential impact on the results of studies because of the knock-on effect with regard to the generalisability of results. Policymakers may be ill-advised to continue calling for patient preference studies because there needs to be a more in-depth consideration

of the type of patients that should be providing preferences. For example, if the study is being used to predict uptake, perhaps 'potential service users' or 'potential beneficiaries' would be more relevant. Alternatively, if an intervention is being refined, perhaps 'service user' preferences are the most relevant as they have experience of the intervention. The subtle differences here are important if, as suggested in this thesis, results from preference studies differ across these groups.

10.4 Strengths and Limitations

This thesis has several notable strengths. First, it provides evidence on how preferences might differ between user groups with different levels of experience using a DCE. Very few published studies have explored this area to date, and there have been specific calls for further research in this area this year (Levitan et al., 2017). The interest in this subject area is not solely due to academic curiosity but also due to the increased application of data from DCE studies in healthcare decision-making. This can be seen clearly in table 3.1; five of the ten examples of DCEs being formally incorporated into economic evaluations have been published in the past three years. In addition the most recent example, which elicits preferences from both a patient and a general public sample, cites a publication that was based on Chapter Three of this thesis when explaining their recruitment strategy (Mott & Najafzadeh, 2016). The authors state that the choice of whose preferences to elicit is important to consider (Watson et al., 2017). Another strength is that this thesis did not stick with the 'patient vs. general public' dichotomy but instead put forward a framework that could potentially help to explain differences within patient and general public samples (section 5.2). The WLM intervention used as a case study in this thesis enabled this framework to be put into practice, but there are many other ways in which the framework could be utilised in future. For example, the framework could be readily applied in other health contexts and its use does not need to be restricted to preference studies. Finally, another strength is that the online survey used in the empirical work was novel in a number of ways (see Chapter Six). For example, the ranking data was used in one of the rationality tests in the DCE, as well as in the CV task. In addition, both absolute and relative values were used to describe the weight re-gain attribute in the DCE and respondents viewed this information in the units of measurement that they preferred. The overall face validity of the results and the pass rates for the rationality tests may not have been as encouraging had this not been done. The CV task itself was also novel, requiring respondents to essentially 'self-select' the type of CV question that they faced and using a range of different health outcomes in order to explore the extent to which this influenced WTP.

Section 10.2 discussed the results relating to the four research questions and highlighted some of the limitations of the methodological approaches and the data. In addition, there are several broader limitations that cut across all of the empirical work in this thesis that need to be acknowledged. First, the framework tested in the empirical work of this thesis (section 5.2) could be criticised as overly simplistic, given that experience can differ in many other ways. This could relate to experience of a friend or family member's health issue or treatment, as outlined by Dolan (1999). In other words, experience by proxy is not accounted for in this framework. It is also the case that the framework does not allow for differences in length of experience. For example, in the context of this study, an individual that has struggled with their weight for many years may have different preferences to someone that has only recently become overweight or obese. However, the framework used in this thesis is meant to be pragmatic such that it can be tested, and incorporating more complexity into the framework would reduce its practicality. That being said, it is not necessarily the case that this framework could be readily applied in all other health settings. For example, in clinical settings it may be challenging to recruit individuals in the PSU group as it may not be feasible (or ethical) to approach individuals that require surgery after it is known that surgery is required, but before the treatment has been delivered. It may also be challenging to apply the framework in other public health settings where interventions are provided en masse and not consumed by individuals in the same way as the intervention in this study (e.g. advertising campaigns for smoking cessation). In such cases, it is likely to be difficult to identify the four user groups in the framework, as individuals in all groups may have experience of the intervention.

It is also the case that the DCE design, in particular the identification of attributes and levels, did not involve any qualitative research with individuals from any of the four user groups. This is a limitation because it means that, although the attributes and levels were informed by a relevant systematic review and expert consultation, the DCE design was ultimately researcher-led. It is therefore possible that there were attributes and/or levels that would have been important to respondents that were not included in the DCE design, which could have influenced the choices that were made during the task. That being said, as explained in section 6.4.1, it could have been challenging to incorporate qualitative feedback in this study given that four user groups with different levels of experience were to be recruited and a single DCE design was required in order to make reasonable comparisons. Nonetheless, the importance of qualitative research to inform DCE design has been stressed recently in the literature (Vass, Rigby et al., 2017) and should be an important consideration for all future DCE studies.

Another limitation is that, whilst Chapter Three highlighted the relevance of the preference debate in health economics, the empirical work did not make a typical 'patient vs. general public' comparison. That is, a random, fully representative sample of the UK general population was not recruited as part of this study. Instead, purposive sampling was used for the online panel to ensure that at least 200 respondents were recruited for the PSU, PB and NU groups. Had the online panel sample been random and fully representative (i.e. a general public sample), it might not have been varied enough to obtain sample sizes that were large enough for each of these groups. However, despite this limitation, there is an increasing emphasis on the consideration of patient preferences in (largely regulatory) decision-making in both the literature (Mühlbacher, 2015; Johnson & Zhou, 2016; Mühlbacher et al., 2016), and in guidance from decision-makers (FDA, 2016; FDA, 2017), which has increased throughout the duration of this work. It is therefore highly likely that comparisons between different types of user group would be of greater interest amongst researchers and policymakers than further 'patient vs. general public' comparisons. Another limitation is that the empirical work in this thesis cannot determine whether differences in preferences are due to differences in experience or differences in the recruitment vehicle (i.e. trial vs. online panel). However, preferences do appear to differ between the three groups recruited via the online panel, providing some evidence to suggest that not all individuals in the online panel are alike. Whilst this is a significant empirical challenge to overcome, section 10.2.4 provided a suggestion as to how future work might overcome this limitation.

Finally, another limitation that applies to the entirety of the empirical work is that the sample sizes could have been larger, allowing more detailed comparisons to be made. However, there was a clear upper limit for the SU group due to the size of the NULevel trial, and larger sample sizes for the online panel would not have made up for this. Given the data requirements for choice modelling, future studies would be well advised to recruit large numbers of respondents in every group of interest when using a DCE (e.g. several hundred), especially if comparisons across samples or subgroups are to be conducted.

10.5 Future Research

10.5.1 Stated Preference Studies

This thesis has provided an insight into how preferences may differ in stated preference studies when the samples recruited differ in relation to their prior experiences of the condition, treatment of the condition, and also in the survey recruitment process. However, the insight provided by this work is not sufficient to draw confident conclusions about what may happen in other contexts. As a result, further research would be beneficial. In particular,

it would be useful to explore whether the differences in WTP estimates and rationality test pass rates between patients in clinical trials and potential patients from online panels that are observed in this study are observed in other contexts (e.g. different types of public health and healthcare interventions). Existing literature does seem to provide some support that WTP estimates are likely to differ between groups (Najafzadeh et al., 2013; Finkelstein et al. 2015), but little research exists in relation to differences in rationality test pass rates. It would also be interesting to examine whether the minor differences observed between the three groups recruited via the online panel would be exacerbated in another context. One might expect to see greater differences in a more clinical context where attributes may relate to adverse events and where process attributes may be less well understood by the average individual. For example, a process attribute for an anticoagulant therapy might be "availability of an antidote in the case of a bleed", whereas one of the process attributes used in the empirical work in this thesis was "availability of a website or mobile phone application (for tracking weight)". This would be particularly important with respect to regulatory decision-making as this focuses on benefit-risk trade-offs, which may be difficult to understand without any prior experience of the health issue or its treatment.

Another avenue for further research would be to examine the external validity of DCEs, which to date has only been examined in a small number of studies (Clark et al., 2014). WLM interventions would be better suited to this than many other healthcare services, given that in future they could realistically be purchased and consumed on the free market (as opposed to being provided via the NHS), much like commercial weight loss programmes. In such a study it may be useful to explore whether, and to what extent, experience affects external validity. This could help to improve the accuracy of future DCE studies.

10.5.2 HTA Decision-Making

In practice, it might be unlikely for HTA decisions to be made based solely on the relative importance of a small range of attributes included in a DCE. Instead, results from DCEs might be used to provide additional information for use in formal economic evaluations, such as those outlined in table 3.1. Therefore, future research that looks at how preferences differ between groups could take a broader approach and examine the potential knock-on effects on economic evaluations resulting from the findings of preference studies. The key question is: could the differences in preferences potentially lead to different conclusions being drawn from an economic evaluation depending on whose preferences are considered? One can speculate that the differences observed in WTP estimates derived from the DCE conducted in this project would not have led to drastically different conclusions being drawn from an

economic evaluation, given the small magnitude of the differences. However, this might not be the case in other studies and other contexts.

The question of whether relative experience (of a health issue and/or its treatment) affects preferences is likely to become increasingly important if calls for a more patient-centric HTA process (Mühlbacher, 2015) are taken seriously by decision-makers. That being said, the current emphasis on using CUA might present stumbling block for meeting such calls from a pragmatic perspective. This is because there are only limited opportunities to incorporate patient preferences within a QALY framework advocated by agencies such as NICE, with perhaps the obvious option (using health state utilities derived from patients) proving to be contentious in some countries such as the UK, as discussed in Chapter Three. Furthermore, as discussed in Chapter Three, it has been suggested for some time that DCEs can be used to incorporate the patient voice into CBA through WTP estimates derived from stated preference studies (McIntosh, 2006). The use of CBA may be particularly appropriate for public health interventions, as the potential wider societal effects of such interventions might warrant a broader economic evaluation. However, estimating WTP in a health context is likely to be contentious because of resistance to putting a monetary value on life (Zweifel et al., 2009).

10.5.3 Regulatory Decision-Making

Given the issues outlined in the previous section, it perhaps makes sense that the incorporation of patient preferences into healthcare decision-making appears to be more advanced in a regulatory context. In the regulatory space trade-offs between risk and benefit have been required for some time, and these could potentially be delivered using stated preference studies (Ho et al., 2015). The implementation of new legislation in the USA over the past two years has highlighted the importance of considering the patient experience during the drug development process. The legislation has led to the FDA producing a plan for issuance of "patient-focused drug development guidance" (FDA, 2017), which will lead to a series of new guidance documentation being finalised between 2019 and 2021. It would seem likely that this guidance will lead to a rise in the collection and incorporation of patient preference data in regulatory decision-making in the US and, as outlined in section 3.4, there is a similar focus in Europe.

A recent editorial has considered the current practice with regard to using DCEs for informing benefit-risk assessments and has raised the issue of whose preferences are most suitable to elicit, as well as highlighting the issue of preference heterogeneity (Vass & Payne, 2017). Whilst the DCE used in the empirical work of this thesis did not contain a risk attribute, it may well be possible that benefit-risk trade-offs would differ depending on the make-up of

respondent samples. As quantitative preference information elicited from patients, using the likes of DCEs, are increasingly being incorporated into regulatory decision-making, it is important for future studies to explore how patients with differing levels of experience might have different preferences in relation to benefit-risk trade-offs.

10.5.4 Multi-Criteria Decision-Analysis

In addition, future research could extend beyond the current status quo in healthcare decision-making. It has recently been highlighted that there is potential for DCEs to be used in healthcare decision-making within a multi-criteria decision-analysis (MCDA) framework (Thokala et al., 2016; Marsh, Goetghebeur et al., 2017). MCDA has been put forward as a potential alternative to the likes of CUA, CBA or traditional risk-benefit assessment. MCDA is an umbrella term representing a highly flexible set of techniques used to assist in decision-making in any potential context. Broadly speaking, MCDA requires a set of criteria that are critical to the decision problem at hand, which are scored based on the available evidence and weighted according to stakeholder preferences. The weights may be derived from patient preferences using a DCE, although other techniques may also be used (Tervonen et al., 2017).

Whether MCDA is likely to be acceptable to decision-makers is another story. The NICE reference case is very specific and aims to maximise comparability between economic evaluations by recommending, for example, a single instrument for generating HSUVs and so forth. It might be the case that agencies such as NICE would wish to employ MCDA in a similar manner. This might involve: the use of a single set of (static) criteria; a standardised approach to scoring; and a single set of criteria weights. The latter would make it very difficult to incorporate patient preferences, as the patient population would differ across economic evaluations. There are also numerous methodological concerns regarding the use of MCDA for HTA that would need to be resolved before this could become a reality (Marsh, Sculpher et al., 2017).

Nonetheless, there is an apparent appetite amongst some researchers in health economics to explore how MCDA might be used for HTA (Marsh, Goetghebeur et al., 2017). If this continues, future research may wish to explore how using different types of respondents to weight the criteria might result in different decisions in a MCDA framework.

10.6 Conclusion

This thesis began by reviewing the relevant economic theory, the use of quantitative preference information in health economics and the ‘preference debate’. This was followed by a systematic review that explored the question of whose preferences are typically elicited in

published DCE studies. Following this, a framework was set out that could be used to classify survey respondents in relation to their experience of a healthcare intervention and/or a health issue. The empirical work tested whether preferences differed between the four different user groups defined within the framework. The results suggest that preferences differ, to varying extents, between the four user groups. The most pronounced differences were between the group recruited via an RCT relative to the three groups recruited from an online panel with respect to the DCE results and the WTP estimates derived from them. No differences in WTP estimates from the CV task between the four groups were observed, though this may have been due to data limitations. In addition, there is evidence to suggest that more experienced individuals are more likely to pass rationality tests.

The findings of this thesis will be particularly useful for researchers that design preference studies using the likes of DCEs. At present, there are very few studies that explore how preferences differ according to the level of experience of the respondents. In addition, there is an increasing interest in the use of patient preferences studies, particularly in the regulatory space. An increase in demand for patient preference studies for use in decision-making will mean that it will become increasingly important to understand how preferences might differ between different types of respondent for both researchers and policymakers. It would seem that experienced respondents (i.e. the SU group in this study) are likely to provide better quality data, but express different preferences to less experienced respondents.

Future research could build upon the empirical work conducted in this thesis. There are many other contexts in which a preference study could be conducted, and different recruitment modes for study participants can provide unique opportunities to explore how preferences differ. In addition, future studies may wish to explore how these differences in preferences could affect the outcomes of HTAs and BRAs. Following on from this, if researchers continue to push for its increased application in healthcare decision-making, researchers might wish to explore this question in the context of MCDA.

References

Al-Janabi, H., Flynn, T.N. & Coast, J. (2012) Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Quality of Life Research*. 21 (1), 167–176.

Allison, D.B., Basile, V.C. & Yuker, H.E. (1991) The measurement of attitudes toward and beliefs about obese persons. *The International Journal of Eating Disorders*. 10 (5), 599–607.

Arrow, K., Solow, R., Portney, P.R., Leamer, E.E., Radner, R. & Schuman, H. (1993) Report of the NOAA Panel on Contingent Valuation. *Federal Register*. 58, 4601-4614.

Avenell, A., Broom, J., Brown, T.J., Poobalan, A., Aucott, L., Stearns, S.C., Smith, W.C.S., Jung, R.T., Campbell, M.K. & Grant, A.M. (2004) Systematic review of the long-term effects and economic consequences of treatments for obesity and implications for health improvement. *Health Technology Assessment*. 8 (21), iii–iv, 1–182.

Barry, M.J. & Edgman-Levitan, S. (2012) Shared decision making—the pinnacle of patient-centered care. *The New England Journal of Medicine*. 366 (9), 780–781.

de Bekker-Grob, E.W., Berlin, C., Levitan, B., Raza, K., Christoforidi, K., Cleemput, I., Pelouchova, J., Enzmann, H., Cook, N. & Hansson, M.G. (2017) Giving patients' preferences a voice in medical treatment life cycle: the PREFER public-private project. *The Patient*. 10 (3), 263–266.

de Bekker-Grob, E.W., Donkers, B., Jonker, M.F. & Stolk, E.A. (2015) Sample size requirements for discrete-choice experiments in healthcare: a practical guide. *The Patient*. 8 (5), 373–384.

de Bekker-Grob, E.W., Ryan, M. & Gerard, K. (2012) Discrete choice experiments in health economics: a review of the literature. *Health Economics*. 21 (2), 145–172.

Benning, T.M., Kimman, M.L., Dirksen, C.D., Boersma, L.J. & Dellaert, B.G.C. (2012) Combining individual-level discrete choice experiment estimates and costs to inform health care management decisions about customized care: the case of follow-up strategies after

breast cancer treatment. *Value in Health*. 15 (5), 680–689.

Blanden, J. & Machin, S. (2004) Educational inequality and the expansion of UK higher education. *Scottish Journal of Political Economy*. 51 (2), 230–249.

Boonen, L.H.H.M., Donkers, B. & Schut, F.T. (2011) Channeling consumers to preferred providers and the impact of status quo bias: does type of provider matter? *Health Services Research*. 46 (2), 510–530.

Böttger, B., Thate-Waschke, I.M., Bauersachs, R., Kohlmann, T. & Wilke, T. (2015) Preferences for anticoagulation therapy in atrial fibrillation: The Patients' view. *Journal of Thrombosis and Thrombolysis*. 40 (4), 406–415.

Boyd, N.F., Sutherland, H.J., Heasman, K.Z., Tritchler, D.L. & Cummings, B.J. (1990) Whose utilities for decision analysis? *Medical Decision Making*. 10 (1), 58–67.

Boyers, D., van der Pol, M., & Watson, V. (2016) Using DCEs to estimate willingness to pay for dental care. *European Health Economics Association Conference, 'Know the Ropes - Balancing Costs and Quality in Health Care'*. Germany: Universität Hamburg. 13-16 July 2016.

Boyers, D. (Forthcoming) *[Working Title] External validity of DCEs: a case study of dental care*. PhD Thesis: University of Aberdeen.

Brazier, J., Deverill, M. & Green, C. (1999) A review of the use of health status measures in economic evaluation. *Journal of Health Services Research & Policy*. 4 (3), 174–184.

Brazier, J., Rowen, D., Karimi, M., Peasgood, T., Tsuchiya, A. & Ratcliffe, J. (2017) Experience-based utility and own health state valuation for a health state classification system: why and how to do it. *The European Journal of Health Economics*.

Brazier, J., Ratcliffe, J., Salomon, J. & Tsuchiya, A. (2016) *Measuring and Valuing Health Benefits for Economic Evaluation*. 2nd Edition. Oxford: Oxford University Press.

Brazier, J., Akehurst, R., Brennan, A., Dolan, P., Claxton, K., McCabe, C., Sculpher, M. & Tsuchyia, A. (2005) Should patients have a greater role in valuing health states? *Applied Health Economics and Health Policy*. 4 (4), 201–208.

Bridges, J.F.P., Hauber, A.B., Marshall, D., Lloyd, A., Prosser, L.A., Regier, D.A., Johnson, F.R. & Mauskopf, J. (2011) Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value in Health.* 14 (4), 403–413.

Brown, D.S., Poulos, C., Johnson, F.R., Chamiec-Case, L. & Messonnier, M.L. (2014) Adolescent girls' preferences for HPV vaccines: a discrete choice experiment. *Advances in Health Economics and Health Services Research*. 2493–121.

Bryan, S. & Dolan, P. (2004) Discrete choice experiments in Health Economics. For better or for worse? *The European Journal of Health Economics*. 5 (3), 199–202.

Buchanan, J. (2015) *Issues related to the economic analysis of genomic diagnostic technologies in the UK National Health Service: an exploration of methods.* PhD Thesis: University of Oxford.

Buchanan, J., Wordsworth, S. & Schuh, A. (2016) Patients' Preferences for Genomic Diagnostic Testing in Chronic Lymphocytic Leukaemia: A Discrete Choice Experiment. *The Patient*. 9 (6), 525–536.

Buchanan, J. & Wordsworth, S. (2015) Welfarism versus extra-welfarism: can the choice of economic evaluation approach impact on the adoption decisions recommended by economic evaluation studies? *PharmacoEconomics*. 33 (6), 571–579.

Burr, J.M., Botello-Pinzon, P., Takwoingi, Y., Hernández, R., Vazquez-Montes, M., Elders, A., Asaoka, R., Banister, K., van der Schoot, J., Fraser, C., King, A., Lemij, H., Sanders, R., Vernon, S., Tuulonen, A., Kotecha, A., Glasziou, P., Garway-Heath, D., Crabb, D., et al. (2012) Surveillance for ocular hypertension: an evidence synthesis and economic evaluation. *Health Technology Assessment*. 16 (29), 1–271, iii–iv.

Burr, J.M., Kilonzo, M., Vale, L. & Ryan, M. (2007) Developing a preference-based

Glaucoma Utility Index using a discrete choice experiment. *Optometry and Vision Science*. 84 (8), 797–808.

Burr, J.M., Mowatt, G., Hernández, R., Siddiqui, M.A.R., Cook, J., Lourenco, T., Ramsay, C., Vale, L., Fraser, C., Azuara-Blanco, A., Deeks, J., Cairns, J., Wormald, R., McPherson, S., Rabindranath, K. & Grant, A. (2007) The clinical effectiveness and cost-effectiveness of screening for open angle glaucoma: a systematic review and economic evaluation. *Health Technology Assessment*. 11 (41), iii–iv, ix–x, 1–190.

CADTH (Canadian Agency for Drugs and Technologies in Health) (2017) *Guidelines for the Economic Evaluation of Health Technologies: Canada.* 4th Edition. Canada: CADTH.

Clark, M.D., Determann, D., Petrou, S., Moro, D. & de Bekker-Grob, E.W. (2014) Discrete choice experiments in Health Economics: a review of the literature. *PharmacoEconomics*. 32 (9), 883–902.

Coast, J., Smith, R.D. & Lorgelly, P. (2008) Welfarism, extra-welfarism and capability: the spread of ideas in Health Economics. *Social Science and Medicine*. 67 (7), 1190–1198.

Cookson, R. (2003) Willingness to pay methods in health care: a sceptical view. *Health Economics*. 12 (11), 891–894.

Crowne, D.P. & Marlowe, D. (1960) A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*. 24349–354.

Danyliv, A., Pavlova, M., Gryga, I. & Groot, W. (2012) Willingness to pay for physician services: comparing estimates from a discrete choice experiment and contingent valuation. *Society and Economy*. 34 (2), 339–357.

Devlin, N.J., Shah, K.K., Feng, Y., Mulhern, B. & van Hout, B. (2017) Valuing health-related quality of life: an EQ-5D-5L value set for England. *Health Economics*.

Diener, A., O'Brien, B. & Gafni, A. (1998) Health care contingent valuation studies: a review and classification of the literature. *Health Economics*. 7 (4), 313–326.

Dolan, P. (2008) Developing methods that really do value the ‘Q’ in the QALY. *Health Economics, Policy, and Law*. 3 (Pt 1), 69–77.

Dolan, P. (1999) Whose preferences count? *Medical Decision Making*. 19 (4), 482–486.

Dolan, P. & Kahneman, D. (2008) Interpretations of utility and their implications for the valuation of health. *The Economic Journal*. 118 (525), 215–234.

Dolan, P., Olsen, J.A., Menzel, P. & Richardson, J. (2003) An inquiry into the different perspectives that can be used when eliciting preferences in health. *Health Economics*. 12 (7), 545–551.

Dombrowski, S.U., Knittle, K., Avenell, A., Araújo-Soares, V. & Sniehotta, F.F. (2014) Long term maintenance of weight loss with non-surgical interventions in obese adults: systematic review and meta-analyses of randomised controlled trials. *BMJ*. 348g2646.

Drummond, M.F., Sculpher, M.J., Claxton, K., Stoddart, G.L. & Torrance, G.W. (2015) *Methods for the economic evaluation of health care programmes*. 4th Edition. Oxford: Oxford University Press.

Estrin, S., Dietrich, M. & Laidler, D.E.W. (2012) *Microeconomics*. 6th Edition. Harlow: Pearson Education.

Evans, E.H., Araújo-Soares, V., Adamson, A., Batterham, A.M., Brown, H., Campbell, M., Dombrowski, S.U., Guest, A., Jackson, D., Kwasnicka, D., Ladha, K., McColl, E., Olivier, P., Rothman, A.J., Sainsbury, K., Steel, A.J., Steen, I.N., Vale, L., White, M., et al. (2015) The NULevel trial of a scalable, technology-assisted weight loss maintenance intervention for obese adults after clinically significant weight loss: study protocol for a randomised controlled trial. *Trials*. 16 (1), 421.

Facey, K.M., Hansen, H.P. & Single, A. (eds.) (2017) *Patient Involvement in Health Technology Assessment*. Singapore: Springer.

FDA (US Food and Drug Administration) (2017) *FDA Plan for Issuance of Patient-Focused Drug Development Guidance Under 21st Century Cures Act Title III Section 3002*. Maryland:

FDA.

FDA (US Food and Drug Administration) (2016) *Patient Preference Information – Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling: Guidance for Industry Food and Drug Administration Staff, and Other Stakeholders*. Maryland: FDA.

Finkelstein, E.A., Bilger, M., Flynn, T.N. & Malhotra, C. (2015) Preferences for end-of-life care among community-dwelling older adults and patients with advanced cancer: A discrete choice experiment. *Health Policy.* 119 (11), 1482–1489.

Flynn, T.N., Huynh, E., Peters, T.J., Al-Janabi, H., Clemens, S., Moody, A. & Coast, J. (2015) Scoring the ICECAP-A capability instrument. Estimation of a UK general population tariff. *Health Economics*. 24 (3), 258–269.

Frew, E.J., Wolstenholme, J.L. & Whynes, D.K. (2004) Comparing willingness-to-pay: bidding game format versus open-ended and payment scale formats. *Health Policy.* 68 (3), 289–298.

Froberg, D.G. & Kane, R.L. (1989) Methodology for measuring health-state preferences—III: Population and context effects. *Journal of Clinical Epidemiology*. 42 (6), 585–592.

Gandjour, A. (2010) Theoretical foundation of patient v. population preferences in calculating QALYs. *Medical Decision Making*. 30 (4), E57–63.

Garcia-Dominguez, J.M., Muñoz, D., Comellas, M., Gonzalbo, I., Lizán, L. & Polanco Sánchez, C. (2016) Patient preferences for treatment of multiple sclerosis with disease-modifying therapies: a discrete choice experiment. *Patient Preference and Adherence.* 101945–1956.

Gerard, K., Salisbury, C., Street, D., Pope, C. & Baxter, H. (2008) Is fast access to general practice all that should matter? A discrete choice experiment of patients' preferences. *Journal of Health Services Research & Policy*. 13 Suppl 23–10.

Germain, D.P. (2010) Fabry disease. *Orphanet Journal of Rare Diseases*. 530.

Ghijben, P., Lancsar, E. & Zavarsek, S. (2014) Preferences for oral anticoagulants in atrial fibrillation: a best-best discrete choice experiment. *PharmacoEconomics*. 32 (11), 1115–1127.

Giles, E.L., Becker, F., Ternent, L., Sniehotta, F.F., McColl, E. & Adams, J. (2016) Acceptability of financial incentives for health behaviours: a discrete choice experiment. *PloS One*. 11 (6), e0157403.

Gold, M.R., Siegel, J.E., Russel, L.B. & Weinstein, M.C. (eds.) (1996) *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press.

Gravelle, H. & Rees, R. (2004) *Microeconomics*. 3rd Edition. Harlow: Pearson Education.

Green, C., Brazier, J. & Deverill, M. (2000) Valuing health-related quality of life. A review of health state valuation techniques. *PharmacoEconomics*. 17 (2), 151–165.

Gujarati, D.N. & Porter, D.C. (2009) *Basic Econometrics*. 5th Edition. New York: McGraw-Hill Irwin.

Hadorn, D.C. (1991) The role of public values in setting health care priorities. *Social Science and Medicine*. 32 (7), 773–781.

Hauber, A.B., González, J.M., Groothuis-Oudshoorn, C.G.M., Prior, T., Marshall, D.A., Cunningham, C., IJzerman, M.J. & Bridges, J.F.P. (2016) Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value in Health*. 19 (4), 300–315.

Herdman, M., Gudex, C., Lloyd, A., Janssen, M.F., Kind, P., Parkin, D., Bonsel, G. & Badia, X. (2011) Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*. 20 (10), 1727–1736.

Hess, S. & Train, K. (2017) Correlation and scale in mixed logit models. *Journal of Choice Modelling*. 23 (Supplement C), 1–8.

H.M. Treasury. (2011) *The Green Book: Appraisal and Evaluation in Central Government*. London: H.M. Treasury.

Ho, M.P., Gonzalez, J.M., Lerner, H.P., Neuland, C.Y., Whang, J.M., McMurry-Heath, M., Hauber, A.B. & Irony, T. (2015) Incorporating patient-preference evidence into regulatory decision making. *Surgical Endoscopy*. 29 (10), 2984–2993.

Hole, A.R. (2008) Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. *Journal of Health Economics*. 27 (4), 1078–1094.

Hole, A.R. (2007a) Fitting mixed logit models by using maximum simulated likelihood. *The Stata Journal*. 7 (3), 388–401

Hole, A.R. (2007b) *WTP: Stata module to estimate confidence intervals for willingness to pay measures*. Available at: https://econpapers.repec.org/software/bocbocode/s456808.htm [Accessed 11th December 2017].

Hughes, D., Charles, J., Dawoud, D., Edwards, R.T., Holmes, E., Jones, C., Parham, P., Plumpton, C., Ridyard, C., Lloyd-Williams, H., Wood, E. & Yeo, S.T. (2016) Conducting economic evaluations alongside randomised trials: current methodological issues and novel approaches. *PharmacoEconomics*. 34 (5), 447–461.

Hughes, D., Waddingham, E., Mt-Isa, S., Goginsky, A., Chan, E., Downey, G.F., Hallgreen, C.E., Hockley, K.S., Juhaeri, J., Lieftucht, A., Metcalf, M.A., Noel, R.A., Phillips, L.D., Ashby, D., Micaleff, A. & PROTECT Benefit-Risk Group (2016) Recommendations for benefit-risk assessment methodologies and visual representations. *Pharmacoepidemiology and Drug Safety*. 25 (3), 251–262.

Hughes, D., Waddingham, E.A.J., Mt-Isa, S., Goginsky, A., Chan, E., Downey, G., Hallgreen, C.E., Hockley, K.S., Juhaeri, J., Lieftucht, A., Metcalfe, M.A., Noel, R.A., Phillips, L., Ashby, D. & Micaleff, A. & PROTECT Benefit-Risk Group (2013) *IMI-PROTECT Benefit-Risk Group Recommendations for the methodology and visualisation techniques to be used in the assessment of benefit and risk of medicines*. Available at: http://www.imi-protect.eu/documents/HughesetalRecommendationsforthemethodologyandvisualisationtechni

questobeusedintheassessmento.pdf [Accessed 11th December 2017]

Irony, T., Ho, M., Christopher, S. & Levitan, B. (2016) Incorporating Patient Preferences into Medical Device Benefit-Risk Assessments. *Statistics in Biopharmaceutical Research*. 8 (3), 230–236.

Jiang, C.C. & Fraenkel, L. (2017) The influence of varying cost formats on preferences. *Medical Decision Making*. 37 (1), 17–26.

Johannesson, M. (1996) The willingness to pay for health changes, the human-capital approach and the external costs. *Health Policy*. 36 (3), 231–244.

Johannesson, M., Jönsson, B. & Borgquist, L. (1991) Willingness to pay for antihypertensive therapy--results of a Swedish pilot study. *Journal of Health Economics*. 10 (4), 461–473.

Johnson, F.R. & Zhou, M. (2016) Patient Preferences in Regulatory Benefit-Risk Assessments: a US Perspective. *Value in Health*. 19 (6), 741–745.

Johnson, R. & Orme, B. (2003) Getting the most from CBC. *Sawtooth Software Research Paper Series.* Washington: Sawtooth Software.

Jonnalagadda, S.S., Skinner, R. & Moore, L. (2004) Overweight athlete: fact or fiction? *Current Sports Medicine Reports*. 3 (4), 198–205.

Kahneman, D. (2006) Determinants of health economic decisions in actual practice: the role of behavioral economics. *Value in Health*. 9 (2), 65–67.

Kaltoft, M.K., Neilsen, J.B., Salkeld, G. & Dowie, J. (2015) Can a discrete choice experiment contribute to person-centred healthcare? *European Journal for Person Centered Healthcare*. 3 (4), 431–437.

King, G. (2013) *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. New Jersey: Princeton University Press.

Kistler, C.E., Hess, T.M., Howard, K., Pignone, M.P., Crutchfield, T.M., Hawley, S.T.,

Brenner, A.T., Ward, K.T. & Lewis, C.L. (2015) Older adults' preferences for colorectal cancer-screening test attributes and test choice. *Patient Preference and Adherence*. 91005–1016.

Klose, T. (1999) The contingent valuation method in health care. *Health Policy*. 47 (2), 97–123.

Lancaster, K.J. (1966) A new approach to consumer theory. *The Journal of Political Economy*. 74 (2), 132–157.

Lancsar, E., Fiebig, D.G. & Hole, A.R. (2017) Discrete choice experiments: a guide to model specification, estimation and software. *PharmacoEconomics*. 35 (7), 697–716.

Lancsar, E. & Louviere, J. (2008) Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *PharmacoEconomics*. 26 (8), 661–677.

Landfeldt, E., Eriksson, J., Ireland, S., Musingarimi, P., Jackson, C., Tweats, E. & Gaudig, M. (2016) Patient, physician, and general population preferences for treatment characteristics in relapsed or refractory chronic lymphocytic leukemia: A conjoint analysis. *Leukemia Research*. 4017–23.

Latner, J.D., O'Brien, K.S., Durso, L.E., Brinkman, L.A. & MacDonald, T. (2008) Weighing obesity stigma: the relative strength of different forms of bias. *International Journal of Obesity*. 32 (7), 1145–1152.

Levitan, B., Hauber, A.B., Damiano, M.G., Jaffe, R. & Christopher, S. (2017) The ball is in your court: agenda for research to advance the science of patient preferences in the regulatory review of medical devices in the United States. *The Patient*. 10 (5), 531–536.

Little, M.H.R., Reitmeir, P., Peters, A. & Leidl, R. (2014) The impact of differences between patient and general population EQ-5D-3L values on the mean tariff scores of different patient groups. *Value in Health*. 17 (4), 364–371.

Llewellyn-Thomas, H., Sutherland, H.J., Tibshirani, R., Ciampi, A., Till, J.E. & Boyd, N.F. (1984) Describing health states. Methodologic issues in obtaining values for health states.

Medical Care. 22 (6), 543–552.

Lloyd, A.J., Gallop, K., Ali, S., Hughes, D. & MacCulloch, A. (2017) Social preference weights for treatments in Fabry disease in the UK: a discrete choice experiment. *Current Medical Research and Opinion*. 33 (1), 23–29.

Louviere, J.J., Flynn, T.N. & Carson, R.T. (2010) Discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling*. 3 (3), 57–72.

Louviere, J.J. & Lancsar, E. (2009) Choice experiments in health: the good, the bad, the ugly and toward a brighter future. *Health Economics, Policy, and Law*. 4 (Pt 4), 527–546.

Mahieu, P-A., Andersson, H., Beaumais, O., Crastes, R. & Wolff, F-C. (2014) *Is choice experiment becoming more popular than contingent valuation? A systematic review in agriculture, environment and health*. French Association of Environmental and Resource Economists (FAERE) Working Papers. 2014.12.

Mangham, L.J. & Hanson, K. (2008) Employment preferences of public sector nurses in Malawi: results from a discrete choice experiment. *Tropical Medicine & International Health*. 13 (12), 1433–1441.

Marsh, K., Goetghebeur, M., Thokala, P. & Baltussen, R. (eds.) (2017) *Multi-Criteria Decision Analysis to Support Healthcare Decisions*. New York: Springer.

Marsh, K., Sculpher, M., Caro, J.J. & Tervonen, T. (2017) The Use of MCDA in HTA: Great Potential, but More Effort Needed. *Value in Health*. In Press, doi:10.1016/j.jval.2017.10.001.

Marshall, H.S., Chen, G., Clarke, M. & Ratcliffe, J. (2016) Adolescent, parent and societal preferences and willingness to pay for meningococcal B vaccine: A discrete choice experiment. *Vaccine*. 34 (5), 671–677.

McAteer, A., Yi, D., Watson, V., Norwood, P., Ryan, M., Hannaford, P.C. & Elliott, A.M. (2015) Exploring preferences for symptom management in primary care: a discrete choice experiment using a questionnaire survey. *The British Journal of General Practice*. 65 (636), e478–88.

McCormack, K., Wake, B., Perez, J., Fraser, C., Cook, J., McIntosh, E., Vale, L. & Grant, A. (2005) Laparoscopic surgery for inguinal hernia repair: systematic review of effectiveness and economic evaluation. *Health Technology Assessment*. 9 (14), 1–203, iii–iv.

McFadden, D. (1973) 'Conditional logit analysis of qualitative choice behavior', in P. Zarembka (ed.) *Frontiers in Econometrics*. New York: Wiley. pp. 105–135.

McIntosh, E., Clarke, P., Frew, E. & Louviere, J. (2010) *Applied Methods of Cost-Benefit Analysis in Health Care*. Oxford: Oxford University Press.

McIntosh, E. (2006) Using discrete choice experiments within a cost-benefit analysis framework: some considerations. *PharmacoEconomics*. 24 (9), 855–868.

McNamee, P., Ternent, L., Gbangou, A. & Newlands, D. (2010) A game of two halves? Incentive incompatibility, starting point bias and the bidding game contingent valuation method. *Health Economics*. 19 (1), 75–87.

Miguel, F.S., Ryan, M. & Amaya-Amaya, M. (2005) 'Irrational' stated preferences: a quantitative and qualitative investigation. *Health Economics*. 14 (3), 307–322.

Mitchell, P.M., Venkatapuram, S., Richardson, J., Iezzi, A. & Coast, J. (2017) Are quality-adjusted life years a good proxy measure of individual capabilities? *PharmacoEconomics*. 35 (6), 637–646.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G. & PRISMA Group (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 339b2535.

Mott, D.J. & Najafzadeh, M. (2016) Whose preferences should be elicited for use in health-care decision-making? A case study using anticoagulant therapy. *Expert Review of Pharmacoeconomics and Outcomes Research*. 16 (1), 33–39.

Mühlbacher, A.C. (2015) Patient-centric HTA: different strokes for different folks. *Expert Review of Pharmacoeconomics and Outcomes Research*. 15 (4), 591–597.

Mühlbacher, A.C., Bridges, J.F.P., Bethge, S., Dintsios, C.-M., Schwalm, A., Gerber-Grote, A. & Nübling, M. (2017) Preferences for antiviral therapy of chronic hepatitis C: a discrete choice experiment. *The European Journal of Health Economics*. 18 (2), 155–165.

Mühlbacher, A.C. & Juhnke, C. (2013) Patient preferences versus physicians' judgement: does it make a difference in healthcare decision making? *Applied Health Economics and Health Policy*. 11 (3), 163–180.

Mühlbacher, A.C., Juhnke, C., Beyer, A.R. & Garner, S. (2016) Patient-focused benefit-risk analysis to inform regulatory decisions: the European Union perspective. *Value in Health*. 19 (6), 734–740.

Najafzadeh, M., Gagne, J.J., Choudhry, N.K., Polinski, J.M., Avorn, J. & Schneeweiss, S.S. (2014) Patients' preferences in anticoagulant therapy: discrete choice experiment. *Circulation: Cardiovascular Quality and Outcomes*. 7 (6), 912–919.

Najafzadeh, M., Johnston, K.M., Peacock, S.J., Connors, J.M., Marra, M.A., Lynd, L.D. & Marra, C.A. (2013) Genomic testing to determine drug response: measuring preferences of the public and patients using discrete choice experiment (DCE). *BMC Health Services Research*. 13454.

Neumann, P.J., Sanders, G.D., Russell, L.B., Siegel, J.E. & Ganiats, T.G. (eds.) (2016) *Cost-Effectiveness in Health and Medicine*. 2nd Edition. New York: Oxford University Press.

NICE (National Institute for Health and Care Excellence) (2014) *Developing NICE guidelines: the manual*. London: NICE.

NICE (National Institute for Health and Care Excellence) (2013) *Guide to the methods of technology appraisal*. London: NICE.

Obse, A., Ryan, M., Heidenreich, S., Normand, C. & Hailemariam, D. (2016) Eliciting preferences for social health insurance in Ethiopia: a discrete choice experiment. *Health Policy and Planning*. 31 (10), 1423–1432.

OECD (The Organisation for Economic Co-operation and Development) (2017) *Health at a Glance 2017: OECD Indicators*. Paris: OECD Publishing. doi:10.1787/health_glance-2017-en

Okumura, K., Inoue, H., Yasaka, M., Gonzalez, J.M., Hauber, A.B., Levitan, B., Yuan, Z. & Briere, J.-B. (2015) Comparing patient and physician risk tolerance for bleeding events associated with anticoagulants in atrial fibrillation—evidence from the United States and Japan. *Value in Health Regional Issues*. 6 (Supplement C), 65–72.

Oldridge, N., Furlong, W., Perkins, A., Feeny, D. & Torrance, G.W. (2008) Community or patient preferences for cost-effectiveness of cardiac rehabilitation: does it matter? *European Journal of Cardiovascular Prevention and Rehabilitation*. 15 (5), 608–615.

Olsen, J.A. & Smith, R.D. (2001) Theory versus practice: a review of 'willingness-to-pay' in health and health care. *Health Economics*. 10 (1), 39–52.

Papageorgiou, K., Vermeulen, K.M., Schroevers, M.J., Stiggelbout, A.M., Buskens, E., Krabbe, P.F.M., van den Heuvel, E. & Ranchor, A.V. (2015) Do individuals with and without depression value depression differently? And if so, why? *Quality of Life Research*. 24 (11), 2565–2575.

Pedersen, L.B., Bech, M., Kløgaard, M., Søgaard, R., Veldwijk, J., de Wit, G.A. & Lambooij, M. (2015) How discrete choice experiments contribute to person-centered healthcare. A commentary on Kaltoft et al. (2015): "Can a discrete choice experiment contribute to person-centred healthcare?" *European Journal for Person Centered Healthcare*. 3 (4), 438–441.

Peeters, Y. & Stiggelbout, A.M. (2010) Health state valuations of patients and the general public analytically compared: a meta-analytical comparison of patient and population health state utilities. *Value in Health*. 13 (2), 306–309.

Pearmain, D., Swanson, J., Kroes, E., & Bradley, M. (1991) *Stated preference techniques: a guide to practice.* 2nd Edition. London: Steer Davies Gleave and Hague Consulting Group.

Pharmaceutical Benefits Board. (2003) *General guidelines for economic evaluations from the Pharmaceutical Benefits Board*. Sweden: Pharmaceutical Benefits Board.

van der Pol, M., Shiell, A., Au, F., Johnston, D. & Tough, S. (2008) Convergent validity between a discrete choice experiment and a direct, open-ended method: comparison of preferred attribute levels and willingness to pay estimates. *Social Science and Medicine*. 67 (12), 2043–2050.

van der Pol, M., Shiell, A., Au, F., Jonhston, D. & Tough, S. (2010) Eliciting individual preferences for health care: a case study of perinatal care. *Health Expectations*. 13 (1), 4–12.

Quaife, M., Eakle, R., Cabrera Escobar, M.A., Vickerman, P., Kilbourne-Brook, M., Mvundura, M., Delany-Moretlwe, S. & Terris-Prestholt, F. (2017) Divergent preferences for HIV prevention: a discrete choice experiment for multipurpose HIV prevention products in South Africa. *Medical Decision Making*. In Press, doi:10.1177/0272989X17729376.

Quaife, M. (2017) *Eliciting preferences to understand the impact of new HIV prevention products: Choice experiments with men, women, adolescent girls and female sex workers in South Africa.* PhD Thesis: London School of Hygiene and Tropical Medicine.

Quaife, M., Eakle, R., Cabrera, M., Vickerman, P., Tsepe, M., Cianci, F., Delany-Moretlwe, S. & Terris-Prestholt, F. (2016) Preferences for ARV-based HIV prevention methods among men and women, adolescent girls and female sex workers in Gauteng Province, South Africa: a protocol for a discrete choice experiment. *BMJ Open*. 6 (6), e010682.

Ramsay, C., Clarkson, J., Duncan, A., Lamont, T., Heasman, P., Boyers, D., Goulao, B., et al. (Forthcoming) Improving the Quality of Dentistry (IQuaD): A cluster factorial randomised controlled trial comparing the effectiveness and cost benefit of oral hygiene advice and/or periodontal instrumentation to routine care for the prevention and management of periodontal disease in dentate adults attending dental primary care. *Health Technology Assessment.*

Reed Johnson, F., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D.A., Bresnahan, B.W., Kanninen, B. & Bridges, J.F.P. (2013) Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value in Health*. 16 (1), 3–13.

Regier, D.A. (2008) *Discrete choice experiments informing cost benefit analysis: a Bayesian approach with an application to genetic testing*. PhD Thesis: University of Aberdeen.

Regier, D.A., Friedman, J.M., Makela, N., Ryan, M. & Marra, C.A. (2009) Valuing the benefit of diagnostic testing for genetic causes of idiopathic developmental disability: willingness to pay from families of affected children. *Clinical Genetics*. 75 (6), 514–521.

Reynolds, W.M. (1982) Development of reliable and valid short forms of the Marlowe-Crowne social desirability scale. *Journal of Clinical Psychology*. 38 (1), 119–125.

Rose, J.M., & Bliemer, M.C.J. (2012) *Ngene Version 1.1.1*. Available from: http://www.choice-metrics.com/download.html. [Accessed 11th December 2017].

Rosser, R. & Kind, P. (1978) A scale of valuations of states of illness: is there a social consensus? *International Journal of Epidemiology*. 7 (4), 347–358.

Rowen, D., Azzabi Zouraq, I., Chevrou-Severac, H. & van Hout, B. (2017) International regulations and recommendations for utility data for health technology assessment. *PharmacoEconomics*. 35 (Suppl 1), 11–19.

Ryan, M. (2004a) A comparison of stated preference methods for estimating monetary values. *Health Economics*. 13 (3), 291–296.

Ryan, M. (2004b) Discrete choice experiments in health care. *BMJ*. 328 (7436), 360–361.

Ryan, M., Bate, A., Eastmond, C.J. & Ludbrook, A. (2001) Use of discrete choice experiments to elicit preferences. *Quality in Health Care*. 10 Suppl 1i55–60.

Ryan, M. & Gerard, K. (2003) Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Applied Health Economics and Health Policy*. 2 (1), 55–64.

Ryan, M., Gerard, K. & Amaya-Amaya, M. (eds.) (2007) *Using Discrete Choice Experiments to Value Health and Health Care*. Dordrecht: Springer.

Ryan, M., Scott, D.A. & Donaldson, C. (2004) Valuing health care using willingness to pay: a comparison of the payment card and dichotomous choice methods. *Journal of Health Economics*. 23 (2), 237–258.

Ryan, M. & Watson, V. (2009) Comparing welfare estimates from payment card contingent valuation and discrete choice experiments. *Health Economics*. 18 (4), 389–401.

Ryan, M., Watson, V. & Entwistle, V. (2009) Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. *Health Economics*. 18 (3), 321–336.

Ryan, M., Yi, D., Avenell, A., Douglas, F., Aucott, L., van Teijlingen, E. & Vale, L. (2015) Gaining pounds by losing pounds: preferences for lifestyle interventions to reduce obesity. *Health Economics, Policy, and Law*. 10 (2), 161–182.

Santos, I., Sniehotta, F.F., Marques, M.M., Carraça, E.V. & Teixeira, P.J. (2017) Prevalence of personal weight control attempts in adults: a systematic review and meta-analysis. *Obesity Reviews*. 18 (1), 32–50.

Sen A. (1985) *Commodities and Capabilities*. Amsterdam: North-Holland.

Shah, K.K., Tsuchiya, A. & Wailoo, A.J. (2015) Valuing health at the end of life: a stated preference discrete choice experiment. *Social Science and Medicine*. 124, 48–46.

Smith, R.D. (2007) Use, option and externality values: are contingent valuation studies in health care mis-specified? *Health Economics*. 16 (8), 861–869.

Smith, R.D. & Sach, T.H. (2010) Contingent valuation: what needs to be done? *Health Economics, Policy, and Law*. 5 (Pt 1), 91–111.

Sossong, B., Felder, S., Wolff, M. & Krüger, K. (2017) Evaluating the consequences of rheumatoid arthritis. *The European Journal of Health Economics*. 18 (6), 685–696.

Stamuli, E. (2011) Health outcomes in economic evaluation: who should value health? *British Medical Bulletin*. 97 (1), 197–210.

Stewart, J.M., O'Shea, E., Donaldson, C. & Shackley, P. (2002) Do ordering effects matter in willingness-to-pay studies of health care? *Journal of Health Economics*. 21 (4), 585–599.

Stiggelbout, A.M. & de Vogel-Voogt, E. (2008) Health state utilities: a framework for studying the gap between the imagined and the real. *Value in Health*. 11 (1), 76–87.

Swait, J. & Louviere, J. (1993) The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research*. 30 (3), 305–314.

Terris-Prestholt, F., Quaife, M. & Vickerman, P. (2016) Parameterising user uptake in economic evaluations: the role of discrete choice experiments. *Health Economics*. 25 Suppl 1116–123.

Tervonen, T., Gelhorn, H., Sri Bhashyam, S., Poon, J.-L., Gries, K.S., Rentz, A. & Marsh, K. (2017) MCDA swing weighting and discrete choice experiments for elicitation of patient benefit-risk preferences: a critical assessment. *Pharmacoepidemiology and Drug Safety*. 26 (12), 1483–1491.

Thokala, P., Devlin, N., Marsh, K., Baltussen, R., Boysen, M., Kalo, Z., Longrenn, T., Mussen, F., Peacock, S., Watkins, J. & Ijzerman, M. (2016) Multiple Criteria Decision Analysis for Health Care Decision Making--An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value in Health*. 19 (1), 1–13.

Tilling, C., Devlin, N., Tsuchiya, A. & Buckingham, K. (2010) Protocols for time tradeoff valuations of health states worse than dead: a literature review. *Medical Decision Making*. 30 (5), 610–619.

Tinelli, M., Ryan, M. & Bond, C. (2016) What, who and when? Incorporating a discrete choice experiment into an economic evaluation. *Health Economics Review*. 6 (1), 31.

Torrance, G.W. (1986) Measurement of health state utilities for economic appraisal. *Journal of Health Economics*. 5 (1), 1–30.

Tsuchiya, A. & Dolan, P. (2005) The QALY model and individual preferences for health states and health profiles over time: a systematic review of the literature. *Medical Decision*

Making. 25 (4), 460–467.

Tsuchiya, A. & Watson, V. (2017) Re-Thinking 'The Different Perspectives That can be Used When Eliciting Preferences in Health'. *Health Economics*. In Press, doi:10.1002/hec.3480.

Ubel, P.A., Loewenstein, G. & Jepson, C. (2003) Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Quality of Life Research*. 12 (6), 599–607.

Utz, K.S., Hoog, J., Wentrup, A., Berg, S., Lämmer, A., Jainsch, B., Waschbisch, A., Lee, D.-H., Linker, R.A. & Schenk, T. (2014) Patient preferences for disease-modifying drugs in multiple sclerosis therapy: a choice-based conjoint analysis. *Therapeutic Advances in Neurological Disorders*. 7 (6), 263–275.

Vale, L. (2004) *The Progressive Economic Evaluation of a Surgical Technology: A Case Study of Surgical Repair of Inguinal Hernia*. PhD Thesis: University of Aberdeen.

Varian, H.R. (2014) *Intermediate Microeconomics: A Modern Approach*. 9th (International Student) Edition. New York: W. W. Norton & Company.

Vass, C.M., Wright, S., Burton, M. & Payne, K. (2017) Scale heterogeneity in healthcare discrete choice experiments: a primer. *The Patient*. In Press, doi:10.1007/s40271-017-0282-4.

Vass, C., Rigby, D. & Payne, K. (2017) The role of qualitative research methods in discrete choice experiments. *Medical Decision Making*. 37 (3), 298–313.

Vass, C.M. & Payne, K. (2017) Using discrete choice experiments to inform the benefit-risk assessment of medicines: are we ready yet? *PharmacoEconomics*. 35 (9), 859–866.

Veldwijk, J., Lambooij, M.S., de Bekker-Grob, E.W., Smit, H.A. & de Wit, G.A. (2014) The effect of including an opt-out option in discrete choice experiments. *PloS one*. 9 (11), e111805.

Versteegh, M.M. & Brouwer, W.B.F. (2016) Patient and general public preferences for health

states: A call to reconsider current guidelines. *Social Science and Medicine*. 165, 66–74.

Watson, A.J., Cook, J., Hudson, J., Kilonzo, M., Wood, J., Bruhn, H., Brown, S., Buckley, B., Curran, F., Jayne, D., Loudon, M., Rajagopal, R., McDonald, A. & Norrie, J. (2017) A pragmatic multicentre randomised controlled trial comparing stapled haemorrhoidopexy with traditional excisional surgery for haemorrhoidal disease: the eTHoS study. *Health Technology Assessment*. 21 (70), 1–224.

WHO Expert Consultation (2004) Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *The Lancet*. 363 (9403), 157–163.

Whynes, D.K., Wolstenholme, J.L. & Frew, E. (2004) Evidence of range bias in contingent valuation payment scales. *Health Economics*. 13 (2), 183–190.

de Wit, G.A., Busschbach, J.J. & De Charro, F.T. (2000) Sensitivity and perspective in the valuation of health status: whose values count? *Health Economics*. 9 (2), 109–126.

Zweifel, P., Breyer, F. & Kifmann, M. (2009) 'Economic Valuation of Life and Health', in *Health Economics*. Heidelberg: Springer. pp. 17–74.

Appendix A. Supplementary Materials for Chapter Four

A-1 EuHEA Conference Paper

Title: Whose preferences matter in discrete choice experiments?

David John Mott[1] (PhD Student)

Supervisors: Dr Laura Ternent[1], Professor Luke Vale[1] & Professor John Wildman[1,2]

[1] Health Economics Group, Institute of Health & Society, Newcastle University, Newcastle upon Tyne, UK

[2] Business School, Newcastle University, Newcastle upon Tyne, UK

Correspondence: d.j.mott@newcastle.ac.uk

Abstract

The normative debate surrounding whose preferences should be used in health state valuation has been widely covered with both discussion and empirical papers. The importance of this debate lies in how the use of different perspectives for health state utility values affects the results of economic evaluations, and hence decision-making. Following the same logic, the use of more general preference data from different populations is also likely to have an effect on decision-making.

This paper will contribute to the ongoing preference debate by considering an increasingly popular methodology used in health economics to elicit preferences, the discrete choice experiment (DCE). Using descriptive statistics from a review of recently conducted DCEs to inform this paper, the choice of whose preferences are typically elicited will be investigated. It is expected that the majority of DCE studies elicit the preferences of patients rather than the general public, in contrast to the dominance of public preferences in health state valuation due to methodological arguments that feed into guidance documents e.g. those from the National Institute for Health and Care Excellence (NICE).

This paper aims to highlight the need for more discussion on whose preferences to elicit when using DCEs by raising various normative arguments that support the use of different perspectives. If DCEs are to be used in decision-making, as is often claimed, this area deserves greater attention.

1. Introduction

In health economics, whose preferences are used as the basis of health state valuations within cost-utility analysis (CUA) for health technology assessment is a subject of debate (Brazier *et al.*, 2005). CUA is perhaps the most prevalent form of economic evaluation in health care, which involves a comparison of the costs of alternative health programmes within the health care sector with the benefits, which are reported in the form of the incremental health improvement attributable to the programme (Drummond *et al.*, 2005, p. 137). The most commonly used measure of incremental health improvement is the quality-adjusted life year (QALY); a combined measure of both quality and quantity of life. QALYs are calculated by adjusting the length of time affected through the health outcome by the utility value of the resulting health status (Drummond *et al.*, 2005, p. 14). Utility values typically vary between 0 (which represents being dead) and 1 (which represents full health); hence 1 QALY is equivalent to a year in full health.

In the UK the National Institute for Health and Care Excellence (NICE) advocates the use of CUA for health technology assessment. The reference case suggests that when undertaking a CUA the utility values used should be provided by a sample of the general population rather than patients that have experienced the health states (NICE, 2013). The normative debate as to whose preferences should be used for health state valuation has been widely covered within the literature in both discussion and empirical papers (Boyd *et al.*, 1990; Dolan, 1999; Ubel *et al.*, 2003; Brazier *et al.*, 2005; Dolan, 2008; Dolan and Kahneman, 2008; Gandjour, 2010; Weyler and Gandjour, 2011) but no clear consensus has been reached due to its subjective nature.

Despite the advocacy of CUA in health technology assessment, this method of economic evaluation is unable to take into account the wider societal benefits associated with health care interventions. In the case of public health interventions, where the benefits are likely to go beyond the individual receiving the intervention, the use of CUA may not be appropriate. This has been identified as an issue by NICE, which states that more focus should be placed on cost-benefit analysis (CBA) over CUA when it comes to the appraisal of public health programmes (NICE, 2012). CBA is a form of economic evaluation that is grounded in welfare economic theory, measuring all of the relevant societal costs and benefits of an intervention in commensurate units (Drummond *et al.*, 2005, p. 16). The overall net benefit is calculated and used to determine whether or not resources should be allocated to providing the intervention.

A significant difficulty with the use of CBA is the valuation of all of the societal benefits in monetary terms as market prices do not always exist. For example, the monetary benefit of a reduced waiting time for an NHS-provided surgery is not known because in the UK patients using the NHS do not pay for this directly. Stated preference methods can be used to simulate markets such that values representing individuals' willingness-to-pay (WTP) for various outcomes can be inferred for use in CBA (willingness-to-accept, WTA, can also be estimated). In the UK, the HM Treasury Green Book advocates the use of stated preference methods where appropriate (HM-Treasury, 2011).

A commonly used stated preference method in health economics is the discrete choice experiment (DCE). DCE methodology was first developed in mathematical psychology and has since been used in a range of different fields such as market research, transport economics and environmental economics (Ryan and Farrar, 2000). In a typical DCE, participants are given scenarios where they must choose between hypothetical options; each option is presented as a set of common attributes with differing levels. For example, respondents may have to choose between two hypothetical GP appointments where the common attributes are "distance", "waiting time" and "personal cost", but the levels of each option differ. Respondents are expected to choose the option that they prefer, which is interpreted as the option that will maximise their utility. When a number of respondents choose between the various scenarios, researchers can examine the responses to determine the relative importance of the attributes and the levels of those attributes. Furthermore, when an attribute relating to personal cost is included WTP estimates can be calculated, which could be used in CBAs. Consequently, results from DCE studies could potentially be used to aid the design or determine the provision of health care.

It has been acknowledged that DCE methodology has improved significantly in recent years (Louviere and Lancsar, 2009) and within health economics DCEs have become increasingly popular over time (Ryan and Gerard, 2003; de Bekker-Grob *et al.*, 2012; Clark *et al.*, 2014). They have been used to elicit preferences towards a wide variety of issues from different perspectives including, but not limited to, preferences of the general public towards priority setting (Lancsar *et al.*, 2011; Brazier *et al.*, 2013) and health professionals' preferences for various jobs (Huicho *et al.*, 2012; Pedersen and Gyrd-Hansen, 2013).

The most common use of DCEs in health economics overall is to elicit the preferences of patients or the general public towards a selection of (hypothetical) health care services, which are typically aimed at addressing a specific health issue. A common claim within such studies is that the results of the DCE will provide relative preference information with respect to the

health care service in question, which could later be used in the design, or in determining the provision, of said health care services. Due to the relative popularity of DCE studies of this nature, the focus of this paper will be to look at whose preferences are typically being elicited within this particular section of the literature. A significant proportion of the arguments made within the preference debate in the health state valuation literature appear to be valid when considering the use of such DCE results. This paper aims to use these to put forward the question: whose preferences should be elicited if results from these DCEs are to be used in decision-making?

This paper will be set out as follows. Section 2 will provide an overview of the existing preference debate in the health state valuation literature to provide a background for the subsequent discussion. Section 3 will examine the current trends in DCE studies and raise the issues surrounding whose perspective may be more appropriate in various circumstances. Section 4 raises some further issues and ideas for future research, and Section 5 concludes.

2. The preference debate in health state valuation

In 1996 the Washington Panel on Cost-effectiveness in Health and Medicine provided guidance with regards to a variety of methodological issues in the economic evaluation of health care interventions (Gold *et al.*, 1996). One suggestion by the panel was that the utility values used to calculate QALYs should be based on general public preferences rather than patient preferences. This advice was very influential and was reflected in economic guidelines produced in England and Wales as well as elsewhere (Brazier *et al.*, 2005). Despite the impact on economic guidelines, the debate as to whether patients should have a role in valuing health states due to their relative experience continued.

According to Dolan (1999), the principal reason for the Washington Panel's recommendation was that since the general public bears the costs associated with resource-allocation decisions, they should also have some say in the determination of the benefits. In addition to this, general public values may be less likely to be affected by self-interest and strategic biases than those of patients; this was argued using the notion of the general public acting under a "veil of ignorance" (Gold *et al.*, 1996). In other words if the general public do not know whether they will experience the health state in future, there is little incentive to respond dishonestly because they are "blind" to their own self-interest.

In contrast, the main argument for using patient preferences is that patients have experience of their condition whereas the general public would have to imagine themselves experiencing the health state, which may be problematic (Dolan, 2008). It may be very difficult for an

individual to imagine themselves in a particular health state, and it could be the case that their valuation of the state is significantly affected due to this. A typical example of this difficulty is the situation where a patient may feasibly be able to adapt to a change in health state, yet a member of the general public may not be able to envisage such adaptation (Menzel *et al.*, 2002). Furthermore, patients are considered the winners and losers of any final policy decision and the public may not actually want to be involved in the decision-making process (Brazier *et al.*, 2007, pp. 114-7).

Even the theoretical foundation of this debate in preference-utilitarian theory and welfare economics has been discussed (Gandjour, 2010), yet this arguably only matters if the values are found to differ in practice. When empirically tested it is generally found that discrepancies do indeed exist between the two populations and that typically patients value health states that they have experienced higher than the general public (Boyd *et al.*, 1990; Menzel *et al.*, 2002; Ubel *et al.*, 2003). A number of reasons as to why discrepancies may occur have been suggested such as: the issue of adaptation to health states by patients; the problem of insufficient information being provided to the general public; and the fact that the general public may be less willing to state a timescale after which they would be unwilling to endure the state any further (Menzel *et al.*, 2002; Ubel *et al.*, 2003; Weyler and Gandjour, 2011).

Clearly the choice of whose preferences should be elicited in health state valuation matters due to a variety of normative arguments and the fact that the utility values have been shown to differ in practice. It seems sensible to assume that several of these arguments may be valid in other areas within health economics where preferences are elicited. Thus, the choice of perspective in DCEs in health economics is of interest.

3. A commentary on current DCE trends

3.1. The literature search

A literature search was undertaken in January 2014 using the Medline database to identify all of the studies in health economics that used a DCE between 2009 and 2013 as these years were not covered by a published systematic review at the time. The search terms used were taken from two previous systematic reviews of the DCE literature in health economics (Ryan and Gerard, 2003; de Bekker-Grob *et al.*, 2012). To test the search strategy beforehand, the search was performed for the years 2001 to 2008 to try and identify the papers from the de Bekker-Grob *et al.* (2012) review and all of the papers were successfully identified. Following this, the search was performed again for the years 2009 to 2013 and a total of 1,136 journal articles were initially identified. Figure 1 illustrates how the number of papers was subsequently narrowed down from this point. It was decided that the most recent 3 years of DCE studies would be sufficient to reflect current practice due to the large number of studies over this time period. Further to this, as previously mentioned, the only studies that were included were those eliciting the preferences of patients or the general public for health care services. This resulted in a total of 121 studies being included in the final analysis.

Key information was extracted from the 121 included studies such as whose preferences were being elicited, the types of attribute used in the paper and the way in which the results were presented. For the purpose of this exercise a patient population was defined as a clear group of individuals, all of whom had experience of a *condition* that the hypothetical service(s) described in the DCE tasks were addressing (or were acting as a proxy on behalf of those that had such experience). As a result, any DCEs surrounding screening, vaccinations or preventative services in general were typically considered to use a public population, even if only a subsection of the general public were included in the sample (e.g. those most at risk). Where a health care service was not clearly directed at a particular disease, the population were only defined as patients if they were recruited whilst waiting for, or shortly after receiving, an equivalent service. In other words the criteria for a population to be defined as patients in this paper was that they must have clearly had some experience of either the condition or the service that relates to the topic of the DCE task.

It should be made clear at this point that the use of the terms “patient perspective” and “general public perspective” in this paper refer to the characteristics of the sample that is responding to the DCE task. Within a DCE task, respondents may be asked to consider a

perspective other than their own; this has not been investigated in this paper but will be discussed briefly in Section 4.

Figure 1 – PRISMA flow diagram detailing literature search

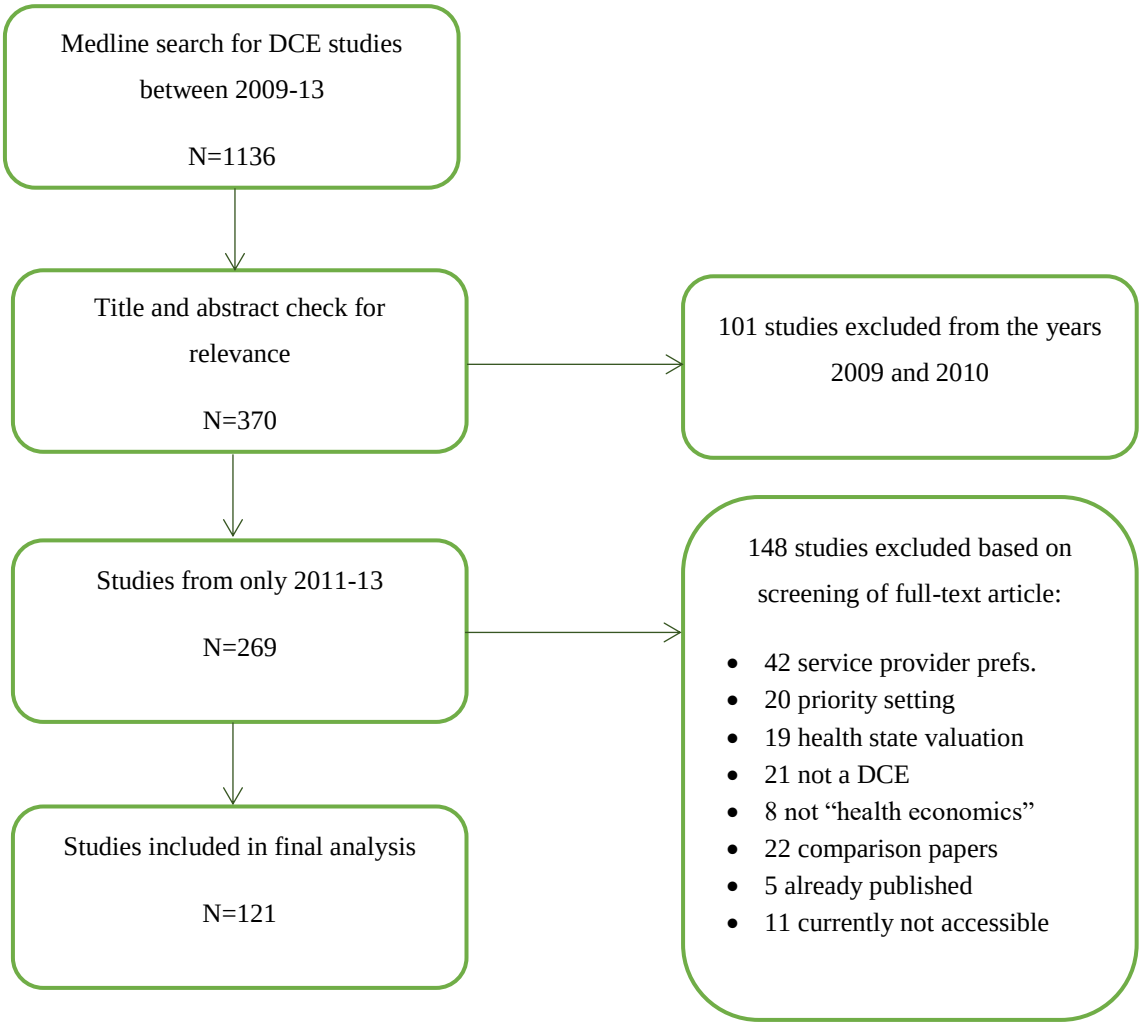


Table 1 – Definitions of the three different categories of DCE study included in this paper

Category	Definition
A	Studies with a DCE that contains attributes describing the consumer experience of a health care service **but does not** contain any attributes relating to health outcomes.
B	Studies with a DCE that contains attributes describing the consumer experience of a health care service **as well as** attributes describing the outcome(s) of the service.

C	Studies with a DCE that contains attributes describing the outcomes of a health care service **but does not** contain any attributes relating the consumer experience of the service.

When looking at DCE studies that elicit preferences towards health care services, there appears to be three distinct categories under which they may fall. Table 1 defines these three categories. The reason that these have been identified is that they are essentially asking for three different things. Category A studies will identify the relative importance of various attributes relating to how a health care service is being delivered; things such as the waiting time, the distance, the personnel involved and the frequency of the service. It makes sense that the results from such a study could be used to design a service that would be most convenient for patients and maximise uptake rates. Category C studies do something very different; they identify the relative importance of the various possible health outcomes from a health care service. In other words, they can tell researchers what the most important health outcomes are so that services could be designed to focus on the most important outcomes. Category B studies contain both consumer experience and outcome attributes, and hence do something slightly different again. In these DCEs, respondents are essentially being made to consider whether they would be willing to sacrifice some of their health to receive a service that is more convenient or preferable in some way. Researchers have the ability to compute marginal rates of substitution within DCEs and hence in studies such as these, consumer experience attributes could be valued in terms of health outcomes.

Similar distinctions between the types of DCE studies that look at health care services already appear to have been made (de Bekker-Grob *et al.*, 2012; Clark *et al.*, 2014), but very little discussion exists surrounding whose preferences might be most suitable to elicit in such studies. Bryan and Dolan (2004) raised a number of issues that should be discussed surrounding the use of DCEs in health economics. One area that they addressed were "normative issues", of which they made the point that if DCE results are to be used in policymaking, consideration needs to be given towards whose preferences are relevant for which policies. This paper will try to add to their points by considering current practice.

3.2. Trends across study categories

Table 2 shows how the studies were split across the different perspectives as well as the categories defined in Table 1. The clearly dominant perspective is that of the patient, with a total of 82 (68%) studies using this perspective. DCEs that look only at consumer experiences

(category A) or at both consumer experiences and outcomes (category B) were most the common (38% and 54.5% of all studies, respectively), whereas those that focused only on outcomes (category C) were particularly uncommon (7.5% of all studies). Within each category the choice of perspective varied slightly, however the patient perspective was consistently the most popular.

Table 2 – DCE studies split across category and perspective

	Category A	Category B	Category C	Total
Public	20 (43.5%)	17 (25.8%)	2 (22%)	**39 (32%)**
Patient	26 (56.5%)	49 (74.2%)	7 (78%)	**82 (68%)**
Total	**46**	**66**	**9**	**121**

Given that there is a dominance of patient preferences in current practice, it seems appropriate to consider what this actually would mean if the results of the DCE studies were to be used in decision-making. For category A studies it would mean that patients (i.e. the consumer) are typically the ones shaping health care services in terms of considering the importance of the various attributes of their delivery. This seems reasonable given their relative experience of their condition and/or similar services although, much like in the preference debate in health state valuation, it could be argued that public values should be used. One point that could be raised is that designing a service to meet patient preferences can have resource implications and hence there may be opportunity costs, which the patients may not necessarily incur. For category C studies it would mean that patients are typically the ones that are informing service providers and decision-makers of the relative importance of the various health outcomes. Once again, this seems fairly reasonable given their experience of the condition because the general public may not be able to understand the severity of various outcomes, or be fully able to appreciate the fact that people can adapt to their health conditions.

For category B studies, the most common type, the choice of perspective becomes slightly more interesting. As it stands, patients would typically be the ones that are providing relative preference information towards the non-health attributes of health care services, having been made to consider the importance of the health outcome(s) of the service simultaneously. In such DCEs, patients are essentially valuing health because they (unless they have strictly dominant preferences towards an outcome attribute) are trading off health for non-health

improvements. It seems somewhat contradictory that while public preferences are usually recommended for use in the valuation of health states, patient preferences are overwhelmingly being elicited to value health care services, even when the outcome of the service is included in the valuation task. It's quite plausible that patients, due to their experience, care strongly about non-health attributes, whereas a member of the general public may have a narrower, health-focused point of view as argued by Bryan and Dolan (2004). If the latter is found to be true, it may be hard to distinguish whether this finding is due to a lack of understanding of the DCE tasks on the part of the general public, or whether this is a reflection of their true preferences. Regardless, the choice of perspective in this category appears to be far more significant than the others because two important and very different elements of health care have to be considered simultaneously by the respondents, and there is arguably more reason to believe that this may vary across the two populations.

The point could be made that if the results are only to be used in the design of services, this issue is slightly less significant because the relative importance of the non-health attributes is what matters. On the other hand, we cannot be sure that the relative importance of the non-health attributes is entirely unaffected by the inclusion of the outcome attribute(s). That is to say, we cannot be sure that the same conclusions would be reached if an identical DCE had been given out to the same respondents but without the outcome attribute(s). This is known as the independence of irrelevant alternatives (IIA) assumption, which is implicit in the multinominal logit (MNL) model that is often used when analysing data from DCEs. The IIA assumption can be relaxed by using alternative models, but the MNL was one of the more commonly used models between 2001 and 2008 in DCE studies in health economics (de Bekker-Grob *et al.,* 2012).

3.3. Trends across countries

A further point raised by Bryan and Dolan (2004) was that DCE studies eliciting patient preferences are arguably more applicable to private health insurance schemes, rather than the predominantly tax-based systems found in the UK and many other European countries. Clearly the dominant perspective of the most recent DCE studies eliciting preferences for health care services overall is that of the patient, however it could be the case that there is significant variation across the various countries. 74.4% of the 121 studies included in this paper came from only six countries; USA, United Kingdom, Australia, The Netherlands, Canada and Germany. Table 3 illustrates the proportions of the two perspectives according to country and study category.

Despite the variation in the health care systems of these six countries, each country elicited preferences from patients more often than the general public. As health care in the USA is largely provided through private health insurance schemes, this seems to follow the point made by Bryan and Dolan (2004). On the other hand, 38.5% of the studies from the USA used a general public perspective and only the Netherlands had a higher proportion (40%). Furthermore only 27.3% of the studies from the UK adopted a general public perspective despite the UK health care system being publicly funded.

The other four countries sit between the UK and the USA elsewhere along the spectrum between publicly funded health care and private health insurance schemes, utilising both in varying capacities. Due to the differing nature of the health care systems, it may be helpful to see how the choice in perspective matches up with government expenditure on health care.

Table 3 – DCE studies from six countries split across category and perspective

		Public	%	Patient	%	Total
USA	**All**	**10**	**38.5%**	**16**	**61.5%**	**26**
	Cat A	6	75.0%	2	25.0%	8
	Cat B	4	23.5%	13	76.5%	17
	Cat C	0	0.0%	1	100.0%	1
United Kingdom	**All**	**6**	**27.3%**	**16**	**72.7%**	**22**
	Cat A	3	33.3%	6	66.7%	9
	Cat B	2	28.6%	5	71.4%	7
	Cat C	1	16.7%	5	83.3%	6
Australia	**All**	**3**	**23.1%**	**10**	**76.9%**	**13**
	Cat A	0	0.0%	4	100.0%	4
	Cat B	3	37.5%	5	62.5%	8
	Cat C	0	0.0%	1	100.0%	1
The Netherlands	**All**	**4**	**40.0%**	**6**	**60.0%**	**10**
	Cat A	2	50.0%	2	50.0%	4

	Cat B	2	33.3%	4	66.7%	6
	Cat C	0	-	0	-	0
Canada	**All**	**3**	**30.0%**	**7**	**70.0%**	**10**
	Cat A	2	40.0%	3	60.0%	5
	Cat B	1	20.0%	4	80.0%	5
	Cat C	0	-	0	-	0
Germany	**All**	**0**	**0.0%**	**9**	**100.0%**	**9**
	Cat A	0	0.0%	2	100.0%	2
	Cat B	0	0.0%	7	100.0%	7
	Cat C	0	-	0	-	0
Total		**26**	**29%**	**64**	**71%**	**90**

Table 4 shows the percentage of health care expenditure in each country that is from the government, and the percentage of the DCE studies included in this paper that used a public perspective. The countries are ranked based on these figures, and there are significant differences between the two sets of rankings. This illustrates that there no obvious link between the way in which health care is funded and the choice of perspective utilised in DCE studies eliciting preferences towards health care services. It appears, from this evidence, that the type of health care system in place is not a driving factor when researchers decide whose preferences to elicit in such DCE studies.

Table 4 – A comparison between government expenditure on health and DCE perspectives

	Government expenditure on health*		Public perspective DCE studies	
Country	**Percentage**	**Relative Rank**	**Percentage**	**Relative Rank**
United Kingdom	84.1%	1	27.3%	4

The Netherlands	79%	2	40%	1
Germany	76.9%	3	0%	6
Canada	70.6%	4	30%	3
Australia	68%	5	23.1%	5
USA	47.7%	6	38.5%	2

*"General government expenditure on health as a percentage of total expenditure on health" from WHO (2012) World Health Statistics Part III (actual figures from 2009).

3.4. Trends across studies with cost attributes

The importance of perspective can be argued further when considering how DCE results could be used in determining the provision of health care services. Providing that they include a cost attribute, DCEs can be used to generate willingness-to-pay (WTP) estimates, which can then be incorporated into economic evaluations. Table 5 is identical to Table 2 but the studies that did not include a cost attribute in the DCE task have been removed.

Table 5 - DCE studies with cost attributes split across category and perspective

	Category A	Category B	Category C	Total
Public	14	11	0	**25**
Patient	12	30	1	**43**
Total	**26**	**41**	**1**	**68**

The majority (56.2%) of DCE studies included a cost attribute, however it appears more common that a cost attribute would be included in a DCE with a public perspective (25/39=64.1%) than in a DCE with a patient perspective (43/82=52.4%). In fact, despite there being more DCE studies that only look at consumer experiences (category A) with a patient perspective in total, there were more public perspective DCE studies with a cost attribute in this category. This trend is not repeated in the DCE studies that consider both consumer experiences and outcomes (category B) however, where the majority (30/41=73%) of the studies with cost attributes used the patient perspective. In the DCE studies that only consider outcomes (category C) only one study included a cost attribute, and this used the patient perspective.

One of the main uses for WTP estimates in health economics is for use in CBAs, which weigh up the societal costs and benefits of a health care service. Given this societal perspective it seems somewhat contradictory, albeit unsurprising, that the majority of DCE studies with cost attributes use the patient perspective (43/68=63.2%). It would seem that the majority of the WTP estimates that are being calculated from DCE studies looking at health care services are not appropriate for use in a traditional, welfarist CBA. On the other hand, it should be noted that it is possible that WTP estimates from patients could be incorporated into other economic evaluations that adopt a narrower perspective (but not as narrow as a traditional CUA), although these are rare.

4. Further discussion

Section 3 has shown that a number of arguments exist to suggest that the choice of perspective between a patient and general public sample in a DCE study is important if the results are to be used in decision-making. Little research has been undertaken to examine the differences in preferences that may exist by undertaking a DCE study with both a patient and a general public population. One study that did do this in the context of preferences for health care services is Najafzadeh *et al.* (2013). The study, based in Canada, elicited preferences towards hypothetical genomic testing for guiding cancer treatment (described as an aggressive but curable cancer). The patient sample was made up of current and former cancer patients (n=38) and a selection of the general public sample were matched to the patient sample using propensity scoring (n=533 to n=83). The authors found that the two samples expressed different preferences regarding the value of the various aspects of genomic testing, with patients caring more about the sensitivity of the tests and being more willing to opt-out of a test if it was not sensitive enough. The differences in preferences lead to large differences in WTP estimates, both for individual attribute levels as well as genomic tests overall. While empirical evidence on the differences between patient and general public preferences is scarce, the fact that they do vary in the case of Najafzadeh *et al.* (2013) is unsurprising given the differences in experiences between the two samples.

Further research into the differences between patient and general public preferences may be useful for a number of reasons. For example, it could be that general public preferences are consistently different, and if it appears to be due to a lack of experience of the condition and/or similar health care services, there may be a strong argument for the use of patient preferences in decision-making. This could ultimately mean that patients become a driving force behind the way in which services are designed (which can have resource implications), or even more significantly, whether certain services are provided at all. If a similar case to

that of the preference debate in health state valuation is made for general public preferences to be used instead, further research in this area may help to aid the design of DCEs such that the general public have the best possible opportunity to engage and express their preferences in a more accurate manner. This is because research in this area may help to highlight the difficulties that the general public experience when responding to DCE surveys.

Further consideration could be given with regards to the choice of perspective beyond the somewhat narrow definitions of a "patient sample" and "general public sample" used in this paper. Dolan *et al.* (2003) developed a conceptual framework of the different perspectives that could be adopted by respondents in studies that elicit preferences for use in informing resource allocation decisions in health care. They suggest that there are three ways in which a respondent may be asked to think about the exercise at hand; on a personal level (only concerning oneself), a societal level (concerning everybody else in society), or a socially inclusive level (concerning everybody else in society as well as oneself). Further to this, the authors make the point that respondents will either have some uncertainty surrounding whether they may need the health care in future (referred to as the *ex ante* context) or they may be told that they will need it (referred to as the *ex post* context). This results in a total of six different perspectives that can be adopted. Further work will be done to examine the wider variety of such perspectives, although some difficulties can be foreseen such as limited reporting of the wording used in the DCE task. It is important to note however, that these are perspectives that respondents are being asked to adopt, and systematic differences between a patient sample and a general public sample are likely to still exist even if they are asked to *adopt* the same perspective within the DCE tasks.

5. Conclusion

This paper has provided a summary of the arguments that could be put forward surrounding the elicitation of patient or general public preferences towards hypothetical health services using DCEs. Current practice shows that the patient perspective is more popular in DCE studies, however when considering how the results may be used in decision-making a number of normative issues arise, many of which have been discussed in the past as part of the preference debate in health state valuation.

Further research into how the results from DCEs may be affected by the status of the respondent would appear to be useful in determining the importance of these issues in this context; however it seems that further consideration and discussion would be beneficial if DCEs are to play a role in decision-making.

References

Boyd, N.F., Sutherland, H.J., Heasman, K.Z., Tritchler, D.L. and Cummings, B.J. (1990) 'Whose utilities for decision analysis?', *Medical Decision Making*, 10(1), pp. 58-67.

Brazier, J., Akehurst, R., Brennan, A., Dolan, P., Claxton, K., McCabe, C., Sculpher, M. and Tsuchyia, A. (2005) 'Should patients have a greater role in valuing health states?', *Appl Health Econ Health Policy*, 4(4), pp. 201-8.

Brazier, J., Ratcliffe, J., Salomon, J. and Tsuchyia, A. (2007) *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford: Oxford University Press.

Brazier, J., Rowen, D., Mukuria, C., Whyte, S., Keetharuth, A. and Hole, A.R. (2013) 'Eliciting societal preferences for burden of illness, therapeutic improvement and end of life for value based pricing: a report of the main survey', *EEPRU*.

Bryan, S. and Dolan, P. (2004) 'Discrete choice experiments in health economics', *The European Journal of Health Economics, formerly: HEPAC*, 5(3), pp. 199-202.

Clark, M., Determann, D., Petrou, S., Moro, D. and de Bekker-Grob, E. (2014) 'Discrete Choice Experiments in Health Economics: A Review of the Literature', *PharmacoEconomics*, pp. 1-20.

de Bekker-Grob, E.W., Ryan, M. and Gerard, K. (2012) 'Discrete choice experiments in health economics: a review of the literature', *Health Economics*, 21(2), pp. 145-172.

Dolan, P. (1999) 'Whose preferences count?', *Med Decis Making*, 19(4), pp. 482-6.

Dolan, P. (2008) 'Developing methods that really do value the 'Q' in the QALY', *Health Econ Policy*
Law, 3(Pt 1), pp. 69-77.

Dolan, P. and Kahneman, D. (2008) 'Interpretations Of Utility And Their Implications For The Valuation Of Health*', *The Economic Journal*, 118(525), pp. 215-234.

Dolan, P., Olsen, J.A., Menzel, P. and Richardson, J. (2003) 'An inquiry into the different perspectives that can be used when eliciting preferences in health', *Health Economics*, 12(7), pp. 545-551.

Drummond, M., Sculpher, M., Torrance, G., O'Brien, B. and Stoddart, G. (2005) *Methods for the economic evaluation of health care programme*. 3rd edn. Oxford: Oxford University Press.

Gandjour, A. (2010) 'Theoretical foundation of patient v. population preferences in calculating QALYs', *Medical Decision Making*.

Gold, M.R., Siegel, J.E., Russell, L.B. and Weinstein, M.C. (1996) *Cost-effectiveness in health and medicine*. Oxford: Oxford University Press.

HM-Treasury (2011) *The Green Book: Appraisal and Evaluation in Central Government*. London: HM Treasury.

Huicho, L., Miranda, J.J., Diez-Canseco, F., Lema, C., Lescano, A.G., Lagarde, M. and Blaauw, D. (2012) 'Job preferences of nurses and midwives for taking up a rural job in Peru: a discrete choice experiment', *PLoS One*, 7(12), p. e50315.

Lancsar, E., Wildman, J., Donaldson, C., Ryan, M. and Baker, R. (2011) 'Deriving distributional weights for QALYs through discrete choice experiments', *Journal of health economics*, 30(2), pp. 466-478.

Louviere, J.J. and Lancsar, E. (2009) 'Choice experiments in health: the good, the bad, the ugly and toward a brighter future', *Health Economics, Policy and Law*, 4(04), pp. 527-546.

Menzel, P., Dolan, P., Richardson, J. and Olsen, J.A. (2002) 'The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis', *Soc Sci Med*, 55(12), pp. 2149-58.

Najafzadeh, M., Johnston, K.M., Peacock, S.J., Connors, J.M., Marra, M.A., Lynd, L.D. and Marra, C.A. (2013) 'Genomic testing to determine drug response: measuring preferences of the public and patients using Discrete Choice Experiment (DCE)', *BMC Health Serv Res*, 13, p. 454.

NICE (2012) *Methods for the development of NICE public health guidance*. London: National Institute for Health and Care Excellence.

NICE (2013) *Guide to the methods of technology appraisal*. London: National Institute for Health and Care Excellence.

Pedersen, L.B. and Gyrd-Hansen, D. (2013) 'Preference for practice: a Danish study on young doctors' choice of general practice using a discrete choice experiment', *Eur J Health Econ*.

Ryan, M. and Farrar, S. (2000) 'Using conjoint analysis to elicit preferences for health care', *BMJ: British Medical Journal*, 320(7248), p. 1530.

Ryan, M. and Gerard, K. (2003) 'Using discrete choice experiments to value health care programmes: current practice and future research reflections', *Applied health economics and health policy*, 2(1), pp. 55-64.

Ubel, P., Loewenstein, G. and Jepson, C. (2003) 'Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public', *Quality of Life Research*, 12(6), pp. 599-607.

Weyler, E.J. and Gandjour, A. (2011) 'Empirical Validation of Patient versus Population Preferences in Calculating QALYs', *Health services research*, 46(5), pp. 1562-1574.

WHO (2012) *World Health Statistics: Global Health Indicators (Part III)*.

A-2 Data Extraction Table for 2011 to 2013

Reference Information			Discrete Choice Experiment Information										Output Information					
ID	Lead Author	Journal	Country	Objective[1]	Perspective	Sample Size	Cost	Payment Vehicle	Time	Risk	Process	Outcome	Coeffs[2]	WTP[3]	MRS[4]	ORs[5]	RI %[6]	Uptake
2011-1	Ahmed	Journal of the American Academy of Nurse Practitioners	USA	A	Public	383	1	Personal Cost	1	0	1	0	1	1	0	0	0	0
2011-2	Bech	Health Economics	DK	A	Public	1053	1	Personal Cost	1	0	1	0	1	1	0	0	0	0
2011-3	Bijlenga	Health Economics	NL	B	Public	88	1	Personal Cost	1	0	1	1	1	1	0	0	0	0
2011-4	Bogelund	Current Medical Research & Opinion	DK	B	Patient	270	1	Personal Cost	0	0	1	1	0	1	0	0	0	0
2011-5	Boonen	Health Services Research	NL	A	Public	1875	1	Copayment, Discount	0	0	1	0	1	1	1	0	0	1
				A	Public	1907	1	Copayment, Discount	0	0	1	0	1	1	1	0	0	1
2011-6	Brown	Haemophilia	USA	B	Patient	53	1	Personal Cost	0	1	1	1	1	1	0	0	1	0
2011-7	Casciano	The International Journal of Clinical Practice	Multi	B	Patient	14033	0	N/A	0	1	1	1	0	0	0	0	1	0
2011-8	Damen	Journal of Plastic, Reconstructive & Aesthetic Surgery	NL	B	Patient	272	0	N/A	1	1	1	1	1	0	0	0	0	0
2011-9	Darba	Osteoporosis International	ESP	A	Patient	166	1	Personal Cost	0	0	1	0	1	1	0	0	0	0
2011-10	Flood	Vaccine	USA	B	Public	544	0	N/A	0	1	1	1	0	0	0	0	1	0
2011-11	Flood	Clinical Pediatrics	USA	B	Public	451	0	N/A	0	0	1	1	0	0	0	0	1	0

Reference Information			Discrete Choice Experiment Information										Output Information					
ID	Lead Author	Journal	Country	Objective[1]	Perspective	Sample Size	Cost	Payment Vehicle	Time	Risk	Process	Outcome	Coeffs[2]	WTP[3]	MRS[4]	ORs[5]	RI %[6]	Uptake
2011-12	Goto	Value in Health	JP	B	Patient	600	0	N/A	0	1	1	1	1	0	0	0	0	0
2011-13	Guo	Value in Health	CA	B	Patient	194	0	N/A	0	1	1	1	1	0	0	0	1	0
2011-14	Hauber	Journal of Dermatological Treatment	USA	B	Patient	419	1	Personal Cost	0	1	1	1	1	1	0	0	0	0
2011-15	Howard	Value in Health	AUS	A	Patient	130	1	Personal Cost	0	1	0	1	1	0	0	0	0	0
2011-16	Kruk	Health Services Research	LR	A	Public	1431	1	Personal Cost	1	0	1	0	1	0	0	0	0	0
2011-17	Laver	BMC Geriatrics	AUS	B	Patient	21	1	Personal Cost	0	0	1	1	1	0	0	0	0	0
2011-18	Lloyd	International Journal of Technology Assessment in Health Care	UK	B	Patient	200	0	N/A	1	0	1	1	1	0	0	0	0	0
2011-19	Lloyd	Clinical Therapeutics	UK	B	Patient	252	1	Personal Cost	0	0	1	1	0	1	0	1	0	1
2011-20	Mheen	Medical Decision Making	NL	A	Patient	308	0	N/A	1	1	1	0	1	0	0	0	1	0
2011-21	Mohamed	Haemophilia	USA	B	Patient	147	0	N/A	0	1	1	1	1	0	0	0	1	0
2011-22	Musters	Human Reproduction	NL	B	Patient	206	1	Personal Cost	0	0	1	1	1	0	0	0	0	0
2011-23	Oteng	Health Services Research	CA	B	Public	1157	1	Personal Cost	0	1	1	1	1	1	0	0	0	0

Reference Information			Discrete Choice Experiment Information										Output Information					
ID	Lead Author	Journal	Country	Objective[1]	Perspective	Sample Size	Cost	Payment Vehicle	Time	Risk	Process	Outcome	Coeffs[2]	WTP[3]	MRS[4]	ORs[5]	RI %[6]	Uptake
2011-24	Park	Telemedicine & e-Health	SK	A	Patient	118	1	Personal Cost	0	0	1	0	1	1	0	0	1	1
2011-25	Poulos	Social Science & Medicine	VN	B	Public	299	1	Personal Cost	0	1	1	1	0	1	0	0	0	1
2011-26	Robertson	Journal of Health Services Research & Policy	UK	A	Patient	2181	0	N/A	0	0	1	0	1	0	0	1	0	0
2011-27	Scalone	Sexually Transmitted Diseases	Multi	B	Patient	154	1	Personal Cost	0	1	1	1	1	1	0	0	0	1
2011-28	Schaarschmidt	Archives of Dermatology	DE	B	Patient	163	1	Personal Cost	1	1	1	1	1	0	0	0	1	0
2011-29	Schwappach	Pharmacoepidemiology & Drug Safety	DE	A	Patient	1000	0	N/A	0	0	1	0	1	0	0	1	0	1
2011-30	Scotland	Birth	UK	B	Patient	730	0	N/A	1	0	1	1	1	0	1	0	0	0
2011-31	Stockwell	Vaccine	USA	B	Public	258	0	N/A	0	0	1	1	0	0	0	0	1	0
2011-32	Swinburn	BJU International	UK	C	Patient	332	0	N/A	0	0	0	1	1	0	1	1	0	0
2011-33	van Gils	Patient Preference & Adherence	NL	A	Patient	46	1	Copayment, Bonus	1	0	1	0	1	0	0	1	0	0
2011-34	van Houtven	Medical Decision Making	USA	B	Patient	570	0	N/A	0	1	1	1	1	0	0	0	0	0
2011-35	Sanders	Spinal Cord	USA	B	Patient	66	1	Personal Cost	0	0	1	1	0	0	0	0	1	1
2011-36	Hong	Journal of the American Pharmacists Association	USA	A	Public	355	1	Personal Cost	0	0	1	0	1	1	0	0	1	0

Reference Information			Discrete Choice Experiment Information										Output Information					
ID	Lead Author	Journal	Country	Objective[1]	Perspective	Sample Size	Cost	Payment Vehicle	Time	Risk	Process	Outcome	Coeffs[2]	WTP[3]	MRS[4]	ORs[5]	RI %[6]	Uptake
2011-37	Fegert	Expert Reviews of Pharmacoeconomics & Outcomes Research	DE	B	Patient	121	0	N/A	0	0	1	1	1	0	0	0	0	0
2011-38	Yi	European Journal of Pain	UK	A	Patient	124	0	N/A	1	0	1	0	1	0	0	0	0	1
2011-39	Waltzman	Annals of Plastic Surgery	USA	A	Public	111	0	N/A	0	0	1	0	1	0	0	0	1	0
2011-40	Bridges	Health Policy and Planning	SA	B	Public	640	0	N/A	0	0	1	0	0	0	0	1	0	0
2011-41	Waschbush	Journal of Clinical Child & Adolescent Psycology	USA	B	Patient	183	1	Personal Cost	0	1	1	1	1	0	0	0	1	0
2011-42	Laver	Journal of Rehabilitation Medicine	AUS	B	Patient	50	1	Personal Cost	0	0	1	1	1	1	1	0	0	0
2011-43	Mohamed	Pharmacoeconomics	USA	C	Patient	138	0	N/A	0	0	0	1	1	0	0	0	1	0
2012-1	Bowen	Health Technology Assessment	UK	C	Public	213	0	N/A	1	0	0	1	1	0	1	0	0	0
2012-2	Bridges	Lung Cancer	UK	C	Patient	89	0	N/A	0	0	0	1	0	0	0	1	1	0
2012-3	Bridges	Patient	SA	B	Public	645	0	N/A	1	0	1	0	1	0	0	0	1	0
2012-4	Burnett	Arthritis Care & Research	CA	B	Patient	105	1	Personal Cost	0	0	1	1	1	0	0	0	1	0
2012-5	Cheng	BMC Medical Research Methodology	CA	A	Patient	468	1	Personal Cost	0	0	1	0	1	0	0	0	1	0
2012-6	Damman	Medical Decision Making	NL	A	Public	714	0	N/A	1	1	1	0	1	0	0	1	1	0

Reference Information			Discrete Choice Experiment Information										Output Information					
ID	Lead Author	Journal	Country	Objective[1]	Perspective	Sample Size	Cost	Payment Vehicle	Time	Risk	Process	Outcome	Coeffs[2]	WTP[3]	MRS[4]	ORs[5]	RI %[6]	Uptake
				A	Public	589	0	N/A	1	0	1	0	1	0	0	1	1	0
2012-7	Gerard	Value in Health	UK	A	Patient	451	0	N/A	0	0	1	0	1	0	0	0	0	1
2012-8	Gidengil	Vaccine	USA	A	Public	558	1	Personal Cost	0	1	1	0	1	1	0	0	0	0
2012-9	Hancock-Howard	Birth Defects Research (Part A)	CA	A	Public	175	1	Personal Cost	0	0	1	0	1	1	0	0	0	0
2012-10	Kauf	Patient	Multi	B	Patient	284	0	N/A	1	1	1	0	1	0	0	0	1	0
2012-11	King	British Journal of Cancer	AUS	C	Patient	357	0	N/A	0	0	0	1	1	0	0	0	0	0
2012-12	Laba	BMC Family Practice	AUS	B	Public	161	1	Personal Cost	0	1	1	1	0	0	0	0	1	0
2012-13	Landfeldt	Human Reproduction	SE	A	Patient	294	1	Personal Cost	0	0	1	0	1	1	0	0	1	0
2012-14	Manjunath	Epilepsy	USA	B	Patient	193	1	Personal Cost	0	0	1	1	0	1	0	0	1	1
2012-15	Marti	European Journal of Health Economics	CH	B	Patient	131	1	Personal Cost	0	1	1	1	1	1	0	0	0	0
2012-16	Miners	Sexually Transmitted Infections	UK	A	Patient	3358	0	N/A	1	0	1	0	0	0	0	1	0	0
2012-17	Mohamed	Journal of Medical Economics	USA	A	Patient	318	0	N/A	0	0	1	0	0	0	0	0	1	0
2012-18	Morton	American Journal of Kidney Disease	AUS	B	Patient	178	0	N/A	1	0	1	1	0	0	1	1	0	0

Reference Information			Discrete Choice Experiment Information										Output Information					
ID	**Lead Author**	**Journal**	**Country**	**Objective[1]**	**Perspective**	**Sample Size**	**Cost**	**Payment Vehicle**	**Time**	**Risk**	**Process**	**Outcome**	**Coeffs[2]**	**WTP[3]**	**MRS[4]**	**ORs[5]**	**RI %[6]**	**Uptake**
2012-19	Naik-Panvelkar	Journal of Asthma	AUS	B	Patient	80	1	Personal Cost	0	0	1	1	1	1	0	0	0	0
2012-20	Paczkowski	Plos One	ET	A	Public	1006	1	Personal Cost	0	0	1	0	1	0	0	0	0	0
2012-21	Philips	Health Policy	???	A	Patient	350	0	N/A	1	0	1	0	1	0	0	0	1	0
2012-22	Pignone	Journal of General Internal Medicine	USA	B	Public	104	1	Personal Cost	0	0	1	1	0	0	0	0	1	0
2012-23	Rennie	Value in Health	UK	A	Public	473	1	Personal Cost	0	0	1	0	1	1	0	0	0	0
2012-24	Schiotz	Patient Education & Counseling	DK	A	Patient	2187	1	Personal Cost	0	0	1	0	0	1	0	0	0	0
2012-25	Schmieder	Journal of the American Academy of Dermatology	DE	B	Patient	163	1	Personal Cost	0	1	1	1	1	0	0	0	1	0
2012-26	Sung	Journal of Clinical Epidemiology	CA	A	Patient	274	0	N/A	0	1	1	0	1	0	0	0	0	1
2012-27	Goodall	Health Policy	AUS	A	Patient	161	0	N/A	0	0	1	0	1	0	0	0	0	0
2012-28	Tinelli	BMC Dermatology	UK	C	Patient	183	1	Personal Cost	0	0	0	1	1	1	0	0	0	0
2012-29	Tong	The Annals of Thoracic Surgery	USA	B	Patient	224	0	N/A	0	1	1	1	0	0	0	0	1	0
2012-30	Wong	Journal of Medical Economics	USA	B	Patient	272	0	N/A	0	1	1	1	0	0	0	0	1	0
2012-31	Yeo	Diabetic Medicine	UK	A	Patient	160	0	N/A	1	0	1	0	1	0	1	0	0	0

Reference Information			Discrete Choice Experiment Information										Output Information					
ID	**Lead Author**	**Journal**	**Country**	**Objective[1]**	**Perspective**	**Sample Size**	**Cost**	**Payment Vehicle**	**Time**	**Risk**	**Process**	**Outcome**	**Coeffs[2]**	**WTP[3]**	**MRS[4]**	**ORs[5]**	**RI %[6]**	**Uptake**
2013-1	Alayi-Goebbels	Value in Health	NL	B	Patient	273	1	Personal Cost	0	0	1	1	1	1	0	0	0	0
2013-2	Augustovski	Value in Health	AR	B	Patient	240	1	Personal Cost	0	1	1	1	1	0	0	0	0	0
2013-3	Best	Perspectives on Sexual & Reproductive Health	USA	A	Public	266	0	N/A	0	1	1	0	1	0	0	0	0	0
2013-4	Boeri	Journal of Health Economics	UK	B	Public	493	1	Personal Cost	0	0	1	0	1	1	0	0	0	0
2013-5	Brenner	Journal of General Internal Medicine	Multi	B	Public	306	0	N/A	0	1	1	0	0	0	0	0	1	0
2013-6	Carroll	Prenatal Diagnosis	UK	B	Public	106	1	Personal Cost	1	0	1	1	1	1	0	0	0	0
2013-7	Chu	Geriatrics Gerontology International	CN	A	Patient	1540	1	Personal Cost	0	0	1	0	1	1	0	0	0	0
2013-8	Cunningham	Journal of Abnormal Child Psychology	CA	A	Patient	1059	0	N/A	0	0	1	0	1	0	0	0	0	1
2013-9	Cunningham	Journal of Health Communication: International Perspectives	CA	A	Public	1129	0	N/A	0	0	1	0	1	0	0	0	1	1
2013-10	de Bekker-Grob	British Journal of Cancer	NL	B	Public	459	1	Personal Cost	0	1	1	0	1	1	1	0	0	0
2013-11	Deal	Value in Health	CA	B	Patient	361	1	Personal Cost	0	1	1	1	1	1	0	0	0	0
2013-12	Dixon	Health Expectations	UK	A	Patient	77	0	N/A	0	0	1	0	1	0	0	0	0	0

Reference Information			Discrete Choice Experiment Information										Output Information					
ID	**Lead Author**	**Journal**	**Country**	**Objective[1]**	**Perspective**	**Sample Size**	**Cost**	**Payment Vehicle**	**Time**	**Risk**	**Process**	**Outcome**	**Coeffs[2]**	**WTP[3]**	**MRS[4]**	**ORs[5]**	**RI %[6]**	**Uptake**
2013-13	Dwight Johnson	General Hospital Psychiatry	USA	A	Patient	63	1	Personal Cost	0	0	1	0	0	0	0	1	0	0
2013-14	Gelhorn	Diabetes, Obesity & Metabolism	UK	C	Patient	100	0	N/A	0	1	0	1	1	0	0	0	1	0
2013-15	Glenngard	Nordic Journal of Psychiatry	Multi	B	Patient	285	1	Personal Cost	0	1	1	1	1	1	0	0	0	0
2013-16	Gonzalez	Headache	USA	C	Patient	510	0	N/A	0	1	0	1	1	0	0	0	1	0
2013-17	Hall	Medical Decision Making	AUS	A	Patient	72	0	N/A	0	0	1	0	1	0	0	0	0	0
				A	Patient	96	0	N/A	0	0	1	0	1	0	0	0	0	0
2013-18	Hauber	Osteoarthritis & Cartilage	UK	C	Patient	294	0	N/A	0	1	0	1	0	0	1	0	1	0
2013-19	Hauber	Patient Preference & Adherence	USA	B	Patient	1114	1	Personal Cost	0	1	1	1	0	1	0	0	1	1
2013-20	Kjaer	European Journal of Health Economics	GL	A	Public	194	1	Tax Increase	0	0	1	0	1	1	0	0	0	0
2013-21	Knox	Social Science & Medicine	AUS	B	Public	527	1	Personal Cost	0	0	1	1	1	1	0	0	0	0
2013-22	Kobayashi	International Journal of Health Care Quality Assurance	JP	A	Public	111	1	Personal Cost	0	0	1	0	1	0	0	0	1	0
2013-23	Krucien	Thorax	FR	B	Patient	121	1	Personal Cost	0	0	1	1	1	0	0	0	0	0
2013-24	Laba	BMC Musculoskeletal Disorders	AUS	B	Patient	188	1	Personal Cost	0	0	1	1	0	1	0	1	0	1

Reference Information			Discrete Choice Experiment Information										Output Information					
ID	Lead Author	Journal	Country	Objective[1]	Perspective	Sample Size	Cost	Payment Vehicle	Time	Risk	Process	Outcome	Coeffs[2]	WTP[3]	MRS[4]	ORs[5]	RI %[6]	Uptake
2013-25	Lathia	Support Care Cancer	CA	B	Patient	88	1	Personal Cost	0	0	1	0	1	0	0	0	0	1
2013-26	Llewellyn	BMJ Open	UK	A	Public	233	0	N/A	0	0	1	0	0	0	0	1	0	0
2013-27	Lynn	Health Expectations	UK	B	Patient	146	1	Personal Cost	0	0	1	1	1	0	0	0	0	0
2013-28	Mattsson	Maturitas	SE	B	Patient	423	1	Personal Cost	0	0	1	1	0	1	0	0	0	0
2013-29	Meghani	BMC Medical Informatics & Decision Making	USA	B	Patient	241	1	Personal Cost	0	0	1	1	1	0	0	0	1	0
2013-30	Miller	Journal of Medical Economics	USA	B	Patient	205	1	Personal Cost	0	0	1	0	1	1	0	0	0	0
2013-31	Milte	Journal of Rehabilitation Medicine	AUS	B	Patient	87	0	N/A	0	1	1	1	1	0	1	0	0	0
2013-32	Mohamed	Diabetes & Metabolism	Multi	B	Patient	383	1	Personal Cost	0	0	1	1	1	0	0	0	0	0
2013-33	Moia	Internal Emergency Medicine	IT	A	Patient	255	1	Personal Cost	0	1	1	0	1	1	0	0	0	0
2013-34	Muhlbacher	International Journal of Integrated Medicine	DE	A	Patient	110	0	N/A	0	0	1	0	1	0	0	0	0	0
2013-35	Muhlbacher	Health Economics Review	DE	B	Patient	218	0	N/A	0	1	1	1	1	0	0	0	0	0
2013-36	Pignone	JAMA Internal Medicine	Multi	C	Public	302	0	N/A	0	1	0	0	0	0	0	0	1	0
2013-37	Poulos	Arthritis Care & Research	USA	B	Patient	296	0	N/A	0	1	1	0	1	0	0	0	0	0

Reference Information			Discrete Choice Experiment Information										Output Information					
ID	Lead Author	Journal	Country	Objective[1]	Perspective	Sample Size	Cost	Payment Vehicle	Time	Risk	Process	Outcome	Coeffs[2]	WTP[3]	MRS[4]	ORs[5]	RI %[6]	Uptake
2013-38	Sadique	Plos One	UK	A	Public	369	1	Personal Cost	0	1	1	0	1	1	0	0	0	1
2013-39	Schaarschmidt	Journal of the European Academy of Dermatology and Venereology	DE	B	Patient	163	1	Personal Cost	0	0	1	1	0	0	0	0	1	0
2013-40	Shingler	Journal of Medical Economics	UK	B	Patient	99	0	N/A	0	0	1	1	0	0	0	1	0	0
2013-41	Terris-Prestholt	Plos One	SA	B	Public	1017	1	Personal Cost	0	0	1	0	1	0	0	0	1	1
2013-42	Tsunematsu	Asian Pac Journal of Cancer	JP	A	Public	993	1	Personal Cost	0	0	1	0	0	0	0	0	1	0
2013-43	Veldwijk	BMC Public Health	NL	B	Patient	1250	1	Personal Cost	0	0	1	1	1	0	0	0	0	1
2013-44	Whitty	Plos One	AUS	A	Patient	91	1	Personal Cost	0	0	1	0	1	0	0	0	0	0
2013-45	Zimmermann	Journal of Affective Disorders	DE	B	Patient	227	0	N/A	0	1	1	1	1	0	0	0	1	0
2013-46	Ahn	Telemedicine & e-Health	SK	A	Public	400	1	Personal Cost	0	0	1	0	1	1	0	0	1	1
2013-47	Parkinson	Applied Health Economics and Health Policy	AUS	B	Public	175	1	Personal Cost	0	0	1	1	1	1	0	0	0	0
					Public	179	1	Personal Cost	0	0	1	1	1	1	0	0	0	0
					Public	181	1	Personal Cost	0	0	1	1	1	1	0	0	0	0
2013-48	Wittink	Patient	USA	A	Public	86	0	N/A	0	1	1	0	1	0	0	0	0	0

Reference Information			Discrete Choice Experiment Information										Output Information					
ID	**Lead Author**	**Journal**	**Country**	**Objective[1]**	**Perspective**	**Sample Size**	**Cost**	**Payment Vehicle**	**Time**	**Risk**	**Process**	**Outcome**	**Coeffs[2]**	**WTP[3]**	**MRS[4]**	**ORs[5]**	**RI %[6]**	**Uptake**
2013-49	Pisa	Patient	DE	B	Patient	300	0	N/A	0	0	1	1	1	0	0	0	1	0

1 = Included, 0 = Not Included

[1]For information on objectives, see Appendix A.1; [2]Coeffs = Coefficients (regression output); [3]WTP = Willingness to Pay; [4]MRS = Marginal rates of substitution; [5]ORs = Odds Ratios; [6]RI %; Relative Importance

A-3 Data Extraction Table for 2014 to March 2017

No.	Lead Author	Year	Journal	Quote on Sample Recruitment Strategy	Classification
5	Alten	2016	Patient Pref Adherence	"The following inclusion criteria were to be met by patients: subjects had to be diagnosed with RA, at least 18 years of age, show a sufficient level of proficiency in the German language, and be treated with at least one DMARD [disease-modifying anti rheumatic drug]."	Patient
11	Benning	2014	Value Health	"On this platform, 1277 individuals from the general Dutch population in the age category 55 to 75 years started our online survey in the period mid-May to the end of May 2012. … Finally, the data of 631 Dutch respondents in the age category 55 to 75 years who have no experience with CRC were used in the analysis."	General Public
14	Bewtra	2015	Am J Gastroenterol	“Patients were eligible if they were ≥18 years of age with an ICD-9 code for UC (556.0–556.6 and 556.8–556.9) or CD (555.0–555.2, 555.9) and an outpatient gastroenterology visit in the prior 2 years."	Patient
21	Brown	2014	Advances in Health Economics and Health Services Research	"The survey was fielded to a U.S. sample of 307 girls aged 13–17 years who had not yet received an HPV vaccine in June 2008."	General Public
23	Buchanan	2016	Patient	"The sampling population selected for this DCE was UKCLL patients as it is recommended that DCEs are under-taken in populations with experience in the area of interest"	Patient
43	Determann	2014	PLoS One	"A market research company (Flycatcher) was hired to administer the online questionnaire to a representative sample of the general adult population of the Netherlands."	General Public
59	Garcia - Dominguez	2016	Patient Pref Adherence	"Patients aged 18 years or older with a diagnosis of RRMS or SPMS (MS type was reported by the patient and not con-firmed by any other data source) were included in the study."	Patient
77	Hauber	2015	Diabetes Therapy	"To qualify for inclusion in this study, patients had to (1) be aged 18 years or older; (2) have a self-reported physician diagnosis of T2DM; (3) be currently taking one OAD or no OAD to treat their T2DM; and (4) be not currently taking injectable T2DM treatments (e.g., insulin and glucagon-like 1 receptor agonist)."	Patient
89	Holmes	2016	British Journal of Clinical Pharmacology	"Community-dwelling, hypertensive patients recruited from nine European countries were invited to complete a discrete choice experiment (DCE)"	Patient

No.	Lead Author	Year	Journal	Quote on Sample Recruitment Strategy	Classification
105	Kistler	2015	Patient Pref Adherence	"Two team members enrolled a volunteer sample of community-dwelling older adults (i.e., older adults not residing in a long-term care facility or other institution) aged 70–90 years, with no personal history of CRC or inflammatory bowel disease."	General Public
108	Knight	2015	Medical Decision Making	"Participants were eligible for the survey if they were 50 years of age or older and a United States resident."	General Public
111	Kromer	2015	PLoS One	"Inclusion criteria were age 18 years and moderate-to-severe psoriasis according to the criteria of the Committee for Medicinal Products for Human Use (CMPH), i.e., PASI 10 in the course of the disease, involvement of head, palms or plantar surfaces, or psoriatic arthritis according to Classification of Psoriatic Arthritis (CASPAR) criteria with any skin involvement"	Patient
122	Lehman	2016	Radiotherapy & Oncology	“Patients were eligible if they were diagnosed with Stage III NSCLC and were to receive definitive CRT. Patients were identified at the weekly Lung Cancer Multi-disciplinary Meeting"	Patient
123	Levitan	2015	Psychiatric Services	"Respondents who were at least18 years of age and who had a self-reported physician diagnosis of schizophrenia were identified via a pre-screening survey and were recruited through Kantar Health’s online patient panel in May 2012."	Patient
126	Louder	2016	Am Health Drug Benefits	"For participation in the study, patients had to be aged 21 to 80 years at the time of survey administration, be currently enrolled in a fully insured Humana commercial health plan with medical and pharmacy benefits, and to have had at least 2 RA-related medical claims in the previous 12 months, at least 30 days apart, as identified by International Classification of Diseases, Ninth Revision, Clinical Modification diagnosis code 714.0 (rheumatoid arthritis)."	Patient
137	Medina-Lara	2014	Malaria Journal	"Eight different choice questions were randomly allocated and presented to each household head in a community survey of health-seeking behaviour for treating fever in adults and children in the Zomba District of Malawi."	General Public
149	Mohamed	2016	Patient	"To be included in the study, respondents were required to be aged C18 years, have a self-reported physician diagnosis of CF, or be the parent of a paediatric patient(aged 6–17 years) with CF, and the patient must have stated that they have tested positive for P. aeruginosa in their lung culture at least twice a year (self-reported)."	Patient

No.	Lead Author	Year	Journal	Quote on Sample Recruitment Strategy	Classification
171	Noble	2015	Haemotologica	"Patients were eligible if they met the following criteria: Histologically confirmed cancer; Receiving ongoing treatment for cancer (chemotherapy and or radiotherapy); Radiologically confirmed symptomatic deep vein thrombosis (DVT) and/or pulmonary embolus (PE); Receiving anticoagulation for their diagnosed DVT and/or PE; Receiving anticoagulation for no more than six months."	Patient
173	O'Hara	2016	Can J Surg	"We recruited adult patients with end-stage shoulder OA waiting for TSA surgery to complete the DCE questionnaire. All participants were recruited from a single surgeon's practice located in metropolitan Ontario. "	Patient
181	Posnett	2015	Patient Pref Adherence	"A prospective, internet-based, double-blind survey recruited adult participants with self-reported OAK in one or both knees. Previous studies have shown that self-reported physician-level diagnoses of osteoarthritis (OA) are reliable, with a concordance of 86.9% with primary-care records."	Patient
205	Tokes	2014	J Gen Intern Med	"Participants were screened to meet key target criteria: 18–65 years old; Chinese, Korean, or Vietnamese ethnicity; with a doctor's diagnosis of CHB at least 6 months previously; on current antiviral treatment for CHB treatment (Treated) or naïve to CHB antiviral treatment (Treatment-naïve); and with no participation in CHB studies in the last 6 months."	Patient
208	Uemura	2016	BMC Urol	"Inclusion criteria were as follows: (1) aged 20 years or older, (2) diagnosed with CRPC, (3)chemotherapy-naïve, (4) able to read and understand Japanese, and (5) provided written informed consent."	Patient
209	Ultz	2014	Therapeutic Advances in Neurological Disorders	"Patients with relapsing–remitting MS (n = 156) completed a questionnaire assessing treatment preferences,"	Patient

A-4 References of Included Studies

AHMED, A. & FINCHAM, J. E. 2011. Patients' view of retail clinics as a source of primary care: boon for nurse practitioners? J Am Acad Nurse Pract, 23, 193-9.

AHN, J., SHIN, J., LEE, J., SHIN, K. & PARK, H. 2014. Consumer preferences for telemedicine devices and services in South Korea. Telemed J E Health, 20, 168-74.

ALAYLI-GOEBBELS, A. F., DELLAERT, B. G., KNOX, S. A., AMENT, A. J., LAKERVELD, J., BOT, S. D., NIJPELS, G. & SEVERENS, J. L. 2013. Consumer preferences for health and nonhealth outcomes of health promotion: results from a discrete choice experiment. Value Health, 16, 114-23.

ALTEN, R., KRÜGER, K., RELLECKE, J., SCHIFFNER-ROHE, J., BEHMER, O., SCHIFFHORST, G. & NOLTING, H.-D. 2016. Examining patient preferences in the treatment of rheumatoid arthritis using a discrete-choice approach. Patient Prefer Adherence. 102217–2228.

AUGUSTOVSKI, F., BERATARRECHEA, A., IRAZOLA, V., RUBINSTEIN, F., TESOLIN, P., GONZALEZ, J., LENCINA, V., SCOLNIK, M., WAIMANN, C., NAVARTA, D., CITERA, G. & SORIANO, E. R. 2013. Patient preferences for biologic agents in rheumatoid arthritis: a discrete-choice experiment. Value Health, 16, 385-93.

BECH, M., KJAER, T. & LAURIDSEN, J. 2011. Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. Health Econ, 20, 273-86.

BENNING, T.M., DELLAERT, B.G.C., SEVERENS, J.L. & DIRKSEN, C.D. 2014. The effect of presenting information about invasive follow-up testing on individuals' noninvasive colorectal cancer screening participation decision: results from a discrete choice experiment. Value Health. 17 (5), 578–587.

BEST, C., TANNER, A. E., HENSEL, D. J., FORTENBERRY, J. D. & ZIMET, G. D. 2014. Young women's contraceptive microbicide preferences: associations with contraceptive behavior and sexual relationship characteristics. Perspect Sex Reprod Health, 46, 15-22.

BEWTRA, M., FAIRCHILD, A.O., GILROY, E., LEIMAN, D.A., KERNER, C., JOHNSON, F.R. & LEWIS, J.D. 2015. Inflammatory Bowel Disease Patients' Willingness to Accept Medication Risk to Avoid Future Disease Relapse. Am. J. Gastroenterol. 110 (12), 1675–1681.

BIJLENGA, D., BONSEL, G. J. & BIRNIE, E. 2011. Eliciting willingness to pay in obstetrics: comparing a direct and an indirect valuation method for complex health outcomes. Health Econ, 20, 1392-406.

BOERI, M., LONGO, A., GRISOLIA, J. M., HUTCHINSON, W. G. & KEE, F. 2013. The role of regret minimisation in lifestyle choices affecting the risk of coronary heart disease. J Health Econ, 32, 253-60.

BOGELUND, M., VILSBOLL, T., FABER, J., HENRIKSEN, J. E., GJESING, R. P. & LAMMERT, M. 2011. Patient preferences for diabetes management among people with type 2 diabetes in Denmark - a discrete choice experiment. Curr Med Res Opin, 27, 2175-83.

BOONEN, L. H., DONKERS, B. & SCHUT, F. T. 2011. Channeling consumers to preferred providers and the impact of status quo bias: does type of provider matter? Health Serv Res, 46, 510-30.

BOWEN, A., HESKETH, A., PATCHICK, E., YOUNG, A., DAVIES, L., VAIL, A., LONG, A., WATKINS, C., WILKINSON, M., PEARL, G., LAMBON RALPH, M. & TYRRELL, P. 2012. Clinical effectiveness, cost-effectiveness and service users' perceptions of early, well-resourced communication therapy following a stroke: a randomised controlled trial (the ACT NoW Study). Health Technol Assess, 16, 1-160.

BRENNER, A., HOWARD, K., LEWIS, C., SHERIDAN, S., CRUTCHFIELD, T., HAWLEY, S., REULAND, D., KISTLER, C. & PIGNONE, M. 2014. Comparing 3 values clarification methods for colorectal cancer screening decision-making: a randomized trial in the US and Australia. J Gen Intern Med, 29, 507-13.

BRIDGES, J. F., MOHAMED, A. F., FINNERN, H. W., WOEHL, A. & HAUBER, A. B. 2012. Patients' preferences for treatment outcomes for advanced non-small cell lung cancer: a conjoint analysis. Lung Cancer, 77, 224-31.

BRIDGES, J. F., SEARLE, S. C., SELCK, F. W. & MARTINSON, N. A. 2012. Designing family-centered male circumcision services: a conjoint analysis approach. Patient, 5, 101-11.

BRIDGES, J. F., SELCK, F. W., GRAY, G. E., MCINTYRE, J. A. & MARTINSON, N. A. 2011. Condom avoidance and determinants of demand for male circumcision in Johannesburg, South Africa. Health Policy Plan, 26, 298-306.

BROWN, D.S., POULOS, C., JOHNSON, F.R., CHAMIEC-CASE, L. & MESSONNIER, M.L. 2014. Adolescent girls' preferences for HPV vaccines: a discrete choice experiment. Adv Health Econ Health Serv Res. 2493–121.

BROWN, T. M., PASHOS, C. L., JOSHI, A. V. & LEE, W. C. 2011. The perspective of patients with haemophilia with inhibitors and their care givers: preferences for treatment characteristics. Haemophilia, 17, 476-82.

BUCHANAN, J., WORDSWORTH, S. & SCHUH, A. 2016. Patients' Preferences for Genomic Diagnostic Testing in Chronic Lymphocytic Leukaemia: A Discrete Choice Experiment. Patient. 9 (6), 525–536.

BURNETT, H. F., REGIER, D. A., FELDMAN, B. M., MILLER, F. A. & UNGAR, W. J. 2012. Parents' preferences for drug treatments in juvenile idiopathic arthritis: a discrete choice experiment. Arthritis Care Res (Hoboken), 64, 1382-91.

CARROLL, F. E., AL-JANABI, H., FLYNN, T. & MONTGOMERY, A. A. 2013. Women and their partners' preferences for Down's syndrome screening tests: a discrete choice experiment. Prenat Diagn, 33, 449-56.

CASCIANO, R., MALANGONE, E., RAMACHANDRAN, A. & GAGLIARDINO, J. J. 2011. A quantitative assessment of patient barriers to insulin. Int J Clin Pract, 65, 408-14.

CHENG, J., PULLENAYEGUM, E., MARSHALL, D. A., MARSHALL, J. K. & THABANE, L. 2012. An empirical comparison of methods for analyzing correlated data from a discrete choice survey to elicit patient preference for colorectal cancer screening. BMC Med Res Methodol, 12, 15.

CHU, L. W., SO, J. C., WONG, L. C., LUK, J. K., CHIU, P. K., CHAN, C. S., KWAN, F. S., CHAU, J., HUI, E., WOO, J. & MCGHEE, S. M. 2014. Community end-of-life care among Chinese older adults living in nursing homes. Geriatr Gerontol Int, 14, 273-84.

CUNNINGHAM, C. E., CHEN, Y., DEAL, K., RIMAS, H., MCGRATH, P., REID, G., LIPMAN, E. & CORKUM, P. 2013. The interim service preferences of parents waiting for children's mental health treatment: a discrete choice conjoint experiment. J Abnorm Child Psychol, 41, 865-77.

CUNNINGHAM, C. E., WALKER, J. R., EASTWOOD, J. D., WESTRA, H., RIMAS, H., CHEN, Y., MARCUS, M., SWINSON, R. P., BRACKEN, K. & THE MOBILIZING MINDS

RESEARCH, G. 2014. Modeling mental health information preferences during the early adult years: a discrete choice conjoint experiment. J Health Commun, 19, 413-40.

DAMEN, T. H., DE BEKKER-GROB, E. W., MUREAU, M. A., MENKE-PLUIJMERS, M. B., SEYNAEVE, C., HOFER, S. O. & ESSINK-BOT, M. L. 2011. Patients' preferences for breast reconstruction: a discrete choice experiment. J Plast Reconstr Aesthet Surg, 64, 75-83.

DAMMAN, O. C., SPREEUWENBERG, P., RADEMAKERS, J. & HENDRIKS, M. 2012. Creating compact comparative health care information: what are the key quality attributes to present for cataract and total hip or knee replacement surgery? Med Decis Making, 32, 287-300.

DARBA, J., RESTOVIC, G., KASKENS, L., BALBONA, M. A., CARBONELL, A., CAVERO, P., JORDANA, M., PRIETO, C., MOLINA, A. & PADRO, I. 2011. Patient preferences for osteoporosis in Spain: a discrete choice experiment. Osteoporos Int, 22, 1947-54.

DE BEKKER-GROB, E. W., BLIEMER, M. C., DONKERS, B., ESSINK-BOT, M. L., KORFAGE, I. J., ROOBOL, M. J., BANGMA, C. H. & STEYERBERG, E. W. 2013. Patients' and urologists' preferences for prostate cancer treatment: a discrete choice experiment. Br J Cancer, 109, 633-40.

DEAL, K., MARSHALL, D., DABROWSKI, D., CHARTER, A., BUKOSKI, M. & MOAYYEDI, P. 2013. Assessing the value of symptom relief for patients with gastroesophageal reflux disease treatment: willingness to pay using a discrete choice experiment. Value Health, 16, 588-98.

DETERMANN, D., KORFAGE, I.J., LAMBOOIJ, M.S., BLIEMER, M., RICHARDUS, J.H., STEYERBERG, E.W. & DE BEKKER-GROB, E.W. 2014. Acceptance of vaccinations in pandemic outbreaks: a discrete choice experiment. PloS one. 9 (7), e102505.

DIXON, S., NANCARROW, S. A., ENDERBY, P. M., MORAN, A. M. & PARKER, S. G. 2015. Assessing patient preferences for the delivery of different community-based models of care using a discrete choice experiment. Health Expect, 18, 1204-14.

DWIGHT JOHNSON, M., APESOA-VARANO, C., HAY, J., UNUTZER, J. & HINTON, L. 2013. Depression treatment preferences of older white and Mexican origin men. Gen Hosp Psychiatry, 35, 59-65.

FEGERT, J. M., SLAWIK, L., WERMELSKIRCHEN, D., NUBLING, M. & MUHLBACHER, A. 2011. Assessment of parents' preferences for the treatment of school-age children with ADHD: a discrete choice experiment. Expert Rev Pharmacoecon Outcomes Res, 11, 245-52.

FLOOD, E. M., RYAN, K. J., ROUSCULP, M. D., BEUSTERIEN, K. M., BLOCK, S. L., HALL, M. C. & MAHADEVIA, P. J. 2011. A survey of children's preferences for influenza vaccine attributes. Vaccine, 29, 4334-40.

FLOOD, E. M., RYAN, K. J., ROUSCULP, M. D., BEUSTERIEN, K. M., DIVINO, V. M., BLOCK, S. L., HALL, M. C. & MAHADEVIA, P. J. 2011. Parent preferences for pediatric influenza vaccine attributes. Clin Pediatr (Phila), 50, 338-47.

GARCIA-DOMINGUEZ, J.M., MUÑOZ, D., COMELLAS, M., GONZALBO, I., LIZÁN, L. & POLANCO SÁNCHEZ, C. 2016. Patient preferences for treatment of multiple sclerosis with disease-modifying therapies: a discrete choice experiment. Patient Prefer Adherence. 101945–1956.

GELHORN, H. L., STRINGER, S. M., BROOKS, A., THOMPSON, C., MONZ, B. U., BOYE, K. S., HACH, T., LUND, S. S. & PALENCIA, R. 2013. Preferences for medication attributes among patients with type 2 diabetes mellitus in the UK. Diabetes Obes Metab, 15, 802-9.

GERARD, K., TINELLI, M., LATTER, S., BLENKINSOPP, A. & SMITH, A. 2012. Valuing the extended role of prescribing pharmacist in general practice: results from a discrete choice experiment. Value Health, 15, 699-707.

GIDENGIL, C., LIEU, T. A., PAYNE, K., RUSINAK, D., MESSONNIER, M. & PROSSER, L. A. 2012. Parental and societal values for the risks and benefits of childhood combination vaccines. Vaccine, 30, 3445-52.

GLENNGARD, A. H., HJELMGREN, J., THOMSEN, P. H. & TVEDTEN, T. 2013. Patient preferences and willingness-to-pay for ADHD treatment with stimulants using discrete choice experiment (DCE) in Sweden, Denmark and Norway. Nord J Psychiatry, 67, 351-9.

GONZALEZ, J. M., JOHNSON, F. R., RUNKEN, M. C. & POULOS, C. M. 2013. Evaluating migraineurs' preferences for migraine treatment outcomes using a choice experiment. Headache, 53, 1635-50.

GOODALL, S., KING, M., EWING, J., SMITH, N. & KENNY, P. 2012. Preferences for support services among adolescents and young adults with cancer or a blood disorder: a discrete choice experiment. Health Policy, 107, 304-11.

GOTO, R., TAKAHASHI, Y. & IDA, T. 2011. Changes in smokers' attitudes toward intended cessation attempts in Japan. Value Health, 14, 785-91.

GUO, N., MARRA, C. A., FITZGERALD, J. M., ELWOOD, R. K., ANIS, A. H. & MARRA, F. 2011. Patient preference for latent tuberculosis infection preventive treatment: a discrete choice experiment. Value Health, 14, 937-43.

HALL, J., KENNY, P., HOSSAIN, I., STREET, D. J. & KNOX, S. A. 2014. Providing Informal Care in Terminal Illness: An Analysis of Preferences for Support Using a Discrete Choice Experiment. Med Decis Making, 34, 731-45.

HANCOCK-HOWARD, R. L., UNGAR, W. J., MARSHALL, D., EINARSON, A. & KOREN, G. 2012. Public preferences for counseling regarding antidepressant use during pregnancy: a discrete choice experiment. Birth Defects Res A Clin Mol Teratol, 94, 532-9.

HAUBER, A. B., ARDEN, N. K., MOHAMED, A. F., JOHNSON, F. R., PELOSO, P. M., WATSON, D. J., MAVROS, P., GAMMAITONI, A., SEN, S. S. & TAYLOR, S. D. 2013. A discrete-choice experiment of United Kingdom patients' willingness to risk adverse events for improved function and pain control in osteoarthritis. Osteoarthritis Cartilage, 21, 289-97.

HAUBER, A. B., GONZALEZ, J. M., SCHENKEL, B., LOFLAND, J. H. & MARTIN, S. 2011. The value to patients of reducing lesion severity in plaque psoriasis. J Dermatolog Treat, 22, 266-75.

HAUBER, A. B., HAN, S., YANG, J. C., GANTZ, I., TUNCELI, K., GONZALEZ, J. M., BRODOVICZ, K., ALEXANDER, C. M., DAVIES, M., IGLAY, K., ZHANG, Q. & RADICAN, L. 2013. Effect of pill burden on dosing preferences, willingness to pay, and likely adherence among patients with type 2 diabetes. Patient Prefer Adherence, 7, 937-49.

HAUBER, A.B., TUNCELI, K., YANG, J.-C., GANTZ, I., BRODOVICZ, K.G., ALEXANDER, C.M., DAVIES, M.J. & RADICAN, L. 2015. A survey of patient preferences for oral antihyperglycemic therapy in patients with type 2 diabetes mellitus. Diabetes Ther., 6 (1), 75–84.

HOLMES, E.A.F., MORRISON, V.L. & HUGHES, D.A. 2016. What influences persistence with medicines? A multinational discrete choice experiment of 2549 patients. Br. J. Clin. Pharmacol., 82 (2), 522–531.

HONG, S. H., LIU, J., WANG, J., BROWN, L. & WHITE-MEANS, S. 2011. Conjoint analysis of patient preferences on Medicare medication therapy management. J Am Pharm Assoc (2003), 51, 378-87.

HOWARD, K., SALKELD, G., PIGNONE, M., HEWETT, P., CHEUNG, P., OLSEN, J., CLAPTON, W. & ROBERTS-THOMSON, I. C. 2011. Preferences for CT colonography and colonoscopy as diagnostic tests for colorectal cancer: a discrete choice experiment. Value Health, 14, 1146-52.

KAUF, T. L., MOHAMED, A. F., HAUBER, A. B., FETZER, D. & AHMAD, A. 2012. Patients' willingness to accept the risks and benefits of new treatments for chronic hepatitis C virus infection. Patient, 5, 265-78.

KING, M. T., VINEY, R., SMITH, D. P., HOSSAIN, I., STREET, D., SAVAGE, E., FOWLER, S., BERRY, M. P., STOCKLER, M., COZZI, P., STRICKER, P., WARD, J. & ARMSTRONG, B. K. 2012. Survival gains needed to offset persistent adverse treatment effects in localised prostate cancer. Br J Cancer, 106, 638-45.

KISTLER, C.E., HESS, T.M., HOWARD, K., PIGNONE, M.P., CRUTCHFIELD, T.M., HAWLEY, S.T., BRENNER, A.T., WARD, K.T. & LEWIS, C.L. 2015. Older adults' preferences for colorectal cancer-screening test attributes and test choice. Patient preference and adherence. 91005–1016.

KJAER, T., BECH, M., KRONBORG, C. & MORKBAK, M. R. 2013. Public preferences for establishing nephrology facilities in Greenland: estimating willingness-to-pay using a discrete choice experiment. Eur J Health Econ, 14, 739-48.

KNIGHT, S.J., MOHAMED, A.F., MARSHALL, D.A., LADABAUM, U., PHILLIPS, K.A. & WALSH, J.M.E. 2015. Value of Genetic Testing for Hereditary Colorectal Cancer in a Probability-Based US Online Sample. Med Decis Mak.

KNOX, S. A., VINEY, R. C., GU, Y., HOLE, A. R., FIEBIG, D. G., STREET, D. J., HAAS, M. R., WEISBERG, E. & BATESON, D. 2013. The effect of adverse information and positive promotion on women's preferences for prescribed contraceptive products. Soc Sci Med, 83, 70-80.

KOBAYASHI, M., MANO, T. & YAMAUCHI, K. 2013. Patients' preference on selecting a medical institution. Int J Health Care Qual Assur, 26, 341-52.

KROMER, C., SCHAARSCHMIDT, M.-L., SCHMIEDER, A., HERR, R., GOERDT, S. & PEITSCH, W.K. 2015. Patient Preferences for Treatment of Psoriasis with Biologicals: A Discrete Choice Experiment. PloS one. 10 (6), e0129120.

KRUCIEN, N., GAFNI, A., FLEURY, B. & PELLETIER-FLEURY, N. 2013. Patients' with obstructive sleep apnoea syndrome (OSAS) preferences and demand for treatment: a discrete choice experiment. Thorax, 68, 487-8.

KRUK, M. E., ROCKERS, P. C., TORNORLAH VARPILAH, S. & MACAULEY, R. 2011. Population preferences for health care in liberia: insights for rebuilding a health system. Health Serv Res, 46, 2057-78.

LABA, T. L., BRIEN, J. A. & JAN, S. 2012. Understanding rational non-adherence to medications. A discrete choice experiment in a community sample in Australia. BMC Fam Pract, 13, 61.

LABA, T. L., BRIEN, J. A., FRANSEN, M. & JAN, S. 2013. Patient preferences for adherence to treatment for osteoarthritis: the MEdication Decisions in Osteoarthritis Study (MEDOS). BMC Musculoskelet Disord, 14, 160.

LANDFELDT, E., JABLONOWSKA, B., NORLANDER, E., PERSDOTTER-EBERG, K., THURIN-KJELLBERG, A., WRAMSBY, M. & STROM, O. 2012. Patient preferences for characteristics differentiating ovarian stimulation treatments. Hum Reprod, 27, 760-9.

LATHIA, N., ISOGAI, P. K., WALKER, S. E., DE ANGELIS, C., CHEUNG, M. C., HOCH, J. S. & MITTMANN, N. 2013. Eliciting patients' preferences for outpatient treatment of febrile neutropenia: a discrete choice experiment. Support Care Cancer, 21, 245-51.

LAVER, K., RATCLIFFE, J., GEORGE, S., BURGESS, L. & CROTTY, M. 2011. Is the Nintendo Wii Fit really acceptable to older people? A discrete choice experiment. BMC Geriatr, 11, 64.

LAVER, K., RATCLIFFE, J., GEORGE, S., LESTER, L., WALKER, R., BURGESS, L. & CROTTY, M. 2011. Early rehabilitation management after stroke: what do stroke patients prefer? J Rehabil Med, 43, 354-8.

LEHMAN, M., GORAYSKI, P., WATSON, S., EDELING, D., JACKSON, J. & WHITTY, J. 2016. Patient preferences regarding prophylactic cranial irradiation: A discrete choice experiment. Radiotherapy and Oncology. 121 (2), 225–231.

LEVITAN, B., MARKOWITZ, M., MOHAMED, A.F., JOHNSON, F.R., ALPHS, L., CITROME, L. & BRIDGES, J.F.P. 2015. Patients' Preferences Related to Benefits, Risks, and Formulations of Schizophrenia Treatment. Psychiatric services. 66 (7), 719–726.

LLEWELLYN, C. D., SAKAL, C., LAGARDE, M., POLLARD, A. & MINERS, A. H. 2013. Testing for sexually transmitted infections among students: a discrete choice experiment of service preferences. BMJ Open, 3, e003240.

LOUDER, A.M., SINGH, A., SAVERNO, K., CAPPELLERI, J.C., ATEN, A.J., KOENIG, A.S. & PASQUALE, M.K. 2016. Patient Preferences Regarding Rheumatoid Arthritis Therapies: A Conjoint Analysis. American health & drug benefits. 9 (2), 84–93.

LYNN, F. A., CREALEY, G. E., ALDERDICE, F. A. & MCELNAY, J. C. 2015. Preferences for a third-trimester ultrasound scan in a low-risk obstetric population: a discrete choice experiment. Health Expect, 18, 892-903.

MANJUNATH, R., YANG, J. C. & ETTINGER, A. B. 2012. Patients' preferences for treatment outcomes of add-on antiepileptic drugs: a conjoint analysis. Epilepsy Behav, 24, 474-9.

MARANG-VAN DE MHEEN, P. J., DIJS-ELSINGA, J., OTTEN, W., VERSLUIJS, M., SMEETS, H. J., VREE, R., VAN DER MADE, W. J. & KIEVIT, J. 2011. The relative importance of quality of care information when choosing a hospital for surgical treatment: a hospital choice experiment. Med Decis Making, 31, 816-27.

MARTI, J. 2012. Assessing preferences for improved smoking cessation medications: a discrete choice experiment. Eur J Health Econ, 13, 533-48.

MATTSSON, L. A., ERICSSON, A., BOGELUND, M. & MAAMARI, R. 2013. Women's preferences toward attributes of local estrogen therapy for the treatment of vaginal atrophy. Maturitas, 74, 259-63.

MEDINA-LARA, A., MUJICA-MOTA, R.E., KUNKWENZU, E.D. & LALLOO, D.G. 2014. Stated preferences for anti-malarial drug characteristics in Zomba, a malaria endemic area of Malawi. Malaria journal. 13 (1), 259.

MILLER, P. J., BALU, S., BUCHNER, D., WALKER, M. S., STEPANSKI, E. J. & SCHWARTZBERG, L. S. 2013. Willingness to pay to prevent chemotherapy induced nausea and vomiting among patients with breast, lung, or colorectal cancer. J Med Econ, 16, 1179-89.

MILTE, R., RATCLIFFE, J., MILLER, M., WHITEHEAD, C., CAMERON, I. D. & CROTTY, M. 2013. What are frail older people prepared to endure to achieve improved mobility following hip fracture? A Discrete Choice Experiment. J Rehabil Med, 45, 81-6.

MINERS, A., LLEWELLYN, C., POLLARD, A., LAGARDE, M., RICHARDSON, D., CAIRNS, J., FISHER, M. & SMITH, H. 2012. Assessing user preferences for sexually transmitted infection testing services: a discrete choice experiment. Sex Transm Infect, 88, 510-6.

MOHAMED, A. F., EPSTEIN, J. D. & LI-MCLEOD, J. M. 2011. Patient and parent preferences for haemophilia A treatments. Haemophilia, 17, 209-14.

MOHAMED, A. F., HAUBER, A. B. & NEARY, M. P. 2011. Patient benefit-risk preferences for targeted agents in the treatment of renal cell carcinoma. Pharmacoeconomics, 29, 977-88.

MOHAMED, A. F., KILAMBI, V., LUO, M. P., IYER, R. G. & LI-MCLEOD, J. M. 2012. Patient and parent preferences for immunoglobulin treatments: a conjoint analysis. J Med Econ, 15, 1183-91.

MOHAMED, A. F., ZHANG, J., JOHNSON, F. R., LOMON, I. D., MALVOLTI, E., TOWNSEND, R., OSTGREN, C. J. & PARHOFER, K. G. 2013. Avoidance of weight gain is important for oral type 2 diabetes treatments in Sweden and Germany: patient preferences. Diabetes Metab, 39, 397-403.

MOHAMED, A.F., JOHNSON, F.R., BALP, M.-M. & CALADO, F. 2016. Preferences and Stated Adherence for Antibiotic Treatment of Cystic Fibrosis Pseudomonas Infections. Patient. 9 (1), 59–67.

MOIA, M., MANTOVANI, L. G., CARPENEDO, M., SCALONE, L., MONZINI, M. S., CESANA, G. & MANNUCCI, P. M. 2013. Patient preferences and willingness to pay for different options of anticoagulant therapy. Intern Emerg Med, 8, 237-43.

MORTON, R. L., SNELLING, P., WEBSTER, A. C., ROSE, J., MASTERSON, R., JOHNSON, D. W. & HOWARD, K. 2012. Dialysis modality preference of patients with CKD and family caregivers: a discrete-choice study. Am J Kidney Dis, 60, 102-11.

MUHLBACHER, A. & BETHGE, S. 2013. Preferences of overweight and obese patients for weight loss programmes: a discrete-choice experiment. Int J Integr Care, 13, e034.

MUHLBACHER, A. C., STOLL, M., MAHLICH, J. & NUBLING, M. 2013. Patient preferences for HIV/AIDS therapy - a discrete choice experiment. Health Econ Rev, 3, 14.

MUSTERS, A. M., DE BEKKER-GROB, E. W., MOCHTAR, M. H., VAN DER VEEN, F. & VAN MELLO, N. M. 2011. Women's perspectives regarding subcutaneous injections, costs and live birth rates in IVF. Hum Reprod, 26, 2425-31.

NAIK-PANVELKAR, P., ARMOUR, C., ROSE, J. & SAINI, B. 2012. Patients' value of asthma services in Australian pharmacies: the way ahead for asthma care. J Asthma, 49, 310-6.

NOBLE, S., MATZDORFF, A., MARAVEYAS, A., HOLM, M.V. & PISA, G. 2015. Assessing patients' anticoagulation preferences for the treatment of cancer-associated thrombosis using conjoint methodology. Haematologica. 100 (11), 1486–1492.

O'HARA, N.N., SLOBOGEAN, G.P., MOHAMMADI, T., MARRA, C.A., VICENTE, M.R., KHAKBAN, A. & MCKEE, M.D. 2016. Are patients willing to pay for total shoulder arthroplasty? Evidence from a discrete choice experiment. Canadian journal of surgery. 59 (2), 107–112.

OTENG, B., MARRA, F., LYND, L. D., OGILVIE, G., PATRICK, D. & MARRA, C. A. 2011. Evaluating societal preferences for human papillomavirus vaccine and cervical smear test screening programme. Sex Transm Infect, 87, 52-7.

PACZKOWSKI, M. M., KRUK, M. E., TESSEMA, F., TEGEGN, A. & GALEA, S. 2012. Depressive symptoms and posttraumatic stress disorder as determinants of preference weights for attributes of obstetric care among Ethiopian women. PLoS One, 7, e46788.

PARK, H., CHON, Y., LEE, J., CHOI IE, J. & YOON, K. H. 2011. Service design attributes affecting diabetic patient preferences of telemedicine in South Korea. Telemed J E Health, 17, 442-51.

PARKINSON, B., GOODALL, S. & NORMAN, R. 2013. Measuring the loss of consumer choice in mandatory health programmes using discrete choice experiments. Appl Health Econ Health Policy, 11, 139-50.

PHILIPS, H., MAHR, D., REMMEN, R., WEVERBERGH, M., DE GRAEVE, D. & VAN ROYEN, P. 2012. Predicting the place of out-of-hours care--a market simulation based on discrete choice analysis. Health Policy, 106, 284-90.

PIGNONE, M. P., BRENNER, A. T., HAWLEY, S., SHERIDAN, S. L., LEWIS, C. L., JONAS, D. E. & HOWARD, K. 2012. Conjoint analysis versus rating and ranking for values elicitation and clarification in colorectal cancer screening. J Gen Intern Med, 27, 45-50.

PIGNONE, M. P., HOWARD, K., BRENNER, A. T., CRUTCHFIELD, T. M., HAWLEY, S. T., LEWIS, C. L. & SHERIDAN, S. L. 2013. Comparing 3 techniques for eliciting patient values for decision making about prostate-specific antigen screening: a randomized controlled trial. JAMA Intern Med, 173, 362-8.

PISA, G., FREYTAG, S. & SCHANDRY, R. 2013. Chronic obstructive pulmonary disease (COPD) patients' disease-related preferences : a study using conjoint analysis. Patient, 6, 93-101.

POSNETT, J., DIXIT, S., OPPENHEIMER, B., KILI, S. & MEHIN, N. 2015. Patient preference and willingness to pay for knee osteoarthritis treatments. Patient Prefer Adherence. 9733–744.

POULOS, C., HAUBER, A. B., GONZALEZ, J. M. & TURPCU, A. 2014. Patients' willingness to trade off between the duration and frequency of rheumatoid arthritis treatments. Arthritis Care Res (Hoboken), 66, 1008-15.

POULOS, C., YANG, J. C., LEVIN, C., VAN MINH, H., GIANG, K. B. & NGUYEN, D. 2011. Mothers' preferences and willingness to pay for HPV vaccines in Vinh Long Province, Vietnam. Soc Sci Med, 73, 226-34.

RENNIE, L., PORTEOUS, T. & RYAN, M. 2012. Preferences for managing symptoms of differing severity: a discrete choice experiment. Value Health, 15, 1069-76.

ROBERTSON, R. & BURGE, P. 2011. The impact of patient choice of provider on equity: analysis of a patient survey. J Health Serv Res Policy, 16 Suppl 1, 22-8.

SADIQUE, M. Z., DEVLIN, N., EDMUNDS, W. J. & PARKIN, D. 2013. The effect of perceived risks on the demand for vaccination: results from a discrete choice experiment. PLoS One, 8, e54149.

SANDERS, P. M., IJZERMAN, M. J., ROACH, M. J. & GUSTAFSON, K. J. 2011. Patient preferences for next generation neural prostheses to restore bladder function. Spinal Cord, 49, 113-9.

SCALONE, L., WATSON, V., RYAN, M., KOTSOPOULOS, N. & PATEL, R. 2011. Evaluation of patients' preferences for genital herpes treatment. Sex Transm Dis, 38, 802-7.

SCHAARSCHMIDT, M. L., SCHMIEDER, A., UMAR, N., TERRIS, D., GOEBELER, M., GOERDT, S. & PEITSCH, W. K. 2011. Patient preferences for psoriasis treatments: process characteristics can outweigh outcome attributes. Arch Dermatol, 147, 1285-94.

SCHAARSCHMIDT, M. L., UMAR, N., SCHMIEDER, A., TERRIS, D. D., GOEBELER, M., GOERDT, S. & PEITSCH, W. K. 2013. Patient preferences for psoriasis treatments: impact of treatment experience. J Eur Acad Dermatol Venereol, 27, 187-98.

SCHIOTZ, M., BOGELUND, M., ALMDAL, T. & WILLAING, I. 2012. Discrete choice as a method for exploring education preferences in a Danish population of patients with type 2 diabetes. Patient Educ Couns, 87, 217-25.

SCHMIEDER, A., SCHAARSCHMIDT, M. L., UMAR, N., TERRIS, D. D., GOEBELER, M., GOERDT, S. & PEITSCH, W. K. 2012. Comorbidities significantly impact patients' preferences for psoriasis treatments. J Am Acad Dermatol, 67, 363-72.

SCHWAPPACH, D. L., MULDERS, V., SIMIC, D., WILM, S. & THURMANN, P. A. 2011. Is less more? Patients' preferences for drug information leaflets. Pharmacoepidemiol Drug Saf, 20, 987-95.

SCOTLAND, G. S., MCNAMEE, P., CHEYNE, H., HUNDLEY, V. & BARNETT, C. 2011. Women's preferences for aspects of labor management: results from a discrete choice experiment. Birth, 38, 36-46.

SHINGLER, S. L., SWINBURN, P., ALI, S., PERARD, R. & LLOYD, A. J. 2013. A discrete choice experiment to determine patient preferences for injection devices in multiple sclerosis. J Med Econ, 16, 1036-42.

STOCKWELL, M. S., ROSENTHAL, S. L., STURM, L. A., MAYS, R. M., BAIR, R. M. & ZIMET, G. D. 2011. The effects of vaccine characteristics on adult women's attitudes about vaccination: a conjoint analysis study. Vaccine, 29, 4507-11.

SUNG, L., ALIBHAI, S. M., ETHIER, M. C., TEUFFEL, O., CHENG, S., FISMAN, D. & REGIER, D. A. 2012. Discrete choice experiment produced estimates of acceptable risks of therapeutic options in cancer patients with febrile neutropenia. J Clin Epidemiol, 65, 627-34.

SWINBURN, P., LLOYD, A., ALI, S., HASHMI, N., NEWAL, D. & NAJIB, H. 2011. Preferences for antimuscarinic therapy for overactive bladder. BJU Int, 108, 868-73.

TERRIS-PRESTHOLT, F., HANSON, K., MACPHAIL, C., VICKERMAN, P., REES, H. & WATTS, C. 2013. How much demand for New HIV prevention technologies can we really expect? Results from a discrete choice experiment in South Africa. PLoS One, 8, e83193.

TINELLI, M., OZOLINS, M., BATH-HEXTALL, F. & WILLIAMS, H. C. 2012. What determines patient preferences for treating low risk basal cell carcinoma when comparing surgery vs imiquimod? A discrete choice experiment survey from the SINS trial. BMC Dermatol, 12, 19.

TOKES, K., QUADRI, S., CAHILL, P., CHIU, G., IVANOV, A. & TANG, H. 2014. Disease and treatment perceptions among Asian Americans diagnosed with chronic hepatitis B infection. Journal of general internal medicine. 29 (3), 477–484.

TONG, B. C., HUBER, J. C., ASCHEIM, D. D., PUSKAS, J. D., FERGUSON, T. B., JR., BLACKSTONE, E. H. & SMITH, P. K. 2012. Weighting composite endpoints in clinical trials: essential evidence for the heart team. Ann Thorac Surg, 94, 1908-13.

TSUNEMATSU, M., KAWASAKI, H., MASUOKA, Y. & KAKEHASHI, M. 2013. Factors affecting breast cancer screening behavior in Japan--assessment using the health belief model and conjoint analysis. Asian Pac J Cancer Prev, 14, 6041-8.

UEMURA, H., MATSUBARA, N., KIMURA, G., YAMAGUCHI, A., LEDESMA, D.A., DIBONAVENTURA, M., MOHAMED, A.F., BASURTO, E., MCKINNON, I., WANG, E., CONCIALDI, K., NARIMATSU, A. & AITOKU, Y. 2016. Patient preferences for treatment of castration-resistant prostate cancer in Japan: a discrete-choice experiment. BMC urology. 16 (1), 63.

UTZ, K.S., HOOG, J., WENTRUP, A., BERG, S., LÄMMER, A., JAINSCH, B., WASCHBISCH, A., LEE, D.-H., LINKER, R.A. & SCHENK, T. 2014. Patient preferences for disease-modifying drugs in multiple sclerosis therapy: a choice-based conjoint analysis. Therapeutic advances in neurological disorders. 7 (6), 263–275.

VAN GILS, P. F., LAMBOOIJ, M. S., FLANDERIJN, M. H., VAN DEN BERG, M., DE WIT, G. A., SCHUIT, A. J., STRUIJS, J. N. & VAN DEN BERG, B. 2011. Willingness to participate in a lifestyle intervention program of patients with type 2 diabetes mellitus: a conjoint analysis. Patient Prefer Adherence, 5, 537-46.

VAN HOUTVEN, G., JOHNSON, F. R., KILAMBI, V. & HAUBER, A. B. 2011. Eliciting benefit-risk preferences and probability-weighted utility using choice-format conjoint analysis. Med Decis Making, 31, 469-80.

VELDWIJK, J., LAMBOOIJ, M. S., VAN GILS, P. F., STRUIJS, J. N., SMIT, H. A. & DE WIT, G. A. 2013. Type 2 diabetes patients' preferences and willingness to pay for lifestyle programs: a discrete choice experiment. BMC Public Health, 13, 1099.

WALTZMAN, J. T., SCHOLZ, T. & EVANS, G. R. 2011. What patients look for when choosing a plastic surgeon: an assessment of patient preference by conjoint analysis. Ann Plast Surg, 66, 643-7.

WASCHBUSCH, D. A., CUNNINGHAM, C. E., PELHAM, W. E., RIMAS, H. L., GREINER, A. R., GNAGY, E. M., WAXMONSKY, J., FABIANO, G. A., ROBB, J. A., BURROWS-MACLEAN, L., SCIME, M. & HOFFMAN, M. T. 2011. A discrete choice conjoint experiment to evaluate parent preferences for treatment of young, medication naive children with ADHD. J Clin Child Adolesc Psychol, 40, 546-61.

WHITTY, J. A., STEWART, S., CARRINGTON, M. J., CALDERONE, A., MARWICK, T., HOROWITZ, J. D., KRUM, H., DAVIDSON, P. M., MACDONALD, P. S., REID, C. & SCUFFHAM, P. A. 2013. Patient preferences and willingness-to-pay for a home or clinic based program of chronic heart failure management: findings from the Which? trial. PLoS One, 8, e58347.

WITTINK, M. N., MORALES, K. H., CARY, M., GALLO, J. J. & BARTELS, S. J. 2013. Towards personalizing treatment for depression : developing treatment values markers. Patient, 6, 35-43.

WONG, M. K., MOHAMED, A. F., HAUBER, A. B., YANG, J. C., LIU, Z., ROGERIO, J. & GARAY, C. A. 2012. Patients rank toxicity against progression free survival in second-line treatment of advanced renal cell carcinoma. J Med Econ, 15, 1139-48.

YEO, S. T., EDWARDS, R. T., FARGHER, E. A., LUZIO, S. D., THOMAS, R. L. & OWENS, D. R. 2012. Preferences of people with diabetes for diabetic retinopathy screening: a discrete choice experiment. Diabet Med, 29, 869-77.

YI, D., RYAN, M., CAMPBELL, S., ELLIOTT, A., TORRANCE, N., CHAMBERS, A., JOHNSTON, M., HANNAFORD, P. & SMITH, B. H. 2011. Using discrete choice experiments to inform randomised controlled trials: an application to chronic low back pain management in primary care. Eur J Pain, 15, 531.e1-10.

ZIMMERMANN, T. M., CLOUTH, J., ELOSGE, M., HEURICH, M., SCHNEIDER, E., WILHELM, S. & WOLFRATH, A. 2013. Patient preferences for outcomes of depression treatment in Germany: a choice-based conjoint analysis study. J Affect Disord, 148, 210-9.

Appendix B. Supplementary Materials for Chapter Six

B-1 Full Online Survey

The notes were originally intended for the online programmers at Research Now (http://www.researchnow.com) and aimed to provide clarifications. In addition, please note that the "Online Panel" sample was previously referred to as the "General Population" sample and the "Trial" sample was previously referred to as the "Patient" sample.

Study information page: Survey of attributes influencing individual decisions to participate in weight loss maintenance programmes

We are inviting you to take part in a research project that is investigating factors that determine participation in weight loss maintenance programmes. This project is being carried out by researchers at Newcastle University.

What is the research about?

Programmes to promote weight loss are highly effective, but people tend to struggle to maintain weight loss over the longer-term. It is important that any weight that is lost is not put back on so that the side effects of being overweight (such as increased risk of diabetes, heart disease and some types of cancer) can be avoided. Approaches to help people keep off the weight they have lost have been developed but we need to show how well they work and what features of these approaches help or hinder people using them.

This research aims to identify which characteristics of weight loss maintenance programmes are most important to people when deciding whether or not to participate in them. For example, some programmes may last for a year and provide feedback via text messages, whereas others may be shorter but provide 'face to face' feedback.

What will it involve and what happens to the information that is collected?

If you decide to take part in this study, we will ask you to complete a survey. The survey will take approximately 20-30 minutes to complete. First you will be presented with a series of questions containing descriptions of two different weight loss maintenance programmes. For each of these questions you will be asked which weight loss maintenance programme you would prefer. Following this, we will ask you questions about whether you would like to participate in a weight loss maintenance programme, questions about yourself (e.g. your age, education, and household income, etc.) and your health, as well as questions about your attitudes and beliefs about obesity in general.

All of the information that you provide in the survey will be strictly confidential and your data will be stored safely and securely. Your individual responses will be anonymous and accessed on a secure network by the researchers from Newcastle University responsible for the survey. Individual responses to this survey will not be published however unidentifiable grouped responses may be used in future publications as a result of this study.

You can withdraw from the study at any time. If you have any questions or comments, please contact David Mott using the contact details below.

By ticking the box below "**I agree to participate in this study**" and continue to the survey you are indicating your agreement with the following:

1. I have read the information about this study
2. I understand that my participation in this study is voluntary
3. I understand that I am free to withdraw my participation at any time during completion of the survey
4. I understand that responses I provide to the survey questions will be anonymous, and that no personally identifiable information about me will appear in any report or article based on the findings of this study

☐ **I <u>agree</u> to participate in this study**

☐ **I do <u>not</u> agree to participate in this study**

CONTINUE TO THE SURVEY

If you require further information about this study, please contact David Mott (PhD Student at Newcastle University):

Email: d.j.mott@newcastle.ac.uk

Telephone: 07545 174428

1. Screening Questions

Note

GENERAL POPULATION SAMPLE:

The aim of these questions is to identify whether the respondent is eligible to respond to the survey. In addition, the questions will be used to sort the respondent into one of three groups:

1. Potential Service User (Required n=200 total, n=50 for each block of "Scenario Questions")
2. Potential Beneficiary (Required n=200 total, n=50 for each block of "Scenario Questions")
3. Non-User (Required n=200 total, n=50 for each block of "Scenario Questions")

PATIENT SAMPLE:

All patients are eligible to respond to the survey and do not need to be grouped in any way (hence less questions in this section for them).

GP & P: Q1. What is your age? _____ years old

GP & P: Q2. What is your gender?

□ Male □ Female

GP: Q3. Are you currently pregnant or breastfeeding?

□ Yes □ No

P: Q4. Please enter your NULevel Trial ID: _____

Note

GENERAL POPULATION SAMPLE:

Respondents must be aged 18+ as indicated by **Q1** to continue the survey.

If a respondent is female as indicated by **Q2** then ask **Q3**, otherwise move to **Q5**.

If the answer to **Q3** is "Yes", the respondent is not eligible to continue the survey.

If the answer to **Q3** is "No", move to **Q5**.

PATIENT SAMPLE:

All above information in **Q1, 2 & 4** is required to continue.

GP & P: Q5. What are your preferred units of measurement for height & weight?

□ Metres (m) / Kilograms (kg) □ Feet and Inches (ft & in) / Stones and Pounds (st & lb)

□ Feet and Inches (ft & in) / Pounds (lb)

Note

GENERAL POPULATION AND PATIENT SAMPLE:

Only use the preferred unit of measurement from **Q5** from here on (i.e. questions on weight in this section and formatting of weight re-gain characteristic in Section 2 scenarios).

GP & P: Q6a. What is your height? _____cm or _____feet _____inches

GP & P: Q6b. Are you certain about your height or is this an estimate/guess?

□ Certain □ Estimate/Guess

GP & P: Q7a. What is your **current** weight? _____lbs or _____st_____lb or _____kg

GP & P: Q7b. Are you certain about your current weight or is this an estimate/guess?

□ Certain □ Estimate/Guess

GP & P: Q8. What is your ethnic group?

White **Black**

□ British

□ Irish

□ Any other White background:

□ Caribbean

□ African

□ Any other Black background:

_____

Mixed Race

□ White & Black Caribbean

□ White & Black African

□ White & Asian

□ Any other mixed background:

_____

Asian or Asian British

□ Indian

□ Pakistani

□ Bangladeshi

□ Chinese

□ Any other Asian background:

_____

Other Ethnic Group

□ Arab

□ Any other ethnic group

_____

Other

□ Prefer not to say

Note

GENERAL POPULATION SAMPLE:

If any part of **Q5-Q8** is missing (response of "Prefer not to say" in **Q8** counts as missing info), the respondent is not eligible to continue the survey as this information is needed to calculate BMI.

Calculating BMI

$$BMI = \frac{Weight\ in\ kilograms}{(Height\ in\ metres)^2}$$

Or

$$BMI = \frac{Weight\ in\ pounds}{(Height\ in\ inches)^2} \times 703$$

(1 stone = 14 pounds & 1 foot = 12 inches)

BMI Classifications

For respondents answering **Q8** as "Indian", "Pakistani", "Bangladeshi", "Chinese" or "Any other Asian background":

Underweight – BMI less than 18.5

Normal weight – BMI of 18.5 up to 22.9

Overweight – BMI of 23 up to 27.4

Obese – BMI of 27.5 and above

For all other ethnic groups in **Q8**:

Underweight – BMI less than 18.5

Normal weight – BMI of 18.5 up to 24.9

Overweight – BMI of 25 up to 29.9

Obese – BMI of 30 and above

Grouping according to BMI classification

If they are **underweight**, or **weigh over 175kg/385lbs/27st 7lbs**, they are not eligible to continue the survey.

If they are **normal weight**, they are a non-user and only answer **Q9** (also **Q10** if answer to **Q9** is "no"). This is the end of the screening questions.

If they are **overweight**, they are a potential beneficiary and only answer **Q9** (also **Q10** if answer to **Q9** is "no"). This is the end of the screening questions.

If they are **obese** they answer **Q9:**

If the answer is "no", they are a potential beneficiary and answer **Q10** only. This is the end of the screening questions.

If the answer is "yes" they move to **Q11**. If the answer to **Q11** indicates that in the past 12 months they have maintained 5% weight loss (i.e. weight reported in **Q7a** is at least 5% lower than weight reported in **Q11a**), they move to **Q12.** If not, they are a potential beneficiary and this is the end of screening questions.

If those that move to **Q12** answer "no" they are a potential beneficiary and this is the end of screening questions. If they answer "yes", they move to **Q13**.

If those that move to **Q13** answer "no" they are a potential beneficiary and this is the end of screening questions. If they answer "yes", they are a potential service user and this is the end of the screening questions.

PATIENT SAMPLE:

All of the information in **Q5-8** is required to continue. **Q8** is the end of this section for this sample.

GP: Q9. Have you attempted to lose weight in past 12 months?

□ Yes □ No

GP: Q10. Have you ever attempted to lose weight?

□ Yes □ No

GP: Q11a. What was your highest weight in the past 12 months? _____ lbs

___st____lbs or _____kg

GP: Q11b. Are you certain about your highest weight in the past 12 months or is this an estimate/guess? □ Certain □ Estimate/Guess

GP: Q12. Are you able to use a set of scales to weigh yourself? (It doesn't matter if you don't own a set of scales)

□ Yes □ No

GP: Q13. Do you have access to a mobile phone that can connect to the Internet?

□ Yes □ No

2. Scenario Questions

Note

THIS SECTION IS IDENTICAL FOR BOTH THE GENERAL POPULATION AND PATIENT SAMPLE

Weight Loss Maintenance

A lot of people successfully lose weight in order to feel better and to improve their health. Despite this success it can be difficult to keep the weight off over time. This is particularly true when people reach their target weight and try to remain at this weight - this is called "weight maintenance".

Research suggests that most people that have lost weight tend to put the weight back on within 3-5 years. Programmes have been developed in order to help prevent this happening to people that have lost weight – these are called "weight loss maintenance programmes".

The questions in this section will ask you about your preferences for different kinds of "weight loss maintenance programmes". The programmes described in this section are based on a real weight loss maintenance programme that has been tested. You will be asked to make choices between different versions of this programme, as if you were making the choice to take part in a programme in real life.

It doesn't matter if you do not think you need to take part in a programme, please imagine that you do need a programme (i.e. you have lost weight, and want to keep it off).

Weight Loss Maintenance Programmes

In every weight loss maintenance programme described in this section, if you were to take part you would be provided with a set of scales to weigh yourself with. You would be asked to weigh yourself using these scales every day and to not let anybody else use the scales. The scales automatically send information about your weight to the programme organisers.

Other parts of the programmes vary:

1. The **length** of each programme will vary from 6 to 24 months. You would not be forced to stay in the programme for the full length of time.
2. In most programmes you would be reminded if you forget to weigh yourself. This would only happen if you didn't weigh yourself within a 48 hour period, and you would only receive one reminder. The way in which these **reminders** are sent to you will vary depending on the programme.
3. In every programme you would receive some **feedback** from the programme staff about your progress. The amount of feedback you receive will depend on your progress, but this would normally be around twice a month. The way in which you receive this would vary depending on the programme.
4. Some programmes may provide you with an **online tool** (e.g. website or mobile phone application) to track your weight independently. In some programmes, it may be the case that you only receive reminders to weigh yourself and/or feedback on your progress from these online tools.
5. Remember that it is normal for people to re-gain weight following weight loss. The amount of **weight re-gain** that you would expect to experience would vary depending on the programme from 0% to 100% of your initial weight loss. The amount of weight re-gain is based on you participating in the programme for the full length of time.
6. Each programme would have a different **monthly cost**, which would be paid for by yourself if you were to take part.

Ranking Exercise

The characteristics of the weight loss maintenance programmes described in this section are summarised in the table below, along with the possible options for each characteristic.

Characteristic	**Summary**	**Possible Options**
Length of the Programme	How long the programme will last in total (in months).	6 months; 12 months; 18 months; 24 months
Delivery of Reminders to Weigh Yourself	How you are reminded if you forget to weigh yourself for over 48 hours (if you are reminded at all).	No reminders; via phone call; via text message; via the online tool(s)
Delivery of Feedback from Programme Staff	How you will receive feedback on your progress from programme staff.	Via phone call; via text message; via the online tool(s); face to face
Availability of Online Tools to Track Your Progress	The type of online tool(s) provided, if any, so that you can track your progress independently.	No online tool; website only; mobile phone application only; website & mobile phone application
Weight Re-gain	The amount of weight that you re-gain, as a percentage of the amount that you lost originally.	0%; 10%; 20%; 40%; 60%; 80%; 90%; 100%
Personal Cost (per month)	The cost, to you, of the programme each month.	£0; £10; £20; £30

Q1. Please rank these characteristics in order of importance to you, where 1 represents the "most important" characteristic and 6 represents the "least important" characteristic.

1. 5.

2. 6.

3.

4.

Q2a. Please rank the possible options for **"Length of the Programme"** in order of your preference, where 1 represents the "most preferred" characteristic and 4 represents the "least preferred" characteristic.

1.
2.
3.
4.

Q2b. Please rank the possible options for **"Delivery of Reminders"** in order of your preference, where 1 represents the "most preferred" characteristic and 4 represents the "least preferred" characteristic.

1.
2.
3.
4.

Q2c. Please rank the possible options for **"Delivery of Feedback"** in order of your preference, where 1 represents the "most preferred" characteristic and 4 represents the "least preferred" characteristic.

1.
2.
3.
4.

Q2d. Please rank the possible options for **"Availability of Online Tools"** in order of your preference, where 1 represents the "most preferred" characteristic and 4 represents the "least preferred" characteristic.

1.
2.
3.
4.

Q2e. Please rank the possible options for **"Weight Re-Gain"** in order of your preference, where 1 represents the "most preferred" characteristic and 8 represents the "least preferred" characteristic.

1.
2.
3.
4.
5.
6.
7.
8.

Q2f. Please rank the possible options for **"Personal Cost (per month)"** in order of your preference, where 1 represents the "most preferred" characteristic and 4 represents the "least preferred" characteristic.

1.
2.
3.
4.

Note

These questions should be drag & drop exercises

Practice Scenario

This is a practice scenario to show you how the questions will be set out. If you forget what a characteristic means, you can hover over the name of it e.g. "Length of Programme" to see its definition again.

Imagine that you lost 10% of your current body weight. This would mean that you lost **X** of weight. Normally when people lose weight, they put all of it back on.

Programme A and Programme B could help you to avoid putting this **X** of weight back on by asking you to weigh yourself every day (using scales that would be given to you) and providing you with regular feedback about your weight. There may also be online tools that you can use to track your weight on your own.

Imagine that you are offered the two programmes described below. Try and decide which programme you would want to take part in if this happened in real life. If you would not want to take part in either programme, you can select "No Programme".

	Programme A	Programme B	No Programme
Length of the Programme	18 months	12 months	Most people re-gain 100% of their weight loss without a programme - this would be **X** for you
Delivery of Feedback from Programme Staff	Via phone call	Via text message	
Delivery of Reminders to Weigh Yourself	Via the online tool	Via phone call	
Availability of Online Tools to Track Your Progress	Mobile application only	No online tool	

Weight Re-Gain	40% (this would be **40%·X** for you)	60% (this would be **60%·X** for you)	
Personal Cost	£20 a month	£10 a month	
1. Which option would you choose?	□	□	□
2. Of the remaining two, which option would you choose?	□	□	□

Note

In the actual practice scenario Programme A will consist of the characteristics ranked as most preferred in the ranking scenario, and Programme B will consist of the characteristics ranked as least preferred.

X is 10% of the respondent's current weight, calculated from **Q7** in the screening questions section.

The unit of measurement used to describe **X** will be based on the answer to **Q5** in the screening questions section.

Within the table in each scenario **X** will be multiplied by the percentage in the "Weight Re-Gain" cells for Programmes A and B.

Formatting requests:

Only show question (1) initially i.e. all 3 options.

After the respondent has answered (1) please remove the column that corresponds to the choice made in (1) and then present question (2).

Also, please present the definitions of each attribute from page 2 of this section (not the shorter summaries) when a respondent's mouse hovers over the related cell in the table.

Note

The following 13 questions should be in the same format as the "Practice Scenario", containing all of the text apart from the *italic text* at the very top.

The question IDs e.g. B1-1/B2-1/B3-1/B4-1 should not be visible to respondents, but it would be useful to have this identifier in the data once the survey responses have been collected.

GENERAL POPULATION SAMPLE:

The 13 questions will vary depending on the block that is randomly assigned to the respondent (1, 2, 3 or 4).

PATIENT SAMPLE:

The 13 questions will vary depending on the block that is assigned to the respondent (1, 2, 3 or 4), which has been predetermined based on their NULevel Trial ID (**Q4** in Section 1). The excel spreadsheet "Patients – Assigned Blocks.xlsx" shows which block to use according to each possible NULevel Trial ID.

GENERAL POPULATION AND PATIENT SAMPLE:

The scenario questions for each block can be found in the following files:

Scenario Questions – Block 1.docx

Scenario Questions – Block 2.docx

Scenario Questions – Block 3.docx

Scenario Questions – Block 4.docx

In other words, block 1 questions are the same for a general population respondent or a patient respondent.

Difficulty Question

How difficult did you find the scenario questions to answer?

Very difficult	Moderately difficult	Slightly difficult	Neither difficult nor easy	Slightly easy	Moderately easy	Very easy
□	□	□	□	□	□	□

Interest Question

Would you like a weight loss maintenance programme, similar to those described in the choice scenarios, to be made available? (Please choose one option)

□ [1] Yes – because I would like to use a programme like this

□ [2] Yes – because I would like to use a programme like this & I want it to be available for others to use too

□ [3] Yes – because I might want to use a programme like this in future

□ [4] Yes – because I might want to use a programme like this in future & I want it to be available for others to use too

□ [5] Yes – I would not use it myself, but I want it to be available for others to use

□ [6] No – I do not think a programme like this should be made available

Note

If the answer to this 1, 2, 3 or 4 then the respondents move to **Section 3 – Version 1**.

If the answer to this is 5, respondents move to **Section 3 – Version 2**.

If the answer to this is 6, respondents move to **Section 3 – Version 3**.

3. Willingness to Pay Questions

Note

THIS SECTION IS IDENTICAL FOR BOTH THE GENERAL POPULATION AND PATIENT SAMPLE

VERSION 1

According to your responses earlier in the survey, the following weight loss maintenance programme contains some of your preferred characteristics:

	Your Favourite Programme
Delivery of Feedback	Rank 1 from Q2b in "Ranking Exercise"
Delivery of Reminders	Rank 1 from Q2c in "Ranking Exercise"
Availability of Online Tools	Rank 1 from Q2d in "Ranking Exercise"
Weight Re-gain	0%

Note

In the weight loss maintenance programme described, the levels come from **Q2b to Q2d** in the **"Ranking Exercise"** in **Section 2**. They are the levels assigned the rank **"1"** in these questions (i.e. the most preferred).

Q1a. Would you be **willing to pay** to take part in this weight loss maintenance programme if it was offered to you?

□ Yes □ No

If the answer to Q1a is "Yes" then move to Q1b, if "No" then move to Q9.

Q1b. What is the **maximum amount** you would be willing to pay (per month, for 12 months) to take part in this weight loss maintenance programme?

£_____ a month (for 12 months) □ I don't know

Now move to Q2a.

Q2-Q8. See note

Note

The following questions, **Q2-8**, will follow the same format as **Q1a** and **Q1b**. However, the level of **"Weight Re-Gain"** will be adjusted from 10% to 20% to 40% to 60% to 80% to 90% and, finally, to 100% in **Q8**.

If during any of the part a's, the respondent states that they would **not** be willing to pay for the programme, the respondent will move to **Q9**.

Now move to Q9.

Q9. Do you have any comments?

_____

_____

_____

Now move to Section 4.

3. Willingness to Pay Questions

Note

THIS SECTION IS IDENTICAL FOR BOTH THE GENERAL POPULATION AND PATIENT SAMPLE

VERSION 2

Q1a. Would you be **willing to pay** (as a one-off payment) to enable a weight loss maintenance programme to be provided to those that need it?

□ Yes □ No

If the answer to Q1a is "Yes", move to Q1b. If the answer is "No", move to Q2.

Q1b. What is the **maximum amount** that you would be willing to pay (as a one-off payment) in order to enable a weight loss maintenance programme to be provided to one individual for 12 months?

£_____ (one-off payment) □ I don't know

Q2. Do you have any comments?

_____

_____

_____

_____

Now move to Section 4.

3. Willingness to Pay Questions

Note

THIS SECTION IS IDENTICAL FOR BOTH THE GENERAL POPULATION AND PATIENT SAMPLE

VERSION 3

Q1. Do you have any comments?

_____

_____

_____

_____

Now move to Section 4.

4. About You

Note

THIS SECTION IS IDENTICAL FOR BOTH THE GENERAL POPULATION AND PATIENT SAMPLE

These questions ask about you and the people you live with (your household).

Q1. What is your current marital status?

□ Married □ Co-habiting □ Separated

□ Divorced □ Widowed □ Never Married

Q2. What was your highest level of qualification when you finished your education?

□ Masters/PGCE/PhD □ 1st Degree (e.g. BA) □ HND/HNC

□ A Level □ GCSE/O Level/CSE □ None

□ Other:_____

Q3. How many dependent children (under 16 years) are you responsible for? _____

Q4. What is the age of the youngest dependent child under the age of 16 for which you are responsible? _____ years

Q5. Including dependent children, how many people currently live with you? _____

Q6. What is your current employment status?

□ Full employment □ Part-time employment □ Student

□ Retired □ Homemaker □ Caring for someone

□ Unemployed (not actively seeking work) □ Unemployed (actively seeking work)

□ Other:_____

Q7. Please estimate your annual household income from all sources (before tax and including your partner/spouse and any other employed household members)

□ Under £10,000 □ £10,001-£20,000 □ £20,001-£30,000 □ £30,001-£40,000

□ £40,001-£50,000 □ £50,001-£60,000 □ £60,001-£70,000 □ Above £70,000

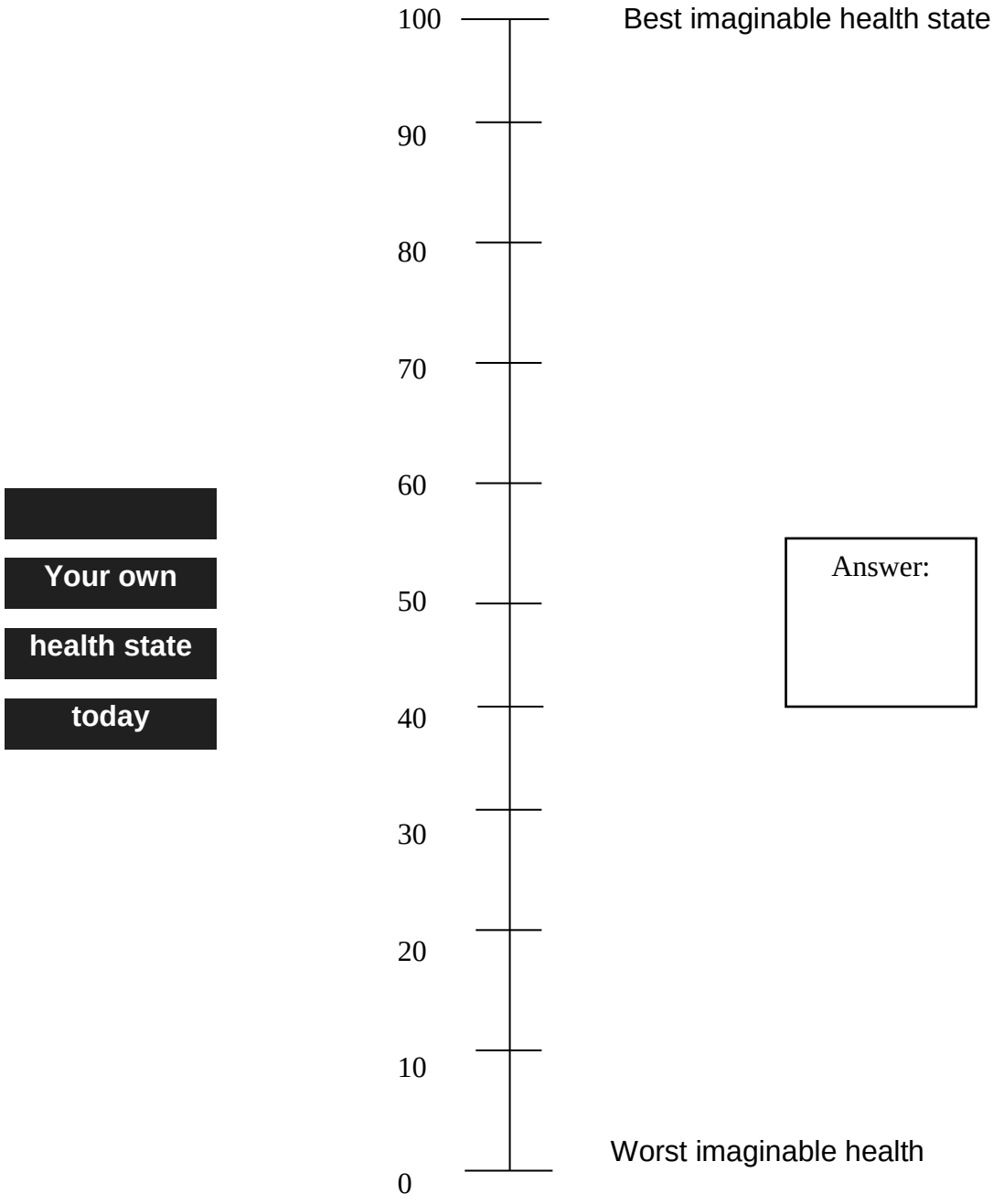
Now move to Section 5.

5. About Your Health

These questions will ask you about your health.

GP & P: Q1. To help people say how good or bad a health state is, we have drawn a scale (rather like a thermometer) on which the best state you can imagine is marked 100 and the worst state you can imagine is marked 0.

We would like you to indicate on this scale how good or bad your own health is today, in your opinion. Please do this by drawing a line from the box below to whichever point on the scale indicates how good or bad your health state is today.



GP & P: Q2. Under each heading, please tick the **ONE** box that best describes your health **TODAY**.

MOBILITY

□ I have no problems in walking about

□ I have slight problems in walking about

□ I have moderate problems in walking about

□ I have severe problems in walking about

□ I am unable to walk about

SELF-CARE

□ I have no problems washing or dressing myself

□ I have slight problems washing or dressing myself

□ I have moderate problems washing or dressing myself

□ I have severe problems washing or dressing myself

□ I am unable to wash or dress myself

USUAL ACTIVITIES (e.g. work, study, housework, family or leisure activities)

□ I have no problems doing my usual activities

□ I have slight problems doing my usual activities

□ I have moderate problems doing my usual activities

□ I have severe problems doing my usual activities

□ I am unable to do my usual activities

PAIN / DISCOMFORT

□ I have no pain or discomfort

□ I have slight pain or discomfort

□ I have moderate pain or discomfort

□ I have severe pain or discomfort

□ I have extreme pain or discomfort

ANXIETY / DEPRESSION

□ I am not anxious or depressed

□ I am slightly anxious or depressed

□ I am moderately anxious or depressed

□ I am severely anxious or depressed

□ I am extremely anxious or depressed

GP & P: Q3. Which of the following categories do you believe you currently fit into?

□ Underweight □ Healthy Weight □ Overweight □ Obese

GP: Q4. Do you regularly play intensive sport or "body build"?

□ Yes □ No

GP: Q5. Have you ever taken part in a weight loss maintenance programme (of any kind)?

□ Yes □ No

Now move to Section 6.

Section 6: Attitudes & Beliefs Surrounding Obese Persons and Related Issues

Note

THIS SECTION IS IDENTICAL FOR BOTH THE GENERAL POPULATION AND PATIENT SAMPLE

	I strongly agree	I moderately agree	I slightly agree	I slightly disagree	I moderately disagree	I strongly disagree
Obesity often occurs when eating is used as a form of compensation for lack of love or attention.						
In many cases, obesity is the result of a biological disorder.						
Obesity is usually caused by overeating.						
Most obese people cause their problem by not getting enough exercise.						
Most obese people eat more than non-obese people.						
The majority of obese people have poor eating habits that lead to their obesity.						
Obesity is rarely caused by a lack of willpower.						

People can be addicted to food, just as others are addicted to drugs, and these people usually become obese.						

	True	False
It is sometimes hard for me to go on with my work if I am not encouraged.		
I sometimes feel resentful when I don't get my way.		
On a few occasions, I have given up doing something because I thought too little of my ability.		
There have been times when I felt like rebelling against people in authority even though I knew they were right.		
No matter who I'm talking to, I'm always a good listener.		
There have been occasions when I took advantage of someone.		
I'm always willing to admit it when I make a mistake.		
I sometimes try to get even rather than forgive and forget.		
I am always courteous, even to people who are disagreeable.		
I have never been irked when people expressed ideas very different from my own.		
There have times when I was quite jealous of the good fortune of others.		
I am sometimes irritated by people who ask favours of me.		
I have never deliberately said something that hurt someone's feelings.		

Q. To what extent do you agree with the following statement?

"Answers to this survey will affect the organisation of/delivery of weight loss maintenance programmes"

Strongly disagree	Moderately disagree	Slightly disagree	Neither agree or disagree	Slightly agree	Moderately agree	Strongly agree
□	□	□	□	□	□	□

Thank you for taking the time to complete this survey.

B-2 Ngene Code

```
;alts = alt1, alt2, alt3
;block = 4
;rows = 40
;eff = (mnl,d)
;cond:
if(alt1.ONLINE=3,alt1.FEEDBACK=[0,1,3]),
if(alt2.ONLINE=3,alt2.FEEDBACK=[0,1,3]),
if(alt1.ONLINE=3,alt1.REMINDERS=[0,1,3]),
if(alt2.ONLINE=3,alt2.REMINDERS=[0,1,3])
;model:
U(alt1) = b2 * LENGTH[6,12,18,24] + b3.dummy[0|0|0] * FEEDBACK[0,1,2,3] +
b4.dummy[0|0|0] * REMINDERS[0,1,2,3] + b5.dummy[0|0|0] * ONLINE[0,1,2,3]
+ b6[(u,-0.001,0)] * OUTCOME[0,10,20,40,60,80,90,100] + b7[(u,-0.001,0)] *
COST[0,10,20,30] + b8 * COST * LENGTH + b9 * OUTCOME * LENGTH /
U(alt2) = asc1 + b2 * LENGTH + b3 * FEEDBACK + b4 * REMINDERS + b5 *
ONLINE + b6 * OUTCOME + b7 * COST + b8 * COST * LENGTH + b9 * OUTCOME
* LENGTH /
U(alt3) = asc2 + b6 * OUTCOME1[100] + b7 * COST1[0] $
```

B-3 DCE Experimental Design

		Alternative 1						Alternative 2					
choice	**block**	**length**	**feedback**	**reminder**	**online**	**outcome**	**cost**	**length**	**feedback**	**reminder**	**online**	**outcome**	**cost**
1	3	18	2	1	1	60	10	12	3	0	3	40	20
2	1	24	0	1	0	100	30	24	2	3	1	0	0
3	3	6	3	3	1	0	0	6	2	0	2	100	30
4	4	6	2	0	1	10	0	18	3	2	2	80	20
5	1	18	1	2	0	40	20	12	3	1	3	60	10
6	4	6	3	2	1	100	0	6	2	1	2	0	30
7	4	18	0	1	2	60	10	12	2	2	0	40	20
8	3	12	2	3	2	80	20	18	3	2	1	20	10
9	4	12	1	0	3	80	10	18	0	3	1	20	20
10	3	18	2	3	1	90	10	12	3	2	0	10	20
11	2	12	2	2	0	80	10	18	3	3	2	20	20
12	3	24	2	1	0	0	0	24	1	2	1	100	30
13	4	24	3	1	0	90	30	6	0	2	1	90	20
14	4	24	1	3	3	0	30	24	0	0	0	100	0
15	1	18	0	3	3	40	20	6	3	1	1	90	0
16	4	24	1	2	2	0	30	24	3	3	3	100	0

		Alternative 1						Alternative 2					
choice	**block**	**length**	**feedback**	**reminder**	**online**	**outcome**	**cost**	**length**	**feedback**	**reminder**	**online**	**outcome**	**cost**
17	2	12	3	0	3	40	20	18	0	3	0	60	10
18	3	24	2	0	2	90	30	12	0	2	0	20	10
19	3	18	3	1	3	20	10	6	1	3	1	90	30
20	2	24	3	3	2	100	0	24	2	1	1	0	30
21	2	18	1	0	2	60	10	12	0	1	3	40	20
22	2	24	0	0	3	10	0	6	2	3	2	10	0
23	1	6	1	1	3	10	30	24	0	0	2	10	20
24	2	12	2	3	0	60	20	18	1	1	1	40	10
25	3	18	3	3	0	20	20	12	1	0	3	60	10
26	4	6	0	0	0	10	30	24	1	2	2	10	0
27	1	12	3	2	2	20	30	12	1	0	0	80	0
28	2	18	2	2	1	80	20	12	1	3	0	20	10
29	3	24	1	2	1	90	10	6	3	0	2	90	0
30	1	6	2	2	2	100	0	6	1	0	1	0	30
31	1	24	1	0	0	10	0	18	0	3	3	80	30
32	1	12	3	0	1	20	20	18	1	3	3	80	10
33	3	12	1	1	2	80	10	24	0	0	3	10	30

		Alternative 1						Alternative 2					
choice	**block**	**length**	**feedback**	**reminder**	**online**	**outcome**	**cost**	**length**	**feedback**	**reminder**	**online**	**outcome**	**cost**
34	4	6	0	0	1	100	30	6	2	2	2	0	0
35	4	12	3	3	3	40	20	18	0	1	2	60	10
36	2	6	0	0	2	0	0	6	3	1	0	100	30
37	2	6	0	3	0	90	0	24	2	2	1	90	0
38	1	6	0	2	2	20	30	24	2	0	0	80	30
39	2	12	0	1	3	60	10	18	3	3	0	40	20
40	1	18	3	0	2	40	20	12	1	1	3	60	10

B-4 Beliefs about Obese Persons Scale Scoring System

Step 1: Multiply the response to the following items by -1 (i.e., reverse the direction of scoring): Item1, Items 3 through Item 6, Item 8

Step 2: Sum the responses to all items.

Step 3: Add 24 to the value obtained in Step 2. This value is the BAOP score.

Higher numbers indicate a stronger belief that obesity is not under the obese person's control.

These measures and additional psychometric information can be found in the following reference: Allison, DB. Handbook of Assessment Methods for Eating Behaviors and Weight-Related Problems. Measures, Theory, and Research. Thousand Oaks, CA: Sage Publications.

B-5 Marlowe Crowne Scale Scoring System

Add 1 point to the score for each true response to statements 5, 7, 9, 10, and 13. Add 0 points to the score for each false response to these statements.

Add 1 point to the score for each false response to statements 1, 2, 3, 4, 6, 8, 11, and 12. Add 0 points to the score for each true response to these statements.

See: Andrews, P. & Meyer, R.G. (2003) Marlowe-Crowne Social Desirability Scale and short Form C: forensic norms. *Journal of clinical psychology*. 59 (4), 483–492.

B-6 Ethics Review – Comments from Reviewers

Faculty of Medical Sciences

Application Case No. 00874

Reviewer One

Name of Reviewer: Mark Pearce

Date: 18/05/2015

Reviewer's Report

1. HAS THE RESEARCHER ADEQUATELY ADDRESSED THE ETHICAL ISSUES RAISED BY THIS PROPOSAL?

No

My only real concerns are for the information sheet

Why is the name of the research survey company crossed out?

I think the section on data anonymity needs to be clearer. What's currently written is true for publications (i.e. aggregated data), but individual level data are to be held by the survey company and, presumably, to be accessed in that way by the researcher. I don't have a problem with them doing this, assuming all data are secure, etc, but this should be made clear in the information sheet

Faculty of Medical Sciences

Application Case No. 00874

Reviewer Two

Name of Reviewer: Ruben Thanacoody

Date: 01/05/2015

Reviewer's Report

1. HAS THE RESEARCHER ADEQUATELY ADDRESSED THE ETHICAL ISSUES RAISED BY THIS PROPOSAL?

YES

However, note that this ethics application relate to the survey for the general public only. An ethics amendment is required for patient in the NULevel trial.

Part 5 od [sic] survey. Please ensure that the note that the questions being asked are based on Latner et al. is included in the online survey. Some people might find the term "fat people" rather derogatory.

I confirm that there are no outstanding ethical issues with this proposal and I recommend it for ethical approval for the period stated in the application.

B-7 Ethical Approval Letter

David Mott

Institute of Health & Society

Faculty of Medical Sciences
Newcastle University
The Medical School
Framlington Place
Newcastle upon Tyne
NE2 4HH United Kingdom

FACULTY OF MEDICAL SCIENCES: ETHICS COMMITTEE

Dear David,

Title: General public preferences towards weight loss maintenance programmes: a discrete choice experiment
Application No: 00874 2015
Start date to end date: 30-04-2015 to 01-10-2016

On behalf of the Faculty of Medical Sciences Ethics Committee, I am writing to confirm that the ethical aspects of your proposal have been considered and your study has been given ethical approval.

The approval is limited to this project: **00874/2015**. If you wish for a further approval to extend this project, please submit a re-application to the FMS Ethics Committee and this will be considered.

During the course of your research project you may find it necessary to revise your protocol. Substantial changes in methodology, or changes that impact on the interface between the researcher and the participants must be considered by the FMS Ethics Committee, prior to implementation.*

At the close of your research project, please report any adverse events that have occurred and the actions that were taken to the FMS Ethics Committee.*

Best wishes,
Yours sincerely

K. Sutherland

Kimberley Sutherland
On behalf of Faculty Ethics Committee

cc.
Professor Daniel Nettle, Chair of FMS Ethics Committee
Ms Lois Neal, Assistant Registrar (Research Strategy)

*Please refer to the latest guidance available on the internal Newcastle web-site.

tel: +44 (0) 191 222 6000
fax: +44 (0) 191 222 6621

www.ncl.ac.uk
The University of Newcastle upon Tyne trading as Newcastle University

The Queen's Anniversary Prizes
for Higher and Further Education
2009

Appendix C. Supplementary Materials for Chapter Seven

C-1 Conditional Logit Models (Main Effects)

Variables	(1) All	(2) Treatment	(3) Control
Alternative-Specific Constant	-0.209	-0.219	-0.194
	(0.222)	(0.327)	(0.305)
Length (months)	0.00924	0.0173	0.00204
	(0.00771)	(0.0113)	(0.0107)
Feedback[1]			
via Phone Call	0.0722	-0.0749	0.198
	(0.142)	(0.212)	(0.193)
via the Online Tool(s)	-0.250*	-0.536**	-0.00521
	(0.146)	(0.220)	(0.198)
via Text Message	0.532***	0.396**	0.670***
	(0.132)	(0.193)	(0.183)
Reminders[2]			
via Text Message	0.400***	0.364*	0.441**
	(0.129)	(0.192)	(0.175)
via Phone Call	0.401***	0.300	0.494**
	(0.151)	(0.226)	(0.204)
via the Online Tool(s)	0.553***	0.620***	0.506**
	(0.149)	(0.225)	(0.202)
Online Tool[3]			
App Only	0.183	0.161	0.198
	(0.139)	(0.205)	(0.191)
Website Only	0.0464	-0.0579	0.146
	(0.151)	(0.224)	(0.207)
App & Website	0.221*	0.0136	0.393**
	(0.127)	(0.191)	(0.173)
Outcome (% weight re-gain)	-0.0253***	-0.0250***	-0.0258***
	(0.00184)	(0.00267)	(0.00257)
Cost (£ per month)	-0.0723***	-0.0700***	-0.0754***
	(0.00560)	(0.00801)	(0.00793)
Sample Size (Observations)	3,390	1,560	1,830
Log Likelihood	-923.9	-430.4	-487.1

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1;
[1] Base level "face to face" feedback; [2] Base level "no reminders"; [3] Base level "no online tool"

C-2 Conditional Logit Models (Main Effects & Attribute Interactions)

Variables	(1) All	(2) Treatment	(3) Control
Alternative-Specific Constant	-0.294	-0.248	-0.328
	(0.266)	(0.393)	(0.366)
Length (months)	-0.0310*	-0.0257	-0.0327
	(0.0182)	(0.0260)	(0.0260)
Feedback[1]			
via Phone Call	0.00771	-0.113	0.121
	(0.146)	(0.219)	(0.199)
via the Online Tool(s)	-0.330**	-0.569**	-0.108
	(0.155)	(0.231)	(0.211)
via Text Message	0.449***	0.353*	0.563***
	(0.136)	(0.200)	(0.189)
Reminders[2]			
via Text Message	0.416***	0.370*	0.457***
	(0.129)	(0.191)	(0.176)
via Phone Call	0.376**	0.301	0.446**
	(0.151)	(0.226)	(0.205)
via the Online Tool(s)	0.635***	0.707***	0.574***
	(0.156)	(0.233)	(0.212)
Online Tool[3]			
App Only	0.0691	0.0690	0.0759
	(0.145)	(0.212)	(0.202)
Website Only	0.00547	-0.114	0.117
	(0.155)	(0.230)	(0.211)
App & Website	0.186	-0.0322	0.368**
	(0.130)	(0.195)	(0.177)
Outcome (% weight re-gain)	-0.0353***	-0.0331***	-0.0372***
	(0.00480)	(0.00681)	(0.00683)
Cost (£ per month)	-0.103***	-0.0952***	-0.111***
	(0.0143)	(0.0205)	(0.0202)
Interaction: Length*Outcome	0.000441**	0.000507	0.000347
	(0.000216)	(0.000314)	(0.000304)
Interaction: Length*Cost	0.00132**	0.00157*	0.00103
	(0.000640)	(0.000947)	(0.000889)
Interaction: Outcome*Cost	0.000166	-7.88e-05	0.000391
	(0.000186)	(0.000273)	(0.000257)
Sample Size (Observations)	3,390	1,560	1,830
Log Likelihood	-920.4	-428.6	-484.8

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1;
[1] Base level "face to face" feedback; [2] Base level "no reminders"; [3] Base level "no online tool"

C-3 Mixed Logit Models (Main Effects & Demographic Interactions)

	All Participants		Treatment Arm		Control Arm	
Variables	Coeff.	Std. Dev.	Coeff.	Std. Dev.	Coeff.	Std. Dev.
Alternative-Specific Constant	2.702	2.437***	4.587	-2.594**	21.393***	8.150***
	(4.136)	(0.530)	(3.713)	(1.080)	(6.295)	(2.016)
Length (months)	0.012	0.088***	0.047	0.162***	-0.038	0.352***
	(0.017)	(0.022)	(0.034)	(0.043)	(0.050)	(0.086)
Feedback[1]						
via Phone Call	0.621**	-0.532	0.461	0.951	2.481**	4.095***
	(0.290)	(0.467)	(0.507)	(0.596)	(1.201)	(1.335)
via the Online Tool(s)	-0.170	1.843***	-0.885	-0.363	0.856	5.938***
	(0.328)	(0.496)	(0.541)	(0.527)	(0.929)	(1.456)
via Text Message	1.080***	0.246	1.038**	1.026	4.765***	2.371***
	(0.260)	(0.358)	(0.471)	(0.714)	(1.252)	(0.738)
Reminders[2]						
via Text Message	0.746***	0.161	0.900*	0.600	1.791**	3.052***
	(0.244)	(0.401)	(0.480)	(0.638)	(0.782)	(0.926)
via Phone Call	0.859***	0.874***	0.721	-1.314**	4.738***	0.381
	(0.308)	(0.319)	(0.564)	(0.593)	(1.327)	(0.790)
via the Online Tool(s)	1.363***	0.128	1.905***	0.671	5.406***	-3.419***
	(0.338)	(0.336)	(0.614)	(0.506)	(1.527)	(0.944)
Online Tool[3]						
App Only	0.280	0.730*	-0.135	-1.186**	2.579**	2.729***
	(0.261)	(0.378)	(0.506)	(0.488)	(1.183)	(1.049)
Website Only	0.084	0.390	0.055	-1.425**	1.672*	4.689***
	(0.274)	(0.329)	(0.529)	(0.609)	(0.963)	(1.244)
App & Website	0.475*	-1.558***	-0.058	-2.150***	3.381***	-5.994***
	(0.278)	(0.435)	(0.498)	(0.669)	(1.113)	(1.568)
Outcome (% weight re-gain)	-0.071***	0.043***	-0.099***	0.071***	-0.263***	0.242***
	(0.010)	(0.006)	(0.022)	(0.019)	(0.058)	(0.054)
Cost (£ per month)	-0.217***	0.212***	-0.238***	0.362***	-1.002***	0.720***
	(0.027)	(0.031)	(0.054)	(0.094)	(0.225)	(0.167)
ASC Interactions						
Age 35-44	-2.362**		2.165		-5.546	
	(1.159)		(2.269)		(4.148)	
Age 45-54	-2.739		0.918		-6.803*	
	(1.868)		(2.285)		(3.589)	
Age 55+	-2.418*		3.631		-8.372**	
	(1.433)		(3.610)		(3.311)	
Female	0.977		1.282		7.707***	

Variables	All Participants Coeff.	Std. Dev.	Treatment Arm Coeff.	Std. Dev.	Control Arm Coeff.	Std. Dev.
	(0.814)		(1.558)		(2.665)	
Postgraduate Qualification	-4.241***		-8.859***		-8.349**	
	(1.138)		(2.608)		(3.487)	
Undergraduate Qualification	-2.603**		-11.010***		-1.461	
	(1.286)		(3.240)		(3.441)	
HND/HNC Qualification	-3.199**		-10.968***		-16.870***	
	(1.510)		(3.630)		(5.041)	
A Level Qualification	-0.782		-0.624		-21.204***	
	(1.372)		(3.750)		(6.159)	
Full-Time Employed	1.976		4.986**		3.161	
	(1.475)		(2.353)		(4.585)	
Part-Time Employed	0.491		11.054***		0.020	
	(1.582)		(3.855)		(4.633)	
Retired	0.465		1.192		17.430**	
	(1.792)		(2.582)		(7.977)	
Income £10-20k	1.317					
	(4.294)					
Income £20-30k	1.172		-0.368		-20.290**	
	(4.335)		(1.845)		(8.371)	
Income £30-50k	-2.426		-7.534***		-15.752***	
	(4.174)		(2.479)		(5.768)	
Income £40-50k	0.140		-0.804		-19.238***	
	(4.415)		(1.978)		(6.984)	
Income £50-60k	0.582		-14.434***		-15.539**	
	(4.346)		(5.093)		(7.436)	
Income £60-70k	0.016		-0.582		-15.350**	
	(4.273)		(2.000)		(7.062)	
Income over £70k	0.648		-4.601*		-24.884***	
	(4.506)		(2.378)		(8.376)	
Observations	3,390		1,560		1,830	
Log Likelihood	-689.9		-314.1		-332.7	

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; [1] Base level “face to face” feedback; [2] Base level “no reminders”; [3] Base level “no online tool”

Appendix D. Supplementary Materials for Chapter Eight

D-1 Conditional Logit Models (Main Effects)

Variables	(1) Potential Service User	(2) Potential Beneficiary	(3) Non-User
Alternative-Specific Constant	0.886***	1.056***	0.803***
	(0.148)	(0.138)	(0.144)
Length (months)	-0.0204***	-0.0182***	-0.0203***
	(0.00521)	(0.00491)	(0.00506)
Feedback[1]			
via Phone Call	0.102	0.173*	0.0376
	(0.0981)	(0.0932)	(0.0979)
via the Online Tool(s)	0.0758	0.0876	0.232**
	(0.0977)	(0.0926)	(0.0958)
via Text Message	0.0968	0.192**	0.317***
	(0.0909)	(0.0856)	(0.0898)
Reminders[2]			
via Text Message	0.0758	0.177**	0.274***
	(0.0864)	(0.0817)	(0.0857)
via Phone Call	-0.0126	0.0975	0.0847
	(0.0961)	(0.0920)	(0.0967)
via the Online Tool(s)	-0.0625	0.179**	0.217**
	(0.0938)	(0.0883)	(0.0923)
Online Tool[3]			
App Only	0.148	0.0867	0.0510
	(0.0957)	(0.0903)	(0.0951)
Website Only	0.228**	0.0982	0.162
	(0.101)	(0.0963)	(0.100)
App & Website	0.217**	0.204**	0.0915
	(0.0917)	(0.0864)	(0.0914)
Outcome (% weight re-gain)	-0.0131***	-0.00941***	-0.00768***
	(0.000999)	(0.000930)	(0.000958)
Cost (£ per month)	-0.0526***	-0.0623***	-0.0636***
	(0.00324)	(0.00306)	(0.00315)
Sample Size (Observations)	210 (6,300)	243 (7,290)	229 (6,870)
Log Likelihood	-1,927	-2,203	-2,162
AIC	3,880	4,432	4,350
BIC	3,968	4,522	4,439

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1;
[1] Base level "face to face" feedback; [2] Base level "no reminders"; [3] Base level "no online tool"

D-2 Conditional Logit Models (Main Effects & Attribute Interactions)

Variables	(1) Potential Service User	(2) Potential Beneficiary	(3) Non-User
Alternative-Specific Constant	1.168***	0.924***	0.778***
	(0.186)	(0.177)	(0.182)
Length (months)	-0.0123	-0.00594	-0.0244***
	(0.00982)	(0.00899)	(0.00912)
Feedback[1]			
via Phone Call	0.105	0.203**	0.0523
	(0.0997)	(0.0950)	(0.0992)
via the Online Tool(s)	0.137	0.0768	0.229**
	(0.101)	(0.0950)	(0.0984)
via Text Message	0.127	0.215**	0.310***
	(0.0927)	(0.0871)	(0.0916)
Reminders[2]			
via Text Message	0.0705	0.169**	0.275***
	(0.0866)	(0.0822)	(0.0860)
via Phone Call	-0.00282	0.0856	0.0826
	(0.0966)	(0.0931)	(0.0971)
via the Online Tool(s)	-0.0771	0.157*	0.221**
	(0.0944)	(0.0890)	(0.0932)
Online Tool[3]			
App Only	0.156	0.118	0.0551
	(0.0964)	(0.0921)	(0.0961)
Website Only	0.219**	0.128	0.164
	(0.102)	(0.0965)	(0.101)
App & Website	0.205**	0.244***	0.0917
	(0.0929)	(0.0881)	(0.0932)
Outcome (% weight re-gain)	-0.00792***	-0.00826***	-0.00791***
	(0.00261)	(0.00245)	(0.00246)
Cost (£ per month)	-0.0511***	-0.0466***	-0.0598***
	(0.00816)	(0.00759)	(0.00769)
Interaction: Length*Outcome	-0.000221*	-3.02e-05	9.92e-05
	(0.000118)	(0.000109)	(0.000114)
Interaction: Length*Cost	0.000262	-0.000883**	4.01e-05
	(0.000372)	(0.000354)	(0.000369)
Interaction: Outcome*Cost	-0.000123	-6.19e-05	-0.000103
	(0.000119)	(0.000114)	(0.000110)
Sample Size (Observations)	210 (6,300)	243 (7,290)	229 (6,870)
Log Likelihood	-1,923	-2,200	-2,161
AIC	3,878	4,432	4,354
BIC	3,986	4,542	4,463

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1;
[1] Base level "face to face" feedback; [2] Base level "no reminders"; [3] Base level "no online tool"

D-3 Mixed Logit Models (Main Effects & Demographic Interactions)

Variables	Potential Service User (PSU) Coeff.	Potential Service User (PSU) Std. Dev.	Potential Beneficiary (PB) Coeff.	Potential Beneficiary (PB) Std. Dev.	Non-User (NU) Coeff.	Non-User (NU) Std. Dev.
Alternative-Specific Constant	3.933***	2.291***	5.541***	2.437***	4.711**	3.227***
	(1.263)	(0.278)	(1.387)	(0.280)	(1.876)	(0.385)
Length (months)	-0.031***	0.031**	-0.028***	0.051***	-0.036***	0.064***
	(0.008)	(0.013)	(0.008)	(0.011)	(0.009)	(0.011)
Feedback[1]						
via Phone Call	0.181	0.446	0.229	-0.990***	0.140	-0.119
	(0.146)	(0.298)	(0.151)	(0.240)	(0.148)	(0.325)
via the Online Tool(s)	0.207	0.699***	0.236*	-0.409	0.550***	-0.128
	(0.151)	(0.245)	(0.136)	(0.285)	(0.148)	(0.303)
via Text Message	0.189	-0.172	0.339***	-0.301	0.587***	0.187
	(0.132)	(0.303)	(0.127)	(0.251)	(0.140)	(0.269)
Reminders[2]						
via Text Message	0.144	-0.306	0.205	0.827***	0.393***	0.464*
	(0.126)	(0.314)	(0.130)	(0.185)	(0.133)	(0.279)
via Phone Call	-0.039	-0.619***	0.162	0.136	0.079	0.561**
	(0.151)	(0.224)	(0.138)	(0.362)	(0.160)	(0.266)
via the Online Tool(s)	0.020	-0.242	0.389***	0.410*	0.331**	0.339
	(0.144)	(0.296)	(0.140)	(0.219)	(0.151)	(0.274)
Online Tool[3]						
App Only	0.244*	-0.001	0.123	0.090	0.201	-0.290
	(0.135)	(0.260)	(0.131)	(0.248)	(0.140)	(0.329)
Website Only	0.211	0.442	-0.099	-0.449*	0.205	0.588**
	(0.147)	(0.288)	(0.145)	(0.265)	(0.154)	(0.270)
App & Website	0.241*	0.324	0.167	-0.571***	0.028	-0.499*
	(0.132)	(0.285)	(0.127)	(0.221)	(0.139)	(0.286)
Outcome (% weight re-gain)	-0.027***	0.029***	-0.019***	0.023***	-0.017***	0.026***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)
Cost (£ per month)	-0.141***	0.137***	-0.160***	0.131***	-0.197***	0.157***
	(0.014)	(0.013)	(0.014)	(0.012)	(0.018)	(0.014)
ASC Interactions						
Age 25-34	-2.625**		-1.974		-2.343	
	(1.281)		(1.280)		(1.795)	
Age 35-44	-2.559**		-1.063		-2.776	
	(1.264)		(1.322)		(1.743)	
Age 45-54	-2.649**		-2.814**		-1.714	
	(1.318)		(1.212)		(1.762)	
Age 55+	-2.100		-2.162*		-3.021*	

Variables	Potential Service User (PSU) Coeff.	Std. Dev.	Potential Beneficiary (PB) Coeff.	Std. Dev.	Non-User (NU) Coeff.	Std. Dev.
	(1.338)		(1.180)		(1.783)	
Female	0.028		0.003		-0.267	
	(0.505)		(0.510)		(0.615)	
Postgraduate Qualification	0.029		-0.605		-0.610	
	(0.702)		(0.692)		(1.065)	
Undergraduate Qualification	0.172		-0.056		0.094	
	(0.599)		(0.689)		(0.872)	
HND/HNC Qualification	-0.004		-0.532		0.702	
	(0.797)		(0.755)		(1.225)	
A Level Qualification	0.141		-0.158		1.891**	
	(0.716)		(0.621)		(0.841)	
Full-Time Employed	1.222*		-0.217		0.941	
	(0.676)		(0.814)		(0.859)	
Part-Time Employed	1.594**		-0.840		0.497	
	(0.713)		(0.830)		(0.902)	
Retired	0.012		-1.333		1.482	
	(0.959)		(0.829)		(1.091)	
Income £20-30k	-0.348		0.758		-0.566	
	(0.683)		(0.775)		(0.800)	
Income £30-50k	-0.123		1.373*		-0.187	
	(0.716)		(0.813)		(1.065)	
Income £40-50k	-0.206		0.082		-0.058	
	(0.768)		(0.755)		(0.823)	
Income £50-60k	0.954		0.092		-1.575	
	(0.995)		(0.756)		(1.062)	
Income £60-70k	-0.445		0.368		-0.214	
	(0.938)		(0.886)		(1.076)	
Income over £70k	-0.395		0.463		-0.076	
	(0.894)		(0.903)		(0.965)	
Observations	6,300		7,290		6,870	
Log Likelihood	-1499		-1741		-1575	

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1;
[1] Base level “face to face” feedback; [2] Base level “no reminders”; [3] Base level “no online tool”

Appendix E. Future Writing Plan

Paper 1: Whose Preferences are Elicited in Discrete Choice Experiments? A Systematic Review

Based on: Chapter 4

First Target: Value in Health

Second Target: Applied Health Economics & Health Policy

Submission Date: Early 2018

Authors: David Mott, Laura Ternent, Luke Vale

Paper 2: A Framework for Classifying Respondent Samples in Discrete Choice Experiments

Based on: Chapter 5

First Target: Health Economics (Letter)

Second Target: The Patient

Submission Date: Early 2018

Authors: David Mott, Laura Ternent, Luke Vale

Paper 3: Trial Participants' Preferences for the NULevel Weight Loss Maintenance Intervention: A Discrete Choice Experiment

Based on: Chapter 7

First Target: Obesity

Second Target: BMC Public Health

Submission Date: Mid 2018

Authors: David Mott, Laura Ternent, Luke Vale, Falko Sniehotta, Liz Evans, Frauke Becker

Paper 4: How Do Preferences for a Weight Loss Maintenance Intervention Differ According to Experience? A Discrete Choice Experiment

Based on: Chapter 9

First Target: Social Science and Medicine

Second Target: Medical Decision Making

Submission Date: Mid 2018

Authors: David Mott, Laura Ternent, Luke Vale