The increasing incidence of childhood empyema thoracis: epidemiology and clinical aspects

Matthew F Thomas

A thesis submitted for the degree of Doctor of Philosophy at Newcastle University

Submitted: 19th Nov 2012

Viva-voce: 25th Feb 2013

Supervisors

Dr David Spencer

Prof. Steven Rushton

Dr Mark Pearce

School of Biology,

Newcastle University

Abstract

Historically, empyema thoracis has been a major cause of morbidity and mortality in children. It became the focus of considerable attention following its resurgence globally in the 1990's. The factors driving this change remain uncertain. In addition, there remains significant controversy over the best method of management of the condition.

This thesis aimed to define the epidemiology of paediatric empyema thoracis, to understand the factors contributing to the rise in the incidence of the disease. Secondary aims included investigation of the impact of the pneumococcal conjugate vaccine on paediatric empyema and evaluation of the effectiveness of different treatment methodologies in the condition.

A progressive framework of multivariate time series models and wavelet analysis was used to investigate relationships between empyema and pneumonia, activity of *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Streptococcus pyogenes* and *Mycoplasma pneumoniae* and time. The spatial epidemiology of the both conditions in North East England was defined and the impact of the introduction of routine pneumococcal vaccination investigated using an interrupted time series analysis. Multivariate survival models were used to investigate outcomes following different treatment methods.

Hospitalisations due to empyema increased significantly in England between 1997 and 2006, underpinned by an increase in bacterial pneumonia. Isolations of *S. pneumoniae* and *S. pyogenes* were positive predictors of empyema nationally. No spatial variation in the risk of empyema was detected. Introduction of pneumococcal vaccination did not decrease empyema hospitalisations. Children who underwent primary surgical treatment for their empyema had a 40% reduction in hospital stay and a lower risk of readmission or of any complication.

The increase in the incidence of paediatric empyema in England was driven predominantly by an increase in pneumococcal and streptococcal pneumonia. Primary surgery in empyema allowed earlier discharge, but further research is needed to establish which outcomes are most acceptable to patients and their families.

Acknowledgments

Although ultimately the work of one person all theses represent collaborations and this one is no exception. This work would not have been possible without the support and dedication of a number of individuals. I would like to thank the following specifically:

Several members of the Health Protection Agency (HPA) provided vital support for this work. Carmen Sheppard of the Respiratory and Systemic Infection Laboratory, Colindale, London and Malcolm Guiver, Molecular Diagnostic Service, HPA North West provided the pneumococcal serotype data. Similarly, Russell Gorton of the HPA North East provided the national *Streptococcus pneumoniae, Staphlyococcus aureus, Streptococcus pyogenes* and *Mycoplasma pneumoniae* surveillance data collected by the HPA.

My predecessor in the UK-ESPE study David Cliff collected over 200 patient's data and supported me significantly when I became involved in empyema research. Kerry Pollard also recruited patients and collected data from patients with empyema and their families. Steven Beaton and James Paton of Glasgow University administrated the UK-ESPE online database. In March 2010 Clare Simmister joined the UK-ESPE study as a research nurse in Newcastle and since then has been invaluable in supporting the study both in Newcastle and across other sites.

Collaborators at the other UK-ESPE sites included: C Wallis, V Rasiah, R Thursfield – Great Ormond Street Hospital, London; C Murray, J Mercy – Royal Manchester Children's Hospital, Manchester; C Pao – Royal London Hospital, London; R Ross-Russell – Addenbrooke's Hospital, Cambridge; J Bhatt, R Radcliffe, K Ward – Nottingham Children's Hospital; A Thomson – Children's Hospital, Oxford; I Balfour-Lynn – Royal Brompton Hospital, London; W Lenney – University Hospital of North Staffordshire, Stoke-on-Trent; J Clarke, I Brookes – Birmingham Children's Hospital, Birmingham; R Primhak – Sheffield Children's Hospital, Sheffield; E Carroll, J Clark – Alder Hey Children's Hospital, Liverpool; T Hilliard – Bristol Royal Hospital for Children, Bristol; A Cade – Derriford Hospital, Plymouth; A Prendiville, H Soloman, Royal Cornwall Hospital, Truro; S Cunningham, Royal Hospital for Sick Children, Edinburgh.

Mohamed Elemraid also recruited several patients to the UK-ESPE study in Newcastle. More important however was his unbridled enthusiasm for research and optimism which has been a constant source of energy and inspiration over the last three years.

It has been a pleasure to be a part of the newly branded Biological, Clinical and Environmental Systems Modelling research group in Newcastle and the help, companionship and tolerance of appalling jokes provided by all members of the group has been a real source of strength during this period. On a more serious note, I need to acknowledge the explicit help of three individuals. Mark Shirley taught me how to manipulate databases accurately and without his help the analysis of hospital episode statistics data would not have been possible. Andrew Close provided patient tutelage on both the art and science of modelling and its implementation in R as well as proof-reading this thesis. Alasdair Blain designed and implemented the code for the spatial analyses, as well as producing the maps illustrating that particular analysis.

I owe my supervisors David Spencer, Mark Pearce and Stephen Rushton a particular debt of gratitude. I will always be indebted to David for providing the opportunity in the first place, as well as for having the drive and enthusiasm to ensure that this work has come to fruition. Thanks go to Mark for his careful and considered analysis, support and encouragement over the last three years. Finally, Steve has provided me with an all encompassing scientific education and done so with the warmth, wit and patience that only very few possess.

It is a cliché but true nonetheless, that none of this would have been possible without my family. My wife Sally and daughter Martha have endured interrupted family time, postponed holidays and levels of paternal absenteeism that usually result in the involvement of the child support agency. Despite this they still allow me entry to the house and shower me with love and affection. My parents Marian and Simon Thomas have provided emergency child care, proof-reading services and wise counsel throughout the past three years. I cannot thank them enough for everything. I also have to thank the wider Thomas, Macpherson and Parkin families for their love and support.

Finally I am grateful to all the children with empyema and their families enrolled in the UK-ESPE study, without whom this work would not have been possible.

Dedication

Dedicated to MT, ST, MT and ST without whom nothing would have been possible

Statement of works

R Gorton and DA Spencer procured the Hospital Episode Statistics dataset. M Shirley initially established database for interrogating this dataset before passing this database to myself. D Cliff, DA Spencer, JY Paton and SP Rushton conceived the UK-ESPE study and obtained ethical approval for the study to proceed. MA Elemraid, C Simmister, D Cliff and K Pollard supported clinical data collection in Newcastle for the UK-ESPE study. C Sheppard, M Guiver and R George were responsible for the pneumococcal serotyping of culture negative pleural fluid described within. AP Blain researched and implemented the Bayesian spatial modelling, although all interpretation is my own.

All other works detailed within are my own.

Declaration

I declare that the work presented in this thesis is my own, except where explicitly stated in the text.

Glossary

Autocorrelation – statistical property of time and spatial data due to serial dependency often inherent in this form of data. This leads to violation of the assumption of independence required in standard statistical approaches.

Autoregressive model – statistical model containing specific function designed to balance presence of serial dependency within the residuals and which models the residual at time *s* as a function of residual of time *s* – 1.

Auto-regressive moving average model – similar to autoregressive model except that the function to model serial dependency is more complex. The residuals are modelled as a function of the *p* previous time points and white noise.

Conditional autoregressive models (CAR) – form of Bayesian statistical modelling that allows for both spatial heterogeneity and serial dependency to accurately estimate differences between different spatial areas.

Generalised linear models – generic term for a large group of statistical methods that are flexible extensions of standard linear regression.

Generalised least squares model – group of statistical models developed for analysis of time series data.

Homeoscedasity – all variables or observations have equal variance.

Periodicity – biological variability based on a cyclical pattern occurring out with normal seasonal patterns.

Seasonality – biological variability based on a seasonal (within year) pattern.

State-space models – stochastic models which can be used to represent time series with multiple processes contained within them.

Survival analysis – specific group of techniques for analysing data with censored end-points.

Wavelet analysis – Statistical techniques based on transformation of the signal of a time series into a series of functions (wavelets) that represent both the frequency and time-scale within that signal allowing assessment of the presence of different periods or cycles within data and associations between time series.

Table of contents

Table of figures

1 Introduction and scientific background

"So much has been written about acute empyema in the recent literature, that one wonders what there is left to say about it." G Heuer, 1932

1.1 Definitions and epidemiology

1.1.1 Definitions

An empyema is a collection of pus within a naturally occurring anatomical space. The term itself is a compound Greek word consisting of the prefix 'en' meaning 'in' or 'within' and the stem 'pyema' meaning accumulation of pus (Christopoulou-Aletra and Papavramidou, 2008). Empyema thoracis refers to infection and pus formation within the pleural space (the space between the parietal and visceral pleura) in the thorax. Other common forms of empyema include subdural empyema and empyema within the uterus known as pyometra. For simplicity throughout this thesis, empyema refers to empyema thoracis. This is in line with common medical usage where empyema is used almost exclusively to mean empyema thoracis and other forms are referred to by their longer form including their precise anatomical location.

1.1.2 Historical perspectives on epidemiology of empyema

Empyema is an old and historically important condition that has been recognised and treated for millennia. The Hippocratic physicians recorded its management in significant detail (Christopoulou-Aletra and Papavramidou, 2008). Ancient Egyptian artefacts also refer to the occurrence of empyema (Cole, 1937).

While tracking the evolution of the management of empyema in children through time is relatively straightforward, doing the same for the epidemiology of empyema is more problematic. Major advances in the identification of pathogens, variation in diagnostic classifications and a paucity of attention to epidemiology, relative to treatment, make establishing historical trends more challenging. However, a number of aspects are clear.

OT Clagett, an author who published on the epidemiology of empyema both before and after the introduction of antibiotics, estimated the incidence of empyema at around 10% of those with pneumonia in the pre-antibiotic era (Clagett, 1943; Clagett, 1973). He cited a number of large series in support of this, including Penberthy and Benson's (1936) series of 5,868 children with pneumonia in Detroit, Michigan who reported a rate of 7%. Clagett reported that there was no apparent difference in the empyema rate between adults and children but did comment on the phenomenon of annual fluctuations in incidence. Other researchers in the early 20th Century such as Heuer (1932) and Graham (1933) recognised this propensity for sudden increases in the incidence of empyema through examination of the annual variations in mortality and linked it to changes in virulence of the different strains of pathogens causing pneumonia.

Mortality from this condition at that time was much greater than it is now, with rates as high as 50% during the first world war, although some of this mortality was iatrogenic because of the dangerous methods of pleural drainage employed (Graham and Berck, 1933). It is difficult to exclude this iatrogenic component from the historical estimates. Perhaps the most accurate estimate comes from Lionakis *et al* (1958) study, which covered the period after changes in the management of empyema resulted in a significant reduction in iatrogenic deaths but before the widespread introduction of antibiotics and reported a mortality of 17% in children treated. The most recent available estimate of mortality in children is <0.5% (Williams *et al.*, 2011).

Heuer (1932) was the first author to describe trends in mortality in empyema by age and co-morbidity. He found that mortality rates were much greater at the extremes of age than in late childhood and adolescence (>40% in the <2 years and 51-70 year age groups; 2%-10% in the 5 to 10 years and 11 to 20 years age groups).

As well as being associated with high levels of mortality, empyema has been historically associated with significant morbidity. Hospital stay was often prolonged (Heuer (1932), 39 days; Penberthy and Benson (1936), 48.6 days and Lionakis *et al* (1958), 48.5 days). The longest stay in the survey of Lionakis *et al* (1958) survey was 212 days. Interestingly, the length of hospital stay did not decrease significantly following the introduction of antibiotics.

Following the adoption of antibiotics in around 1947 there was a substantial reduction in the incidence of empyema. In a 25 year study of empyema in children from 1932 to 1956 in Atlanta, cases almost ceased in the eight years following the introduction of antibiotics (4 cases 1948-1956 (0.5 cases/yr), 129 cases 1932-1947 (8.5 cases/yr). The reduction was so profound it led the authors to conclude that unless new, resistant bacterial strains were to develop faster than new antibiotics it seemed probable that empyema would no longer be an important thoracic disease (Lionakis *et al.*, 1958). This, coupled with their finding of no decrease in hospital stay following the arrival of antibiotics, led them to speculate that antibiotics do not necessarily alter the clinical course of empyema, but do prevent it arising from pneumonia. Decreases of a similar magnitude were reported by other authors in the early years following the introduction of antibiotics (Schwartz *et al.*, 1939; Thompson *et al.*, 1940).

However this decrease appeared short-lived with a significant peak in the incidence empyema reported in the late 1950's and early 1960's. Studies from five large American centres – Pittsburgh (Kiesewetter WB *et al.*, 1959), Harriet Lane, Baltimore (Ravitch MM and Fein R, 1961), St Louis (Middelkamp JN *et al.*, 1964), Washington (Groff DB *et al.*, 1966) and Duke University, North Carolina (Wolfe WG *et al.*, 1968) highlighted a change in the incidence of empyema at approximately contemporaneous times. For example cases in Baltimore increased from 6 cases/yr in the years 1951-4 to 38 in 1955-58 (Ravitch MM and Fein R, 1961). In contrast to the pre-antibiotic era where pneumococci and streptococci were most frequent, the predominant organism was *Staphylococcus aureus*. Isolates of *Staphylococcus aureus* from children with empyema during this period were frequently multi-drug resistant, leading to speculation that the increase in staphylococcal empyema was related to the introduction of antibiotics (Ravitch MM and Fein R, 1961; Geha AS, 1970). Furthermore, certain phage types or strains of *S. aureus* were more commonly isolated e.g. in the St Louis series, phage type 80/81 accounted for over 80% of *S. aureus* isolates from pleural fluid (Middelkamp JN *et al.*, 1964), suggesting that particular strains that had been circulating may have had a predisposition towards causing empyema.

In the decades following this rise in the incidence of empyema, interest in the condition appeared to wane and only a handful of studies addressing the epidemiology of the condition were published. Fajardo and Chang (1987) recovered *Haemophilus influenzae* from pleural fluid in 15% of children with empyema presenting to American military hospitals between 1978 and 1982, a greater proportion than had been previously reported. However, the widespread introduction of vaccination against *H. influenzae* type B shortly afterwards led to a significant decline in invasive disease due to this pathogen (Lewis and Feigin, 2002). Brook's series of 93 cases of empyema (1990) highlighted the role of anaerobic bacteria in empyema, particularly in association with aspiration pneumonia. No

other significant changes in the epidemiology of the condition were reported until the current interest in empyema began in 1997 in the UK (Playfor *et al.*, 1997; Rees *et al.*, 1997).

1.1.3 Current epidemiology of empyema

In the middle of the 1990's there were several reports noting a rise in incidence of empyema in UK children (Playfor *et al.*, 1997; Rees *et al.*, 1997). Similar rises were noted in a range of countries around the world and are illustrated in **Figure 1.1**.

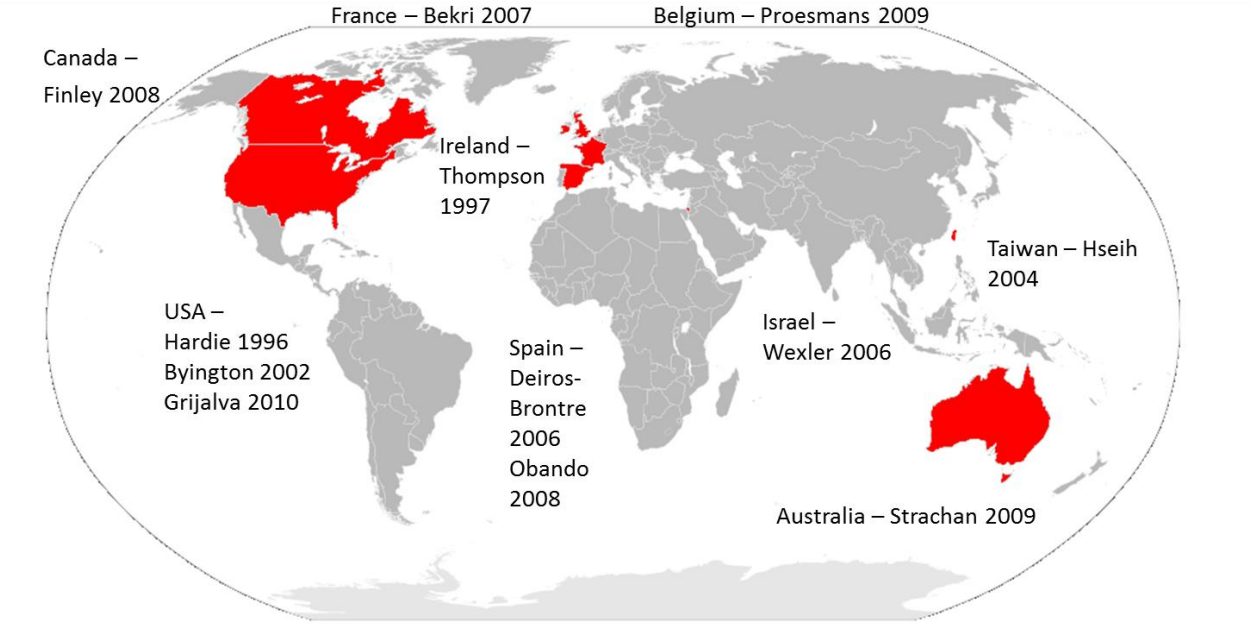


Figure 1.1 Map of countries where an increase in the incidence of paediatric empyema has been reported. *Countries with significant increases in incidence shaded in red* (Grijalva *et al* 2010; Hardie *et al.*, 1996; Thompson *et al.*, 1999; Byington *et al.*, 2002b; Hsieh *et al.*, 2004; Deiros Bronte *et al.*, 2006; Spencer *et al.*, 2006b; Wexler *et al.*, 2006; Bekri *et al.*, 2007; Finley *et al.*, 2008; Obando *et al.*, 2008; Proesmans and De Boeck, 2009; Strachan and Jaffe, 2009).

No hypothesis has been universally accepted for these changes in incidence, although a number have been proposed. These include: an increase in the underlying incidence of pneumonia (Roxburgh *et al.*, 2008); a change in circulating pneumococcal serotypes or increasing pathogen virulence (Byington *et*

al., 2002b; Wexler *et al.*, 2006); a change in causative organism (Schultz *et al.*, 2004) and a change in community antibiotic prescribing (Strachan and Jaffé, 2009).

Empyema is a seasonal condition with peaks in incidence between November and March in the northern hemisphere (Hardie *et al.*, 1996). There has been some suggestion that there are bimodal peaks within the winter season with one in December and a further peak in early spring, although this has not been widely investigated (Spencer *et al.*, 2006a). Three large studies have measured the incidence of empyema in the UK and all have used hospital coding data. Roxburgh *et al* (2008) reported 37 cases of empyema per million in Scotland in 2005. Gupta and Crowley (2006) reported an incidence 26 cases per million in England in 2002/3. Most recently Koshy *et al* (2010) reported a change from 4.7 to a peak of 17.5 per million children and a subsequent fall to 13.7 by 2008. Koshy *et al* acknowledge that their estimates of empyema incidence are lower than those of other studies and cite their usage of more specific diagnostic codes as a potential explanation for the difference.

Empyema occurs at all ages but the natural history of the disease and outcomes vary significantly between adults and children. In childhood, empyema most often occurs in previously healthy children and is associated with very low mortality (<0.5%) in developed countries (Balfour-Lynn *et al.*, 2005; Williams *et al.*, 2011). In contrast in adults, patients frequently have several co-morbidities and mortality in the UK was estimated at 15 % in 2005, with the highest death rate occurring in the elderly (Maskell *et al.*, 2005).

The average length of hospital stay has substantially reduced compared to historical figures (Heuer (1932), 39 days; Penberthy and Benson (1936), 48.6 days and Lionakis *et al* (1958), 48.5 days), with the median between four to seven days post treatment (Balfour-Lynn *et al.*, 2005; Sonnappa *et al.*, 2006; Li and Gates, 2008).

1.2 Microbiology of empyema

The Hippocratic physicians recognised that empyema may have different causes, appearing to differentiate between cases caused by anaerobes and other causes by their relative mortality, and the constitution and smell of the fluid drained from the pleural cavity (Odell, 1994). Within the literature a wide range of bacteria, fungi and viruses have been reported as causes but, non-tuberculous bacteria

are the predominant cause (Chernick *et al.*, 2006). Historically, three species account for the overwhelming majority of cases. These are *S. pneumoniae, S. aureus, and S. pyogenes. S. pneumoniae, S. aureus* and *S. pyogenes* are all colonisers of the human nasopharynx and are, arguably, the leading causes of bacterial pneumonia in children worldwide. The relative contribution of *S. pneumoniae, S. pyogenes* and *S. aureus* to empyema in children in published series from 1950 onwards is shown in **Figure 1.2**.

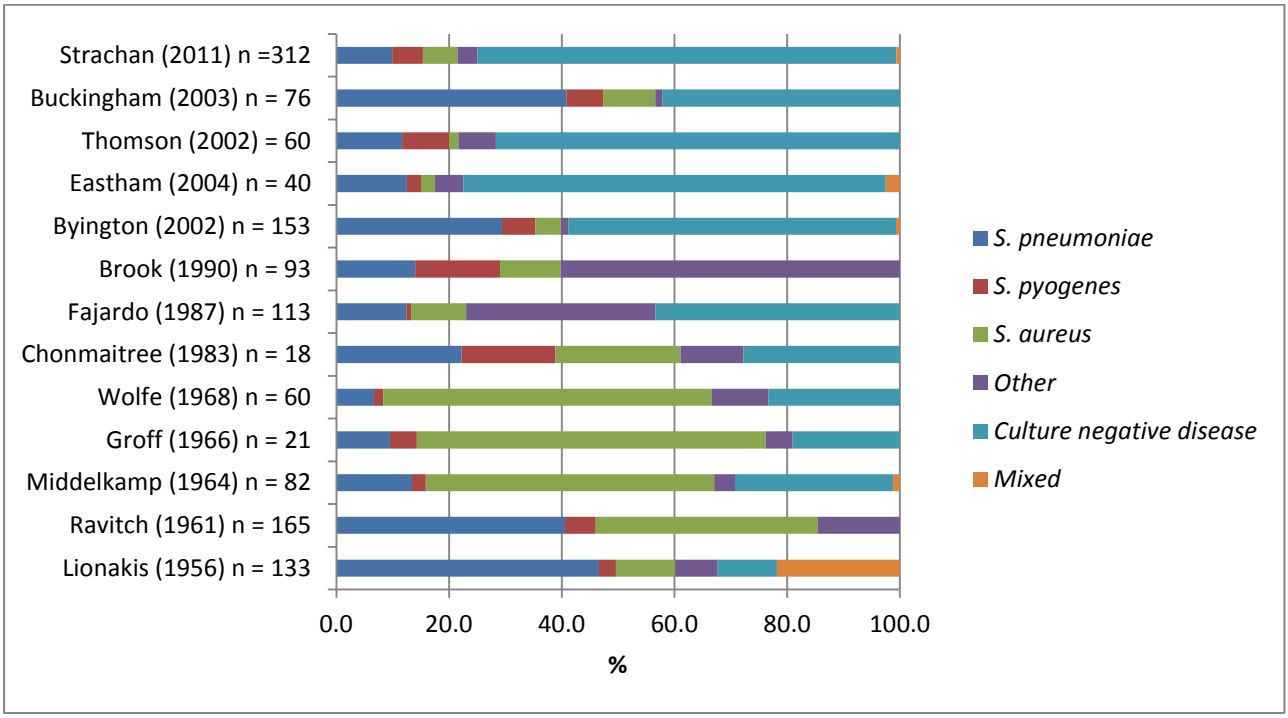


Figure 1.2 Summary of the major bacteriological causes of empyema in children from published series from 1950 onwards. All pathogens were identified by pleural fluid or blood culture.

Overall, *S. pneumoniae* was the leading cause of empyema, accounting for approximately 40% of cases where a pathogen was identified (296 of 757 cases). Staphylococcal empyema was the next most prevalent accounting for 30% of cases (229 cases). *S. pyogenes* was associated with approximately 10% (72 cases). Staphylococcal empyema was predominant in the majority of studies from the 1960's, illustrating staphylococcal dominance at this time. The five studies from 2000 onwards consistently found that the majority of cases were culture negative, but that pneumococcal infection was the dominant organism found. Similarly, the two earliest studies included, Lionakis *et al* (1958) and Ravitch MM and Fein R (1961), found high proportions of pneumococcal disease.

There has been a substantial increase in the proportion of culture-negative disease over time or to put it alternately, empyema where no causative pathogen is isolated using standard culture techniques. Early surveys such as Lionakis *et al* (1958) reported virtually no culture negative cases. In contrast, between 42% and 75% of cases had no identifiable cause in studies published after 2000. This increase in the frequency of culture negative disease has been presumed to be as a consequence of antibiotic treatment prior to sampling (Balfour-Lynn *et al.*, 2005).

A number of studies have used molecular diagnostic techniques in order to address this problem (Boersma *et al.*, 1993; Menezes-Martins *et al.*, 2005; Saglani *et al.*, 2005; Lahti *et al.*, 2006; Le Monnier *et al.*, 2006; Strachan and Jaffé, 2009). These generally utilise PCR for the detection of specific DNA products from bacteria and have been shown to substantially increase the rate of attribution of a definitive cause in empyema. Both broad spectrum techniques such as 16S and single organism specific PCR's such as pneumolysin specific PCR for *S. pneumoniae* have been used. **Table 1.1** illustrates the improvement in detection seen in the two studies that have used 16S in empyema (Saglani *et al.*, 2005; Le Monnier *et al.*, 2006).

Study	**Le Monnier (2006)**		
Result	**Culture (%)**	**16S PCR (%)**	**% Change**
S. pneumoniae	30	47	+18
S. pyogenes	5	9	+4
S. aureus	6	7	+1
Other	8	8	0
Culture negative disease	51	28	-23
Study	**Saglani (2005)**		
Result	**Culture (%)**	**16S PCR (%)**	**% Change**
S. pneumoniae	3	41	+38
S. pyogenes	3	9	+6
S. aureus	9	16	+7
Other	3	9	+6
Culture negative disease	81	34	-56

Table 1.1 Improvement in detection of a pathogen in studies using 16S molecular diagnostics in empyema

Studies using molecular techniques have universally attributed the bulk of culture negative disease to *S. pneumoniae*, but there is a potential reporting bias within this. As all these studies have included assays to detect *S. pneumoniae*, but have not always included assays specific for other pathogens. For example Strachan *et al* (2009) used specific PCR's for *S. pneumoniae, S. aureus, Haemophilus influenzae, Mycoplasma pneumoniae* and *Chlamydia pneumoniae*, but not for *S. pyogenes*. Studies using broad range detection including Le Monnier *et al* (2006) and Saglani *et al* (2005) confirm that the bulk of culture negative disease is due to *S. pneumoniae*.

Brook's (1990) series highlighted the role of anaerobes within paediatric empyema. He detected anaerobic bacteria in 33% of the 72 cases reported and attributed them as the sole cause in 24% of patients. *Bacteriodes* species were the most common anaerobic isolates. No other studies have

reported this level of anaerobic infection and Brook only included cases where a definite cause was identified, potentially biasing the results towards this group. Anaerobic infections are more likely in children who are at risk of aspiration pneumonia, such as those with chronic neurodevelopmental problems and immunocompromised children (Chernick *et al.*, 2006).

Viruses, in particular enteroviruses can be associated with the formation of pleural effusions, but they are only very rarely associated with the development of empyema (Lewis and Feigin, 2002; Chernick *et al.*, 2006). However, there are well-recognised associations between primary viral infection and subsequent bacterial pneumonia and empyema, a classic example of this being the association of the 1918 influenza epidemic and empyema (Dunham *et al.*, 1918).

Fungi are a well recognised cause of empyema, especially in immunocompromised individuals, but as a group they are a much rarer cause in previously healthy individuals. Ko *et al* (2000) studied fungal empyema in 67 adults, *Candida* spp was the most common isolated species, accounting for over 60% of isolates and 85% had some form of predisposing condition. As well as immunocompromise or immunosuppression, abdominal perforation may be significant risk factor for fungal empyema (Ishiguro *et al.*, 2010).

Primary pulmonary tuberculosis (TB) in children is associated with pleural effusion in 2-38% of cases, with the association being greatest in older children and adolescents (Merino *et al.*, 1999). Pleural effusion in TB can look very similar to empyema, particularly as it is commonly associated with consolidation of the underlying lung parenchyma and it is listed as a cause of empyema (Merino *et al.*, 1999; Balfour-Lynn *et al.*, 2005). However, the pathophysiology is different in TB infection and is typically characterised by granulomatous inflammation in contrast to the classical neutrophil led inflammation seen in empyema (Merino *et al.*, 1999). As a consequence, most researchers separate the two conditions by referring to tuberculous and non-tuberculous empyema.

The microbiology of empyema in adults is different to that of children. *S. milleri spp* were the commonest cause of culture positive empyema in adults in the UK accounting for 11% of cases in the largest and best conducted survey yet published (Maskell *et al.*, 2005). The *Streptococcus milleri* group comprises *S. mitis, S. viridans, S. mutans, S. oralis, S. sanguinis* and *S. sobrinis*. They are oropharyngeal commensals and facultive anaerobes. They can cause a range of rare, but serious,

diseases including infective endocarditis, liver abscess and brain abscess (Ruoff, 1988). Their contribution in paediatric empyema is much smaller and they occur only infrequently within published series. Isolation of *S. milleri spp* in paediatric patients is usually confined to children in whom aspiration of food contents is a problem, such as children with severe cerebral palsy.

1.2.1 Pneumococcal serotypes, pneumococcal vaccines and paediatric empyema

S. pneumoniae is not only the leading cause of paediatric empyema, but is also the leading cause of bacterial pneumonia in children (van der Poll and Opal, 2011). It is an obligate human pathogen but frequently exists as a commensal in the nasopharynx in a state referred to as carriage (Lipsitch, 1999). It is a genotypically diverse organism and over 90 different serotypes or groups are recognised (Lipsitch, 1999). Serotyping is based on recognition of the polysaccharide capsule expressed by the organism (Lipsitch, 1999). The capsule is also the most potent virulence factor expressed by *S. pneumoniae* and significant differences in the invasiveness, severity of disease caused and nasopharyngeal carriage are reported between serotypes (van der Poll and Opal, 2011).

Only a small minority of serotypes have been detected in children with empyema. Pneumococcal serotype 1 was the most frequently recovered serotype in the majority of series (Eltringham *et al.*, 2003; Byington *et al.*, 2006). In the most recent data published in the UK, serotype 1 accounted for 47.5 % of cases (n=223), serotype 3 24%, 7F/A 5.4%, 5 4.5% and 19A 4% of children with empyema and culture negative pleural fluid (Sheppard *et al.*, 2011).

Pneumococcal vaccines based on capsular polysaccharides were initially developed in 1945 but did not reach widespread use until the 1970's (MacLeod *et al.*, 1945; Austrian R *et al.*, 1976). Purely polysaccharide based vaccines are poorly immunogenic in young children, however conjugation, a process whereby capsular polysaccharides are linked to a more immunogenic protein, alleviates this problem (Eskola J and Anittila, 1999). A negative feature of conjugation is that the number of specific polysaccharides that can be incorporated onto the vaccine and hence the serotypes protected against are limited. The first commercially available pneumococcal vaccine (PCV-7) stimulated immunity against only seven serotypes (serotypes 4, 6B, 9V, 14, 18C, 19F, and 23F). It was introduced into the routine immunisation schedule for infants in the USA in 2000 and in the UK in 2006. It was replaced in the UK by an expanded valency vaccine (PCV-13) in 2010, providing additional protection against serotypes 1, 3, 5, 6A, 19A and 7F.

The projected impact of PCV-7 on paediatric empyema, extrapolated from the two studies of the pneumococcal serotype distribution carried out pre-vaccine introduction, was of a modest 20% reduction (Eltringham *et al.*, 2003; Fletcher *et al.*, 2006). The only UK data published after the introduction of PCV-7 demonstrated a reduction in incidence of approximately this magnitude (Koshy *et al.*, 2010). This is in contrast to the experience elsewhere. In the USA, where the longest time has elapsed since the introduction of PCV-7, the incidence of paediatric empyema increased substantially in the years following the introduction of PCV-7 (Grijalva *et al.*, 2010; Lee *et al.*, 2010; Li and Tancredi, 2010).

Pneumococcal conjugate vaccines are highly effective in preventing invasive pneumococcal disease caused by serotypes contained within the vaccine. However, concerns have been raised about the possibility of serotype replacement disease. Serotype replacement disease is the process of removal of vaccine serotypes from their ecological niche by vaccination coupled to an expansion of non-vaccine serotypes and concomitant increase in disease due to non-vaccine serotypes. This phenomenon has been observed in both overall invasive pneumococcal disease in England and Wales and in empyema in the USA (Byington *et al.*, 2006; Miller *et al.*, 2011). While serotype replacement disease cannot be responsible for the increase in empyema within the UK, as the increase began around a decade prior to the introduction of PCV-7, its impact on empyema in the UK is likely to be of significance.

1.3 Anatomy and physiology of the pleura and the pathophysiology of empyema

1.3.1 Anatomy and physiology of the pleura

In gross anatomical terms, the pleura consist of two mesothelial layers covered by differing amounts of connective tissue and separated by approximately 10 - 24 µm (Chernick *et al.*, 2006). Within the potential space between the pleura, there is a small volume of fluid. The inner layer is the visceral pleura and the outer layer the parietal pleura. The visceral pleura extend over the entire surface of the lung and the parietal pleura the inner surface of the thoracic cage, mediastinum, and diaphragm (Wang, 1985). The extent of pleura is shown in **Figure 1.3**. The left and right pleural cavities are completely separate, although may be adherent to each other on the anterior and posterior surfaces of the heart (Wang, 1985).

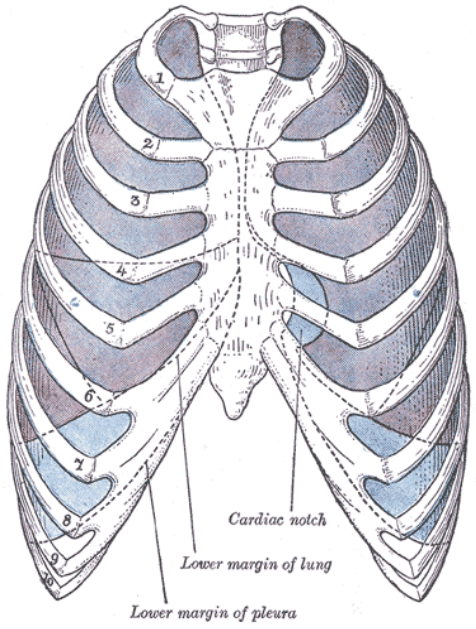


Figure 1.3 The anatomy of the pleura as illustrated in Grey's Anatomy (1918)

Embryologically, the pleura are mesodermal structures that arise in concert with the pericardium and peritoneum from the coloem (Davila and Crouch, 1995). By the seventh week of gestation the developing pleura have been separated from the pericardium by the pleuropericardial membrane and from the peritoneum by the pleuropertioneal membrane (Davila and Crouch, 1995). Mesothelial cells begin to line both the visceral and parietal pleura from around fifth week of gestation as the lung buds

begin to form. As embryogenesis proceeds layers of connective tissue divide and proliferate around the mesothelium, until the pleura are fully formed (Wang, 1985).

There are significant structural differences between the visceral and parietal pleura. The parietal pleura is smooth whereas the visceral pleura is indurated as it adheres too tightly to the surface of the lung, including in the intralobular fissures. The parietal pleura is perfused by the intercostal arteries and pericardiophrenic branch of the internal mammary arteries (Davila and Crouch, 1995). In contrast, the visceral pleura appear to be supplied by both the branches of the bronchial arteries and in some areas the pulmonary arteries (Harley, 1987). There are also differences in the lymphatic drainage between the visceral and parietal pleura. The visceral pleura drains largely to the pulmonary hilum and into the hilar nodes. The parietal pleura drains into the intercostal system anteriorly and into the retroperitoneal system posteriorly (Harley, 1987). There is also differential innervation of the pleura. The innervation of the parietal pleura arises from the intercostal nerves, in a manner similar to that of the lympyhatics. The visceral pleura is supplied by the vagus nerve and by the sympathetic trunks, but is not innervated by pain nerves (Harley, 1987).

At a micro-anatomical level there are further differences between the pleura. The most important is the presence of stomata on the parietal pleura which form an important mechanism in the homeo-static regulation of the pleural space (Wang, 1985). There are differences in the density of microvilli and elastic fibres between the pleura and also between different regions of the pleura (Wang, 1985).

The physiological function of the pleura is to reduce the work of breathing by reducing the friction between the thoracic cage and the lungs and is an adaptation to life on land. All land mammals have a similar anatomical arrangement, except the Asian elephant (*Elepheas maximus)* and African elephant (*Loxodonta africana)* in whom the pleural space is filled with fibrous connective tissue.

The surface area of the pleura is approximate to 4000 cm^2 in a 70 kg man and the estimated volume of pleural fluid in health is 18 mL (equivalent to 0.26 mL/kg) (Zocchi, 2002). Pleural fluid has a low concentration of protein and contains approximately 1700 cells/mm^3 (Zocchi, 2002). The cellular composition is 75% macrophages, 23% lymphocytes and 1% mesothelial cells (Zocchi, 2002).

In order to maintain the function of pleura, the volume and composition of the pleural fluid are under tight homeo-static control (Zocchi, 2002). Three mechanisms are recognised to be important **Figure 1.4**:

1. Starling forces (providing filtration through parietal and absorption through visceral mesothelium)
2. Lymphatic drainage through parietal pleural stomata
3. Electrolyte coupled liquid absorption through mesothelium of both sides and transcytosis

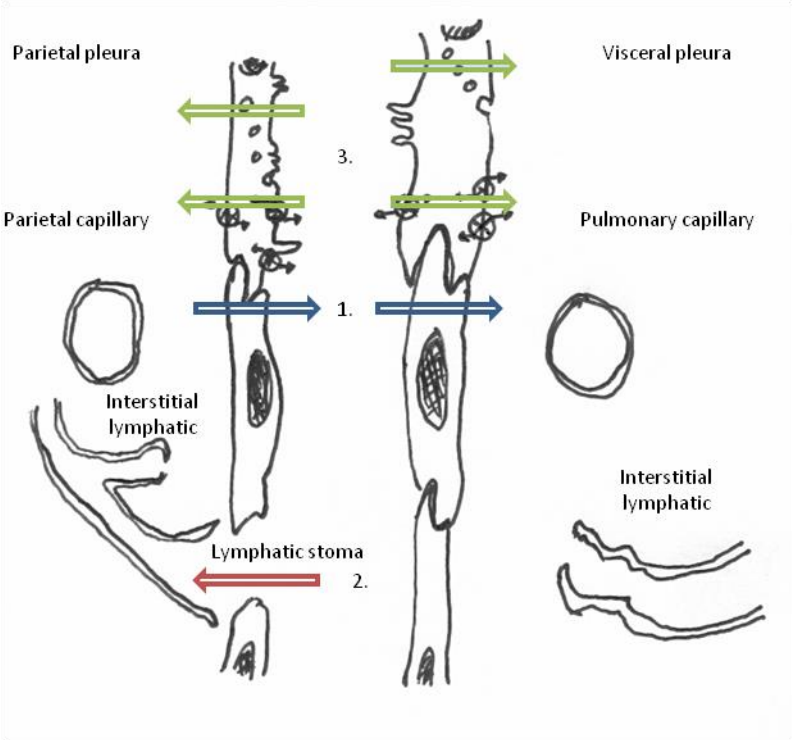


Figure 1.4 Recognised mechanisms for regulation of fluid volume within the pleural space. Blue arrows indicate direction of filtration gradient driven by differences in oncotic pressure (Starling forces) on either side of parietal and visceral pleura. The red arrow indicates reabsorption of fluid into the lymphatic system through the lymphatic stomata of the parietal pleura. Green arrows indicate direct cell transport of liquids and solutes including through transcytosis, which also prevents accumulation of pleural fluid *(adapted from Zocchi, 2002)*.

1.3.2 Pathophysiology of empyema

In up to 95% of cases, the first step in development of an empyema is the development of pneumonia in lung parenchyma adjacent to the pleura (Hamm and Light, 1997). The remaining cases result from direct inoculation of pathogens as a result of penetrating chest trauma or as a complication of thoracic surgery. Once inflammation and infection are established in the pleural space, the subsequent pathophysiological processes appear to be consistent regardless of the initiating event (Hamm and Light, 1997). The development of an empyema can be broken down into four stages which are well described in the literature (Hamm and Light, 1997; Zocchi, 2002; Antony, 2003; Brims *et al.*, 2010).

The *initial or reactive phase*, begins with the spread of inflammatory mediators and a protein rich inflammatory fluid to the extracellular fluid of the sub-mesothelial connective tissue from the infected sub-pleural lung parenchyma. This leads to a reversal of oncotic pressure gradient across the pleural space and the visceral pleura facilitating the accumulation of a transudate-like effusion within the pleural space.

During the *exudative phase* the influx of inflammatory cells and mediators interact with the mesothelial cells, triggering a local pleural immune response with the further release of inflammatory mediators and migration of neutrophils. This immunological activation of the mesothelial monolayer leads to changes in its permeability to both proteins and cells leading to the accumulation of a small neutrophil dominated sterile exudative effusion within the pleural space.

The *fibrinopurulent phase* is catalysed by secondary bacterial invasion of the pleural space. This causes further mesothelial immunological activation, aggravation of mesothelial cell injury and neutrophil migration into the pleural space. The resultant inflammatory milieu creates a relatively pro-fibrinogenic and anti-fibrinolytic environment. As a consequence, the coagulation cascade is triggered and fibrin is deposited on both pleural surfaces. The deposition of fibrin exacerbates the accumulation of fluid within the pleural space by blocking the stomata of the parietal pleura. At the end of the fibrinopurulent phase, frank pus is present within the pleural space.

The *organisational phase* is characterised by the development of dense fibrinous septations between the pleura leading to separate pockets of loculated pus within the pleural space. As part of the healing process, fibroblasts invade the pleural space leading to further fibrosis of the pleural space and

ultimately to the development of a fibrothorax if no intervention occurs. These stages are illustrated in **Figure 1.5**.

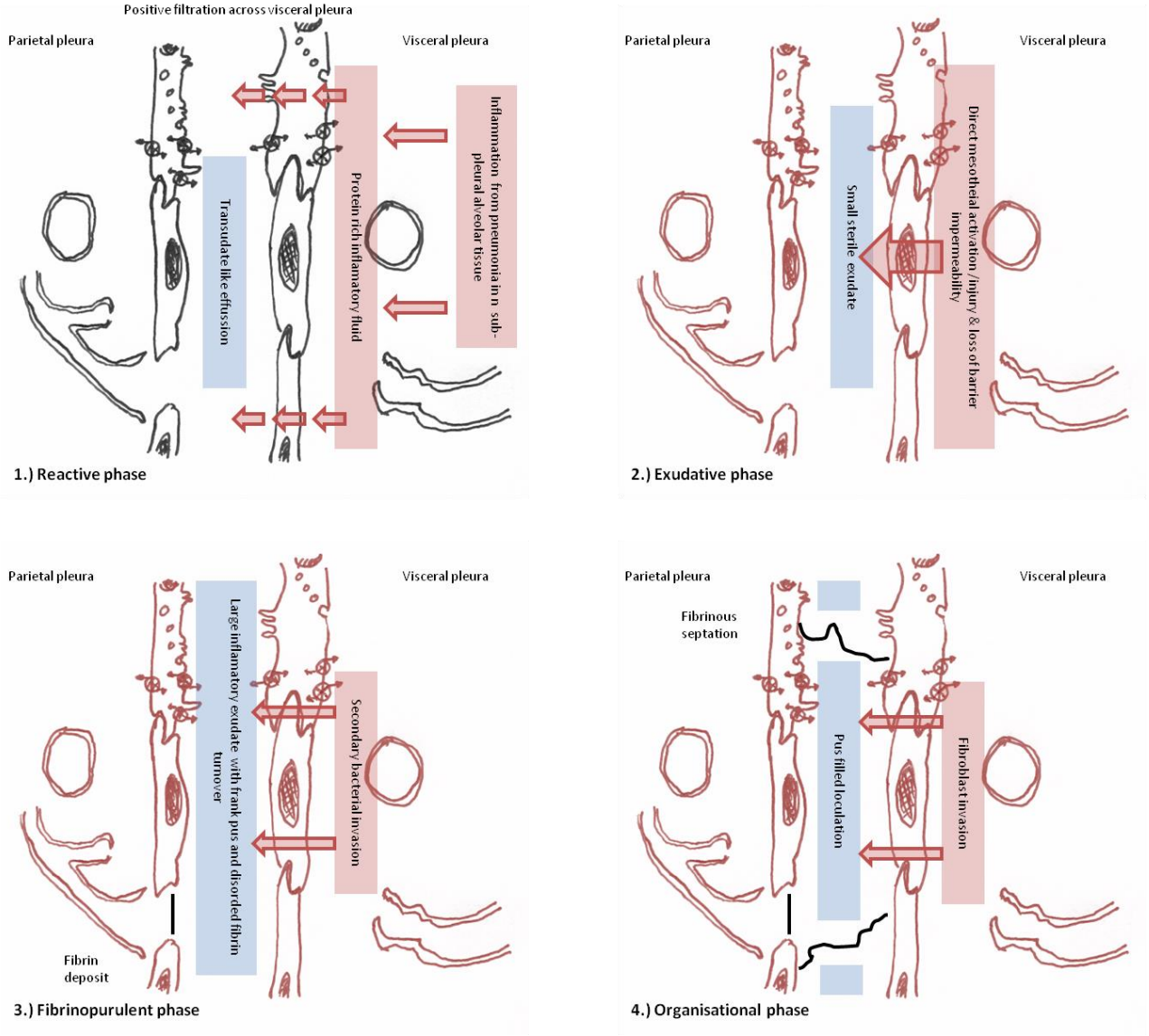


Figure 1.5 Pathophysiological development of an empyema (adapted from illustration of normal physiological regulation of the pleural space, Zocchi, 2002). 1.) Reactive phase 2.) Exudative phase 3.) Fibrinopurulent phase 4.) Organisational phase.

The macroscopic pathological changes in empyema are shown in **Figure 1.6**.

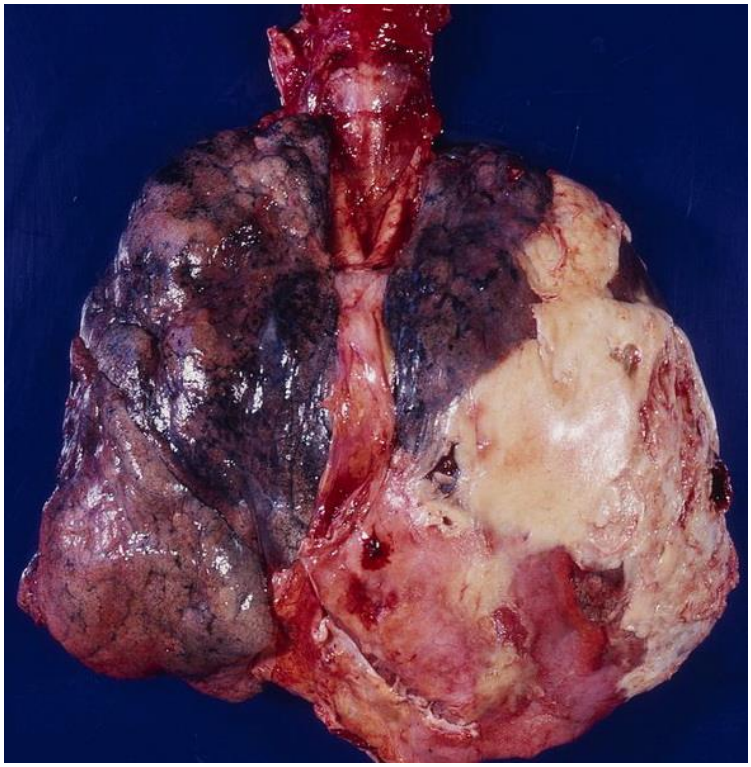


Figure 1.6 A pathology specimen demonstrating an extensive left -sided pleural empyema encasing the left lower lobe, lingular and extending into the left upper lobe. The white debris seen is fibrin deposition and pus formation. The specimen illustrates how an empyema restricts the normal expansion of the lung. Also demonstrated is, the huge expansion in the pleural space illustrated by the contrast between the normal pleura on the right and the left pleura containing the empyema. *(Author – Yale Rosen, Source http://www.flickr.com/photos/pulmonary_pathology/3704699031/)*

1.4 Clinical features & management of empyema

1.4.1 Clinical features

Symptoms of empyema are similar to those of pneumonia and include cough, fever, dyspnoea and chest pain. Dyspnoea, pleuritic and abdominal pain may be more prominent than in pneumonia (Balfour-Lynn *et al.*, 2005). Physical signs seen on examination include reduced chest movement, stony dullness to percussion, reduced breath sounds and a pleural rub (Balfour-Lynn *et al.*, 2005). Compensatory scoliosis may be present.

1.4.2 Investigations

In routine practice, the diagnosis is usually made by plain film chest radiography, although definitive confirmation requires examination of pleural fluid. The presence of significant fluid results in obliteration of the costophrenic or cardiophrenic angles and a meniscus or visible fluid level visible on the radiograph, as shown in **Figure 1.7**. Larger effusions cause mediastinal deviation and compression of the contralateral lung. Thoracic ultrasound is a useful investigation which is able to detect even small volumes of pleural fluid, loculation and abscess cavities without use of ionising radiation (Calder and Owens, 2009). Thoracic computerised tomography (CT) is only required to assess more complicated disease or when conditions such as malignancy are suspected.

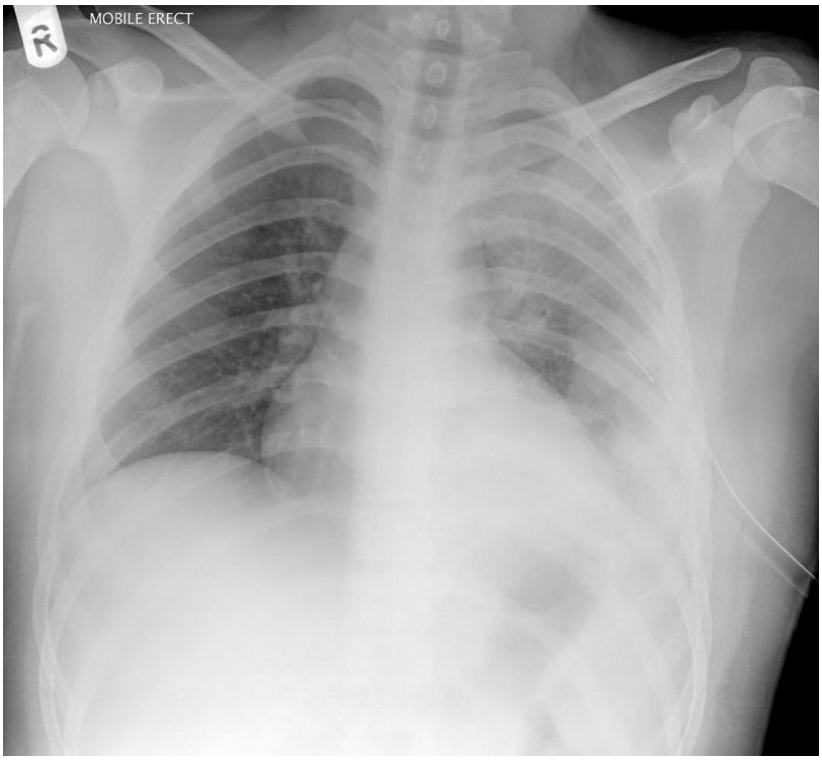


Figure 1.7 Plain radiograph showing pleural empyema in a teenager. Note the obliteration of the cardiophrenic and costophrenic angles in the left hemithorax. A chest drain has been placed for pleural drainage.

In UK patients pleural fluid is usually culture negative using routine bacterial culture, presumably because the majority of patients have received antibiotics prior to sampling (Balfour-Lynn *et al.*, 2005). Advanced molecular techniques including 16S ribosomal gene sequencing, antigen-specific polymerase chain reaction (PCR) based assays and sequential multiplex PCR capable of differentiating between serotypes are increasingly used to identify causative organisms (Saglani *et al.*, 2005; Le Monnier *et al.*, 2006). Raised white cell count, increased protein, reduced glucose and increased lactate dehydrogenase in pleural fluid indicate an exudative fluid within the pleural cavity consistent with empyema (Light *et al.*, 1972).

1.4.3 Historical perspectives on treatment

Treatment of empyema by the Ancient Greeks' can be divided into two stages; initial "conservative" management with medicaments followed by surgery for those not improving (Christopoulou-Aletra and Papavramidou, 2008). Surgery consisted of opening of the chest and direct drainage of the pus. The importance of prompt diagnosis and leaving a hollow read as a form of chest drain *in situ* was recognised (Odell, 1994). These principles still underpin the treatment of empyema (Balfour-Lynn *et al.*, 2005).

Treatment of empyema remained largely unchanged until it became the focus of significant interest following the influenza pandemic of 1918 (Peters, 1989). During the pandemic, exceptionally high mortality of healthy young recruits due to empyema was observed in US army camps (Dunham *et al.*, 1918). The management techniques at the time were mainly focused on open drainage, with an inaccurate understanding of pleural physiology leading to high risk of pneumothorax and unsuccessful re-inflation of infected areas of the lung (Odell, 1994). These inappropriate procedures were undertaken despite the fact that effective closed drainage techniques had been described over forty years previously (Hewett, 1876). The establishment of the US Army Empyema Commission headed by the thoracic surgeon Evarts Graham, recognised these short-comings and proposed a series of recommendations (1) careful avoidance of open pneumothorax during the acute stage (2) prevention of a chronic empyema by rapid sterilisation and obliteration of the infected cavity; and (3) careful attention to the nutrition of the patient (Dunham *et al.*, 1918; Odell, 1994). The adoption of these recommendations drastically reduced mortality from an average of 30.2% (although in some centres using open drainage upwards of 70%) to 3.4% (Odell, 1994).

The next major advance in the management of empyema was the introduction of antibiotics. One of the first reports of the successful use of penicillin in the USA included eight patients with empyema in 1944 (Tillet *et al.*, 1944). A case series by Lionakis *et al* (1958) suggests that the use of antibiotics became routine around 1947.

Despite the advent of antibiotics, surgical pleural drainage remained a key component of empyema treatment. Building on the insights of Graham's Empyema Commission, different strategies have emerged to achieve this which have varied in favour over time (Samson, 1971; Molnar *et al.*, 2004). They have included:

- Thoracentesis
- Intercostal tube drainage
- Rib-resection tube drainage
- Open flap drainage
- Decortication
- 'Empyemectomies'
- Sterilization of the pleural cavity
- Thoracoplasty (Samson, 1971)

Of these, thoracentesis, intercostal tube drainage and a modified form of decortication are in usage in modern paediatric care (Balfour-Lynn *et al.*, 2005). Video-assisted thoracoscopic surgery, known as VATS began to be employed to drain the pleural cavity as part of the wider trend towards minimally invasive surgery in the 1990's. Kern and Rodgers are recognised as the pioneers of this technique, first publishing their results in 1993 (Kern and Rodgers, 1993; Stovroff *et al.*, 1995).

The final major change in the treatment of empyema has been the introduction of intrapleural fibrinolytics to breakdown the fibrinous septations that divide up an organised empyema. Their use was first suggested in 1949, although they were not introduced into routine practice until much later (Kokoska and Chen, 2009). Future therapies may result from an increasing understanding of the

immunobiology of empyema, and are likely to be directed towards disruption of the pathogen triggered immune-mediated inflammatory pathways that drive the development of the condition.

1.4.4 Current management

Empyema is a painful condition due to stimulation of the pain receptors within the parietal pleura and adequate analgesia is very important. Appropriate analgesics include paracetamol, non-steroidal anti-inflammatory drugs and opiates (Balfour-Lynn *et al.*, 2005). Intravenous fluids are often needed initially due to poor fluid intake and increased losses.

Patients should receive high dose antibiotics, but there is little published evidence regarding choice of agent. The British Thoracic Society guidelines suggest that Cefuroxime, Amoxicillin-clavulanate, Penicillin and Flucloxacillin, Ampicillin and Flucloxacillin or Clindamycin are all suitable (Balfour-Lynn *et al.*, 2005). Nosocomial, post-surgical or immunocompromised cases may require broader spectrum treatment to include coverage of Gram negative and anaerobic organisms.(Balfour-Lynn *et al.*, 2005) Antibiotics are usually given intravenously before switching to the oral route following defervescence.

The optimal treatment of empyema remains controversial. The absence of a firm evidence base, differing views on the most important outcome measures and significant regional variation in access to specialised services has resulted in varying approaches to management. It is agreed that the primary aim of treatment is clearance of the pleural cavity. Time to defervescence and length of hospital stay (LOS) are important surrogate measures of a successful response to treatment. There are two major contrasting approaches to management. Some favour early surgical intervention and drainage of the pleural cavity. This is usually, but inaccurately referred to as “decortication”. Other centres prefer an initial medical approach using instillation of fibrinolytics via an intrapleural chest drain, reserving definitive surgery for the approximately 10% of patients who fail to respond.

The presence of fibrinous loculations separating pools of infected fluid within the pleural cavity means that chest drainage alone is unlikely to be effective. Effective clearance of pleural loculations can be achieved either by instillation of intrapleural fibrinolytics via a chest drain or by surgical intervention. The evidence for intrapleural fibrinolysis is limited. One multi-centre RCT supported

this approach reporting a significant reduction in length of stay of 2 days compared to normal saline placebo, but no reduction in time to defervescence (Thomson *et al.*, 2002). The trial used Urokinase given intrapleurally twice daily for three days using 40,000 units in 40 ml of 0.9% saline and a four-hour dwell time. A later RCT found no reduction in length of stay with Streptokinase, another form of fibrinolytic, when compared to saline as a control (Singh *et al.*, 2004).

Different types of chest drain are available but two types are in routine usage, pig-tail or large bore surgical drains. A post-hoc analysis of the Urokinase trial compared pig-tail drains with large bore surgical drains and found that length of stay was 2 days lower with pig-tail drains but this may have been confounded by a centre effect (Thomson *et al.*, 2002).

There are two main surgical techniques for the treatment of empyema:

1. "Mini-thoracotomy". This achieves debridement and evacuation in a similar manner to VATS, but it is an open procedure. This is usually referred to as "decortication", although technically this is not the case, as true decortication involves stripping the parietal pleura, which is not usually required in paediatric cases.

2. Video-assisted thoracoscopic surgery (VATS). This involves thoracoscopic opening of the pleural cavity, debridement of the "rind" formed from fibrinous pyogenic material encasing the affected lung, breakdown of loculations and drainage of pus from the pleural cavity under direct vision.

A chest drain is usually left in situ after either form of surgery, although this can often be removed after a few hours.

The evidence base related to surgical intervention in empyema is weak. The literature is divided between commentators and studies favouring initial medical/non-operative management including Barnes *et al.* (2005); Jaffé and Balfour-Lynn (2005); Khalil *et al.* (2007) and those favouring primary operative intervention which include Coote and Kay (2005); Li and Gates (2008); Shah *et al.* (2011). A substantial number of these publications are retrospective reviews of single centre experiences of a small number of cases are therefore of limited usefulness. Of those published studies which are of

higher quality (RCT's, meta-analyses and large multi-centre retrospective cohort studies) the results are mixed.

A meta-analysis compared primary operative with non-operative management and suggested that early operative intervention was associated with lower mortality, re-intervention rate, LOS, time with tube thoracostomy and duration of antibiotic therapy compared with non-operative management (Avansino *et al.*, 2005). However, this analysis included all forms of surgery and mixed various medical management options including antibiotics alone, making it difficult to have confidence in the author's conclusions.

A large scale retrospective cohort study compared primary operative drainage with medical management reported a reduction in LOS of 4.3 (95% confidence intervals 2.3-6.4) days in favour of operative management, as well as reduced costs and treatment failure rate (Li and Gates, 2008). This analysis had the advantage of using data gathered over one year reducing the potential for temporal trends that could introduce bias into retrospective cohort studies. The authors also took steps to control for potential centre effects. The authors used hospital coding data and reported an unusually low rate of fibrinolysis usage (7%), this raises concerns that the medical management group may not be representative and the procedural coding data may not have been accurate.

Three RCTs have compared VATS with fibrinolysis and drainage as a primary treatment, with mixed results (Kurt *et al.*, 2006; Sonnappa *et al.*, 2006; St. Peter *et al.*, 2009). St Peter *et al* (2009) and Sonnappa *et al* (2006) reported no significant differences in LOS between VATS and fibrinolysis and both sets of authors favoured fibrinolysis over VATS, on grounds of cost and invasiveness. Kurt *et al* (2006) compared primary VATS with initial chest tube drainage and subsequent fibrinolysis if primary chest tube drainage had failed making comparison with the other two trials challenging. They described a reduction in LOS, cost and risk of transfer when the primary drainage procedure was VATS. This is similar to the findings of Avansino *et al* (2005) and Li and Gates (2008), suggesting that chest drainage alone is less successful than primary VATS. Also of note is the large multi-centre retrospective cohort study conducted by Shah *et al* (2011), which compared VATS with fibrinolysis. These authors found no difference in length of stay between VATS and fibrinolysis but did find a lower re-intervention rate in the VATS group.

There is no trial evidence comparing VATS to mini-thoracotomy, although Li and Gates (2008) included a sub analysis comparing the two and found no significant difference in LOS, costs, treatment failure and complications between the two techniques. It is therefore currently not possible to determine whether one technique is superior to the other. Similarly, the benefits of early versus late surgical intervention remain unclear. VATS does require highly specialised and expensive equipment, and operators who are specifically trained in the technique. It is also a far more time consuming procedure than simple decortication.

A summary of the published evidence regarding primary medical management and primary surgical management is shown in **Table 1.2**.

Study	Type of study	Comparison	Results
Avansino (2005)	Meta-analysis	Any surgery vs. CTD or IFT	Primary operative approach (VATS/OT) resulted in lower mortality, less need for re-intervention, shorter LOS, shorter duration of chest tube and antibiotics compared to primary medical therapy
Kurt (2006)	RCT	VATS vs. CTD	VATS lowered LOS and cost and risk of transfer
Sonnappa (2006)	RCT	VATS vs. IFT	No difference between VATS and IFT in terms of LOS, days of CTD and procedure number. IFT less expensive
Li and Gates (2008)	Retrospective cohort review	Any surgery vs. CTD or IFT	Primary operative approach lowered LOS, cost and risk of transfer
St Peter (2009)	RCT	VATS vs. IFT	No difference between VATS and IFT in terms of LOS, febrile days and oxygen requirement
Shah (2011)	Retrospective cohort review	VATS vs. IFT	VATS lowered need for re-intervention; no difference in LOS between VATS and IFT

Table 1.2 Comparison of primary approaches in empyema treatment ***CTD - Chest tube drainage; IFT- Intrapleural fibrinolysis therapy; OT- Open thoracotomy; VATS - Video-assisted thoracoscopy; LOS - Length of stay; RCT - randomised controlled trial*** **Comparison of primary approaches in empyema treatment. Adapted from Anselmo M** ***Can Resp J*** **(2010).**

The most important factor determining optimal outcome is almost certainly the co-ordinated approach of an experienced team of surgeons, anaesthetists and paediatricians rather than the exact surgical approach.

1.5 Study aims, hypotheses and thesis overview

The epidemiology of paediatric empyema has changed over the last twenty years with a substantial rise in the incidence of the condition reported on different continents including Europe (Gupta and Crowley, 2006), North America (Li and Tancredi, 2010), and Asia (Finley *et al.*, 2008). Despite this significant increase in incidence the processes responsible remain unclear.

Historically different pathogens have predominated in childhood empyema (Heuer, 1932; Chonmaitree and Powell, 1983) but recently the majority of cases in the United Kingdom have been caused by *S. pneumoniae* (Eltringham *et al.*, 2003; Saglani *et al.*, 2005). Only a limited range of strains or serotypes appear responsible for cases of pneumococcal empyema (Byington *et al.*, 2002b; Eltringham *et al.*, 2003; Fletcher *et al.*, 2006). At present all pneumococcal vaccines are of limited valency, protecting against only a finite number of serotypes. As a consequence, the introduction of these vaccines has the potential to substantially change the epidemiology of pneumococcal disease by altering the ecology of the species. The introduction of the seven valent pneumococcal conjugate vaccine in the USA was associated with an increase in the incidence of empyema, but this was linked to an expansion in disease due to non-vaccine serotypes (Byington *et al.*, 2006). The impact of the introduction of this vaccine on the epidemiology of paediatric empyema in the UK has yet to be firmly established.

The best method of treatment of empyema in childhood is controversial, with significant variation in approaches between hospitals and there is a paucity of comparative data to guide clinicians (Balfour-Lynn *et al.*, 2005; Thomas *et al.*, 2009). The rise in incidence of empyema has made it increasingly important that different treatment methodologies are compared in an

objective manner for the benefit of, both individual patients and to optimise the use of limited health care resources.

This thesis will describe the epidemiology of paediatric empyema in details in order to understand the factors contributing to the rise in incidence. In addition, it will establish the impact of the introduction of the conjugate pneumococcal vaccine to the routine childhood immunisation schedule on the incidence of paediatric empyema in the UK. Finally, it will evaluate the comparative effectiveness of different treatment methodologies in empyema in children in the UK.

Chapters 2 – 5 are focused around a series of hypotheses in line with these broad aims. These are:

Chapter 2: Patterns in the incidence of paediatric empyema thoracis, the relationship to pneumonia and temporal associations with the activity of common bacterial respiratory pathogens

Key hypotheses:

- *The incidence of empyema has increased*
- *The increase in incidence of empyema has been driven by an increase in pneumonia*
- *The incidence of empyema shows seasonality and cyclicity*
- *The increase in the incidence of empyema is related to the activity of S. pneumoniae, S. aureus, S. pyogenes and M. pneumoniae*
- *Seasonality and cyclicity seen in empyema is associated with the seasonality and cyclicity in the activity of S. pneumoniae, S. aureus, S. pyogenes and M. pneumoniae*

The incidence of paediatric empyema is modeled at local, regional and national level in the UK to establish robustly the presence and the magnitude of any increase in paediatric empyema. Seasonality in empyema incidence is investigated and defined in order to achieve this. The relationship between pneumonia and empyema is then addressed by modeling trends in hospital admissions for both conditions between 1997 and 2006 in England. The epidemiology of the condition is then further investigated by analysing national admissions

data for evidence of cyclicity and relating time trends and cyclicity within admissions data to the activity of common bacterial respiratory pathogens.

Chapter 3: Spatial epidemiology of paediatric pneumonia and empyema

Key hypotheses:

- *There is spatial variation in the risk of pneumonia and empyema.*
- *Spatial variation in risk of pneumonia and empyema will be associated with community level deprivation, ethnicity and proximity to hospital.*

The spatial epidemiology of pneumonia and empyema in the North East of England is defined using specific spatial modelling techniques. The relationship of risk of pneumonia and empyema to community level co-variates including child specific deprivation measures, ethnicity, migration and health-care related factors is then investigated to evaluate possible factors associated with the spatial epidemiology of both conditions.

Chapter 4: The impact of the seven valent pneumococcal conjugate vaccine on paediatric empyema in the North East of England

Key hypothesis:

- *The introduction of the seven valent pneumococcal conjugate vaccine was associated with a reduction in the incidence of paediatric empyema in the North East of England*

The impact of the introduction of PCV-7 on the incidence of empyema in the North East of England is evaluated by use of an interrupted time series approach. This study has been prepared for submission for publication.

Chapter 5: The influence of pathogen and pneumococcal serotype on illness length and the impact of primary pleural drainage methodology on outcomes in empyema

Key hypotheses:

- *Time to disease resolution in empyema is dependent on pathogen and pneumococcal serotype.*
- *Surgical methods of primary pleural drainage in paediatric empyema are associated with shorter hospital stay.*

The impact of pathogen and pneumococcal serotype on illness length are investigated to establish whether the changes in the epidemiology of empyema have impacted on the clinical profile of the disease. Nationally representative data are then used to compare the effectiveness of primary pleural drainage method on LOS and readmission and complication rates in paediatric empyema in the UK.

Chapter 6 is a summary of the previous chapters and conclusions.

2 Patterns in the incidence of paediatric empyema thoracis, the relationship to pneumonia and temporal associations with the activity of common bacterial respiratory pathogens

2.1 Introduction

2.1.1 Incidence of disease

Evaluating temporal trends in disease relies on the interpretation of consistent and comparable measurements of disease across time. The incidence rate (IR) is perhaps the most commonly used measure in this regard (Rothman *et al.*, 2008). It is defined as the number of new cases divided by the person-time period of measurement and can be expressed mathematically as:

$$Incidence\ rate = \frac{number\ of\ new\ cases}{\sum individuals \times time\ spent\ in\ population}$$ (Rothman *et al.*, 2008)

Imprecision in the estimates of incidence rates can occur when there is a lack of certainty over the definition of the population at risk and when recurrent episodes of disease occur in the same individual. However, incidence rates are easy to calculate, easy to interpret and widely reported, making comparisons across studies possible.

Analysis of changes in incidence rates of disease over time can have several difficulties. For example, the changing efficacy of detection of disease can mask true changes in incidence. This is a particular problem in longitudinal studies of infectious diseases based on surveillance, whereby increased awareness of disease and surveillance may lead to increased reporting year on year and artificial inflation of estimates of disease (Miller *et al.*, 2011). Likewise healthcare rarely remains the same over time; diagnostic technologies and methods evolve; disease definitions are refined based on new emerging knowledge e.g. the move from the diagnostic and statistical manual of mental disorders DSM-IV to DSM-V which will lead to major redefinitions of three common and important mental illnesses – attention-deficit hyperactivity disorder, bipolar disorder and major depressive disorder (Allen, 2010). Such changes can introduce bias and misclassification error in longitudinal studies. Furthermore, healthcare systems evolve over time which can lead to difficulties in defining populations at risk and ascertainment of cases as base populations may be different and lack contiguity. All of these factors increase in relevance the longer the time period examined. They are also more likely to be an issue if secondary data sources are used e.g. health-care utilization data.

2.1.2 Incidence of empyema – local, regional, national and international estimates

Empyema exists as part of a spectrum of pleural disease from an uncomplicated pleural effusion to the development of a fibrothorax. As it exists on a spectrum of disease, precisely defining it for the purposes of measuring incidence can be challenging. There is significant variation in definition across the literature (Byington *et al.*, 2002b; Padman *et al.*, 2007). Some of this variation is manifest in the range of terminology used. For example, complicated pneumonia and complicated parapneumonic effusion are terms frequently seen in the literature which can encompass empyema, although they can also refer to patients in the exudative or fibrinopurulent stages of pleural infection or other complications of pneumonia such as necrotic disease e.g. Wexler et al (2006) where complicated pneumonia included presence of pleural effusion (83 %), empyema (52 %), atelectasis (26 %), pneumatocele (19 %), and pneumothorax (10 %).

The traditional and literal definition of empyema is the presence of pus within the pleural space. Therefore, the diagnosis of empyema requires invasive sampling of the pleural space and confirmation of the presence of pus. In children, this is combined with definitive treatment, in order to prevent the need for multiple painful procedures. As a consequence, disease definitions used in epidemiological studies often include treatment criteria, most often a requirement for invasive drainage. However, including treatment criteria within the disease definition complicates matters because significant variations in treatment exist between centers and even within centers (Thomas *et al.*, 2009; Thomas *et al.*, 2011). These variations arise because there is a lack of objective criteria as to which children require drainage and a lack of consensus as to which treatment modality is most effective. The consequence is that using a disease definition which includes a treatment component, e.g. in Spencer *et al* (2006) potentially introduces inaccuracies when comparing incidence between areas and time periods. These variations are likely to be even greater between countries than within countries.

Reports of a change in the incidence of paediatric empyema emerged from the UK in the late 1990's. Two groups published reports of an increase in the number of cases of paediatric empyema presenting to separate hospitals (both tertiary management centres for empyema) in 1997 (Playfor *et al.*, 1997; Rees *et al.*, 1997). However, this apparent increase did not appear to have been maintained when Playfor *et al* (1999) reported their further experience two years later. Shankar et al (2000) did not observe a rise in incidence in their single centre study covering the period 1980 to 1997. In contrast, Spencer *et al* (2006b) reported a six fold rise in cases of surgically managed empyema in from 1995 to 2006 at a further UK hospital.

Three national studies in the UK reported an increase in hospital admissions for empyema; Gupta and Crowley (2006) found that admissions in England rose from 14 per million children (0–14 years) in 1995 to 26 per million in 2003. Similarly, in Scotland admissions increased from <10 per million in 1998 to 37 per million in 2005 (Roxburgh *et al.*, 2008). More recently, Koshy et al (2010) published updated data for England showing an increase in admission rates from 5 to 14 admissions per million children between 1997 and 2008. All three publications used hospital coding data and used two specific ICD-10 codes to identify cases – J86.0 and J86.9 (pyothorax and pyothorax with fistula) although the precise application varied between the three studies. The same method of identification of cases of empyema has recently been shown to have a high predictive value in adults aged from 15-39 in Denmark (96% positive predictive value) but its accuracy in paediatric empyema is unknown (Sogaard *et al.*, 2011).

Following the reports of an increase in incidence of empyema in UK, a number of reports from elsewhere in the world emerged (summarized in **Table 2.1**). Byington et al (2002b) published the first non-UK report of an increase in parapneumonic empyema in Utah, USA. Similar single centre reports showing an increase in empyema have been published in Belgium and Spain (Deiros Bronte *et al.*, 2006; Van Ackere *et al.*, 2009). Large-scale or national-scale studies of coding data have demonstrated increases in Canada, Australia and the USA (Finley *et al.*, 2008; Grijalva *et al.*, 2010; Lee *et al.*, 2010; Li and Tancredi, 2010). Three studies have reported an increase in empyema specifically as a complication of pneumococcal pneumonia (Hsieh *et al.*, 2004; Wexler *et al.*, 2006; Obando *et al.*, 2008) leading to the possibility of the increase in incidence being linked to changes within pneumococci.

Source	Country	Disease Grouping	Estimation Method	Age Group	Time Frame	Change in Incidence
Wexler et al (2006)	Israel	Complicated pneumococcal pneumonia†	Retrospective case note review	0 – 16 yrs	1986 – 1997 First year of rise 1991	5 cases/yr (1986-90) – 12 cases/yr (1991-97)
Finley et al (2008)	Canada	Empyema	Hospital coding data	0 – 19 yrs	1995 – 2003 First year of rise 1996	Age specific data not presented however: Ratio of rate in 1995: rate in 2003 - 2.20 (95% CI 1.56 to 3.10)
Deiros-Bronte et al (2006)	Spain	Parapneumonic effusion	Retrospective case note review	< 15 yrs	1993 – 2003 First year of rise 1995	1993 - 18.1 per 100,000 children 2003 - 42.9 per 100,000
Byington et al (2002)	USA	Parapneumonic empyema	Retrospective case note review	< 19 yrs	1993 – 1999 First year of rise 1995	1993 – 1 per 100,000 (<19yrs) 1999 – 5 per 100,000
Strachan et al (2009)	Australia	Empyema	Hospital coding data	< 19 yrs	1993 – 2005 First year of rise 1996	1993-94 - 4 hospital admissions per million (<19yrs) 2004-05 – 9.4 per million
Grijalva et al (2010)	USA	Empyema	Hospital coding data	<5 yrs	1996 – 2007 First year of rise 1997	1996-98 - 3.5 hospitalisations per 100,000 children 2005-2007 7.0 per 100,000
Hsieh et al (2004)	Taiwan	Complicated pneumonia‡	Retrospective case note review	<15 yrs	1995 – 2002 First year of rise 1997	1995 – 25% of total pneumococcal pneumonia admissions complicated pneumonia 2002 – 70%

Obando et al (2008)	Spain	Pneumococcal parapneumonic empyema	Prospective recruitment, retrospective identification through microbiological isolates	<18 yrs	1998 – 2006 First year of rise 1999	1998 – 5 cases* 2006 – 66 cases*
Van Ackere et al (2009)	Belgium	Complicated parapneumonic effusion	Retrospective case note review	<18 yrs	1993 – 2005 First year of rise 2000	1995-2002 – 20 to55 admissions per 100,000 hospital admissions 2002-05 – 120 to140 admissions per 100,000 hospital admissions
Li et al (2010)	USA	Empyema and complicated pneumonia (empyema, pleural effusion or bacterial pneumonia requiring a chest tube)	Hospital coding data	<18 yrs	1997 – 2006 First year of rise 2000**	1997 empyema hospitalisations 2.2 per 100 000 children, 2006 empyema hospitalisations 3.7 per 100 000.
Lee et al (2010)	USA	Local complications of community acquired pneumonia (>97% empyema)	Hospital coding data	<18 yrs	1997 – 2006 First year of rise 2000**	77.8% increase in local complications 1997 5.4 cases per 100 000 population, 2006 9.6 cases per 100 000 population

† Complicated pneumonia defined as pleural effusion (83 %), empyema (52 %), atelectasis (26 %), pneumatocele (19 %), and pneumothorax (10 %). Study population of four large hospitals in Jerusalem.

‡Complicated pneumonia defined as necrotizing pneumonia and/or empyema

*Refers to figures from Seville and Malaga

** Both studies did not use serial years, instead using 1997, 2000, 2003 and 2006 as estimates so first year of rise not accurately captured

Table 2.1 Global changes in incidence of paediatric empyema

There has been consistency in the timing of the rise in incidence of empyema with >90% of studies reporting the rise as having begun in the five year period between 1995 and 2000 **(Table 2.1**). Also highlighted is the difficulty around precise definition of empyema with studies utilising a range of different definitions. This in turn makes direct comparisons more challenging, although all the studies where comparisons were possible fell within a relatively consistent range of between 1 admission per 100,000 individuals and 10 admissions per 100,000 individuals (Roxburgh *et al.*, 2008; Lee *et al.*, 2010). None of the cited reports come from the developing world. Publication of data on empyema may be less likely from developing countries for several reasons including presence of other health priorities; higher death rates from initial infections; less funding available for health research; less resources available for diagnosis of empyema which may impact on recording, although it may also reflect the absence of an increase in incidence of empyema.

Aside from Playfor *et al* (1999) and Shankar *et al* (2000) three other reports have failed to find an increase in the incidence of paediatric empyema or observed a rise followed by a reduction. Schultz et al (2004) found a halving in cases at one Texas children's hospital after 2000. Buckingham et al (2003) tracked the annual incidence of complicated parapneumonic effusions and saw a progressive increase from 4.5 per 10000 discharges in 1996 to 25.0 in 1999 ($p=0.0001$), then a subsequent decline to 10.1 in 2001 ($p=0.03$) at their centre. Koshy *et al* (2010) documented the substantial rise in empyema in England between 1998 and 2006, but noted a decrease of approximately 20% from 2006 to 2008 which they attributed to the introduction of the pneumococcal conjugate vaccine. However, their estimates of the incidence of empyema were substantially lower than other studies using the same or similar data and their methodology failed to account for seasonal variation in cases numbers which could potentially explain the variability within their results.

2.1.3 Relationship to pneumonia

Given the pathophysiological relationship between the two conditions, one of the first hypotheses proposed to explain the increase in incidence of empyema was a rise in underlying cases of pneumonia (Roxburgh *et al.*, 2008). A number of studies have reported changes in time of both conditions, but only four have addressed this specific hypothesis. Those that have, have reported contradictory results. These studies are listed in **Table 2.2**. Buckingham *et al* (2003) and Deiros Bronte *et al* (2006) retrospectively identified the proportion of pneumonia cases complicated by empyema at single centres and found that there was no variation over time in this, suggesting the changes in incidence empyema were attributable to changes in the incidence of pneumonia. In contrast, Roxburgh *et al* (2008) and Strachan and Jaffe (2009) looked indirectly at the relationship between the two conditions by comparing longitudinal trends in national coding data of both conditions. Roxburgh *et al* (2008) found that both empyema and pneumonia had increased in the 25 years between 1980 and 2005, but that the increase in empyema had occurred much later (>1995) and with a much steeper gradient than the increase in pneumonia. Strachan and Jaffe (2009) found that while empyema had increased significantly, pneumonia had not although there was a trend to an increase, suggesting some variation in the rate of empyema complicating pneumonia. Although not directly addressing the hypothesis of pneumonia driving the increase in empyema, three studies of national coding data in the USA reported an increase in empyema and a decrease in pneumonia after the introduction of the seven valent pneumococcal vaccine, suggesting variation in the rate of empyema complicating pneumonia over time (Li and Gates, 2008; Grijalva *et al.*, 2010; Lee *et al.*, 2010).

There is historical evidence that the incidence of empyema has been previously affected by periodic or cyclic changes in pathogen virulence which appeared to increase the rate of progression from pneumonia to empyema (Heuer, 1932; Middelkamp *et al.*, 1964; Clagett, 1973). For example, Heuer (1932) was the first to relate patterns in empyema incidence and mortality to the perceived virulence of the predominant pathogen causing pneumonia and observed significant annual variations between streptococcal and pneumococcal empyema. However, no authors thus far have formally investigated and quantified periodicity within longitudinal studies of the relationship between empyema and pneumonia.

Study	Country	Data	Methodology of comparison	Conclusions
Buckingham S et al (2001)	USA	Retrospective case note review	Proportion over time	Increase and subsequent decrease proportionally of both pneumonia and empyema
Roxburgh CS et al (2007)	Scotland	Hospital coding data	Indirect comparison of trends	Both conditions increased, but rise in empyema occurred later and not correlated with pneumonia rise
Bueno-Campana M et al (2008)	Spain	Retrospective case note review	Proportion	Increase in pneumonia, but constant proportion of empyema over time
Strachan R et al (2009)	Australia	Hospital coding data	Indirect comparison of trends	Increase in percentage of empyema as proportion of pneumonia
Grijivalva et al (2010) Lee et al (2010) Li et al (2010)	USA	Hospital coding data	No direct comparison	All three studies reported an increase in empyema despite a decrease in pneumonia, but did not formally address the relationship between the two

Table 2.2 Studies addressing the hypothesis that the increase in empyema was driven by an increase in underlying cases of pneumonia.

2.1.4 Seasonality, periodicity and cyclicity in respiratory infections including empyema

Most human respiratory pathogens e.g. Influenza and respiratory infections e.g. pneumonia are seasonal with peaks in frequency in certain months, in addition they also show year on year variations in cyclicity or periodicity (Dowell and Ho, 2004). The mechanisms underpinning seasonality and cyclicity in respiratory infections are uncertain and have perhaps been neglected as a research topic in the past. Fisman (2007) suggests four broad mechanisms that may be relevant. These are:

1) Population behaviours – Seasonal variations in human activities may lead to alterations in disease transmission which support seasonality. Examples include increased viral transmission at the start of the school year in September or the mid-winter peak in pneumococcal disease which appears to coincide with Christmas and New Year holiday period when population mixing between age groups, e.g. children and grandparents, may potentiate transmission.
2) Pathogen-pathogen interactions – The seasonality of some pathogens may allow increases in seasonal transmission of others e.g. seasonal influenza and pneumococcal and meningococcal disease.
3) Environmental effects on pathogens – Certain environmental conditions appear to be favourable to some pathogens e.g. cold dry air appears to increase influenza viral particle's survival in the environment.
4) Environmental effects on hosts – There may be seasonal variation in human susceptibility to pathogens e.g. there is evidence of seasonal variation of immune function linked to photoperiod which may underpin infectious disease seasonality.

Seasonality and cyclicity within empyema were described as early as the 1930's but neither the magnitude nor the mechanisms underpinning them have been explored (Heuer, 1932). Seasonality and cyclicity both pose potential problems to longitudinal estimates of incidence in empyema. For example, if a condition is clustered in the winter, individual seasons are likely to differ in length and timing and therefore may be split across different years unevenly. This may under or overestimate the true incidence dependent on the timing and the subsequent distribution of the season within individual calendar years. The presence of cyclicity may also lead to inaccuracy in the estimation of the change in incidence, in particular if the magnitude of change is estimated by comparison of two arbitrary start and end points.

Both empyema and pneumonia are subject to strong seasonal variations in incidence, with cases predominantly concentrated in the winter months (Hardie *et al.*, 1996; Clark *et al.*, 2007a). Furthermore, there is historical evidence that empyema has been previously affected by periodic or cyclic changes in both pathogen (periods of pneumococcal or staphylococcal dominance) or pathogen virulence (increase in rate of progression from pneumonia to empyema) (Heuer, 1932; Middelkamp *et al.*, 1964; Clagett, 1973; Chonmaitree and Powell, 1983). However, this periodicity or cyclicity has never been quantified (**Figure 2.1**) highlights evidence of possible cyclicity within empyema admissions in a recent study of the incidence of empyema in the USA (Grijalva *et al.*, 2010).

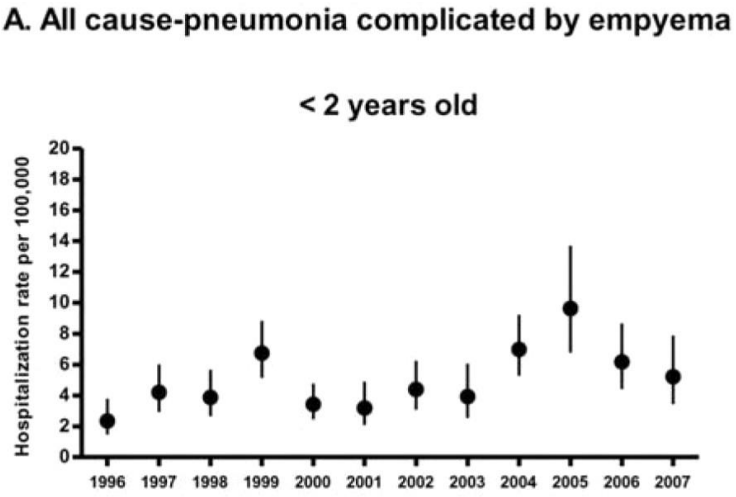


Figure 2.1 Hospital admissions for all cause-pneumonia complicated by empyema from a US in-patient sample, with evidence of potential six year cycle with peaks in 1999 and 2005 as well as upward trend (Grijalva *et al.*, 2010)

2.1.5 Bacterial respiratory pathogens associated with empyema

The microbiological diagnosis of empyema is difficult and a causative agent is only isolated by routine culture in approximately 20-40% of cases (Byington *et al.*, 2002b; Eastham *et al.*, 2004). Molecular diagnostic techniques improve this to upwards of 60% (Saglani *et al.*, 2005; Le Monnier *et al.*, 2006). When summarising all published series of the bacteriological causes of empyema in children from 1950 onwards, three organisms – *Streptococcus pneumonia*e, *Streptococcus pyogenes* and *Staphylococcus aureus* predominate (**Figure 1.2**).

Streptococcus pneumoniae appears consistently as the commonest cause of empyema in most recent epidemiological studies, including those that use molecular testing (Byington *et al.*, 2002b; Eastham *et al.*, 2004). *S. pneumoniae* is an obligate human pathogen and is frequently carried in the nasopharynx (van der Poll and Opal, 2011). Carriage is highest in children under the age of five and associated with

institutional day care (Obaro *et al.*, 1996). Over 90 capsular serotypes are recognised although only a smaller number appear associated with empyema (Byington *et al.*, 2002b). It is the leading bacterial cause of pneumonia in children, accounting for approximately 20-30% of cases (van der Poll and Opal, 2011). Seasonal changes in the incidence of invasive pneumococcal disease are well recognised with peaks predominantly in the winter (Dowell *et al.*, 2003).

Staphylococcus aureus historically was an important cause of empyema. Several more recent studies have also reported it as the leading cause of empyema, particularly the Meticillin resistant strain (MRSA) (Buckingham *et al.*, 2003; Schultz *et al.*, 2004). Similar to *S. pneumoniae* it also carried in the nasopharynx and carriage is asymptomatic. Seasonality has been reported in *S. aureus* carriage, particularly in relation to seasonal upper respiratory tract viral infections (Harrison *et al.*, 1999).

Streptococcus pyogenes is a member of the group A streptococci. Like *S. pneumoniae, S. pyogenes* is an obligate human pathogen with no other biological host recognised. It is an important cause of empyema, although not the most frequent in recent cases series (Bessen, 2009). It can be carried asymptomatically and carriage is associated with both seasonal and epidemic or cyclical changes in prevalence (Lamagni TL *et al.*, 2008).

Mycoplasma pneumoniae is a rare cause of empyema. It has been widely observed to occur in 3-5 yearly epidemics with a low background rate of infectivity. It displays seasonal variation with increases in the summer months. *M. pneumoniae* was first identified in 1944 by Eaton from a patient with atypical pneumonia. It was initially thought to be a virus but reclassified once it's treatability with antibiotics was recognised. It is a ubiquitous human pathogen with evidence of infection in all populations studied thus far (Waites and Talkington, 2004).

It is recognised that different pathogens differ in their potential to progress from pneumonia to empyema and interactions between organisms may therefore be a key dynamic in understanding the recent change in the epidemiology of empyema. For example a leading cause of childhood pneumonia, *Mycoplasma pneumoniae*, is known to occur in 'epidemic-like outbreaks' of an approximate 48 month frequency and is much less likely to be a cause of empyema than the commoner causes *S. pneumoniae* and *S. aureus* (Ali

et al., 1986; Rasmussen JN, 2010) It could be therefore hypothesised that the rate of progression between pneumonia would be negatively correlated with isolations of *M. pneumoniae*. This could be explored by examining the temporal relationship of isolations of *M. pneumoniae* to empyema admissions and to the rate of progression from pneumonia to empyema.

The discussion so far has assumed that the patient is immunologically normal and has no other congenital or acquired abnormalities of the lungs. It is well recognised that a wide variety of other organisms which are not normally regarded as causing empyema may rarely do so in unusual situations. Examples would include children with congenital or acquired immunodeficiency and following trauma to the lungs.

2.1.6 Aims & hypotheses

The aims of this analysis were to investigate longitudinal trends in the incidence of childhood empyema and pneumonia at a regional and national scale in the UK with a view to quantifying the increase in incidence in both condition and the relationships between the two conditions. Longitudinal trends were investigated across three hierarchal populations. This approach was intended to establish whether changes in incidence were comparable, consistent, robust, and a consequence of true change in disease rather than changes in classification or health care systems, as can be a problem in longitudinal studies of hospital coding data.

Both empyema and pneumonia show significant seasonal variation in incidence with concentration of cases within the winter months and therefore robust methods of addressing seasonality were required to accurately assess secular trends in incidence. This provided an opportunity to ascertain the magnitude of seasonal effects in empyema which have not been previously quantified and evaluate whether there were seasonal or cyclical trends beyond the annual cycle, the presence of which may provide information about the underlying mechanisms driving the change in incidence of empyema.

Empyema represents a clinical syndrome with multiple possible pathogenic causes. An important hypothesis therefore in investigating the change in incidence of empyema is whether changes in the activity of one pathogen over others were associated with the observed longitudinal trends in empyema and its relationship to pneumonia. This was investigated by testing the association between the incidence of empyema and activity of the different pathogens over time at a national scale. Different pathogens are also associated with different rates of progression of pneumonia to empyema. To further delineate the relationship between the two conditions, the relationship of the rate of progression of pneumonia to empyema and the activity of different pathogens over time was also investigated on a national scale.

Furthermore, as Fisman (2007) postulates a significant component of seasonality and cyclicity in respiratory infections are pathogen related changes. These factors were investigated in relation to the activity of different pathogens.

Secondary aims included the evaluation of accuracy of hospital coding data in determining the incidence of empyema compared to extant clinical data and estimation of the impact of any increase in empyema on health resources by calculating the costs associated with increasing empyema admissions.

Key hypotheses:

- *The incidence of empyema has increased*
- *The increase in incidence of empyema has been driven by an increase in pneumonia*
- *The incidence of empyema shows seasonality and cyclicity*
- *The increase in the incidence of empyema is related to the activity of S. pneumoniae, S. aureus, S. pyogenes and M. pneumoniae*
- *Seasonality and cyclicity seen within empyema is associated with the seasonality and cyclicity in the activity of S. pneumoniae, S. aureus, S. pyogenes and M. pneumoniae*

2.2 Methods

Data on the incidence of empyema were collated from two sources:

i) Data from the North East tertiary hospital responsible for the management of paediatric empyema
ii) Hospital episode statistics (HES) on paediatric pneumonia and empyema admissions in England and at a regional level

2.2.1 North East empyema data

A historical record of all paediatric empyema patients managed at the Freeman Hospital, Newcastle-upon-Tyne has been maintained since 1995.The Freeman Hospital is the regional cardiothoracic centre, serving the North East of England and the county of Cumbria. Patients with suspected empyema are referred from secondary care paediatric in-patient units across the region and are not directly admitted from primary care. There are no other paediatric empyema management centres within 100 miles by road of the Freeman Hospital. With this relative isolation, referral patterns are stable and there were no changes in the referral arrangements for children with empyema between 1995 and 2011. The UK-ESPE study received a favourable opinion from the Sunderland Regional Ethics Committee in October 2007 (**Appendix Q).** The National Information Governance Board approved the study, with informed consent required for the prospective collection of patient identifiable information but permission was given for the collection and storage of retrospective data.

Case definition

The case definition used was a clinical diagnosis of empyema with evidence of pleural fluid on imaging (either chest x-ray or thoracic ultrasound). All cases referred as suspected empyema were considered, but only those cases that required invasive management (pleural drainage and/or decortication) were included. The proportion of patients referred who required invasive management remained stable throughout the study period (83% 1995/2005 vs. 85% 2008/2009).

Data collection

The period from May 1995 to December 2011 was studied. Data from 1995 to 2005 were collected by retrospective examination of ward admission diaries which were searched to identify any patients admitted with empyema or suspected empyema. Individual case records were then examined to establish whether the case required invasive management and were cross-checked with theatre logs to ensure accuracy. These data have been previously reported (Spencer *et al.*, 2006b). Subsequent data were obtained as part of the UK Enhanced Surveillance of Paediatric Pneumococcal Empyema study (UK-ESPE). Cases from September 2006 to September 2008 were identified in the manner stated previously, applying the same case definition. From September 2008 onwards cases were identified prospectively through daily contact with the cardiothoracic team by a member of the research team. No data were collected for the period January-May 2006 due to a transition between research staff; hence this period was excluded from analyses leaving a total of 195 months available.

Data description

The data from the period 1995-2005 are limited to the number of cases per month and the post-code of each case. Data from 2006 onwards contained detailed demographic information including post-code, microbiological, management and outcome data.

Temperature data

Data were obtained from the Meteorological Office Integrated Data Archive System land surface station database (MIDAS). All available records for the study region were searched and the surface station centre closest to the main centres of population and with the most complete records for the time period was used (Wallington, Grid Ref: 036844). The mean monthly maximum air temperature was calculated and values were available for 193 of 195 months.

2.2.2 National and Regional Hospital episode statistics data

Case definition within coding data

The definition of pneumonia in studies utilising hospital coding data has been variable historically (Roxburgh *et al.*, 2008; Elemraid *et al.*, 2010; Koshy *et al.*, 2010). For this study, one comprised of ICD codes pertaining solely to bacterial pneumonia was used (J13-15 and J18.1). It includes all the codes that cover pneumonia with a specific bacterial diagnosis e.g. J13 Pneumonia due to *Streptococcus pneumoniae* and two more general codes. These are: J15.9 Bacterial pneumonia, unspecified and J18.1 Lobar pneumonia, unspecified (a full list of codes included is in **Table 9.1** in the Appendix). This is a more conservative definition than other studies in the same setting have used, as it excludes any codes relating to viral pneumonia (Roxburgh *et al.*, 2008; Koshy *et al.*, 2010). This is justified as empyema occurs in the vast majority of cases as a progression from bacterial pneumonia rather than viral pneumonia (Chernick *et al.*, 2006). Furthermore, there may be substantial seasonal shifts in both viral pneumonia and viral diagnostic testing in pneumonia in children which may obfuscate the true trends in pneumonia in relation to empyema (Byington *et al.*, 2002a; Chan *et al.*, 2002). It is possible that this is conservative, as it is recognised that a proportion of viral pneumonia may be occult, undiagnosed bacterial infection (Morens *et al.*, 2008). Nevertheless, given the aim of investigating the relationship between empyema and pneumonia it is possible that large numbers of cases of viral pneumonia which have no risk of progression to empyema may distort or obfuscate the true pattern of the relationship.

Two further codes, J15.9 Bacterial pneumonia, unspecified and J18.1 lobar pneumonia were included in the definition. J15.9 includes admissions where clinicians suspected bacterial pneumonia, but specific microbiological investigations had not been carried out, or where the results were unavailable when the admission was coded. Similarly lobar changes on a chest x-ray are likely to be due to bacterial pneumonia and therefore are likely to include admissions for bacterial pneumonia where a precise microbiological diagnosis was unavailable (Virkki *et al.*, 2002).

J86.0 & J86.9 which code for pyothorax with/without fistula respectively were used to identify empyema admissions in UK studies and were therefore used to allow direct comparisons to be made (Gupta and Crowley, 2006; Roxburgh *et al.*, 2008; Koshy *et al.*, 2010).

Hospital Episode Statistics data

HES is a database containing details of all admissions to NHS hospitals in England maintained by The Information Centre for Health and Social Care (*HESonline - Hospital Episode Statistics*, 2010). It contains data on hospital admissions from 1989 onwards and approximately 12 million new records are added yearly. Each record contains clinical information about diagnoses and operations, patient demographics including age, sex and ethnicity, administrative information and geographical information, such as post-code of residence. Clinical information is recorded in the form of codes from the International Classification of Disease (ICD), published by the World Health Organization (WHO, 1990). The tenth iteration was published in 1990 (ICD-10). The ICD-10 classification was adopted by the NHS in April1996 following a two year transition from the ICD-9 system. Up to 14 diagnostic codes were included in an individual patient record up until April 2007 (seven up until April 2002, 20 post 2007). The initial code or diag_01 refers to the primary diagnosis. All other codes represent secondary or subsidiary diagnoses. Codes are attributed by hospital coding departments and are returned by individual hospitals. Coders examine all patients' records and attribute the code based on their assessment of the records and the most appropriate diagnosis. Coders are specifically trained, but are not generally medically qualified. Each record refers to an individual episode of care under a named consultant, or other approved allied healthcare professional. As a result, a patient's single hospital admission can span multiple records if care was transferred between different consultants. The HES data are produced on the timetable of the financial year and consequently admissions that are ongoing at the time of the production of the data may not be included in that year's data, but will be included in the subsequent year.

Data were obtained for all respiratory diagnoses (J00-J99) in children and young people aged up to 18 years of age from April 1997 to April 2006. From this, monthly hospital episodes data were extracted for the period 1st April 1997– 30th March 2006 for all children aged less than 14 years with a recorded diagnosis of bacterial pneumonia or empyema.

The time-period 1st April 1997– 30th March 2006 was selected for two reasons:

- The coding classification fully changed from ICD-9 to ICD-10 in 1997, following a two year transition; hence data were not used from before this period to avoid the potential for differential ascertainment resulting from the change.
- To avoid the potential confounding effects of the seven valent pneumococcal vaccine which was introduced in September 2006.

Data from all years were allocated to a Strategic Health Authority (SHA) using SHA boundaries as from 1st July 2006. The SHA of residence was used for all analyses. A map of SHA boundaries is shown in **Figure 2.2** and total population for each SHA in **Table 2.3**. Data with each disease code were counted; those with an identical HES identification number were amalgamated in order to convert from episodes to hospital admissions. All duplicate records were removed. Readmissions were excluded if they occurred within a month of the original date of admission. A month was judged long enough because of the short average hospital stay in children (median 11 days for empyema from the UK-ESPE study). These counts were aggregated to give total counts of empyema and pneumonia admissions for each month both nationally and for each individual SHA. Admission rates were calculated using Office of National Statistics mid-year population estimates and all rates refer to admissions per million individuals aged between 0 and 14 years (*Mid-year population estimates*, 2011).

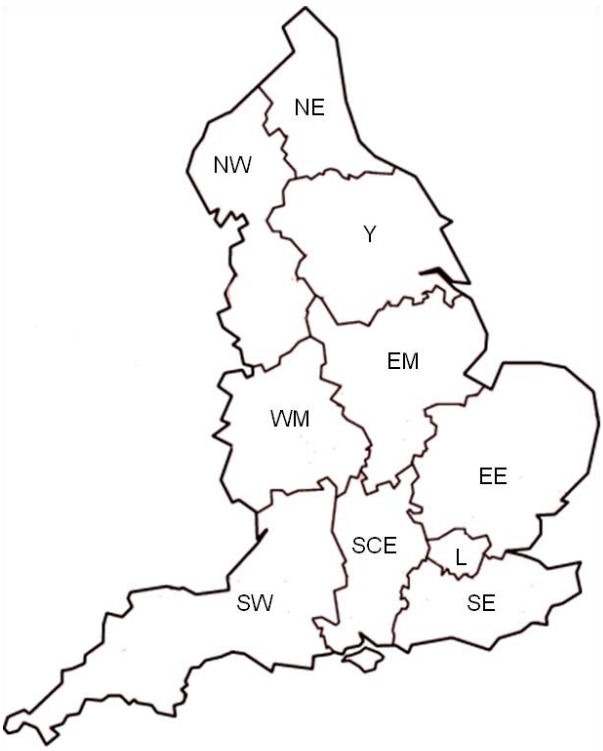


Figure 2.2 Map of Strategic Health Authorities in England in 2006.

SHA	North East (NE)	North West (NW)	Yorkshire (Y)	East Midlands (EM)	West Midlands (WM)
Population / million (2006)	2.5	6.8	5.0	4.2	5.3
SHA	East of England (EE)	London (L)	South East (SE)	South Central (SCE)	South West (SW)
Population / million (2006)	5.4	7.4	4.1	3.9	5.0

Table 2.3 Populations of Strategic Health Authorities in England in 2006.

2.2.3 Microbiological data

As precise attribution of empyema to a pathogen is challenging and thought to be poorly recorded in hospital coding data, an approach using a measure of pathogen activity rather than directly measuring cases of empyema recorded as being caused by a specific organism was used. Four pathogens were investigated. These were *S. pneumoniae, S. pyogenes, S. aureus* and *M. pneumonia*e. *S. pneumoniae, S. pyogenes* and *S. aureus* were the three commonest causes of empyema in a review of studies of paediatric empyema published after 1950. *M. pneumoniae* while a rare but recognised cause of empyema displays significant seasonality and cyclicity with epidemic-like outbreaks occurring regularly. In addition, as it is a well-recognised cause of pneumonia in older children but a rare cause of empyema it may significantly influence the rate of progression from pneumonia to empyema.

Monthly counts of isolations of the four organisms – *S. pneumoniae, S. pyogenes, S. aureus* and *M. pneumoniae* from children aged between 0 and 14 years in England from 1st April 1997 – 30th March 2006 (the same age group, geographical area and time period as the empyema admissions data) were obtained from the Health Protection Agency (HPA) of the UK Government. The term isolation refers to organisms detected by standard culture techniques in routine use in NHS hospitals from patient samples and notified to the HPA. Data for *S. pneumoniae, S. pyogenes* and *S. aureus* were for isolations defined as invasive. Invasive isolations are isolations from normally sterile body sites and include positive samples from blood, pleural fluid and cerebrospinal fluid. Data for *M. pneumonia*e was all detections of *M. pneumoniae* reported to the HPA Centre for Infections. All four sets of data were from routine surveillance carried out by the HPA which is the central body responsible for public health in the UK.

2.2.4 Statistical methods

Five discrete groups of analyses were carried out. Local changes in the incidence of empyema were evaluated using the Freeman hospital North East England data. National trends in empyema were analysed using the HES dataset. Regional variations in the change in empyema and pneumonia were analysed using the HES dataset at SHA level. The relationship of empyema to pneumonia was analysed using the national HES data. Finally the relationships between empyema and the activity of four different

pathogens – *S. pneumoniae, S. aureus, S. pyogenes* and *M. pneumoniae* was investigated using the national HES data and national surveillance data.

Overview

The overall aim of all the analyses was to investigate the changes in incidence in empyema and quantify the relationship between empyema and pneumonia. The outcomes of interest were therefore counts of cases or admissions at specific time points, and therefore in effect time series. Basic frequentist statistical analyses and linear regression rely on the fact that observed data are underpinned by mutually independent random processes. Time series by contrast are generally repeated serial observations and therefore are very unlikely to be independent, violating the assumption of independence required for linear regression and related techniques (Box *et al.*, 2008). Time series analysis therefore requires specific techniques to account for this property.

Several other concepts and definitions are important within time series. Trend refers to the non-random function present over time within a time series. Serial dependence is the concept that at least some observations are statistically dependent on other observations within the same series. The autocorrelation of a time series is the statistical function of the serial dependence present. Seasonality in general refers to known periodic behaviour e.g. annual variation in temperature; cyclicity refers to periodicity which may not be known *a priori*. Finally, stationarity refers to the fact that the stochastic variability (e.g. seasonality) within a time series does not vary with the time point sampled (Diggle, 1990). A range of methodologies have been developed to analyse time series dependent on the properties of series and the aim of the analysis e.g. forecasting, estimation of multi-variate relationships or extrapolation of trends.

Generalised least squares regression models

Generalised least squares (GLS) regression models are a form of extended linear model which allow the modelling of dependence within error structures (Pinhero and Bates, 2000). Historically, they have been developed for time series and spatial data, as both generally have the property of serial dependence (Pinhero and Bates, 2000). They are multivariate and allow the calculation of co-efficients describing the relative effect size of the suggested relationships between variables (Kitagawa, 2010).

Residual autocorrelation structures are used to model serial dependence. They vary in their sophistication but work by assuming the correlation between observations follows a predictable pattern. Zuur (2009) describes three levels of complexity. The simplest residual autocorrelation structure is known as compound symmetry and assumes that the residual correlation does not vary with time between observations. More complex is an auto-regressive model of order 1 (AR1), which models the residual at time *s* as a function of residual of time *s* – 1. The AR1 model can then be extended to a more complex model, the auto-regressive moving average model (ARMA). The ARMA model contains two terms, *p*, the number of auto-regressive parameters and *q*, the number of moving average parameters. The residuals are modelled as a function of the *p* previous time points and white noise. Different combinations of values of *p* and *q* may be need comparing before the optimal structure is found, although values above $p = 3$ and $q = 3$ may be impractical (Zuur *et al.*, 2009).

State-space models

Dynamic or state-space models are stochastic models which can be used to represent time series. State space models have been defined by Kitagwa (2010) in two stages:

$$x_n = F_n x_{n-1} + G_n v_n \ (System\ model)$$

$$y_n = H_n x_n + w_n \ (Observation\ model)$$

y_n is an *l* variate time series. x_n is a *k*-dimensional unobservable vector, referred to as the *state*, v_n is the *m* dimensional system or state noise, w_n the observation noise. F_n, G_n and H_n are the *k* x *k*, *k* x *m* and *l* x *k* matrices. The observation model is a regression model representing the time series y_n and the system model expresses the time change of the regression coefficients. This approach is therefore able to handle non-stationarity within data from example variation in both frequency and amplitude of seasonality within the time series. Lundbye-Christensen et al (2009) demonstrated the usefulness of this approach in modelling hospital admissions with myocardial infarction. Similarly Fanshawe et al (2008) used state-space models to predict spatio-temporal variation in air pollution levels in the North East of England.

Within GLS regression models, the influence of seasonality can be compensated for by the introduction of harmonic parameters, as has been used elsewhere (Jensen *et al.*, 2004), provided the seasonality has stationarity and does not vary in frequency or amplitude over time. If there is variation in seasonality over

time, then a modelling framework robust to time-varying regression coefficients should be used. State-space or dynamic models are able to achieve this.

Wavelet analysis

Wavelet analysis is widely used in a number of fields including ecology, epidemiology and engineering. In infectious disease epidemiology, wavelet analysis has been used to demonstrate the relationship between increasing epidemic phase in measles and their relationship to immunisation rates in the UK (Grenfell *et al.*, 2001), a highly significant association between El Nino, precipitation and dengue epidemics in Thailand (Cazelles *et al.*, 2005) and the presence of ecological interference in epidemics of unrelated human pathogens (Rohani *et al.*, 2003).

Wavelet analysis is based on transformation of the signal of a time series into a series of functions (wavelets) that represent both the frequency and time-scale within that signal. As frequency and time-scale are interdependent, decomposition of a time series necessitates a trade-off between time and frequency resolution and wavelet analysis has been described as the optimal trade-off between time-frequency resolution (Cazelles *et al.*, 2008). Wavelet analysis decomposes a time series into both a time domain and a frequency space, allowing the recognition of the relevant contributions of different periodicities within the time series (Zhang *et al.*, 2009). Wavelet analysis also has the advantage of not requiring stationarity of periodicity within the data and can evaluate the whole extent of a time series (Cazelles *et al.*, 2007).

Wavelet coherence analysis is a method of quantifying statistically the linear relationship between non-stationary time series. Wavelet coherence can be defined mathematically as

$$R_{x,y}(f,\tau) = \frac{\| < W_{x,y}(f,\tau) > \|}{\| < W_{x,x}(f,\tau) > \|^{\frac{1}{2}} \| W_{y,y}(f,\tau) > \|^{\frac{1}{2}}}$$

Where $R_{x,y}(f,\tau)$ is the wavelet coherence between two signals x(t) and y(t), '< >' denotes a smoothing operator in both time and scale. $W_{x,}(f,\tau)$ represents the wavelet co-efficient following wavelet transformation of time series x, with frequency f and τ the time point (Cazelles *et al.*, 2008). In addition the use of complex wavelets allows the calculation of phase difference between two time series.

General principles

Analyses were performed using the R statistical programme (R-Development-Core-Team, 2011), using the nLME (Pinhero *et al.*, 2011), MASS (Venables and Ripley, 2002), sspir (Dethlefsen *et al.*, 2009), WaveletCo (Tian and Cazelles, 2011), Epitools (Aragon, 2010) and Lattice (Sarkar, 2008) packages. A progressive model building approach was used with the simplest acceptable model preferred. The principle of parsimony was adhered to with the exclusion of non-significant terms (Pinhero and Bates, 2000). Residual distributions were plotted and quantified for serial dependence to ensure that model assumptions regarding linearity, homoscedascity, and independence were not violated. Where there were concerns over model assumptions log transformation was used to stabilise the variance and prevent violation of model assumptions (Faraway, 2005). Models were compared by a combination of Akaike Information Criterion (AIC) and the coefficient of determination R^2 (calculated as correlation between observed outcome variable and model fitted values). A significance level of $p<0{\cdot}05$ was used to assess the significance of regression co-efficients.

Local changes in the incidence of empyema (North East data)

The incidence rates of paediatric empyema managed at the Freeman Hospital, Newcastle-upon-Tyne were calculated using the ONS population estimates for the North East SHA and Cumbria PCT for each epidemiological year from 1995 to 2010. Epidemiological years run from 1st of July to the 30th of June the following year (HPA, 2012) and were used to minimise the influence of year to year variation in the timing of peaks in case numbers. 95% confidence intervals for each estimate were derived assuming a Poisson distribution of the residuals (Rothman *et al.*, 2008).

GLS regression models were then used to test the hypotheses:

a) that cases of paediatric empyema in the NE of England increased with time from 1995

b) that monthly cases of paediatric empyema in the NE of England varied with season between 1995 and 2010

The initial model (T0) was defined as:

$$\text{Empyema Cases} \sim \alpha + \beta x \text{Time from 1995} + \beta x \text{Season} + \epsilon$$

where α represented the intercept, β the regression co-efficient and ϵ the error term.

Monthly counts of cases were used in preference to annual counts for two reasons. Firstly to determine the magnitude of seasonal trends and secondly, to increase the power to detect secular trends in the incidence by increasing the number of observed time points. As cases were clustered in the winter and originated from one region, mean monthly air temperature from a central location in the North East was used to model the seasonal trend.

Plotting of the model T0 residuals showed evidence of non-linearity and autocorrelation (Appendix B figures - **Figure 9.1**, **Figure 9.2** and **Figure 9.3**). Count data can be transformed to an approximate

normal distribution by addition of a constant and application of a log transformation (Box and Cox, 1964; Gruttola and Tu, 1994), hence a second model (T1) was specified:

$$\log(\text{Empyema Cases} + 1) \sim \alpha + \beta x \text{Time from } 1995 + \beta x \text{Season} + \epsilon$$

Examination of the residuals of model T1 showed no violations of the assumption of linearity and only borderline evidence of autocorrelation (Appendix B figures - **Figure 9.4, Figure 9.5** and **Figure 9.6).**

Modelling national trends in empyema and pneumonia

National incidence rates for paediatric bacterial pneumonia and paediatric empyema in England were calculated for each epidemiological year from 1997/8 to 2004/5 using total hospital admissions. The ONS population estimates for individuals aged 0-14 years were used as the denominator and confidence intervals estimated assuming a Poisson distribution, as described previously.

Trends in admissions for paediatric bacterial pneumonia and paediatric empyema were investigated using GLS regression models and a progressive modelling framework to test a series of linked hypotheses (**Table 2.4**). These were:

a) that monthly admissions for paediatric bacterial pneumonia and paediatric empyema between April 1997 and April 2006 in England showed seasonal variation (Models M0a/b)
b) that monthly admissions for paediatric empyema were significantly related to monthly admissions for paediatric bacterial pneumonia in England between April 1997 and April 2006 (Model M1)
c) that monthly admissions for paediatric empyema in England increased between April 1997 and April 2006 independently of monthly admissions for paediatric bacterial pneumonia (Model M2)
d) that monthly admissions for paediatric empyema in England increased between April 1997 and April 2006 independently of monthly admissions for paediatric bacterial pneumonia and independent of population size (Model M3)
e) that monthly admissions for paediatric empyema in England increased between April 1997 and April 2006 independently of monthly admissions for paediatric bacterial pneumonia and independent of seasonal variation (Model M4)

Monthly counts of admissions of either pneumonia or empyema were the initial outcome variable for models, however as with the NE data evaluation of the residuals of models M0a/b showed significant autocorrelation in both models and evidence of non-linearity with increasing dispersion of the residuals as the fitted values increased (Appendix C **Figure 2.29-2.34**). Log-transformation was therefore used to transform the data to an approximate normal distribution. This was then used for all subsequent models that included count data.

Name	Model structure
M0a	Empyema admissions ~ $\alpha + \beta_1 * \sin(\frac{2\pi * Time}{12}) + \beta_2 * \cos(\frac{2\pi * Time}{12}) + \epsilon$
M0b	Pneumonia admissions ~ $\alpha + \beta_1 * \sin(\frac{2\pi * Time}{12}) + \beta_2 * \cos(\frac{2\pi * Time}{12}) + \epsilon$
M1	Log(Empyema admissions+1) ~ $\alpha + \beta_1$ (Pneumonia admissions + 1) + ϵ
M2	Log(Empyema admissions+1) ~ $\alpha + \beta_1$ *Log(Pneumonia admissions + 1) + β_2*Time + ϵ
M3	Log(Empyema admissions+1) ~ $\alpha + \beta_1$ *Log(Pneumonia admissions + 1) + β_2*Time +β_3*Population Size +ϵ
M4	Log(Empyema admissions+1) ~ $\alpha + \beta_1$ *Log(Pneumonia admissions + 1) + β_2*Time + $\beta_3 * \sin(\frac{2\pi * Time}{12}) + \beta_4 * \cos(\frac{2\pi * Time}{12}) + \epsilon$

Table 2.4 Modelling framework for assessing trends in national empyema and pneumonia admissions

Definition of periodicity and cyclicity within national empyema admissions

The results of models M0-M4 demonstrated that there were significant seasonal trends in national empyema admissions but that these trends did not appear to be an exact 12 month cycle. Wavelet periodicity analysis was therefore used to define the periodicity within the empyema admission time series with precision and these estimates were then used to specify seasonal harmonics of the same period to optimise the regression model M4. Although seasonality has been reported previously in paediatric empyema (Hardie *et al.*, 1996) the periodicity and cyclicity within empyema admissions has not been previously precisely measured Establishment of the presence of significant periodicity and cyclicity within empyema admissions would provide hypothetical evidence of possible epidemiological mechanisms important in explaining any increase in incidence of empyema e.g. presence of cycles of similar duration to those of particular respiratory pathogens.

Optimal periodicity was defined by analysis of the average wavelet power spectrum against period and the periods that had the greatest average power were recorded. As these estimates represented the average power, the ranges of periods that corresponded with the highest integer value of the average power were used. The highest and lowest integer values of this range and gradations of 0.25 in between were then used as the period *n* for the seasonal harmonics in Model M5:

$$\text{Log(Empyema admissions+1)} \sim \alpha + \beta x\ \text{*Log(Pneumonia admissions} + 1) + \beta x\ \text{Time} + \beta x\text{*sin}(\frac{2\pi * Time}{n}) + \beta x\text{*cos}(\frac{2\pi * Time}{n}) + \epsilon$$

Model optimisation was judged on three characteristics: the model AIC, R^2 and the significance of the included explanatory terms with the optimal model having the lowest AIC, highest R^2 and all terms significant.

A dynamic model was then specified to test whether allowing for time variation in the seasonal harmonics (i.e. allowing for non-stationarity of the seasonal effect) significantly increased the ability of the model to explain the observed trends. The model structure was:

$$\mu_t = \alpha + \beta_t + \gamma P_t + B_t \cos(\omega t) + C_t \sin(\omega t) + \epsilon$$

Where μ_t was the log of empyema admissions in month t, α, β and γ the regression co-efficients, B and C the time-varying regression co-efficients, P the log of pneumonia admissions in month t, ω referred to the optimal seasonal harmonic $\frac{2\pi * Time}{n}$ and ϵ the error component.

Regional trends in pneumonia and empyema admissions

Regional trends in empyema and pneumonia were investigated at the level of the SHA. Population adjusted admissions were calculated using ONS population estimates for each SHA. These were used instead of counts of cases to account for the difference in population size between SHAs. The correlation between population adjusted admissions for pneumonia and empyema in each SHA was assessed by calculation of Spearman's correlation co-efficient. A progressive linear and linear mixed effect modelling approach was then used to test the hypotheses:

a) that population adjusted admissions for paediatric bacterial pneumonia and paediatric empyema varied significantly between SHAs between April 1997 and April 2006 (model S0a/b)
b) that population adjusted admissions for paediatric bacterial pneumonia and paediatric empyema increased significantly in each SHA between April 1997 and April 2006 (model S1a/b)
c) that there was a significant relationship between the monthly paediatric bacterial pneumonia and paediatric empyema population adjusted admissions in each SHA between April 1997 and April 2006 (model S2)
d) that there was significant variation in both the magnitude of the change in pneumonia and empyema population adjusted admissions and the rate of change between SHAs (model S3a/b)

The modelling framework used is listed in **Table 2.5**. The initial models S0a/b indicated that there were significant differences in population adjusted admissions for both empyema and pneumonia between SHAs. Therefore in order to assess trends across both conditions at this regional level, methods of handling the structure introduced by the *a priori* differences between SHAs were required. Mixed effects models flexibly represent the covariance structure of data grouped in this manner and allow quantification of the relative effect attributable to unmeasured variation between SHAs (Pinhero and Bates, 2000).

The hypothesis that populations adjusted admissions increased across all SHAs for both pneumonia and empyema was tested by specifying a model including the SHA as the sole random effect (model S1a/b). As previous analysis had demonstrated the presence of significant seasonality in both conditions, seasonal harmonics were included in the model, to test that the longitudinal trend was independent of this seasonality. The relationship of empyema to pneumonia was then tested (model S2). The hypothesis that there was significant variation in both the magnitude of the change in admission rate and the rate of change between SHAs was tested using models that contained random effects for variation in occurrences

of disease between different SHAs (intercept) and models that contained random effects for both variation in occurrences of disease (intercept) and variation in the rate of change of occurrences of the disease between SHAs (gradient) were compared by use of ANOVA (models S1a/b and models S3a/b). Model residuals were examined to ensure that assumptions of linearity and independence were met and are shown in Appendix E.

Name	Model structure	Model Type
S0a	Log(Pop. adjusted pneumonia admissions+1) ~ $\alpha + \beta_1 * SHA + \epsilon$	Linear
S0b	Log(Pop. adjusted empyema admissions+1) ~ $\alpha + \beta_1 * SHA + \epsilon$	Linear
S1a	Log(Pop. adjusted pneumonia admissions+1) ~ $\alpha + \beta_1 * Time + \beta_2 * \sin(\frac{2\pi * Time}{11.75}) + \beta_3 * \cos(\frac{2\pi * Time}{11.75}) + \epsilon^{1\|SHA}$	Mixed effects
S1b	Log(Pop. adjusted empyema admissions+1) ~ $\alpha + \beta_1 * Time + \beta_2 * \sin(\frac{2\pi * Time}{11.75}) + \beta_3 * \cos(\frac{2\pi * Time}{11.75}) + \epsilon^{1\|SHA}$	Mixed effects
S2	Log(Pop. adjusted empyema admissions+1) ~ $\alpha + \beta_1 * Time + \beta_2 * \sin(\frac{2\pi * Time}{11.75}) + \beta_3 * \cos(\frac{2\pi * Time}{11.75}) + \beta_4 * Log(Pop.\ adjusted\ pneumonia\ admissions+1) + \epsilon^{1\|SHA}$	Mixed effects
S3a	Log(Pop. adjusted pneumonia admissions+1) ~ $\alpha + \beta_1 * Time + \beta_2 * \sin(\frac{2\pi * Time}{11.75}) + \beta_3 * \cos(\frac{2\pi * Time}{11.75}) + \epsilon^{Time\|SHA}$	Mixed effects
S3b	Log(Pop. adjusted empyema admissions+1) ~ $\alpha + \beta_1 * Time + \beta_2 * \sin(\frac{2\pi * Time}{11.75}) + \beta_3 * \cos(\frac{2\pi * Time}{11.75}) + \epsilon^{Time\|SHA}$	Mixed effects

Table 2.5 Modelling framework for assessing trends in regional empyema and pneumonia admissions.

Relationship between empyema and pneumonia in England

The ratio of empyema to pneumonia admissions in England were calculated for each month between April 1997 and April 2006. Trends in the ratio of empyema to pneumonia admissions were then investigated using GLS regression models and a progressive modelling framework to test a series of linked hypotheses (**Table 2.6**). These were:

a) that the monthly ratio of empyema to pneumonia admissions in England between April 1997 and April 2006 showed seasonal variation (Model R0)
b) that the monthly ratio of empyema to pneumonia admissions in England significantly increased between April 1997 and April 2006 (Model R1)
c) that the monthly ratio of empyema to pneumonia admissions in England significantly increased between April 1997 and April 2006 independent of seasonal variation (Model R2)

Name	Model structure
R0	Ratio of empyema to pneumonia admissions ~ $\alpha + \beta_1 * \sin(\frac{2\pi * Time}{12}) + \beta_2 * \cos(\frac{2\pi * Time}{12}) + \epsilon$
R1	Ratio of empyema to pneumonia admissions ~ $\alpha + \beta_1$ Time + ϵ
R2	Ratio of empyema to pneumonia admissions ~ $\alpha + \beta_1$ Time + $\beta_2 * \sin(\frac{2\pi * Time}{12}) + \beta_3 * \cos(\frac{2\pi * Time}{12}) + \epsilon$

Table 2.6 Modelling framework for assessing trends in monthly ratio of empyema to pneumonia admissions.

Model diagnostic plots are shown in Appendix F.

Definition of periodicity and cyclicity within ratio of empyema to pneumonia admissions

Similar to the analysis of empyema admissions, the results of models R0-2 demonstrated that there were significant seasonal trends in national empyema admissions but that these trends did not appear to fit an exact 12 month cycle. Wavelet periodicity analysis was therefore used again to define with precision the periodicity within the ratio time series and these estimates were then used to specify seasonal harmonics of the same period to optimise the regression model R2. Optimal periodicity and model optimisation were defined as previously. The same approach as that used for the national empyema admissions data was used to investigate periodicity and cyclicity within the ratio of empyema admissions to pneumonia admissions over time. Wavelet periodicity analysis was used to define the presence of cyclicity and refine the GLS regression models with seasonal harmonics of varying lengths. Model optimisation was judged using the same three characteristics – model AIC, R^2 and significance of terms.

The highest and lowest integer values of the range of periods with the highest average power and gradations of 0.25 in between were then used as the period *n* for the seasonal harmonics in model R3:

$$\text{Ratio} \sim \alpha + \beta_1 * \text{Time} + \beta_2 * \sin(\frac{2\pi * Time}{n}) + \beta_3 * \cos(\frac{2\pi * Time}{n}) + \epsilon$$

As a second peak of average power was identified within the ratio time series the process was then repeated using model R4 constructed as:

$$\text{Ratio} \sim \alpha + \beta_1 * \text{Time} + \beta_2 * \sin(\frac{2\pi * Time}{n}) + \beta_3 * \cos(\frac{2\pi * Time}{n}) + \beta_4 * \sin(\frac{2\pi * Time}{q}) + \beta_5 * \cos(\frac{2\pi * Time}{q}) + \epsilon$$

where the term q was the range of highest average power in the second peak divided into 0.25 gradations.

As with national empyema admissions a dynamic model (R5) was used to test whether allowing time-variation of the seasonal harmonics improved model function in comparison to model R3/4. The model structure was:

$$\mu_t = \alpha + \beta_t + B_t \cos(\omega t) + C_t \sin(\omega t) + \epsilon$$

Where μ_t is the ratio of admissions in month t, α, β and are the regression co-efficients, B and C are time-varying regression co-efficients, ω refers to the seasonal harmonic $\frac{2\pi * Time}{n}$ and ϵ is the error component.

The relationship between national empyema admissions, the ratio of empyema to pneumonia admissions and isolations of *S. pneumoniae, S. aureus, S. pyogenes* and *M. pneumoniae*

The relationship between admissions for paediatric empyema and isolations of *S. pneumoniae, S. aureus, S. pyogenes* and *M. pneumoniae* were investigated using GLS regression models and a progressive modelling framework to test a series of linked hypotheses. These were:

a) that monthly admissions for paediatric empyema between April 1997 and April 2006 in England were temporally associated with isolations of *S. pneumoniae, S. aureus, S. pyogenes* and *M. pneumoniae* from the same population over the same time frame (Model P0)
b) that the monthly ratio of empyema to pneumonia admissions in England between April 1997 and April 2006 were temporally associated with isolations of *S. pneumoniae, S. aureus, S. pyogenes* and *M. pneumoniae* from the same population over the same time frame (Model S0)

Each organism was assumed to be an independent predictor of the monthly counts of admissions and of the ratio of empyema to pneumonia admissions. As the outcome variables were therefore counts, log transformation was applied as previously. The basic models were therefore:

Name	Model structure
P0	Log(Empyema admissions+1) ~ $\alpha + \beta_1$* Log(Isolations of *S. pneumoniae*+1) + β_2* Log(Isolations of *S. pyogenes*+1) + β_3* Log(Isolations of *S. aureus*+1) + β_4* Log(Isolations of *M. pneumoniae*+1) + ϵ
S0	Ratio of empyema to pneumonia admissions ~ $\alpha + \beta_1$* Log(Isolations of *S. pneumoniae*+1) + β_2* Log(Isolations of *S. pyogenes*+1) + β_3* Log(Isolations of *S. aureus*+1) + β_4* Log(Isolations of *M. pneumoniae*+1) + ϵ

Table 2.7 Modelling framework for investigating relationship between national empyema admissions, the ratio of empyema to pneumonia and isolations of *S. pneumoniae, S. aureus, S. pyogenes* and *M. pneumoniae*.

As isolations of the four pathogens are dependent on infective transmission, there was significant autocorrelation present within the series. This was confirmed by examination of the residuals from models P0 and S0 without the inclusion of a residual autocorrelation structure (Appendix I **Figure 9.49** and **Figure 9.49**).

In order to account for this, residual autocorrelative structures were included in both models (P0 and S0). Zuur *et al* (2009) suggest using a pragmatic approach to this problem by incorporating progressively more complex structures until a balance between complexity and model performance is achieved. This was assessed by comparison of the model AIC and where applicable the value of Phi and Theta (measures of correlation between residuals) for each model. Models that significantly improved AIC (a reduction of greater than three) and that minimised Phi and Theta were preferred (Zuur *et al.*, 2009). The correlation structures included were compound symmetry, AR1 and ARMA with values of p and q of integers between 0 and 3. All combinations of p and q were examined as until the optimal combination was found.

Coherence between national empyema admissions and the ratio of empyema to pneumonia admissions and isolations of *S. pneumoniae, S. aureus, S. pyogenes* and *M. pneumoniae*

Wavelet coherence analysis allows assessment of the relationship between two time series dynamically and identifies coherence between series at different periods i.e. at what times and frequencies do the series become statistically related. Phase difference identifies the direction of any relationship. While multivariate time series regression using GLS has the ability to identify linear relationships between counts of admissions and isolations of the four pathogens over time, coherence analysis allows identification of relationships at different periods or cycles between the two i.e. as part of a dynamic process. Identification of periodic relationships between admissions and individual pathogens would provide significant additional information about the role of that pathogen in the epidemiology of empyema that has not been previously determined.

The wavelet coherence and phase difference between monthly isolations of *S. pneumoniae, S. pyogenes, S. aureus* and *M. pneumoniae* and monthly empyema admissions were calculated to further investigate the relationship between respiratory pathogens and national empyema admissions,. In addition, the wavelet coherence and phase difference between isolations of respiratory pathogens and the ratio of pneumonia to empyema admissions per month were calculated to investigate the relationship between isolations of individual pathogens and changes in the rate of progression of pneumonia to empyema. The same hypotheses were considered:

a) that monthly admissions for paediatric empyema between April 1997 and April 2006 in England were temporally associated with isolations of *S. pneumoniae, S. aureus, S. pyogenes* and *M. pneumoniae* from the same population over the same time frame
b) that the monthly ratio of empyema to pneumonia admissions in England between April 1997 and April 2006 were temporally associated with isolations of *S. pneumoniae, S. aureus, S. pyogenes* and *M. pneumoniae* from the same population over the same time frame

Each relationship was assessed separately by calculation of the cross-wavelet coherence and phase difference between the outcome (either monthly counts of number of admissions for empyema or the monthly ratio of empyema to pneumonia) and the predictors (counts of isolations of each of the pathogens separately). The “Morlet wavelet” was used as the mother wavelet for all wavelet transformations as the Morelet wavelet allows the calculation of phase difference (Zhang *et al.*, 2009).

2.3 Results

2.3.1 Local changes in the incidence of empyema

The incidence of paediatric empyema in NE England increased from a baseline of 13.5 (95% CI 5.83-26.6) cases per million children (0-14 yrs) in the epidemiological year 1995/6 to 51.43 (95% CI 33.59-75.35) in 2010/11. The annual incidence for each year is shown in **Figure 2.3**. The incidence of empyema peaked in 2007/8 at 66.57 cases per million children before declining gradually between 2008 and 2011. The same data per month are shown in **Figure 2.4**.

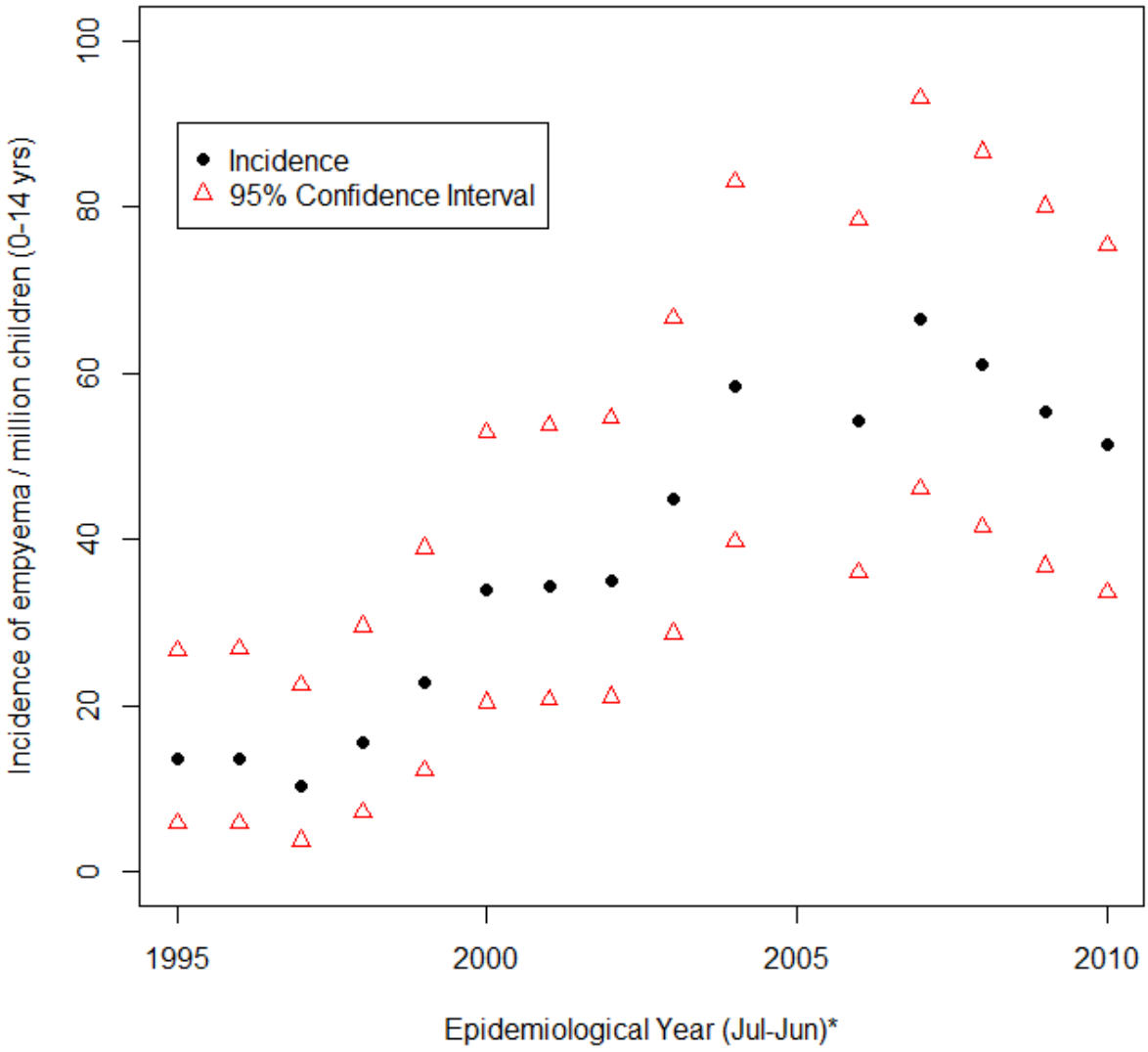


Figure 2.3 Annual incidence of paediatric empyema in the population served by the Freeman hospital, Newcastle-upon-Tyne (*data unavailable for full year of 2005 therefore not included).

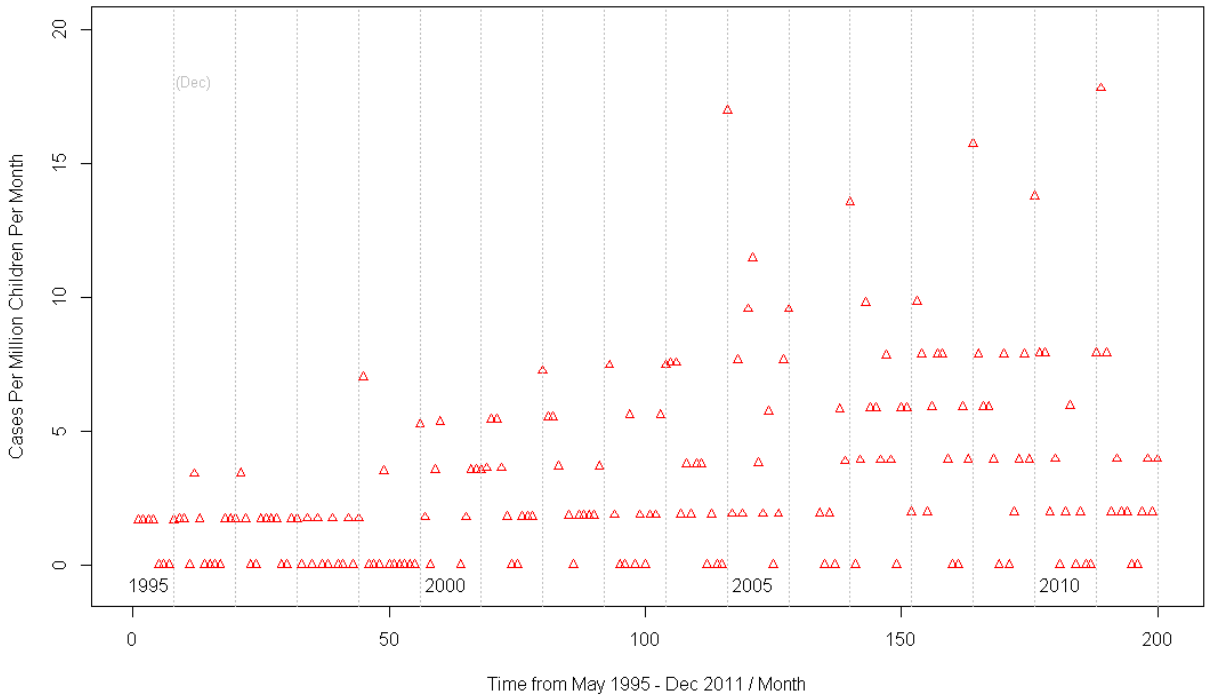


Figure 2.4 Monthly incidence of paediatric empyema in NE England demonstrating both the significant seasonal variation in cases of empyema with peaks consistently in the winter months and the increase in incidence over time. ***Red triangle indicates monthly count of cases per million children. Grey dashed line indicates month of December.***

T1 modelled the relationship between log transformed cases of paediatric empyema per month and time since 1995 and mean monthly air temperature. The regression co-efficients and model parameters are shown in **Table 2.8**. The increase in monthly cases from 1995 was statistically significant ($p<0.001$) and the back-transformed co-efficient of the increase was positive suggesting that cases increased on average by 0.00388 per month of the study period (equivalent to 0.0466 cases per year per million children). Mean monthly air temperature was also a significant predictor of the number of monthly cases, with an increase in cases associated with lower monthly temperatures (regression coefficient -0.0529, $p<0.001$). **Figure 2.5** shows the model fit which was moderate (R^2: 0.5403).

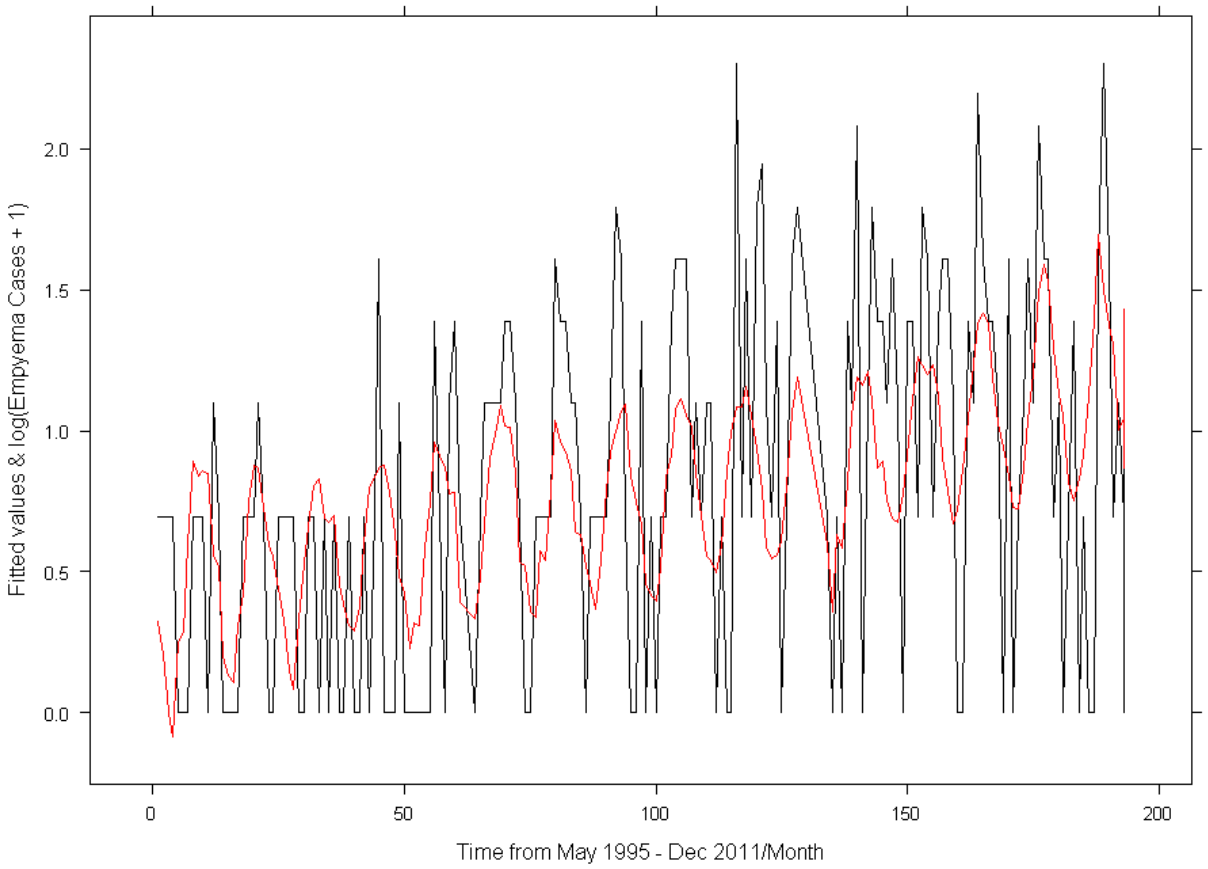


Figure 2.5 Log transformed monthly incidence of paediatric empyema in NE England between 1995 and 2011 *(black line)* and incidence predicted by model T1 *(red line)* (model R^2: 0.5403).

Variable	**Co-efficient**	**SE**	**t-value**	**p-value**
(Intercept)	1.0653	0.1229	8.6662	<0.001
Time from study start (Month)	0.0039	0.0006	5.9963	<0.001
Mean maximum monthly temperature (^{o}C)	-0.0529	0.0079	-6.6822	<0.001
Model structure: Log(Empyema Cases +1) ~ α +βxTime + βxTemperature + ϵ				
Residual SE: 0.523; Range of residuals: -2.672 to 2.333				
DF: 193 total; 190 residual				
AIC: 327.9389 ; R^2 : 0.5402741				

Table 2.8 Regression co-efficients and model parameters for model T1 assessing relationship between empyema cases in the North East of England and trends over time and temperature.

2.3.2 National and regional changes in the incidence of empyema

National data

There were a total 59,178 pneumonia admissions and 1,789 empyema admissions between April 1997 and April 2006. The relative contribution of different HES codes are shown in **Table 2.9** and the change year to year in contribution by codes is shown in **Figure 2.6** and **Figure 2.7**. Lobar pneumonia was the identifying code in over 85% of admissions, confirming that less than 15% of pneumonia admissions within HES data have a specific microbiological cause. The relative proportions of the pneumonia codes assigned remained consistent throughout the study period. In contrast, the proportion of empyema cases coded with the J86.0 pyothorax with fistula code halved from 16.4% in 1998 to 8.7% in 2005 (Chi-squared 4.695, p=0.0302).

Code	**n**	**%**
J13 - Pneumonia due to Streptococcus pneumoniae	1469	3.19
J14 - Pneumonia due to Haemophilus influenzae	311	0.68
J151-9* - other bacterial pneumonia	4095	8.89
J181 - Lobar pneumonia, unspecified	40187	87.25
Total	46062	100
	n	%
J860 - Pyothorax with fistula	144	9.14
J869 - Pyothorax without fistula	1432	90.86
Total	1576	100

Table 2.9 Overall contribution of codes to pneumonia and empyema admissions for HES dataset (* see Appendix A for specific codes from J151-9)

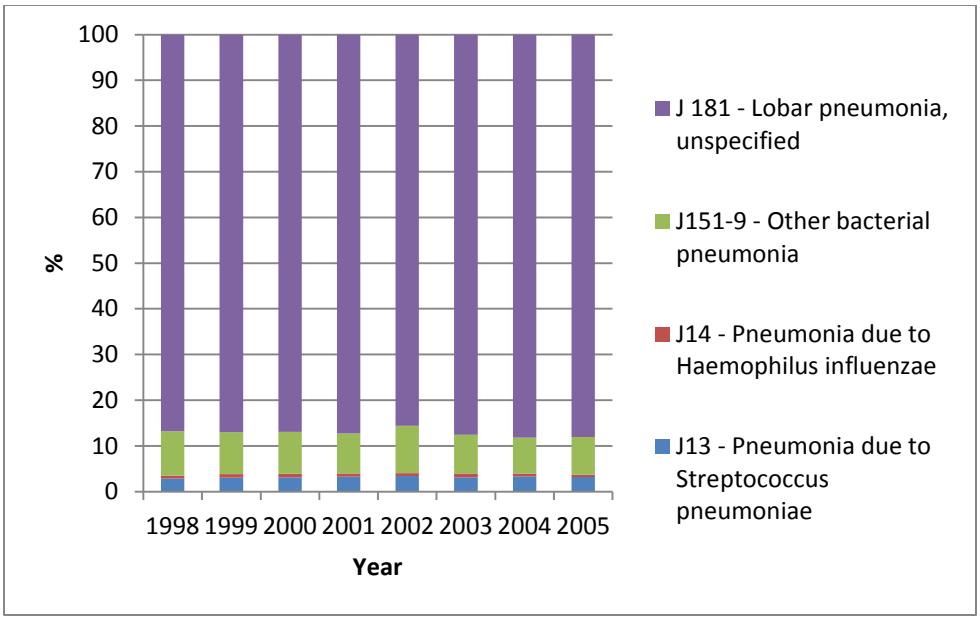


Figure 2.6 Proportion of pneumonia admissions by codes and by year (showing complete years only).

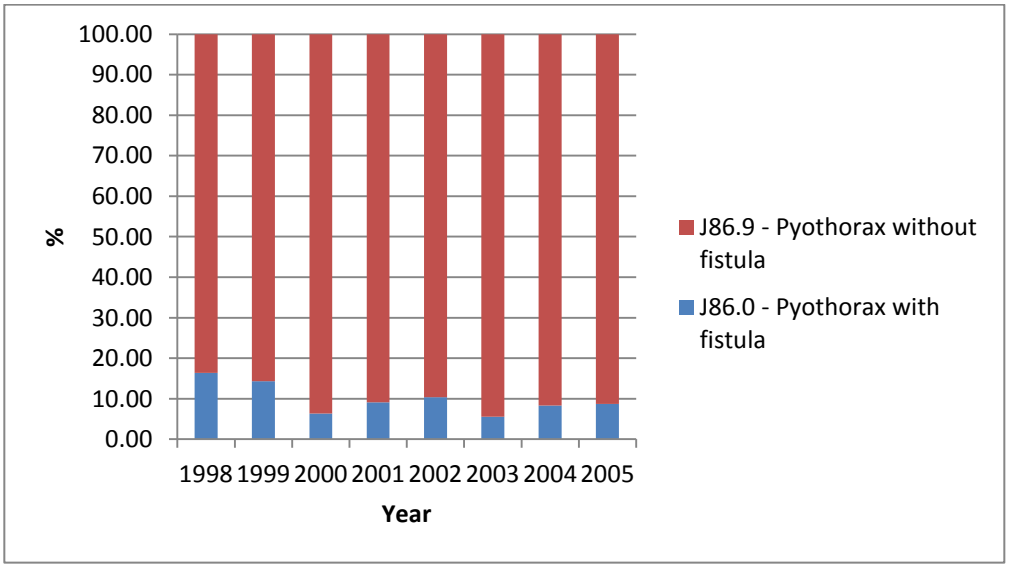


Figure 2.7 Proportion of empyema admissions coded as J860 or J869 by year.

Modelling national trends in empyema and pneumonia

The incidence of empyema as measured by admissions increased from 9.68 (95% CI 7.79-11.89) admissions per million children (0-14 years) in the epidemiological year 1997/8 to 37.47 (95% CI 33.6 - 41.66) in 2004/5. This was equivalent to an increase of 287% (Incidence rate ratio: 3.87, 95% CI: 3.063 - 4.933) or 253 extra cases per annum. Similarly pneumonia (bacterial) increased from 421.8 (95% CI 408.77 - 435.13) admissions per million children to 838.89 ((95% CI 820.17- 857.92). The median pneumonia admission rate over the whole time period was 67.7 per million individuals aged 0-14 years/month (range: 25.8 – 174.8 and the median empyema admission rate 2.5 per million 0-14 yrs/month (range: 0.42 – 13.1). The admission rates for both bacterial pneumonia and empyema increased from a median of 30.6 and 0.48 admissions per month per million individuals aged 0-14 years in 1997-1998 to 76.7 and 3.31 in 2005-2006 respectively. The monthly admission rates for both conditions from 1997 to 2006 are shown in **Figure 2.8**. Both conditions demonstrated seasonal variation with peaks in winter months and a progressive increase in admission rate was observed over the time period.

Models M0a and M0b (**Table 2.10**) confirmed that both empyema and pneumonia admissions displayed seasonal trends, as all four harmonic terms were significant ($p<0.001$). The seasonal trend explained 56% and 62% of the variation in empyema and pneumonia admissions respectively in these initial models. However, examination of the residuals (Appendix C **Figures 7.7** to **7.12**) revealed significant autocorrelation in both models and evidence of non-linearity, with increasing dispersion of the residuals as the fitted values increased. These findings suggest that the estimates of the significance of the co-efficients may be inaccurate.

Monthly admissions for paediatric empyema were significantly related to monthly admissions for paediatric pneumonia – co-efficient 1.163, $p<0.001$ (Model M1, **Table 2.10)** and explained approximately 85% of the variation in empyema admissions. On considering the model diagnostic plots, there was evidence of significant autocorrelation (Appendix C **Figure 9.15**) and clustering of the residuals suggesting that further explanatory or confounding variables may be missing from the model.

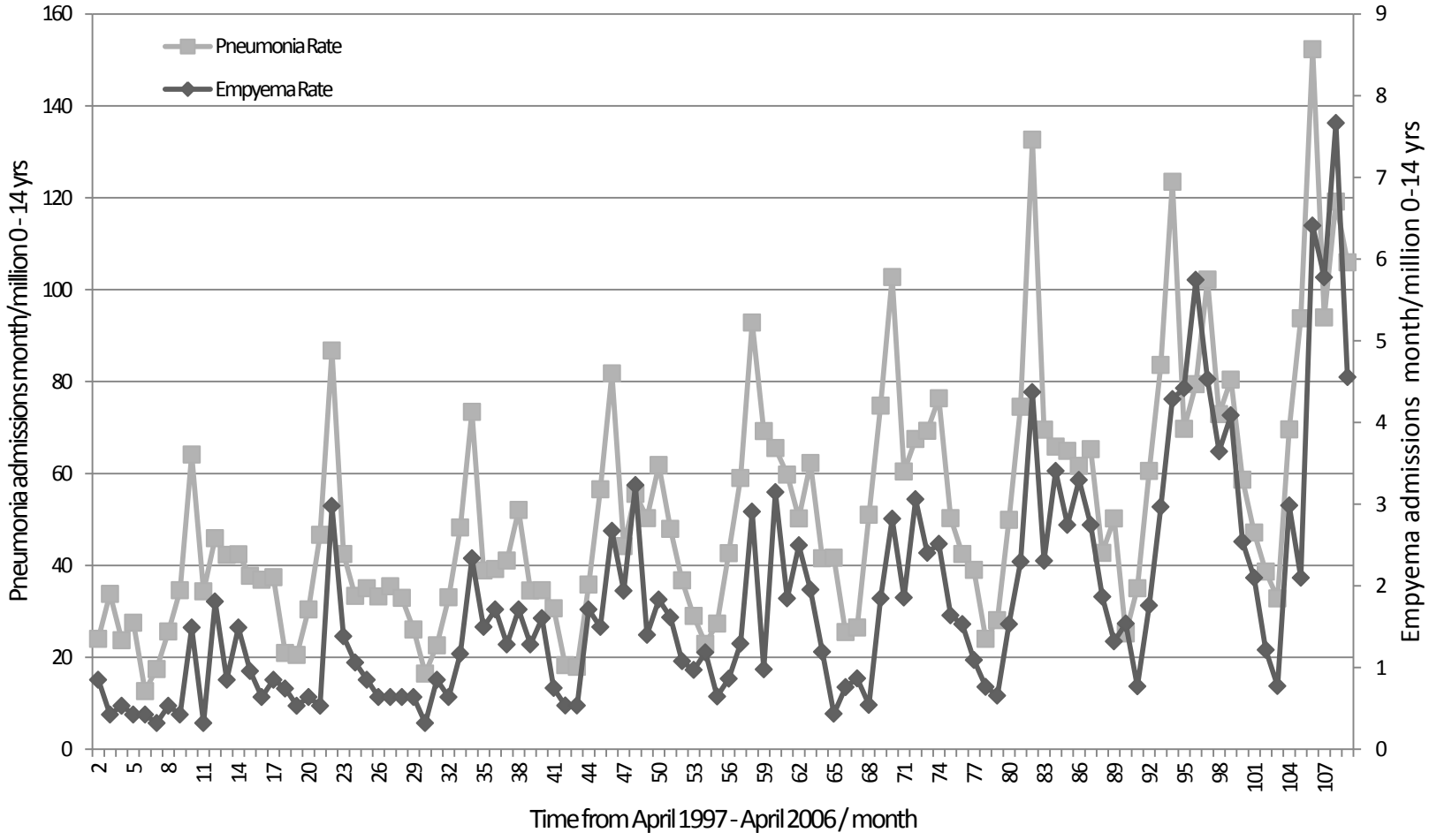


Figure 2.8 Rates of pneumonia and empyema admission over time from April 1997 to April 2006 in England.

Variable	Co-efficient	SE	t-value	p-value	Residual SE	Range of residuals	DF total	DF residual	AIC	R^2
Model M0a: Empyema admissions ~ $\alpha + \beta_1 * \sin(\frac{2\pi * Time}{12}) + \beta_2 * \cos(\frac{2\pi * Time}{12}) + \epsilon$										
(Intercept)	16.564	1.018	16.269	<0.001	10.581	-2.160 to 4.010	108	105	814.045	0.564
Time from study start*$\cos(\frac{2\pi * Time}{12})$	8.037	1.440	5.581	<0.001						
Time from study start*$\sin(\frac{2\pi * Time}{12})$	-6.092	1.449	-4.231	<0.001						
Model M0b: Pneumonia admissions ~ $\alpha + \beta_1 * \sin(\frac{2\pi * Time}{12}) + \beta_2 * \cos(\frac{2\pi * Time}{12}) + \epsilon$										
(Intercept)	476.0464	18.301	26.011	<0.001	190.195	-1.902 to 3.871	108	103	1420.727	0.622
Time from study start*$\cos(\frac{2\pi * Time}{12})$	128.6495	25.882	4.971	<0.001						
Time from study start*$\sin(\frac{2\pi * Time}{12})$	-166.757	25.882	-6.443	<0.001						
Model M1: Log(Empyema admissions+1) ~ $\alpha + \beta_1$*Log(Pneumonia admissions+1) + ϵ										
(Intercept)	-4.395	0.428	-10.274	<0.001	0.358	-2.710 to 2.184	108	106	97.019	0.848
Log(Pneumonia admissions + 1)	1.163	0.0705	16.498	<0.001						
Model M2: Log(Empyema admissions+1) ~ $\alpha + \beta_1$ *Log(Pneumonia admissions+1) + β_2*Time + ϵ										
(Intercept)	-3.486	0.444	-7.85549	<0.001	0.329	-3.100 to 1.965	108	105	91.431	0.875
Log(Pneumonia admissions+1)	0.962	0.079	12.189	<0.001						
Time from study start (Month)	0.0056	0.0012	4.551	<0.001						
Variable	**Co-efficient**	**SE**	**t-value**	**p-value**	**Residual**	**Range of**	**DF**	**DF**	**AIC**	**R^2**

					SE	**residuals**	**total**	**residual**		
Model M3: Log(Empyema admissions+1) ~ $\alpha + \beta_1$*Log(Pneumonia admissions +1)+ β_2*Time +β_3*Population Size +ϵ										
(Intercept)	-9.456	8.972	-1.054	0.294	0.300	-2.941 to 1.980	108	102	109.320	0.900
Log(Pneumonia admissions+1)	0.629	0.112	5.604	<0.001						
Time from study start (Month)	0.012	0.004	2.720	<0.001						
Population size (Individuals 0-14 years)	0.000001	0.000001	0.869	0.387						
Time from study start*cos($\frac{2\pi*Time}{12}$)	0.274	0.057	4.833	<0.001						
Time from study start*sin($\frac{2\pi*Time}{12}$)	-0.070	0.053	-1.325	0.188						
Model M4: Log(Empyema admissions+1) ~ $\alpha + \beta_1$ *Log(Pneumonia admissions + 1) + β_2 Time + β_3*sin($\frac{2\pi*Time}{12}$) + β_4*cos($\frac{2\pi*Time}{12}$) +ϵ										
(Intercept)	-1.678	0.628	-2.673	0.009	0.300	-3.028 to 2.090	108	103	82.18054	0.900
Log(Pneumonia admissions+1)	0.0082	0.0013	6.316	<0.001						
Time from study start (Month)	0.640	0.111	5.737	<0.001						
Time from study start*cos($\frac{2\pi*Time}{12}$)	0.260	0.054	4.792	<0.001						
Time from study start*sin($\frac{2\pi*Time}{12}$)	-0.071	0.052	-1.361	0.176						

Table 2.10 Summary of models of longitudinal trends, seasonality and relationship between pneumonia and empyema admissions in England.

Monthly admissions for paediatric empyema increased independent of monthly pneumonia admissions, population size and seasonal variation (Models M2, M3 and M4, **Table 2.10**) – co-efficient 0.640, $p<0.001$ (Model M4). Model M4 provided the best model fit with an R^2 of 0.900 and a reduction in AIC from 91.431 to 82.181 when compared to M2. Both the trend related to pneumonia admissions and the trend over time remained significant, but only the cosine seasonal harmonic was significant ($p<0.001$). This suggested two significant findings. Firstly that there is evidence of a seasonal trends in empyema above that related to pneumonia and secondly that such trends were not perfectly captured by an exact 12 month cycle. The standardised residuals when compared to fitted values no longer showed any evidence of non-linearity allowing confidence in the model validity (Appendix C **Figure 9.20**). Autocorrelation of marginal significance was present only at lags of 3 and 6 months suggesting that the estimates of the significance of the co-efficients were likely to be reliable (Appendix C **Figure 9.21**). There remained some overdispersion of the residuals (Appendix C **Figure 9.19**), but this was minimal. The model fit against the observed values is shown in **Figure 2.9**.

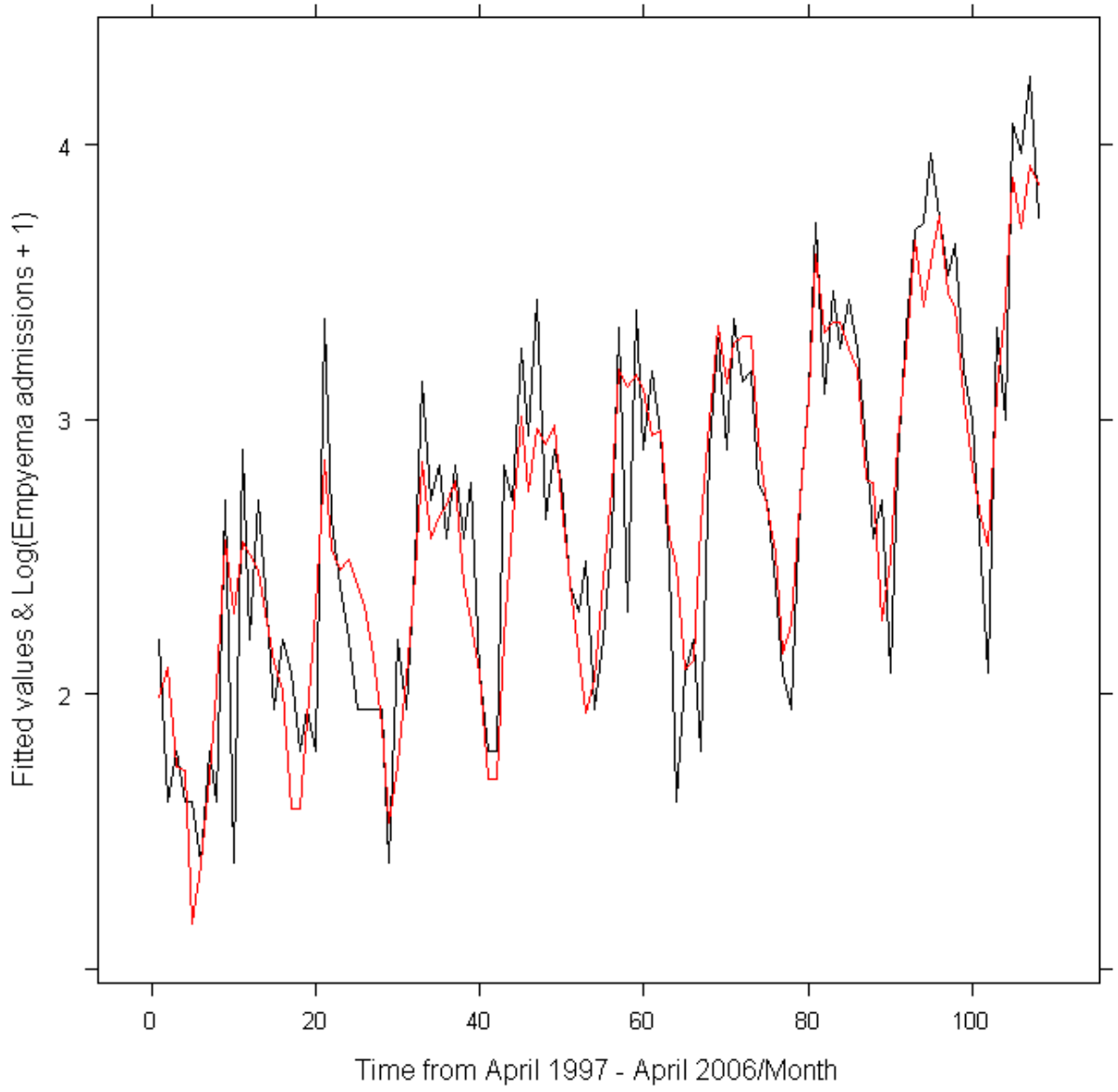


Figure 2.9 Log transformed monthly counts of paediatric empyema in England between 1997 and 2006 against those predicted by model M4 (model R^2:0.900). ***Black line indicates log(admissions+1)*** **and** ***red line indicates predicted values of model*****.**

Definition of periodicity and cyclicity within national empyema admissions

Figure 2.10 shows the wavelet periodicity analysis of national empyema cases. Analysis of the power spectrum against period indicated that periods of between 8 (average power 1.82) and 12 months (average power 1.94) had the highest average power (**Figure 2.11**). As a result of a series of models (M5) constructed as:

$$\text{Log(Empyema admissions}+1) \sim \alpha + \beta x \text{ *Log(Pneumonia admissions} + 1) + \beta x \text{ Time} + \beta x\text{*sin}(\frac{2\pi * Time}{n}) + \beta x\text{*cos}(\frac{2\pi * Time}{n}) + \epsilon$$

where n varied between 8 and 12 at intervals of 0.25 then tested. Three models had significantly lower AIC and greater R^2 than the other models tested – those at lags of 11.5, 11.75 and 12 months (**Figure 2.12**). Of these only in the model with a lag at 11.75 months were all parameters significant. Regression co-efficients and model parameters for this model (M5) are shown in **Table 2.11** and observed vs. fitted values in **Figure 2.13**. There was no evidence of non-linearity or autocorrelation in the residuals of model M5 indicating the model is valid (Appendix D **Figures 7.22 – 7.25)**. The results of the model optimisation and significance of the seasonal harmonics within model M5 indicated that the strongest signal within empyema admissions was of a season/cycle of 11.75 months duration. This reflects the aggregate within the time series and it is possible that other periods were relevant, for example within **Figure 2.11** there is an indication of an increase in power at 35 months. The figure of 11.75 months is close to 12 months and may reflect annual climatic variation with some inter-year variation in exact timing or may be a function of the unequal distribution of days within the months of the year.

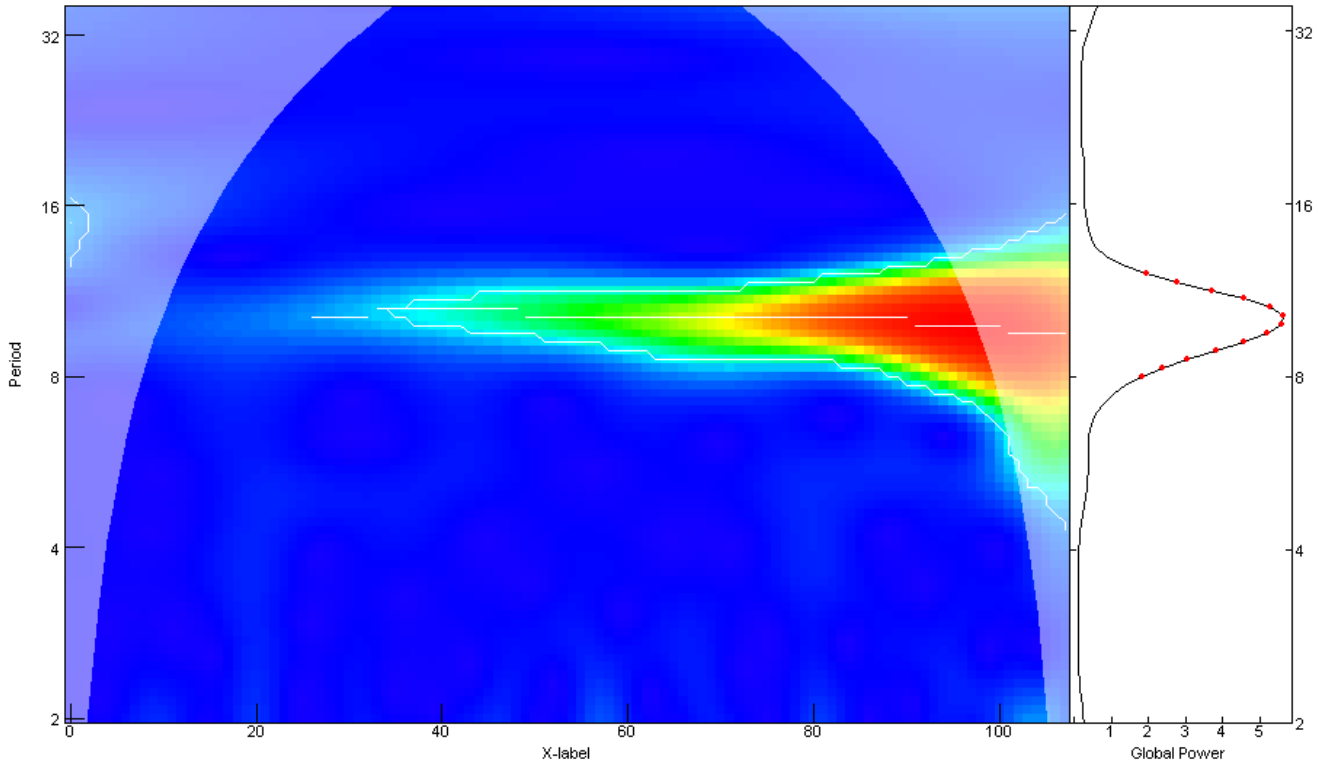


Figure 2.10 Wavelet periodicity analysis of national empyema cases. Wavelet periodicity analysis of national empyema cases. The x-axis represents the time period of April 1997 to April 2006 as months from 1-108. Periodicity is on the left y axis and is the period of the time series i.e. when the series is decomposed into a frequency signal where the peaks of activity occur as measured by the power of the wavelet effect – these are plotted as a heat map with red indicating high power. The same plot is then represented adjacent without the heat map. Dark blue cone indicates area where estimates are not at risk of edge effects. The effect is concentrated at a period of 11 months.

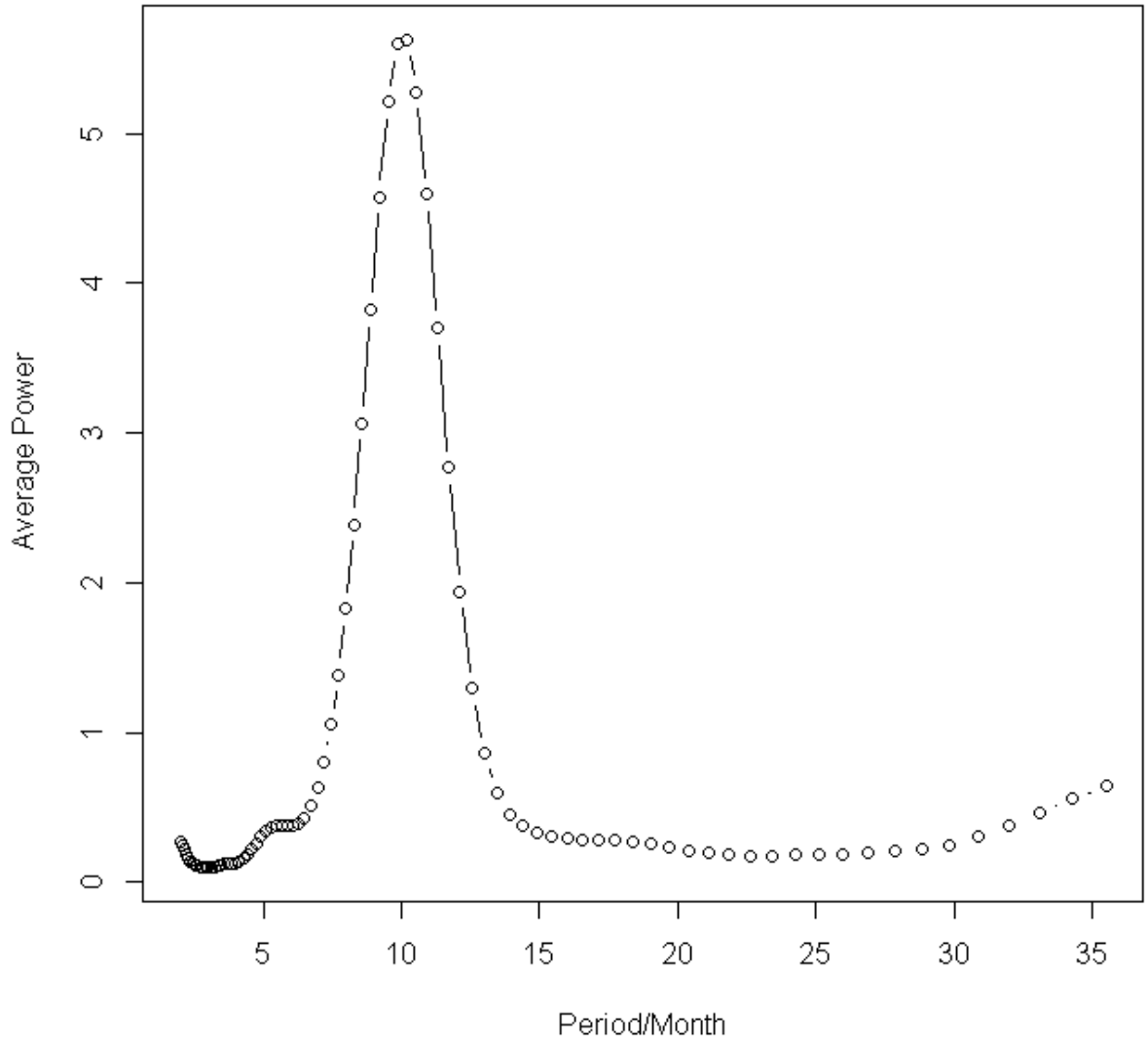


Figure 2.11 Average power vs. period for wavelet periodicity analysis of national empyema cases. Periods between 8 and 12 months had the highest average power.

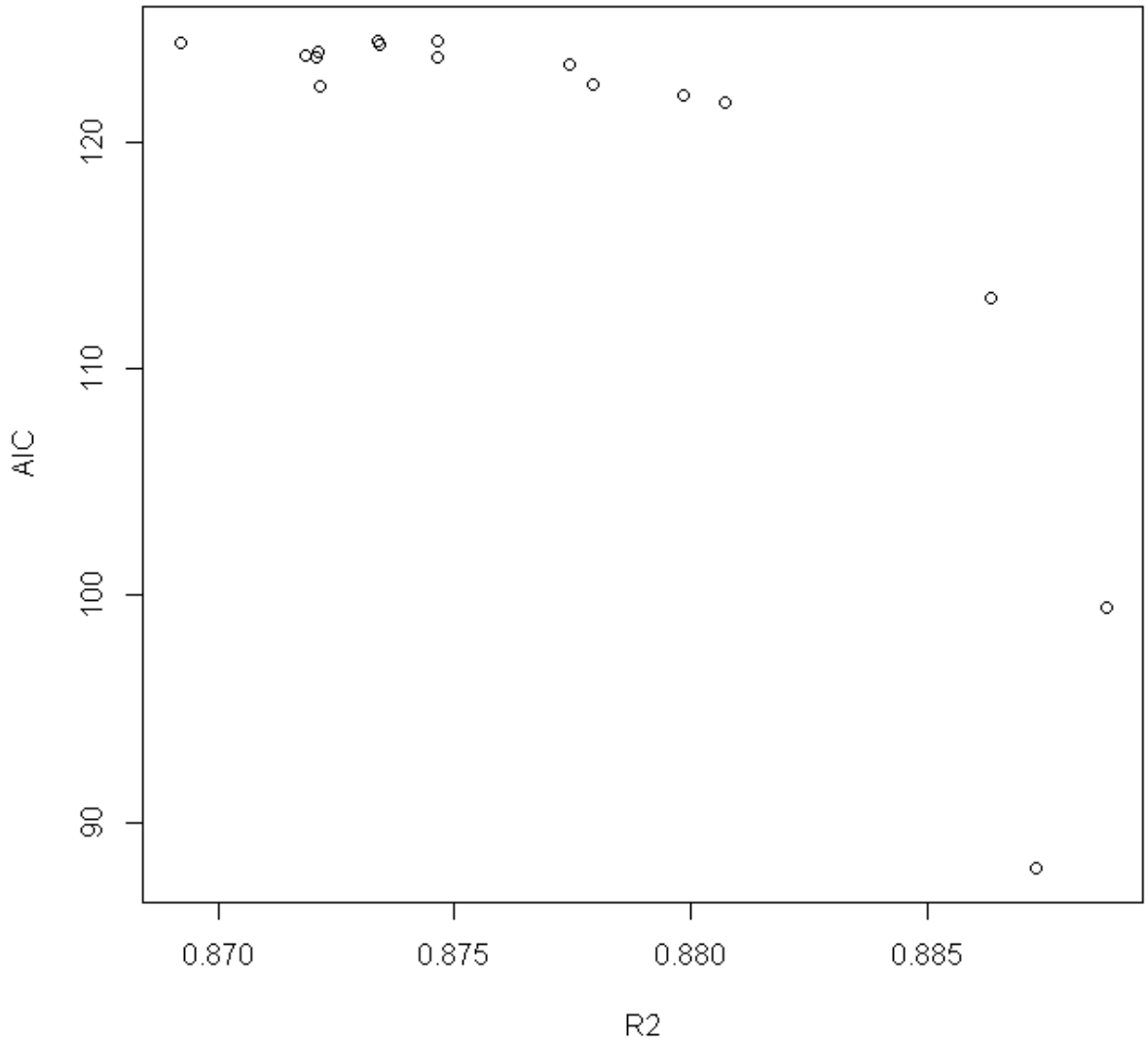


Figure 2.12 AIC vs. R^2 for sequential model with increasing seasonal components in order to estimate optimum model performance as balance between lowest AIC and greatest R2 post wavelet periodicity analysis.

Model M5				
Variable	**Co-efficient**	**SE**	**t-value**	**p-value**
(Intercept)	-2.326	0.602	-3.868	0.0002
Log(Pneumonia admissions+1)	0.0074	0.0013	5.658	<0.0001
Time from study start (Month)	0.754	0.107	7.038	<0.001
Time from study start*cos($\frac{2\pi*Time}{11.75}$)	0.174	0.060	2.864	0.0051
Time from study start*sin ($\frac{2\pi*Time}{11.75}$)	0.118	0.0420	2.816	0.0058
Residual SE: 0.306; Range of residuals: -2.929 to 2.125				
DF: 108 total; 103 residual				
AIC: 87.015; R^2: 0.894				

Table 2.11 Regression co-efficients and model parameters for model M5 assessing association between pneumonia admissions and season.

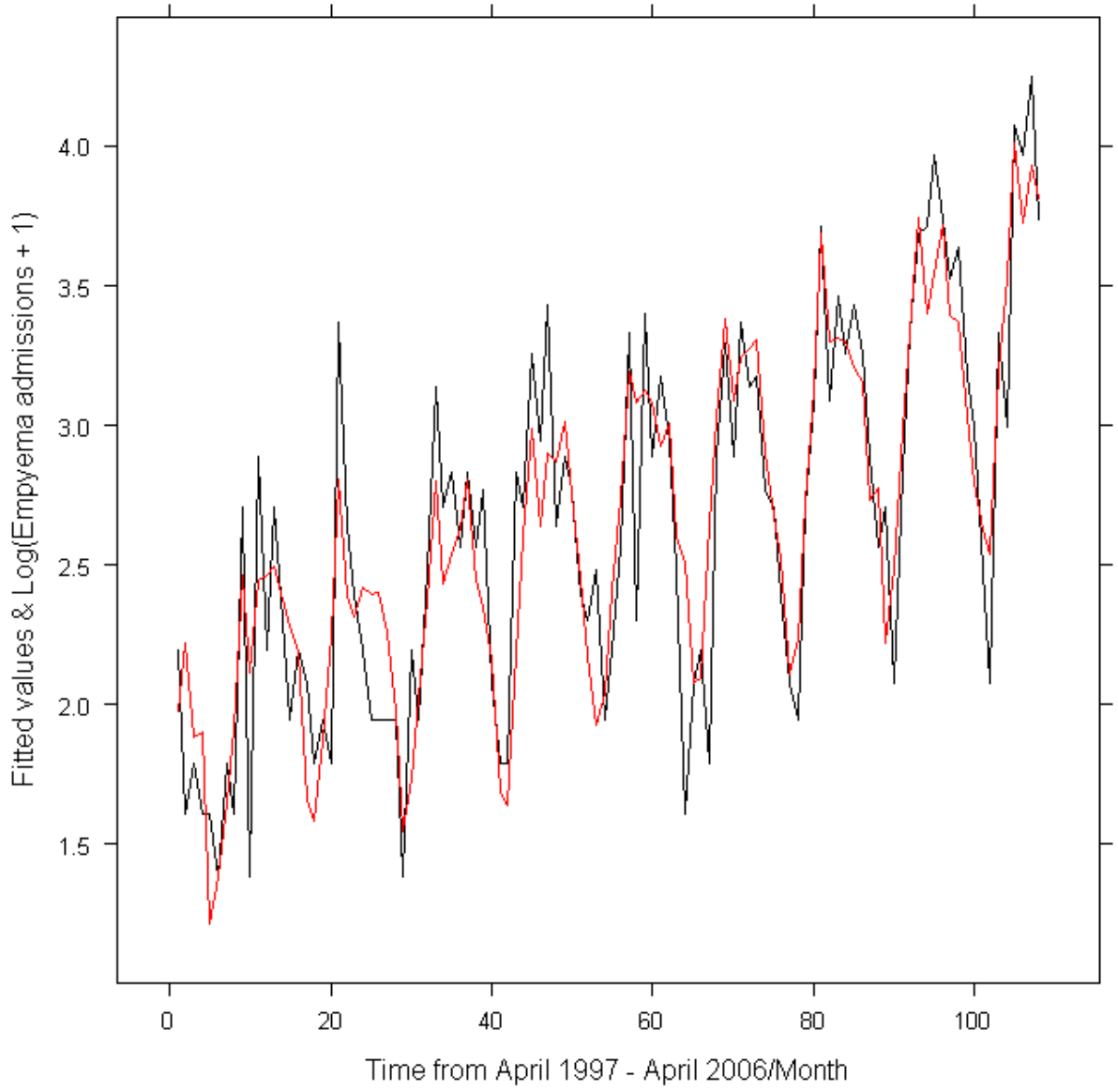


Figure 2.13 Log transformed monthly counts of paediatric empyema in England between 1997 and 2006 against those predicted by model M5 (R^2 0.894). ***Black line indicates log(observed admissions+1) and red line indicates predicted values of model***

The dynamic model (M6) did not improve on the ability of Model M5 to explain the observed trends and had a lower R^2 0.781 (M5 R^2 0.894) and a higher AIC 392 (M5 AIC 87) suggesting that non-stationarity of the seasonal effects was minimal.

Regional trends in pneumonia and empyema admissions

The maximum and minimum monthly admissions by SHA for bacterial pneumonia, empyema and the correlation co-efficient between the conditions are shown in **Table 2.12**. Both conditions were significantly correlated in all SHA's, but the degree of correlation varied between 0.414 – 0.753. There were significant differences in population adjusted admissions for pneumonia and empyema between SHAs (**Table 2.13**).

Both conditions increased across all SHAs - co-efficient for pneumonia: 0.0105 (95% CI 0.00671 to 0.0142), empyema: 0.0094 (95% CI 0.0068 to 0.00120) (Models S1a/b, **Table 2.14**). The rate of increase was greatest in the North East for both pneumonia and empyema admissions, whereas the lowest rate of increase in pneumonia admissions was in the East of England and the lowest rate of increase in empyema admissions was in the South Central SHA (Range for pneumonia – 0.0025 to 0.021, range for empyema – 0.0049 to 0.017). **Figure 2.14** compares the relative rate of increase (as derived from the random effects) across all SHAs for both conditions.

Population adjusted pneumonia admissions were a significant predictor of empyema admissions at a regional level – co-efficient 0.380, $p<0.001$ (Models S2, **Table 2.14**). There was significant variation in both the magnitude of the increase and the rate of increase in population adjusted pneumonia and empyema admissions between SHAs – pneumonia model likelihood ratio 110.93, $p<0.001$, empyema model likelihood ratio 23.94, $p<0.001$ (Models S3a/b, **Table 2.14** and **Table 2.15**). The distribution of residuals and residual correlation were normal for all models (Appendix E, **Figures 2.57 – 2.68**)

SHA	Min & max pneumonia admission rate (million 0-14 yr olds/month)	Min and max empyema admission rate (month/million 0-14 yr olds/month)	Spearman's correlation co-efficient
North West SHA	4.4 - 124.9	0 - 11.3	**0.732 (p<0.001)**
East Of England SHA	10.9 - 112.6	0 - 8.9	**0.482 (p<0.001)**
West Midlands SHA	15.3 - 159.4	0 - 14.2	**0.642 (p<0.001)**
Yorkshire SHA	12.3 - 224.9	0 - 8.7	**0.602 (p<0.001)**
East Midlands SHA	7.6 - 150.2	0 - 5.2	**0.414 (p<0.001)**
South East Coast SHA	6.7 - 148.6	0 - 10.6	**0.446 (p<0.001)**
South Central SHA	12.2 - 158.5	0 - 9.8	**0.447 (p<0.001)**
London SHA	16.8 - 137.4	0 - 7.4	**0.535 (p<0.001)**
North East SHA	0 - 249.5	0 - 24.6	**0.753 (p<0.001)**
South West SHA	14.8 - 203.4	0 - 8.2	**0.491 (p<0.001)**

Table 2.12 Minimum and maximum monthly pneumonia and empyema hospitalisations by SHA and correlation between both measures by SHA.

Pneumonia model					Empyema model			
SHA	**Co-efficient**	**SE**	**t-value**	**p-value**	**Co-efficient**	**SE**	**t-value**	**p-value**
EE	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
EM	0.053	0.082	0.644	0.519	**-0.268**	**0.092**	**-2.923**	**0.004**
L	0.097	0.082	1.177	0.240	0.036	0.092	0.396	0.692
NE	-0.080	0.082	-0.979	0.328	**0.268**	**0.092**	**2.917**	**0.004**
NW	**-0.290**	**0.082**	**-3.532**	**<0.001**	-0.101	0.092	-1.101	0.271
SC	0.070	0.082	0.847	0.397	-0.013	0.092	-0.146	0.884
SEC	-0.108	0.082	-1.319	0.187	**-0.216**	**0.092**	**-2.351**	**0.019**
SW	**0.310**	**0.082**	**3.780**	**<0.001**	**0.192**	**0.092**	**2.089**	**0.037**
Residual standard error: 0.603 on 1070 degrees of freedom					Residual standard error: 0.674 on 1070 degrees of freedom			
Multiple R-squared: 0.0810, Adjusted R-squared: 0.0733					Multiple R-squared: 0.0512, Adjusted R-squared: 0.043			
F-statistic: 10.48 on 9 and 1070 DF, p-value:<0.001					F-statistic: 6.409 on 9 and 1070 DF, p-value: <0.001			
Range of residuals: -3.674 to 1.850					Range of residuals: -1.077 to 2.167			

Table 2.13 Population adjusted admissions for paediatric bacterial pneumonia and paediatric empyema varied significantly between SHAs between April 1997 and April 2006.

Variable	Co-efficient	SE	DF	t-value	p-value	Random effects	St. dev of residuals random effects	Range of residuals	AIC	Log-likelihood
Model S1a: Log(Pop. adjusted pneumonia admissions+1) ~ $\alpha + \beta_1$*Time + β_2*sin($\frac{2\pi*Time}{11.75}$) + β_3*cos($\frac{2\pi*Time}{11.75}$) +$\epsilon^{1\|SHA}$										
(Intercept)	3.257	0.063	1067	51.388	<0.001	1|SHA	0.414	-6.727 to 3.071	1228.911	-608.455
Time from study start (Month)	0.010	0.00041	1067	24.478	<0.001					
Time from study start*cos($\frac{2\pi*Time}{11.75}$)	0.412	0.018	1067	23.072	<0.001					
Time from study start*sin ($\frac{2\pi*Time}{11.75}$)	-0.034	0.018	1067	-1.929	0.054					
Model S1b: Log(Pop. adjusted empyema admissions+1) ~ $\alpha + \beta_1$*Time + β_2*sin($\frac{2\pi*Time}{11.75}$) + β_3*cos($\frac{2\pi*Time}{11.75}$) +$\epsilon^{1\|SHA}$										
(Intercept)	0.310	0.060	1067	5.152	<0.001	1|SHA	0.570	-2.726 to 3.284	1907.765	-947.883
Time from study start (Month)	0.009	0.00056	1067	16.223	<0.001					
Time from study start*cos($\frac{2\pi*Time}{11.75}$)	0.294	0.025	1067	11.956	<0.001					
Time from study start*sin ($\frac{2\pi*Time}{11.75}$)	0.055	0.025	1067	2.244	0.025					

Model S2: Log(Pop. adjusted empyema admissions+1) ~ $\alpha + \beta_1$*Time + β_2*sin($\frac{2\pi * Time}{11.75}$) + β_3*cos($\frac{2\pi * Time}{11.75}$) + β_4*Log(Pop. adjusted pneumonia admissions+1) + $\epsilon^{1\|SHA}$										
(Intercept)	-0.927	0.142	1066	-6.524	0.000	1|SHA	0.548	-2.900 to 2.991	1827.886	-906.943
Time from study start (Month)	0.005	0.00067	1066	7.910	<0.001					
Log(Pop. adjusted pneumonia admissions +1)	0.380	0.040	1066	9.486	<0.001					
Time from study start*cos($\frac{2\pi * Time}{11.75}$)	0.137	0.029	1066	4.763	<0.001					
Time from study start*sin ($\frac{2\pi * Time}{11.75}$)	0.068	0.024	1066	2.884	0.004					
Model S3a: Log(Pop. adjusted empyema admissions+1) ~ $\alpha + \beta_1$*Time + β_2*sin($\frac{2\pi * Time}{11.75}$) + β_3*cos($\frac{2\pi * Time}{11.75}$) + β_4*Log(Pop. adjusted pneumonia admissions+1) + $\epsilon^{1\|SHA}$										
(Intercept)	3.257	0.141	1067	23.038	<0.001	Time | SHA	Time: 0.0059 Residual: 0.375	-5.952 to 3.726	1046.963	-515.481
Time from study start (Month)	0.010	0.0019	1067	5.192	<0.001					
Time from study start*cos($\frac{2\pi * Time}{11.75}$)	0.412	0.016	1067	25.502	<0.001					
Time from study start*sin ($\frac{2\pi * Time}{11.75}$)	-0.034	0.016	1067	-2.132	0.033					

Model S3b: Log(Pop. adjusted empyema admissions+1) ~ $\alpha + \beta_1$*Time + β_2*sin($\frac{2\pi*Time}{11.75}$) + β_3*cos($\frac{2\pi*Time}{11.75}$) +$\epsilon^{Time|SHA}$										
(Intercept)	0.310	0.067	1067	4.616	<0.001	Time | SHA	Time: 0.0038 Residual: 0.558	-2.990 to 3.214	1882.601	-933.300
Time from study start	0.009	0.0013	1067	6.840	<0.001					
Time from study start*cos($\frac{2\pi*Time}{11.75}$)	0.294	0.024	1067	12.200	<0.001					
Time from study start*sin ($\frac{2\pi*Time}{11.75}$)	0.055	0.024	1067	2.290	0.022					

Table 2.14 Summary of models examining regional trends in pneumonia and empyema admissions in England between 1997 and 2006.

Model	Random effects	ANOVA	DF	AIC	BIC	Log-likelihood	Likelihood Ratio	p-value
Pop. adjusted pneumonia admission models								
S1a	1 | SHA	S1a vs. S3a	6	1228.911	1258.797	-608.455	185.948	<0.001
S3a	Time | SHA		8	1046.963	1086.811	-515.481		
Pop. adjusted empyema admission models								
S1b	1 | SHA	S1b vs. S3b	6	1907.765	1937.651	-947.883	29.164	<0.001
S3b	Time | SHA		8	1882.601	1922.449	-933.300		

Table 2.15 Summary of ANOVA comparing models S1 and S3 to establish random effects structure.

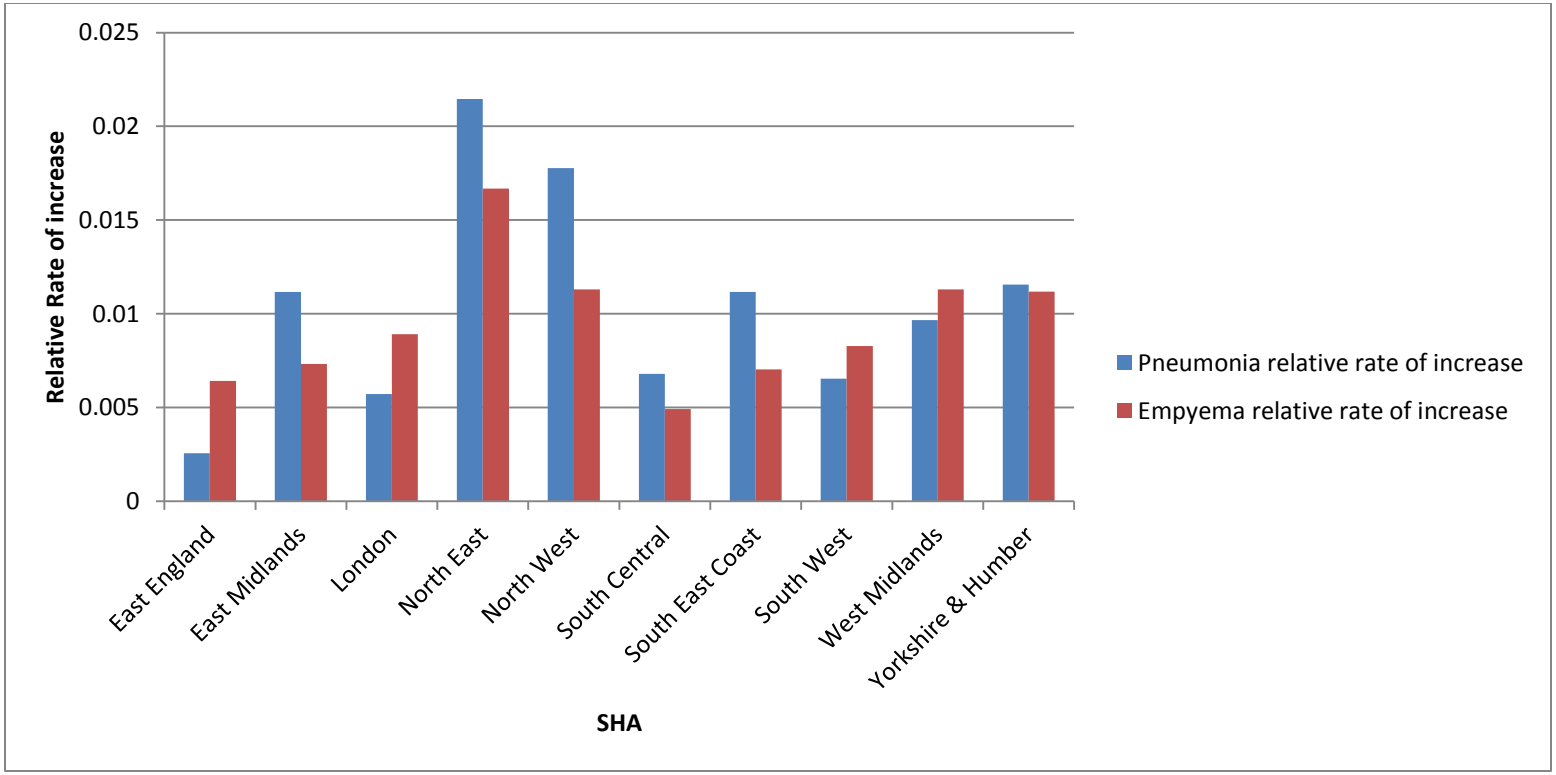


Figure 2.14 Relative rates of increase in population adjusted admissions for bacterial pneumonia and empyema by SHA in England from 1997- 2006.

2.3.3 Relationship between pneumonia and empyema admissions

The ratio of monthly empyema admissions to pneumonia admissions in England increased from a median of 2.04 % in 1997/1998 to 4.31 % in 2005/2006. The variation in this ratio over time is shown in **Figure 2.15** and demonstrates seasonality, as well as the increase over time. A summary of models R0-R2 is presented in **Table 2.16**.

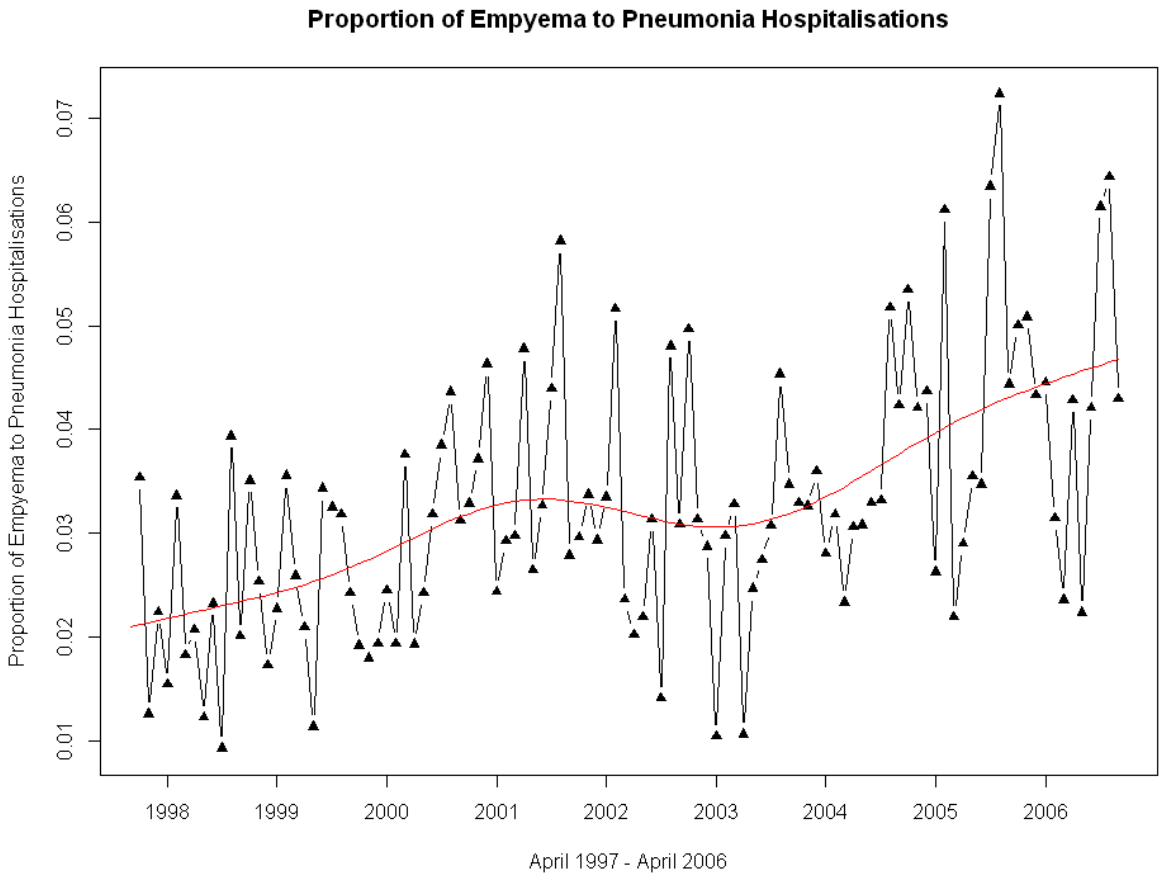


Figure 2.15 Proportion of empyema to pneumonia hospitalisations in England from September 1997 to April 2006 with general additive smoother to highlight trend.

Variable	Co-efficient	SE	t-value	p-value	Residual SE	Range of residuals	DF total	DF residual	AIC	R^2
Model R0: Ratio of empyema to pneumonia admissions ~ $\alpha + \beta_1 * \sin(\frac{2\pi * Time}{12}) + \beta_2 * \cos(\frac{2\pi * Time}{12}) + \epsilon$										
(Intercept)	0.0326	0.0011	28.479	<0.001	0.012	-2.284 to 2.895	108	105	-611.985	0.361
Time from study start*$\cos(\frac{2\pi * Time}{12})$	0.00637	0.0016	3.937	<0.001						
Time from study start*$\sin(\frac{2\pi * Time}{12})$	-0.00077	0.0016	-0.479	0.6331						
Model R1: Ratio of empyema to pneumonia admissions ~ $\alpha + \beta_1$ Time + ϵ										
(Intercept)	0.021	0.00212	10.110	<0.001	0.011	-2.244 to 2.874	108	105	-634.386	0.509
Time from study start (Month)	0.00021	0.00037	6.084	<0.001						
Model R2: Ratio of empyema to pneumonia admissions ~ $\alpha + \beta_1$ Time + $\beta_2 * \sin(\frac{2\pi * Time}{12}) + \beta_3 * \cos(\frac{2\pi * Time}{12}) + \epsilon$										
(Intercept)	0.022	0.0020	10.999	<0.001	0.010	-2.209 to 2.666	108	105	-626.467	0.988
Time from study start (Month)	0.00020	0.00003	6.457	<0.001						
Time from study start*$\cos(\frac{2\pi * Time}{12})$	0.0062	0.0014	4.489	<0.001						
Time from study start*$\sin(\frac{2\pi * Time}{12})$	-0.000021	0.0014	-0.015	0.9879						

Table 2.16 Summary of models of seasonality and longitudinal trends in the monthly ratio of empyema to pneumonia in England 1997 – 2006.

There was partial evidence of significant seasonal variation in the monthly ratio of empyema to pneumonia admissions as only one harmonic term was significant (Model R0, **Table 2.16**). Furthermore, there was significant autocorrelation in the residuals (Appendix F, **Figure 9.39**) and some fanning in the distribution of the residuals around the fitted values of 0.03 to 0.035 (Appendix F, **Figure 9.38**) indicating that there was non-constant variance in the errors and suggesting that significant explanatory variables such as rates of pneumococcal transmission, were not included in the model.

The increase in the monthly ratio of empyema to pneumonia admissions was both significant and independent of seasonal variation – co-efficient 0.00020, $p<0.001$ (Models R1 and R2, **Table 2.16**) equivalent to an increase of 0.021 % per month (95 % CI 0.014 to 0.027). There was no evidence of violation of independence or normality within the final, although some autocorrelation of the residuals remained (Appendix G, **Figure 9.48**). The model fit is shown in **Figure 2.16**.

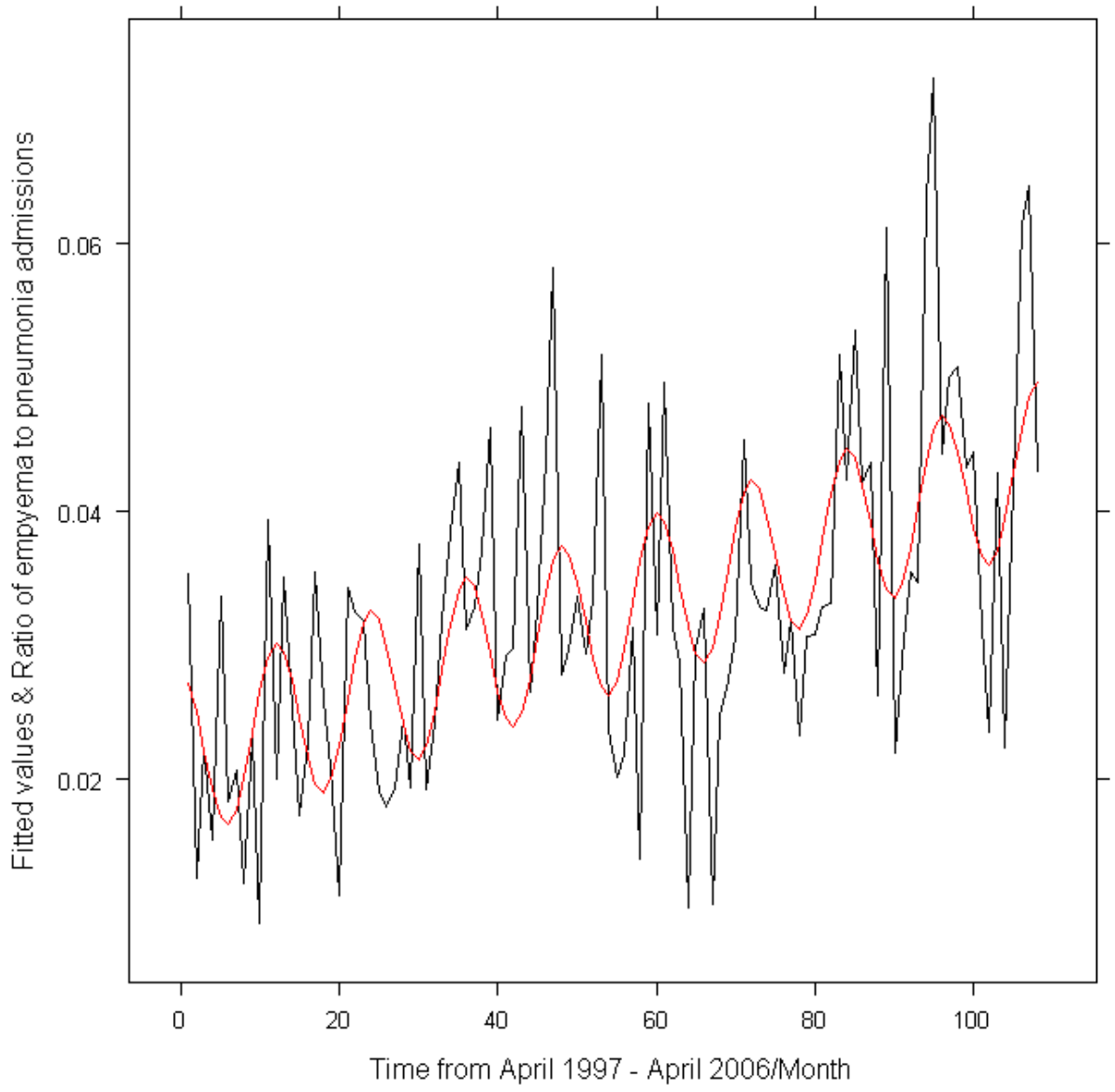


Figure 2.16 Ratio of empyema to pneumonia admissions and estimates from model R2 against time from 1997 to 2006 (R^2 0.616). ***Black line indicates ratio of admissions and red line indicates predicted values of model.***

Definition of periodicity and cyclicity within ratio of empyema to pneumonia admissions

The same approach as that used for the national empyema admissions data was used to define periodicity and cyclicity within the ratio of empyema admissions to pneumonia admissions over time. Wavelet periodicity analysis was used to define the presence of cyclicity and refine the GLS regression models with seasonal harmonics of varying lengths. Model optimisation was judged using the same three characteristics – model AIC, R^2 and significance of terms.

Figure 2.17 shows the wavelet periodicity analysis of the ratio over time. Analysis of the power spectrum against period indicated that periods of between 8 (average power 1.134) and 12 months (average power 1.129 had the highest average power (**Figure 2.18**). Model R3 therefore included seasonal terms where n varied between 8 and 12 at intervals of 0.25. Four models had significantly lower AIC and greater R^2 than the other models tested – those at lags of 11.25, 11.5, 11.75 and 12 months (**Figure 2.19**). Of these, only in the model with a lag at 11.75 months were all parameters significant.

A second peak between 33 and 35 months was also present. Model R4 therefore included terms where q varies between 33 and 35 at intervals of 0.25. There was no significant difference between any of these models in terms of AIC or R^2 (significance defined as difference in R^2 of 1 and AIC of 3 - **Figure 2.20**) and in no model were either of the later seasonal harmonics terms significant. The probable explanation for this, as supported by **Figure 2.18** and the fact that these models all provided a better model fit than R2 (R^2 increased by a minimum of 0.13), is that the period is greater than 36 months. However, as the data series is only 108 months long (equivalent of exactly three 36 month cycles) there were insufficient cycles within the data to establish the correct period.

The final acceptable model R3 is summarised in **Table 2.17** and the model diagnostics included in Appendix G, **Figures 7.46-7.48**. The residual distributions were acceptable, although there was some significant autocorrelation of the residuals at lags 6 and 12 months. The fitted vs. observed values are shown in **Figure 2.21**.

The dynamic model (R5) did not improve on Model R3 and had a lower R^2 0.120 (R3 R^2 0. 0.616) and a higher 371.266 (R3 AIC -624.298) suggesting that non-stationarity of the seasonal effects was minimal.

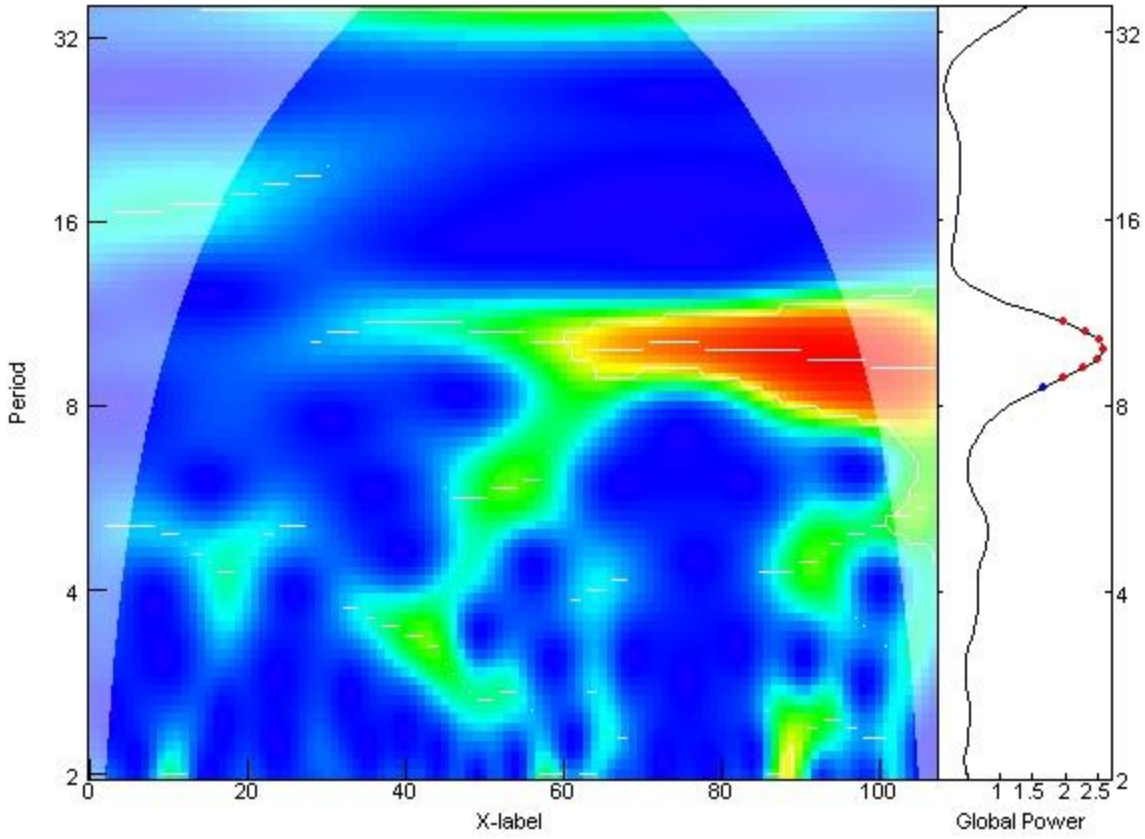


Figure 2.17 Wavelet periodicity analysis of ratio of empyema to pneumonia. The x-axis represents the time period of April 1997 to April 2006 as months from 1-108. Periodicity is on the left y axis and is the period of the time series i.e. when the series is decomposed into a frequency signal where the peaks of activity occur as measured by the power of the wavelet effect – these are plotted as a heat map with red indicating high power. The same plot is then represented adjacent without the heat map. Dark blue cone indicates area where estimates are not at risk of edge effects.

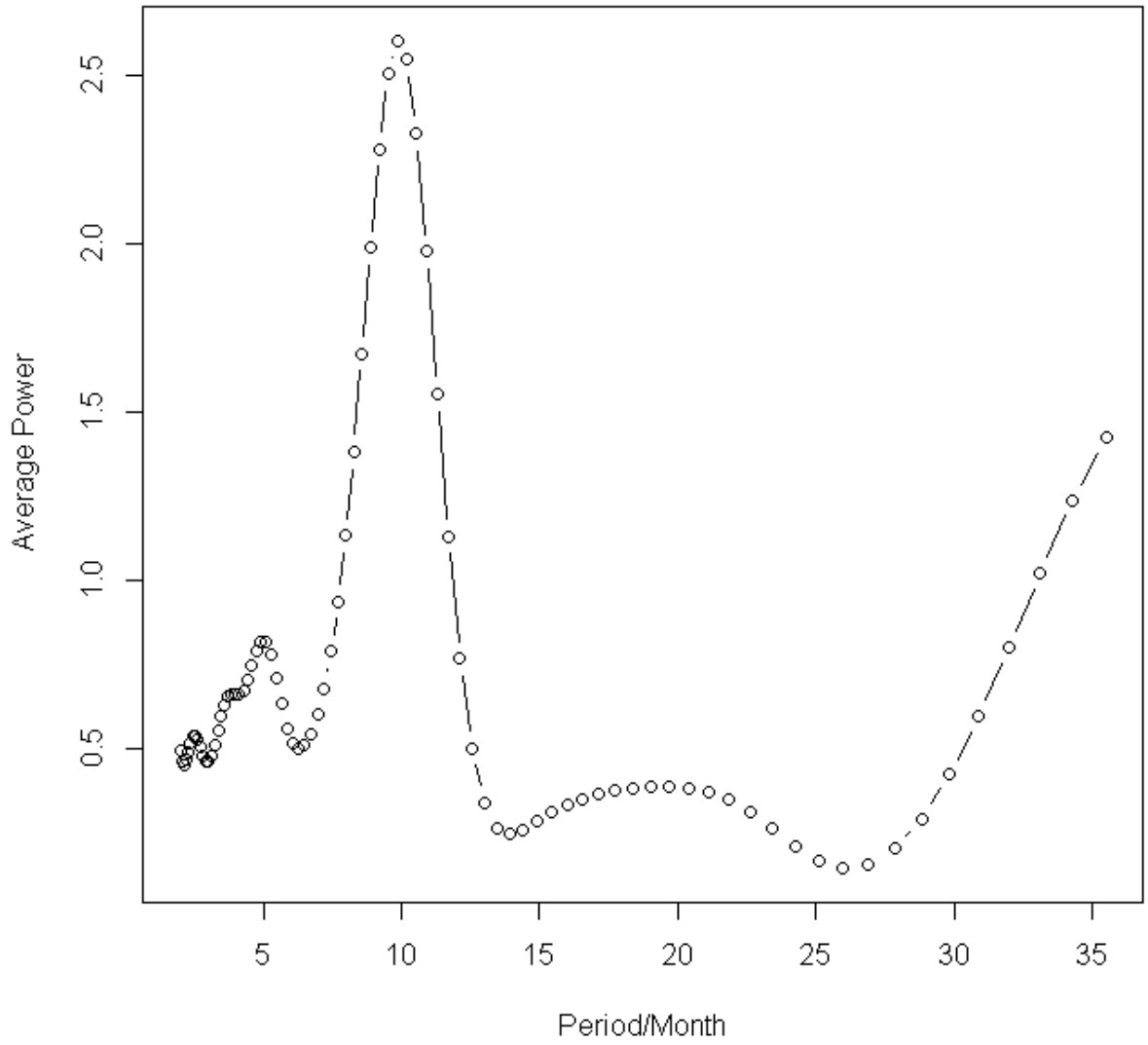


Figure 2.18 Average power vs. period for wavelet periodicity analysis of national empyema cases. Periods between 8 and 12 months and 33 and 35 months had the highest average power.

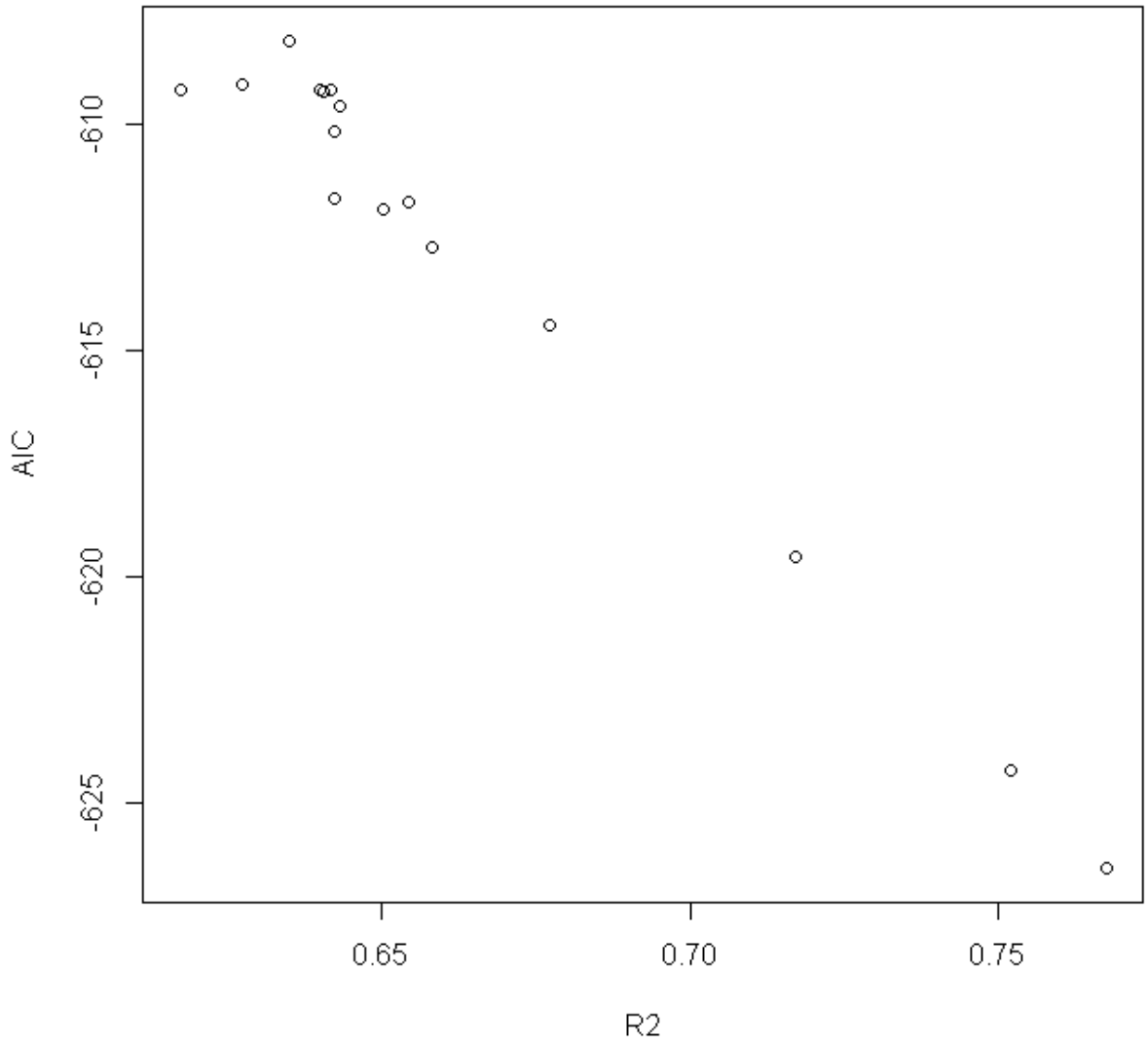


Figure 2.19 AIC vs. R^2 for sequential model with increasing seasonal components in order to estimate optimum model performance as balance between lowest AIC and greatest R2 post wavelet periodicity analysis of the ratio of empyema to pneumonia admissions with lags between 8 and 12 months.

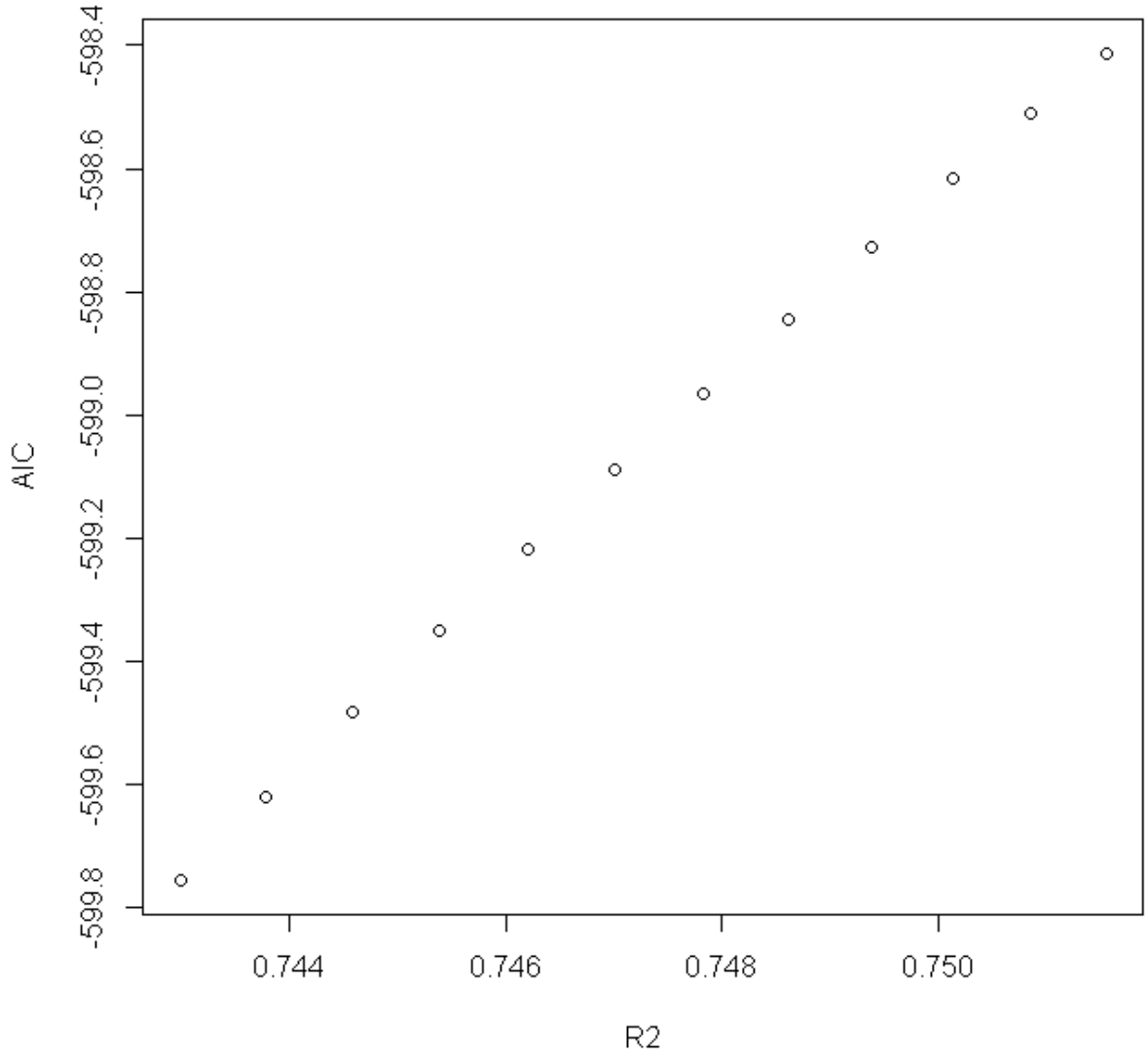


Figure 2.20 AIC vs. R^2 for sequential model with increasing seasonal components in order to estimate optimum model performance as balance between lowest AIC and greatest R2 post wavelet periodicity analysis of the ratio of empyema to pneumonia admissions with lags between 33 and 35 months.

Variable	**Co-efficient**	**SE**	**t-value**	**p-value**
(Intercept)	0.021	0.002	10.761	0.000
Time from study start (Month)	0.0002	0.00003	6.487	0.000
Time from study start*cos($\frac{2\pi*Time}{11.75}$)	0.0039	0.0014	2.826	0.006
Time from study start*sin ($\frac{2\pi*Time}{11.75}$)	0.0042	0.0014	3.046	0.003
Model structure: Ratio of empyema to pneumonia ~ $\alpha + \beta_1$* Time + β_2*sin($\frac{2\pi*Time}{11.75}$) + β_3*cos($\frac{2\pi*Time}{11.75}$) + ϵ				
Residual SE: 0.010; Range of residuals: -2.125 to 2.646				
DF: 108 total; 104 residual				
AIC: -624.298; R^2: 0.605				

Table 2.17 Summary of trends in the monthly ratio of empyema to pneumonia in England from 1997 to 2006 following optimisation of the estimates of the seasonal trends (model R3).

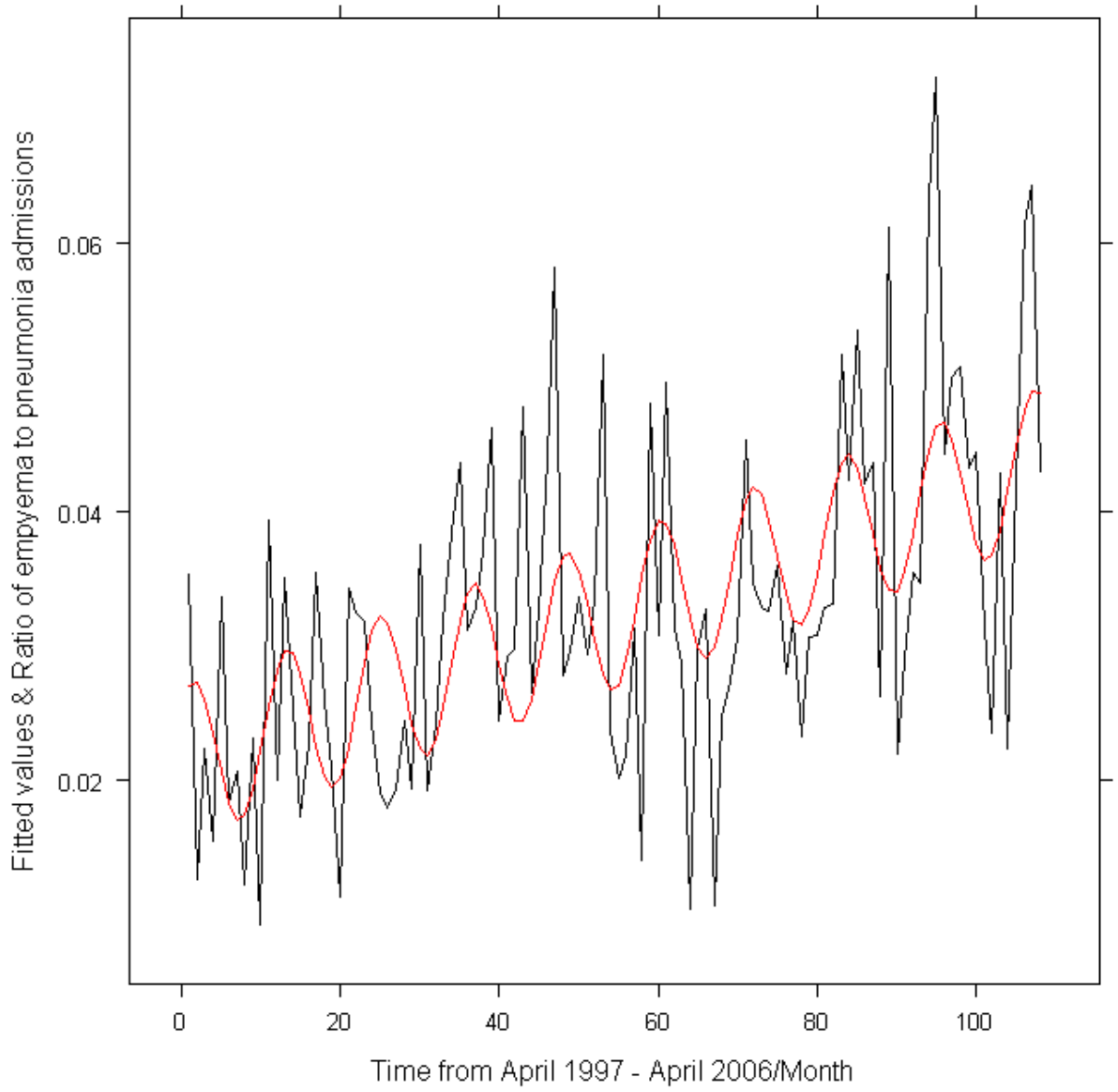


Figure 2.21 Ratio of empyema to pneumonia admissions in England between 1997 and 2006 and estimates from model R3 against time (R^2 0.605). ***Black line indicates ratio of admissions and red line indicates predicted values of model.***

2.3.4 Relationship between empyema admissions, the rate of progression of pneumonia to empyema and S. pneumoniae, S. pyogenes, S. aureus and M. pneumoniae

Trends in microbiological data

Monthly isolations of *S. pneumoniae, S. pyogenes, S. aureus* and *M. pneumoniae* in England from April 1997 to March 2006 from children aged between 0-14 years against time are shown in **Figure 2.22, Figure 2.23, Figure 2.24** and **Figure 2.25** respectively. The distribution *S. pneumoniae* displayed a clear seasonal variation. Seasonal variation was less obvious in the isolations of *S. pyogenes* and *S. aureus*. *M. pneumoniae* displayed cyclical peaks in activity of approximately 3-4 years frequency. *S. pneumoniae, S. pyogenes* and *S. aureus* all increased significantly from 1997 to 2006 (**Table 2.18**).

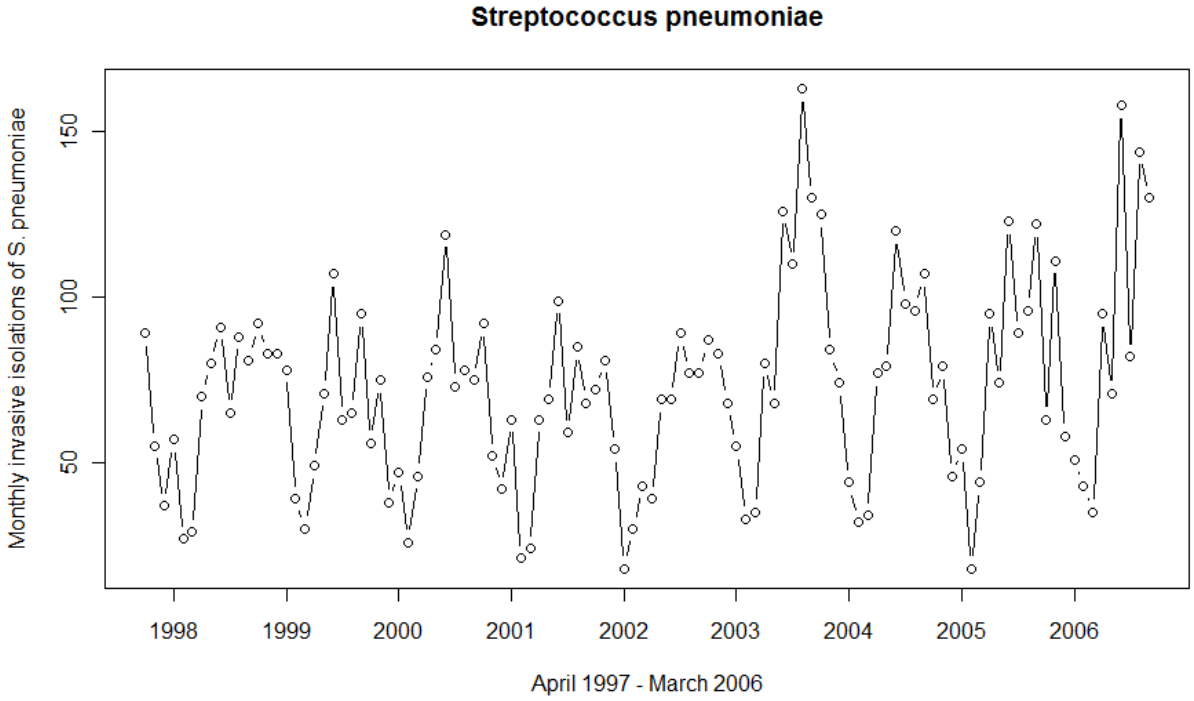


Figure 2.22 Monthly isolations of *S. pneumoniae* in England from April 1997 to March 2006 from children aged between 0-14 years.

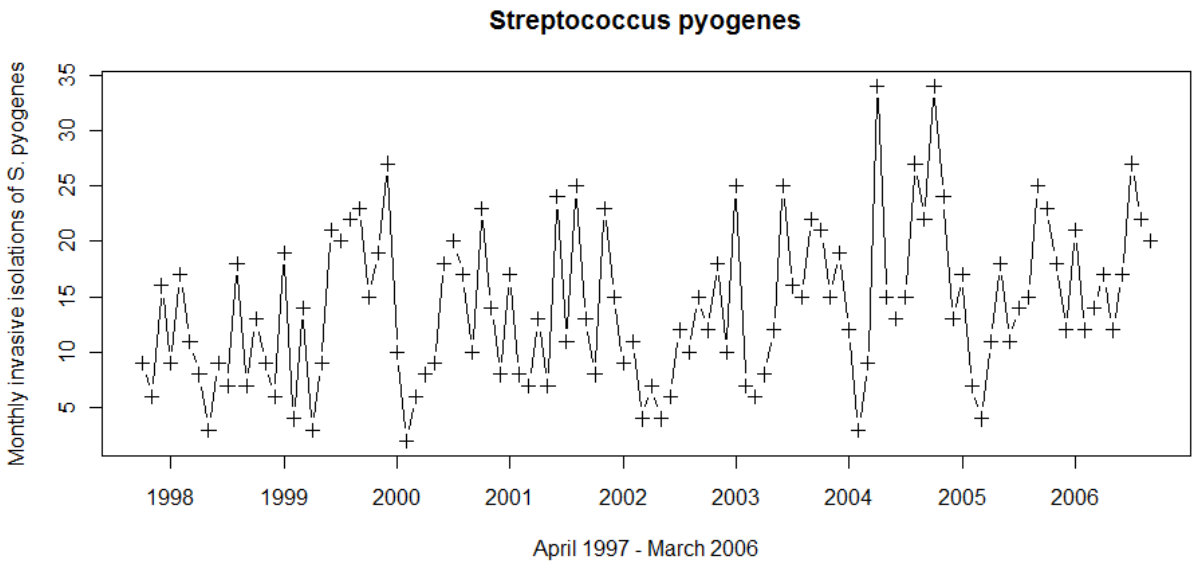


Figure 2.23 Monthly isolations of *S. pyogenes* in England from April 1997 to March 2006 from children aged between 0-14 years.

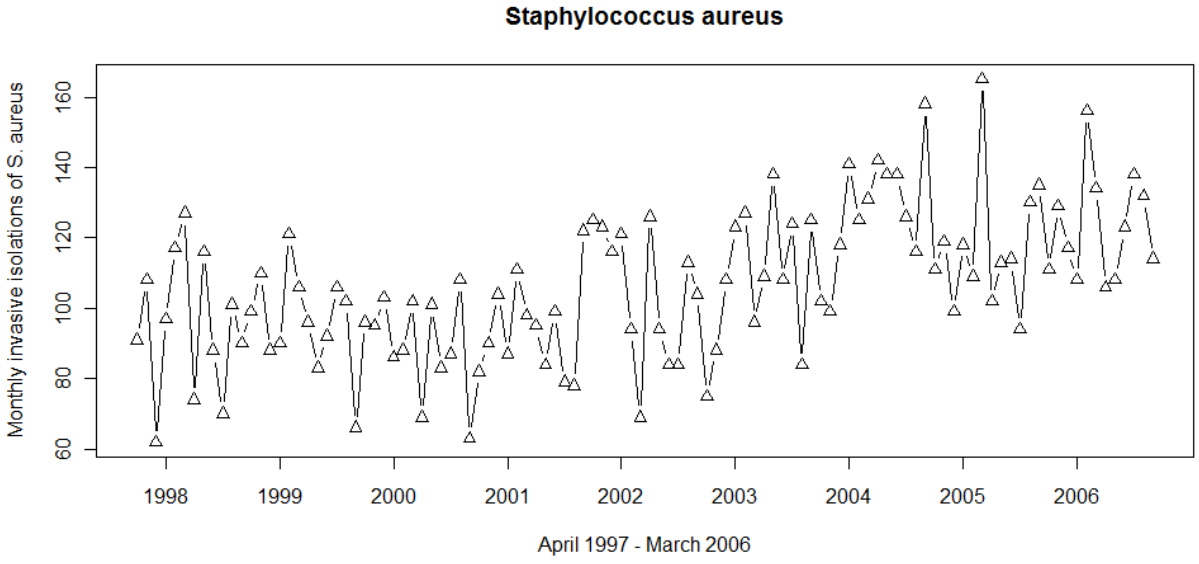


Figure 2.24 Monthly isolations of *S. aureus* in England from April 1997 to March 2006 from children aged between 0-14 years.

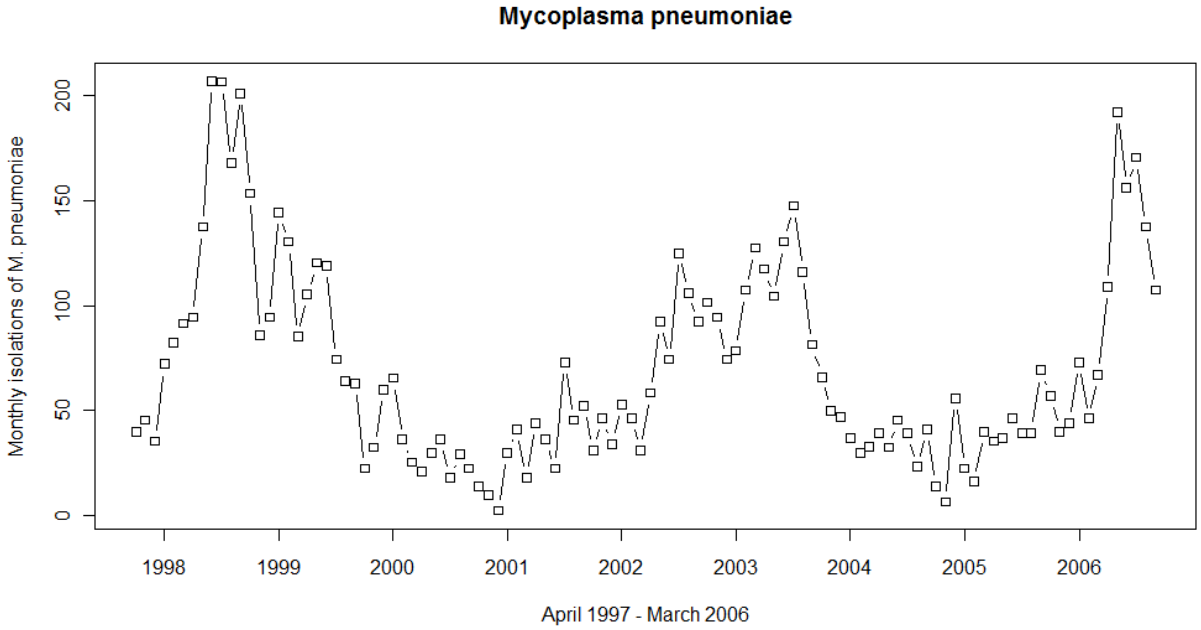


Figure 2.25 Monthly isolations of *M. pneumoniae* in England from April 1997 to March 2006 from children aged between 0-14 years.

Organism	Median monthly count of isolations (Range)	Total number of isolations between April 1997 – 2006	Significant increase over time period?
S. pneumoniae	73 (18-163)	7830	Yes (p=0.03)
S. pyogenes	13 (2-34)	1527	Yes (p<0.001)
S. aureus	106 (62-165)	11487	Yes (p<0.001)
M. pneumoniae	55 (2-207)	7561	No (p=0.39)

Table 2.18 Summary of isolations of *S. pneumoniae*, *S. pyogenes*, *S. aureus* and *M. pneumoniae* from children (0-14 years) in England from April 1997 to April 2006.

The relationship between national empyema admissions, the ratio of empyema to pneumonia admissions and isolations of *S. pneumoniae, S. aureus, S. pyogenes* and *M. pneumoniae*

Model P0 investigated the relationship between national empyema admissions and isolations of *S. pneumonia*e, *S. aureus, S. pyogenes* and *M. pneumonia*. The results of model P0 incorporating the best performing correlation structure are summarised in **Table 2.19** and for the parsimonious model in **Table 2.20**. Model S0 investigated the relationship between the ratio of empyema to pneumonia admissions and isolations of the four pathogens and is summarised in **Table 2.21** (full model) and **Table 2.22** (parsimonious model). The results of fitting progressive correlation structures to each model are summarised in Appendix H - **Table 9.2** and **Table 9.3**. For both models (P0 and S0) application of an auto-regressive moving average model provided the lowest model AIC and value of Phi. For P0 values of p=3 and q=0 provided the best overall performance, although this was a compromise as using values of p=2 and q=2 produced a significantly lower AIC but a high value of Phi. For model S0 values of p=3 and q=0 provided the best compromise between AIC and Phi.

Isolations of *S. pneumoniae* (co-efficient 0.606 95% CI 0.434 - 0.778) and *S. pyogenes* (co-efficient 0.236 95% CI 0.0885 – 0.3835) were significant positive predictors of monthly admissions for empyema suggesting that both pathogens were important causes of paediatric empyema. The magnitude of the association was greater for *S. pneumoniae*. No relationship was seen with isolations of *S. aureus* and *M. pneumoniae*. Predicted values from model P0 and log-transformed monthly empyema admissions over time are shown in **Figure 2.26**. The model R^2 was 0.692. However, examination of the model diagnostic plots indicated the persistence of serial dependence of the residuals and autocorrelation meaning that the validity of the estimates of the significance of the co-efficients cannot be relied upon (**Figure 2.27**, **Figure 9.51**, **Figure 9.52** and **Figure 9.53**).

Only isolations of *S. pyogenes* were a significant predictor of the monthly ratio of empyema to pneumonia admissions (co-efficient: 0.00582 95% CI 0.00151 – 0.0101). Both *S. pneumoniae* and *M. pneumoniae* were negatively associated with the ratio of empyema to pneumonia admissions suggesting that increasing isolations of either pathogen may be associated with a reduction in the relative rate of progression of pneumonia to empyema, although both co-efficients were non-significant. Model fit is shown in **Figure 2.28**. The model R^2 was 0.387. The model diagnostic plots indicated that similar to model S0 significant autocorrelation remained, although it was to a lesser degree than in model P0 (**Figure 9.54**, **Figure 9.55** and **Figure 9.56**).

Variable	**Co-efficient**	**SE**	**t-value**	**p-value**
(Intercept)	-1.119	1.144	-0.978	0.330
Log(Isolations of *S. pneumoniae* + 1)	0.619	0.093	6.665	<0.001
Log(Isolations of *S. pyogenes* + 1)	0.231	0.0766	3.022	0.0032
Log(Isolations of *S. aureus* + 1)	0.129	0.205	0.630	0.530
Log(Isolations of *M. pneumoniae* + 1)	-0.00815	0.0773	-0.105	0.916
Model structure: Log(Empyema admissions+1) ~ $\alpha + \beta_1$* Log(Isolations of *S. pneumoniae*+1) + β_2* Log(Isolations of *S. pyogenes*+1) + β_3* Log(Isolations of *S. aureus*+1) + β_4* Log(Isolations of *M. pneumoniae*+1) + ϵ				
Correlation structure: ARMA, p=3, q=0				
Phi: 0.26, 0.39, 0.18				
Residual SE: 0.555; Range of residuals: -1.987 to 1.836				
DF: total 108; residual				
AIC: 125.57 ; R^2: 0.710				

Table 2.19 Final model (P0) results relating empyema admissions to isolations of *S. pneumoniae*, *S. pyogenes*, *S. aureus* and *M. pneumoniae* from children (0-14 years) in England from April 1997 to April 2006.

Variable	**Co-efficient**	**SE**	**t-value**	**p-value**
(Intercept)	-0.508	0.414	-1.226	0.223
Log(Isolations of *S. pneumoniae* + 1)	0.606	0.0878	6.901	<0.0001
Log(Isolations of *S. pyogenes* + 1)	0.236	0.0753	3.132	0.0023
Model structure: Log(Empyema admissions+1) ~ $\alpha + \beta_1$*Log(Isolations of *S. pneumoniae*+1) + β_2* Log(Isolations of *S. pyogenes*+1) + ϵ				
Correlation structure: ARMA, p=3, q=0				
Phi: 0.26, 0.39, 0.18				
Residual SE: 0.568; Range of residuals: -1.999 to 1.849				
DF: total 108; residual 105				
AIC: 117.32 ; R^2: 0.692				

Table 2.20 Parsimonious model (P0) relating empyema admissions to isolations of *S. pneumoniae* and *S. pyogenes* from children (0-14 years) in England from April 1997 to April 2006.

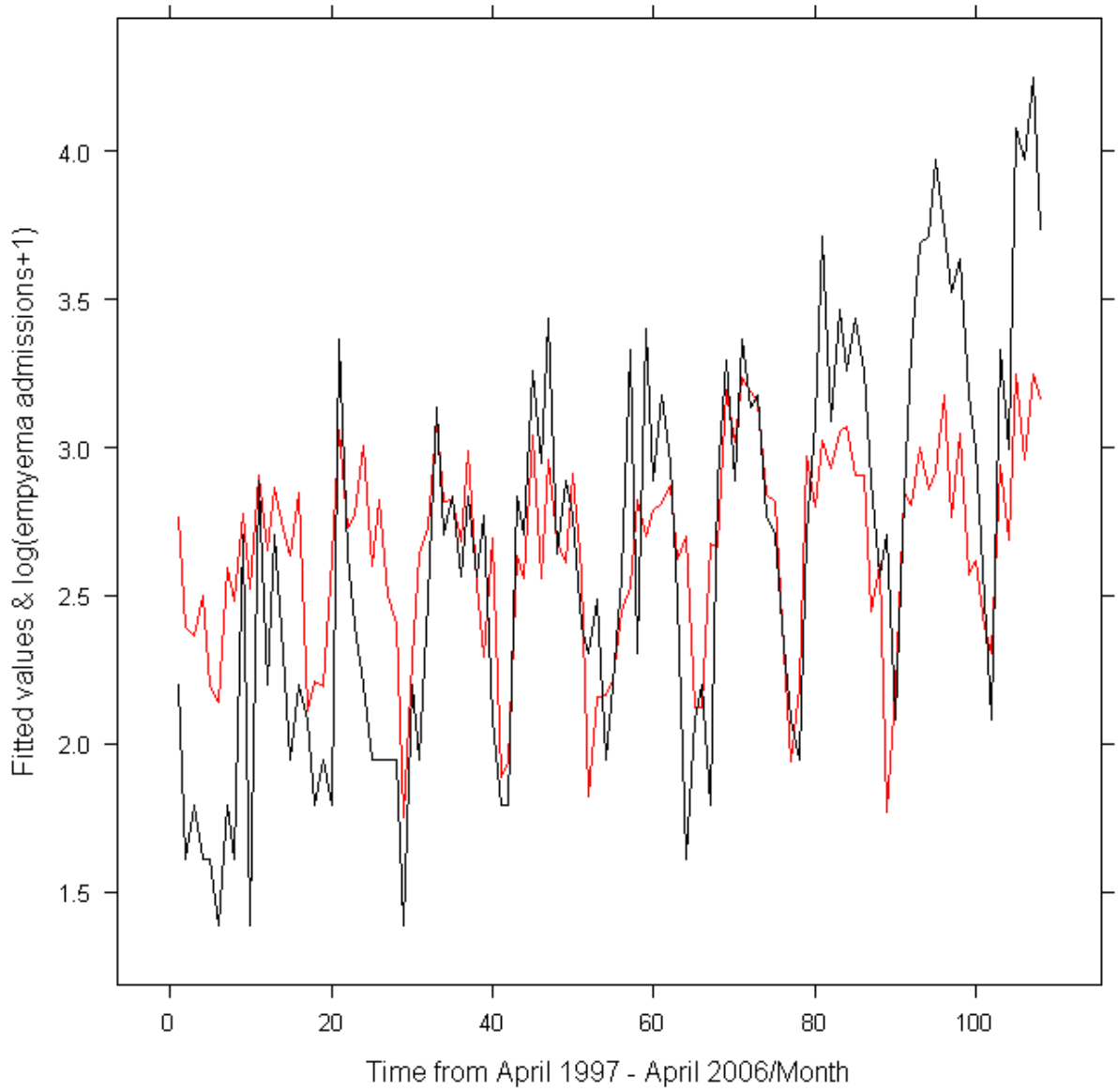


Figure 2.26 Monthly empyema admissions in England from April 1997-April 2006 (log transformed) and parsimonious model predicted monthly admissions using covariates of isolations of *S. pneumoniae* and *S. pyogenes* against time. *Red line – model predictions, black line – empyema admissions*.

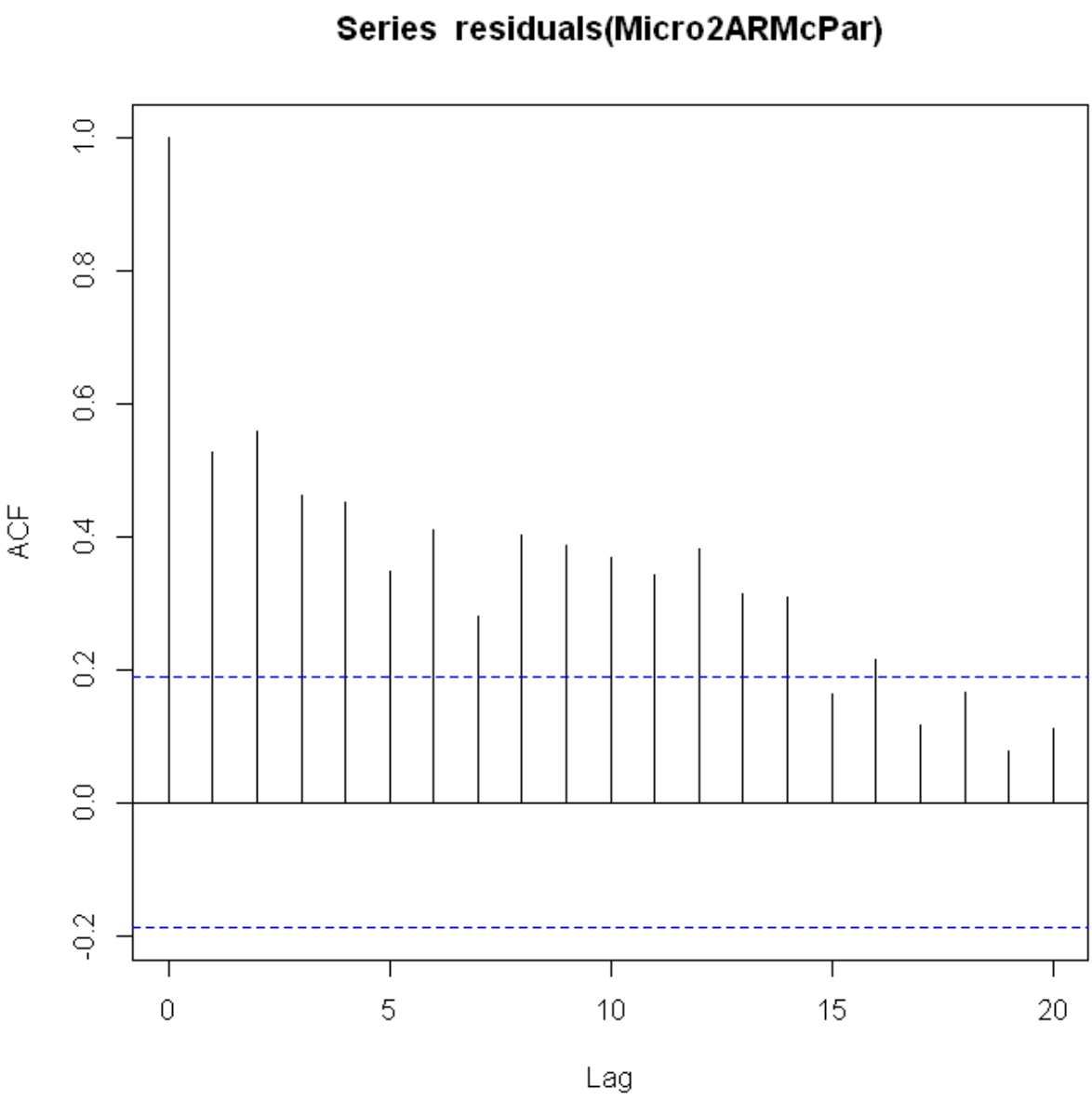


Figure 2.27 Autocorrelation plot of the residuals from model P0 investigating the relationship between empyema admissions and isolations of the four pathogens. The plot highlights the presence of continuing significant auto-correlation. ***Dotted lines indicate significance cut-off.***

Variable	**Co-efficient**	**SE**	**t-value**	**p-value**
(Intercept)	0.0274	0.0317	0.865	0.389
Log(Isolations of *S. pneumoniae* + 1)	-0.00159	0.00297	-0.534	0.595
Log(Isolations of *S. pyogenes* + 1)	0.00632	0.00239	2.650	0.0093
Log(Isolations of *S. aureus* + 1)	0.00142	0.00589	0.241	0.810
Log(Isolations of *M. pneumoniae* + 1)	-0.00274	0.00202	-1.354	0.179
Model structure: Ratio of empyema to pneumonia admissions ~ $\alpha + \beta_1$* Log(Isolations of *S. pneumoniae*+1) + β_2* Log(Isolations of *S. pyogenes*+1) + β_3* Log(Isolations of *S. aureus*+1) + β_4* Log(Isolations of *M. pneumoniae*+1) + ϵ				
Correlation structure: ARMA, p=2, q=0				
Phi: 0.24. 0.22				
Residual SE: 0.012 ; Range of residuals: -2.173 to 3.148				
DF: total 108; residual 103				
AIC: -604.65 ; R^2: 0.407				

Table 2.21 Final model (S0) results relating ratio of empyema to pneumonia admissions to isolations of *S. pneumoniae*, *S. pyogenes*, *S. aureus* and *M. pneumoniae* from children (0-14 years) in England from April 1997 to April 2006.

Variable	**Co-efficient**	**SE**	**t-value**	**p-value**
(Intercept)	0.0176	0.0060	2.917	0.0043
Log(Isolations of *S. pyogenes* + 1)	0.00582	0.0022	2.654	0.0092
Model structure: Ratio of empyema to pneumonia admissions ~ $\alpha + \beta_1$* Log(Isolations of *S. pyogenes*+1) + ϵ				
Correlation structure: ARMA, p=2, q=0				
Phi: 0.24, 0.22				
Residual SE: 0.0120; Range of residuals: -2.184 to 3.219				
DF: total 108; residual 103				
AIC: -636.88; R^2: 0.387				

Table 2.22 Parsimonious model (S0) relating ratio of empyema to pneumonia admissions to isolations of *S. pneumoniae, S. pyogenes, S. aureus* and *M. pneumoniae* from children (0-14 years) in England from April 1997 to April 2006.

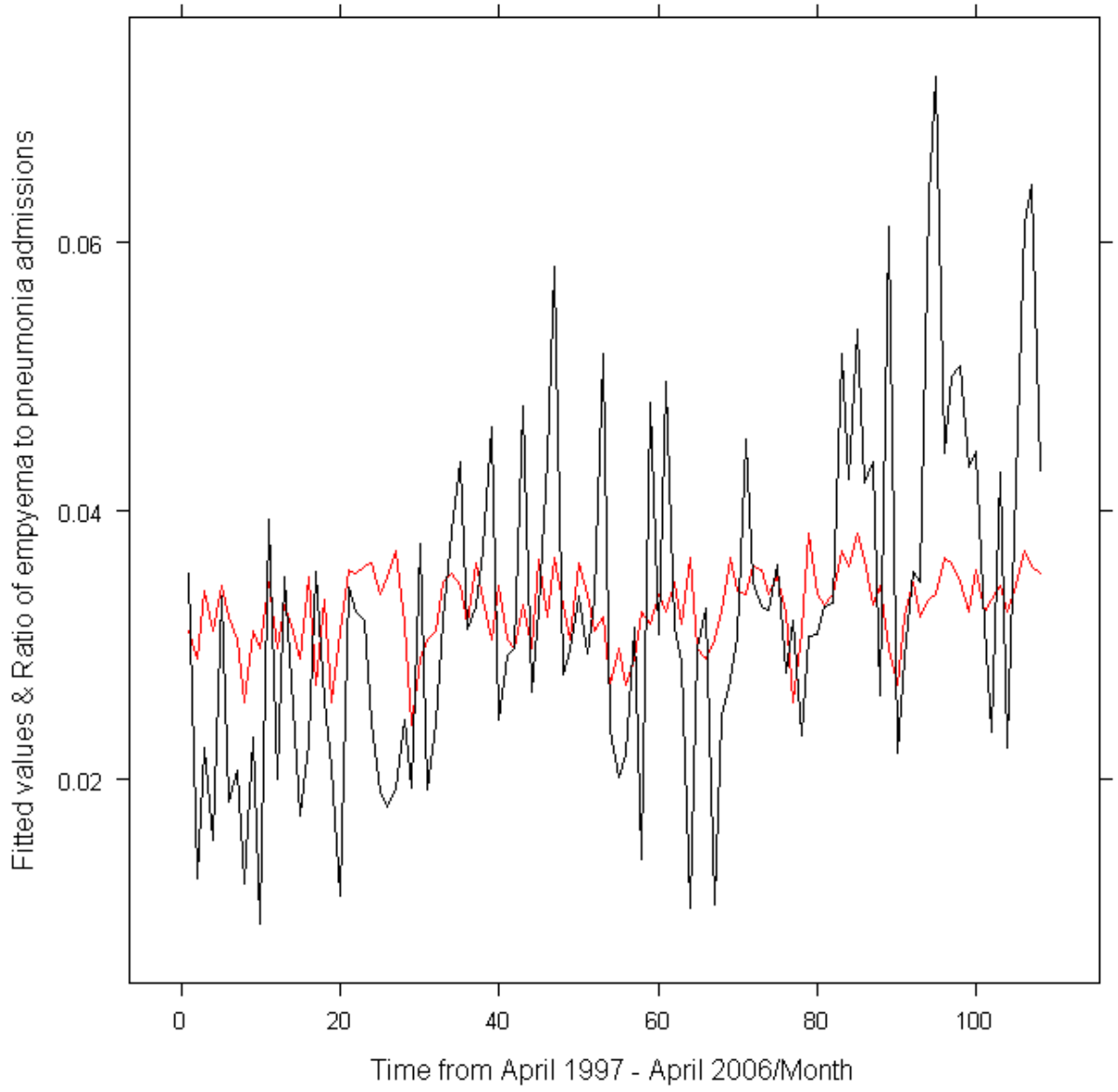


Figure 2.28 Monthly ratio of empyema to pneumonia admissions in England from April 1997-April 2006 and parsimonious model predicted monthly admissions using covariates of isolations of *S. pyogenes* against time. *Red line – model predictions, black line – empyema admissions.*

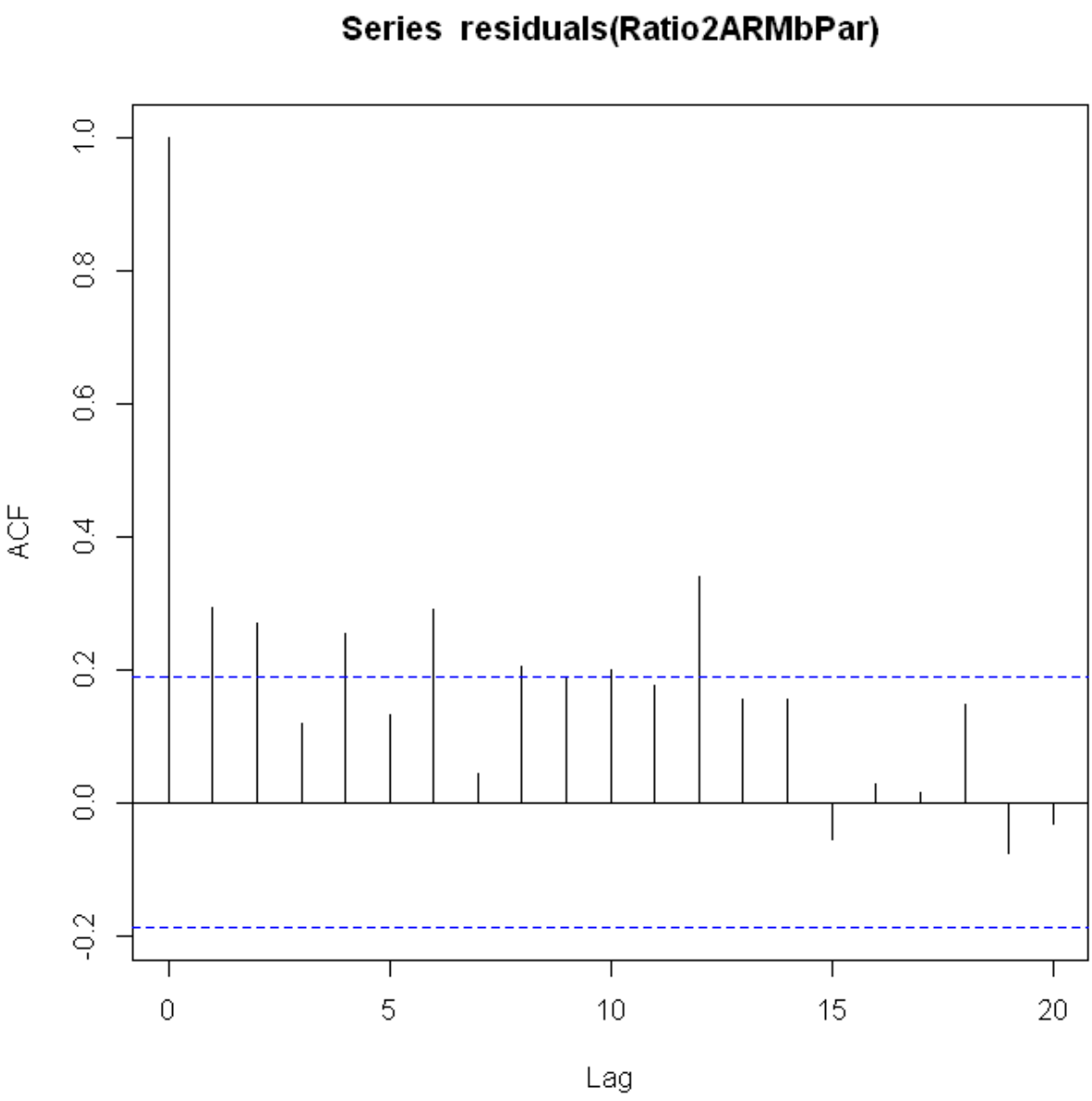


Figure 2.29 Autocorrelation plot of the residuals from model S0 investigating the relationship between the ratio between empyema and pneumonia admissions and isolations of the four pathogens. The plot highlights the presence of continuing significant auto-correlation. ***Dotted lines indicate cut-off for presences of significant autocorrelation. ACF – autocorrelation function (size of autocorrelation, lag indicates number of months between observations when correlation occurs)***

The relationship between periodicity and cyclicity within national empyema admissions and the ratio of empyema to pneumonia admissions and periodicity within isolations of *S. pneumoniae, S. aureus, S. pyogenes* and *M. pneumoniae*

Coherence analysis of empyema admissions and isolations of *S. pneumoniae*, *S. pyogenes*, *S. aureus* and *M. pneumoniae* in England are shown in **Figures 2.30 – 2.33**.

There was evidence of strong in-phase coherence between empyema admissions and isolations of *S. pneumoniae* at a period between 8 and 14 months indicating that seasonal peaks in empyema were temporally correlated with pneumococcal activity (**Figure 2.30**).

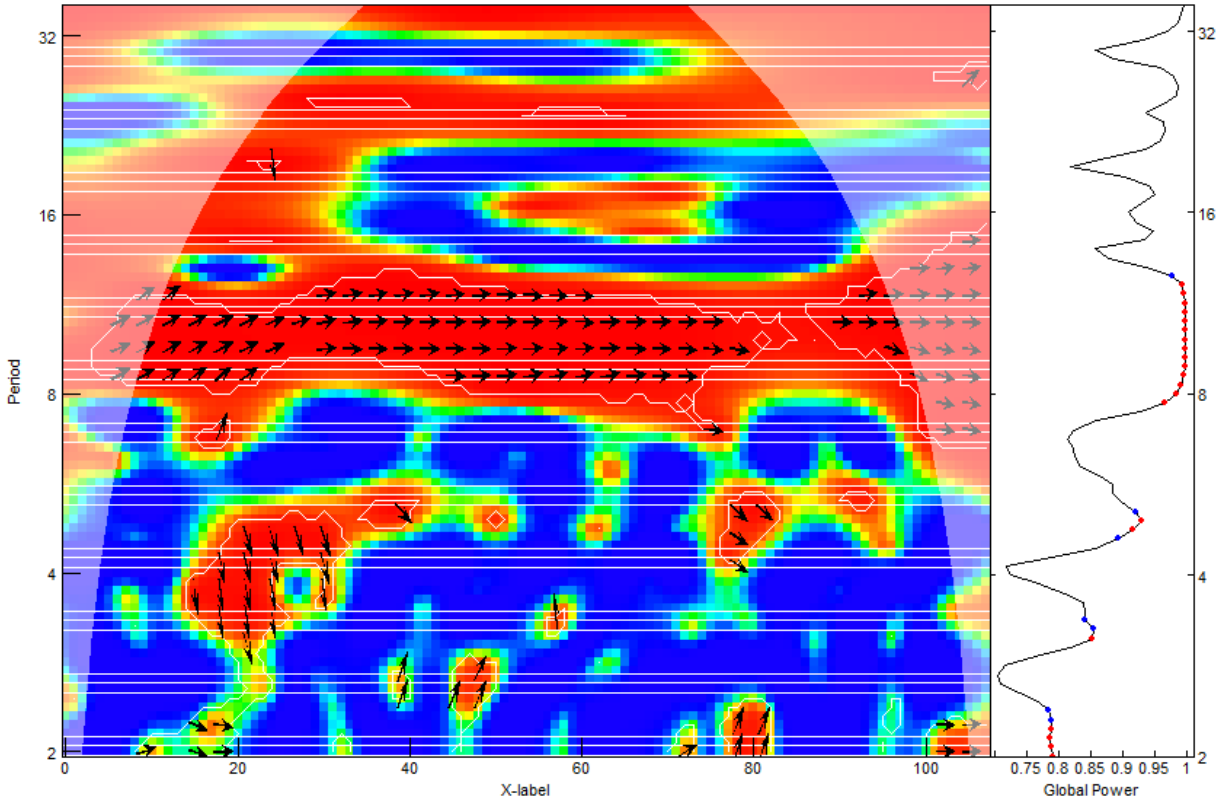


Figure 2.30 Coherence between paediatric empyema admissions and isolations of *S. pneumoniae* between April 1997 and April 2006 from the same age group in England. Periodicity is on the left y axis and is the period of the time series i.e. when the series is decomposed into a frequency signal where the peaks of activity co-incide as measured by the power of the wavelet coherence – these are plotted as a heat map with red indicating high power. The same plot is then represented adjacent

without the heat map. Arrows indicate phase difference, left to right equalling in phase relationship, right to left anti-phase relationship. Cone indicates area stable from edge effects.

There was evidence of coherence between empyema admissions and isolations of *S. pyogenes* at a period between 8 and 14 months indicating that seasonal peaks in empyema were temporally correlated with seasonal *S. pyogenes* activity, although there was some dissonance of phase of uncertain significance (**Figure 2.31**). Furthermore there was evidence of coherence around 18 and 32 months suggesting that other peaks in activity of *S. pyogenes* were likely to be contributing to empyema admissions.

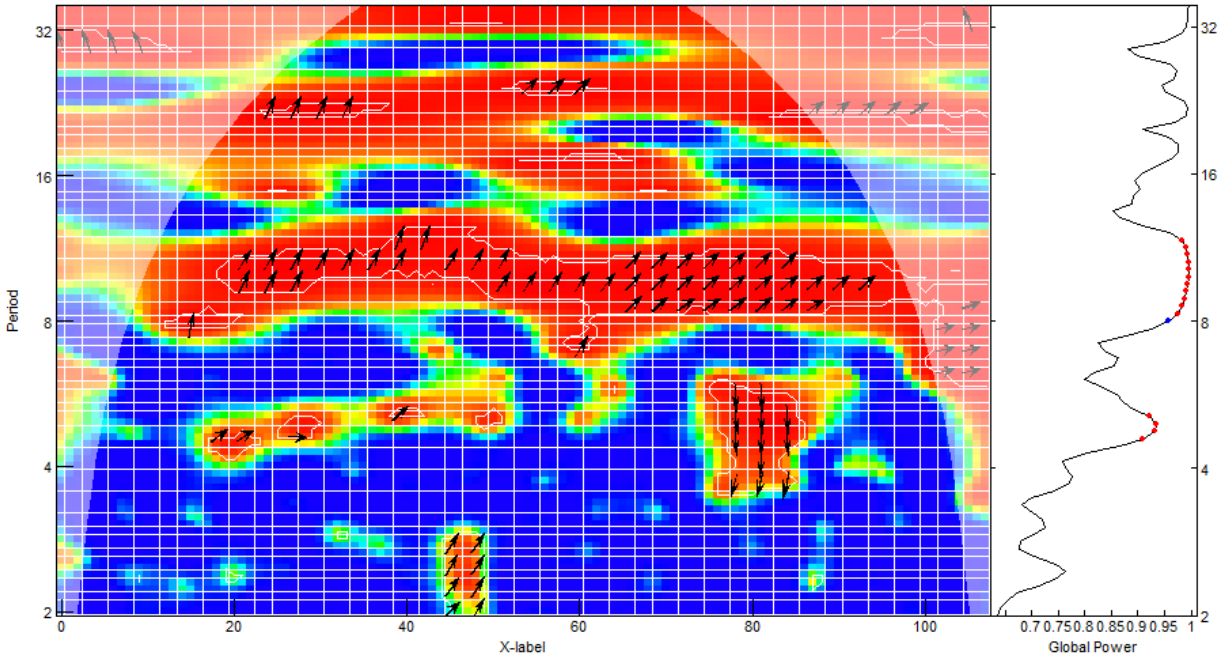


Figure 2.31 Coherence between paediatric empyema admissions and isolations of *S. pyogenes* between April 1997 and April 2006 from the same age group in England. Periodicity is on the left y axis and is the period of the time series i.e. when the series is decomposed into a frequency signal where the peaks of activity co-incide as measured by the power of the wavelet coherence – these are plotted as a heat map with red indicating high power. The same plot is then represented adjacent without the heat map. Arrows indicate phase difference, left to right equalling in phase relationship, right to left anti-phase relationship. Cone indicates area stable from edge effects.

There was only limited evidence of coherence between empyema admissions and isolations of *S. aureus* at a period of approximately 24 months, furthermore the coherence was not in-phase and the phase varied

between periods (**Figure 2.32**). Overall, *S. aureus* activity was not temporally correlated with empyema admissions.

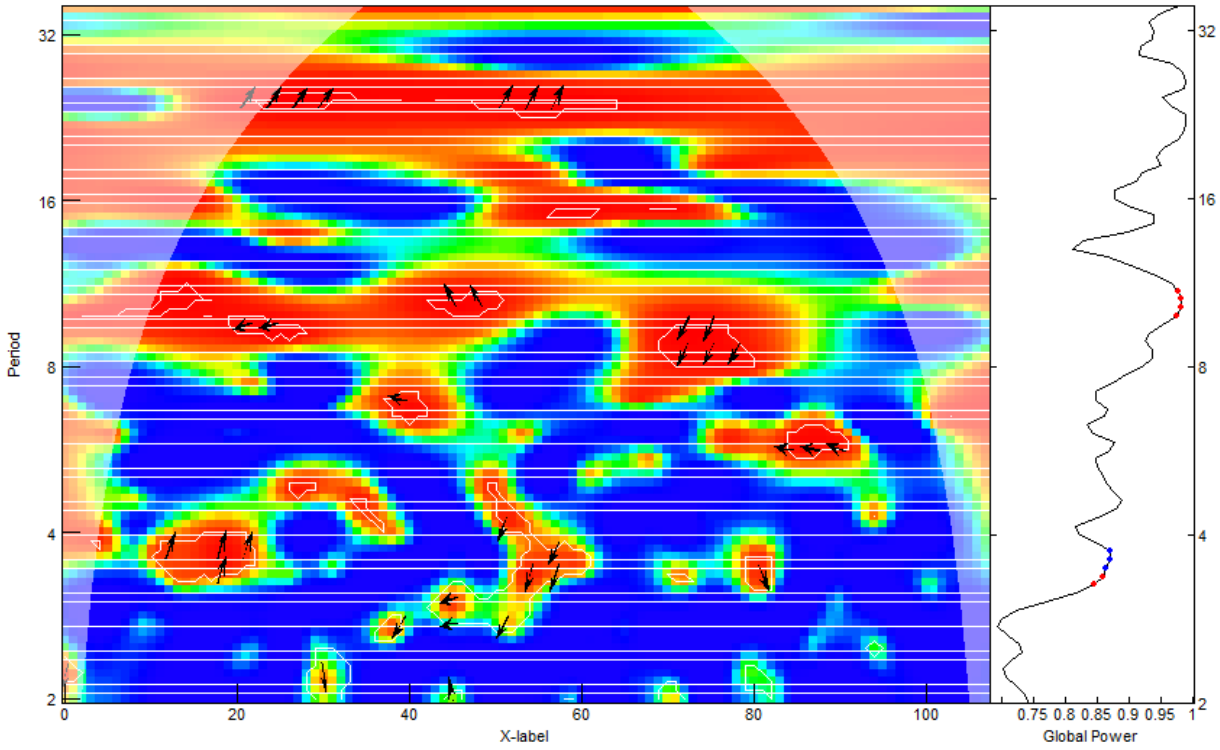


Figure 2.32 Coherence between paediatric empyema admissions and isolations of *S. aureus* between April 1997 and April 2006 from the same age group in England. Periodicity is on the left y axis and is the period of the time series i.e. when the series is decomposed into a frequency signal where the peaks of activity co-incide as measured by the power of the wavelet coherence – these are plotted as a heat map with red indicating high power. The same plot is then represented adjacent without the heat map. Arrows indicate phase difference, left to right equalling in phase relationship, right to left anti-phase relationship. Cone indicates area stable from edge effects.

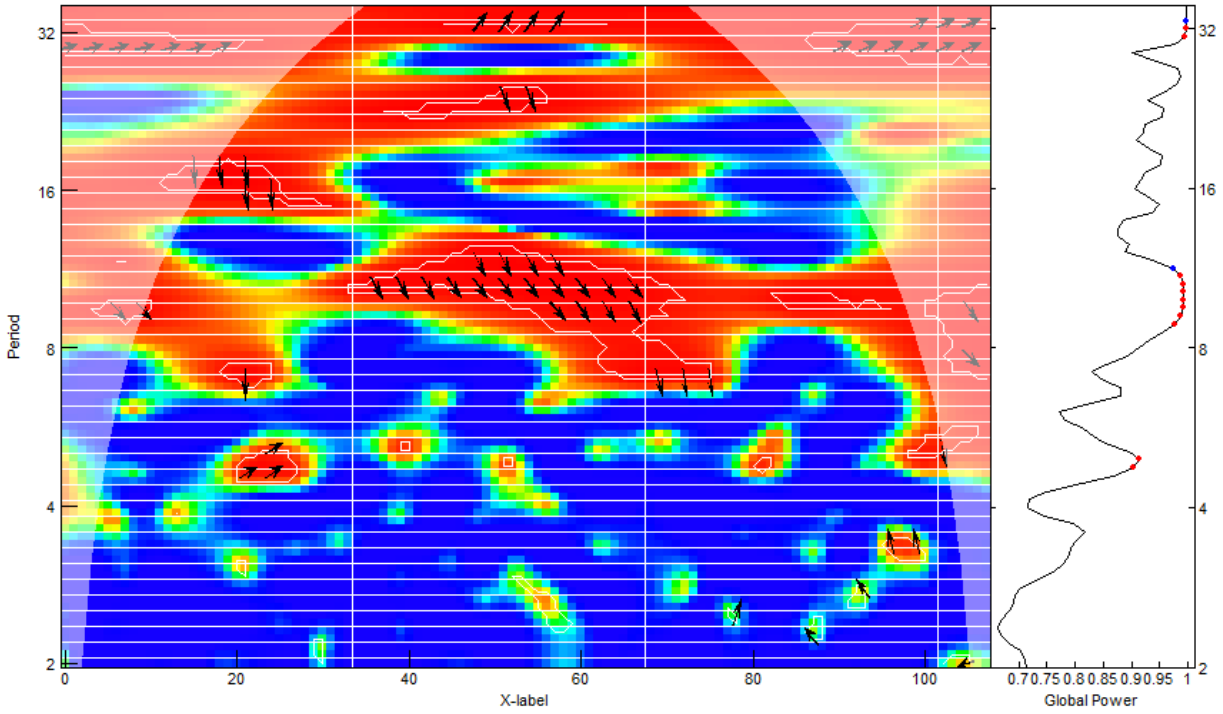


Figure 2.33 Coherence between paediatric empyema admissions and isolations of *M. pneumoniae* between April 1997 and April 2006 from the same age group in England. Periodicity is on the left y axis and is the period of the time series i.e. when the series is decomposed into a frequency signal where the peaks of activity co-incide as measured by the power of the wavelet coherence – these are plotted as a heat map with red indicating high power. The same plot is then represented adjacent without the heat map. Arrows indicate phase difference, left to right equalling in phase relationship, right to left anti-phase relationship. Cone indicates area stable from edge effects.

Similarly, there was only inconsistent evidence of coherence between empyema admissions and *M. pneumoniae* (**Figure 2.33**). The strongest signal was between six and 16 months, however there was substantial variation in this across the whole of the study period with it being strongest between 40 and 60 months and in the final ten months. Furthermore there is significant variation in the phase of the coherence between different periods. In conclusion, this suggests an inconsistent association between mycoplasma activity and empyema.

Coherency analyses of the monthly ratio of empyema to pneumonia admissions and isolations of *S. pneumoniae*, *S. pyogenes*, *S. aureus* and *M. pneumoniae* in England are shown in **Figures 2.34-2.37**.

There was evidence of strong in-phase coherence between the two time series at a period between 8 and 14 months indicating that seasonal increases in the rate of progression from pneumonia to empyema are temporally correlated with pneumococcal activity (**Figure 2.34**).

There was evidence of in-phase coherence between the monthly ratio of empyema to pneumonia admissions and isolations of *S. pyogenes* at a period between 8 and 14 months, although this effect is not as clear in the early part of the study period (first 20 months) (**Figure 2.35**). Furthermore there were periods of consistent coherence at 24 and 32 months, indicating that there is coherence between the rate of progression of pneumonia to empyema and *S. pyogenes* activity at periods greater than the traditional seasonal variation. This coherence suggests that the greater than 12 month periodicity seen in the rate of progression of pneumonia to empyema may be explained by *S. pyogenes*.

There was only limited evidence of coherence between the monthly ratio of empyema to pneumonia admissions and isolations of *S. aureus* at a period around 24 months, furthermore the coherence was not in-phase and the phase varied between periods (**Figure 2.36**).

There was no consistent evidence of coherence between the monthly ratio of empyema to pneumonia admissions and isolations of *M. pneumoniae*. There was also significant phase variation in the coherence that was present. Similar to the relationship between empyema admissions and isolations of *M. pneumoniae*, this suggests an inconsistent association between mycoplasma activity and empyema, which is unlikely to be causal.

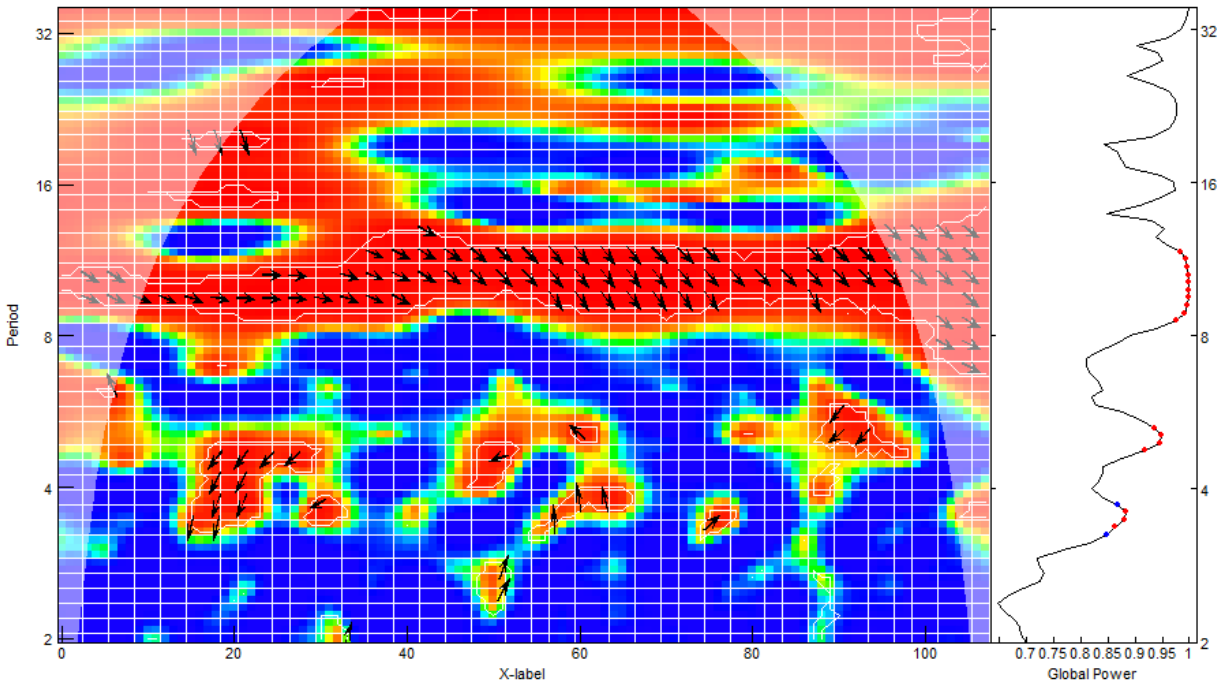


Figure 2.34 Coherence between the monthly ratio of empyema to pneumonia admissions and isolations of *S. pneumoniae* between April 1997 and April 2006 from the same age group in England. Periodicity is on the left y axis and is the period of the time series i.e. when the series is decomposed into a frequency signal where the peaks of activity co-incide as measured by the power of the wavelet coherence – these are plotted as a heat map with red indicating high power. The same plot is then represented adjacent without the heat map. Arrows indicate phase difference, left to right equalling in phase relationship, right to left anti-phase relationship. Cone indicates area stable from edge effects.

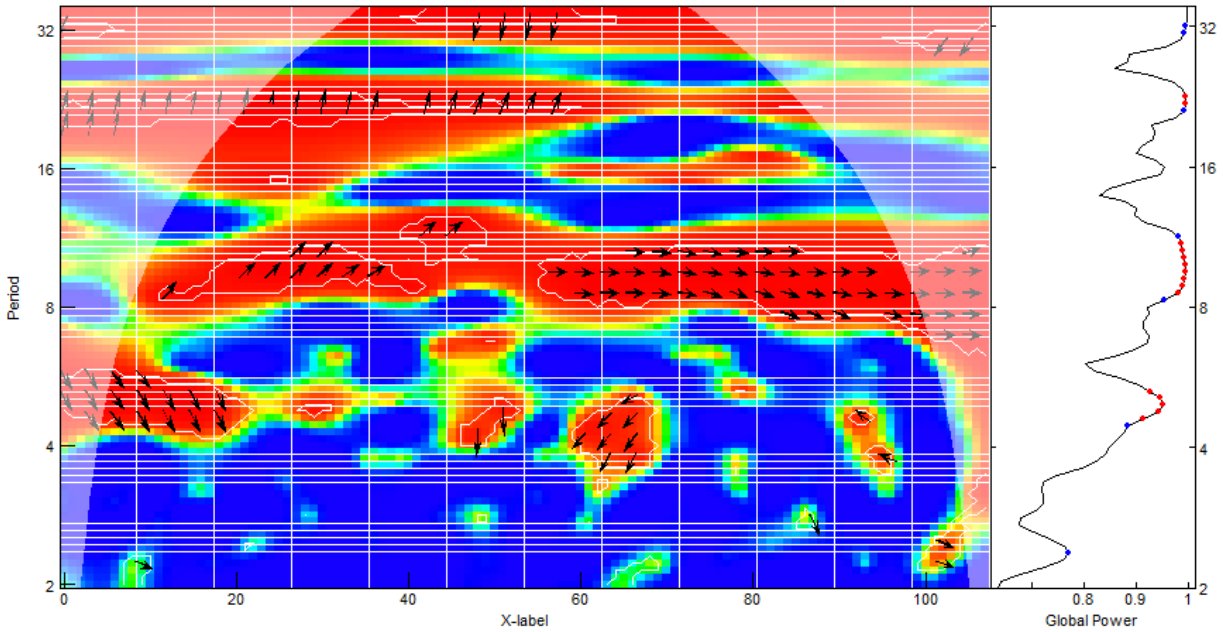


Figure 2.35 Coherence between the monthly ratio of empyema to pneumonia admissions and isolations of *S. pyogenes* between April 1997 and April 2006 from the same age group in England. Periodicity is on the left y axis and is the period of the time series i.e. when the series is decomposed into a frequency signal where the peaks of activity co-incide as measured by the power of the wavelet coherence – these are plotted as a heat map with red indicating high power. The same plot is then represented adjacent without the heat map. Arrows indicate phase difference, left to right equalling in phase relationship, right to left anti-phase relationship. Cone indicates area stable from edge effects.

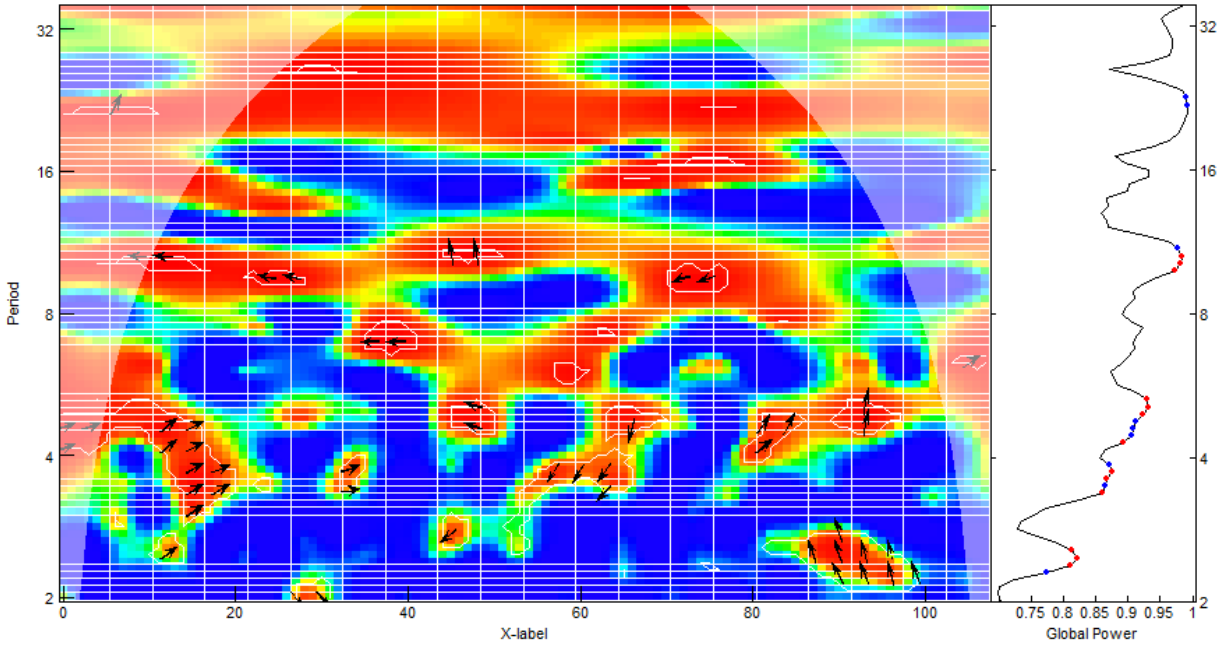


Figure 2.36 Coherence between the monthly ratio of empyema to pneumonia admissions and isolations of *S. aureus* between April 1997 and April 2006 from the same age group in England. Periodicity is on the left y axis and is the period of the time series i.e. when the series is decomposed into a frequency signal where the peaks of activity co-incide as measured by the power of the wavelet coherence – these are plotted as a heat map with red indicating high power. The same plot is then represented adjacent without the heat map. Arrows indicate phase difference, left to right equalling in phase relationship, right to left anti-phase relationship. Cone indicates area stable from edge effects.

.

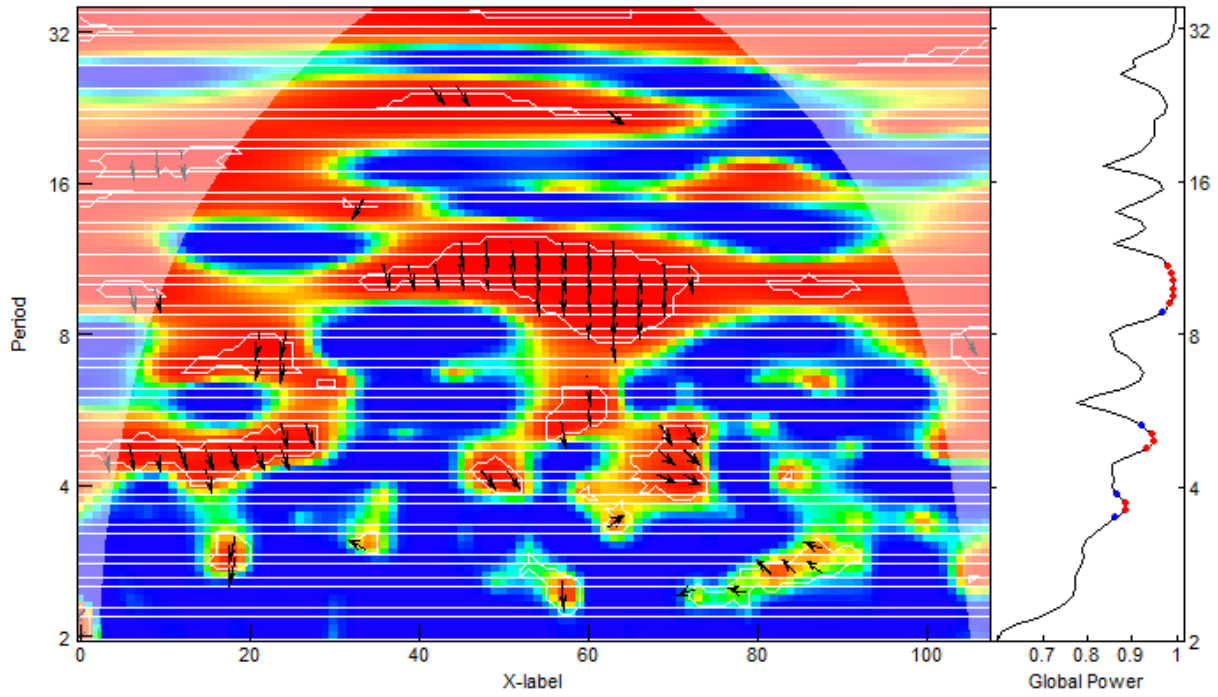


Figure 2.37 Coherence between the monthly ratio of empyema to pneumonia admissions and isolations of *M. pneumoniae* between April 1997 and April 2006 from the same age group in England. Periodicity is on the left y axis and is the period of the time series i.e. when the series is decomposed into a frequency signal where the peaks of activity co-incide as measured by the power of the wavelet coherence – these are plotted as a heat map with red indicating high power. The same plot is then represented adjacent without the heat map. Arrows indicate phase difference, left to right equalling in phase relationship, right to left anti-phase relationship. Cone indicates area stable from edge effects.

2.4 Discussion

The increase in empyema over the study period was consistent at local, regional and national levels. In England, empyema increased by 287% (Incidence rate ratio: 3.87, 95% CI: 3.063-4.933) between 1997 and 2005, equivalent to 253 extra cases per annum. Empyema admissions showed significant seasonal variation.

Pneumonia and empyema admissions were significantly correlated and pneumonia admissions alone explained approximately 85% of the variation in empyema admissions suggesting that the bulk of the increase in empyema was explained by an increase in pneumonia. However, the median ratio of empyema to pneumonia admissions per month doubled over the study period from 2.04 % to 4.13 % and this trend was both significant and independent of seasonal changes implying that while there was an increase in base-line levels of pneumonia, the rate of progression from pneumonia to empyema substantially increased. Furthermore, the rate of progression varied seasonally (11.75 months) and there was evidence of a greater than 36 month cycle.

Admission rates for pneumonia and empyema increased across all ten of the SHAs in England, although the relative rate of increase varied significantly between each SHA. Also, there was marked variation in the correlation between the two conditions between SHAs, suggesting that processes determining the rate of progression from pneumonia to empyema vary regionally.

Isolations of *S. pneumoniae* and *S. pyogenes* were positively associated with numbers of admissions to hospital with empyema in England and both increased between 1997 and 2006. Isolations of *S. pyogenes* were also a significant predictor of the monthly ratio of pneumonia to empyema admissions between 1997 and 2006. Isolations of *S. aureus* and *M. pneumoniae* were not significantly associated with either empyema admissions or the ratio of empyema to pneumonia admissions.

Strong coherence was seen between pneumococcal isolations and both empyema admissions and the rate of progression between pneumonia and empyema on a seasonal basis, strongly suggesting that the

observed increase in empyema has been pneumococcal in origin. Similarly there was clear coherence between *S. pyogenes* isolations and both between the monthly ratio of empyema to pneumonia admissions. Interestingly this was at a traditional seasonal periodicity but also at higher periodicities which overlapped with the cyclicity seen in empyema admissions and the rate of progression between the two conditions. This suggests that cyclical changes in *S. pyogenes* activity increase both the number of empyema admissions but also increase the rate of progression of pneumonia to empyema. Neither *S. aureus* nor *M. pneumoniae* activity appeared consistent with empyema admissions or the rate of progression from pneumonia to empyema.

2.4.1 Strengths and limitations

This analysis has several significant strengths. The consistency of results across local, regional and national populations and between different methods of data collection allows confidence that the trends described are robust. Furthermore, by using a framework for addressing and investigating seasonality the estimates of the change in incidence of empyema are of greater precision than have been previously derived. This analysis also represents the first attempt to examine the change in incidence of empyema in relation to different pathogens at national level. The data are comprehensive, at a sufficient scale and cover a long enough time period to allow accurate judgement of trends in relation to the four pathogens and empyema. Furthermore, the results of the analysis of the relationship between empyema and the different pathogens are consistent across two different analytical processes (GLS multi-variate time series regression and wavelet coherence analysis).

However, there are several limitations both within the data and the modelling framework. There have been legitimate concerns over the appropriateness of using HES data for time series analyses due to possible changes in the data collection systems used in clinical coding between institutions and within the same institution over time that could lead to bias. However, HES data are widely used for research into many aspects of healthcare including the burden of specific illnesses (Melegaro *et al.*, 2006). The upward trend in admissions for both conditions is consistent through the time period both nationally and regionally suggesting that the increase in admissions for both diseases is genuine and not a function of unrecognised changes in the systematic recording of data.

Total childhood hospital admissions in England increased from 1996/7 to 2006/7 by 18%. This may be a consequence of the increased use of Accident and Emergency departments as primary care units. As such, it is possible that some of the increase in pneumonia admissions may not reflect a true increase in cases of disease (Callery *et al.*, 2010). However, due to the severity of empyema, cases will almost invariably require hospital admission, and it is therefore much less likely to be affected by these variations.

The final models for the NE data series and the ratio of empyema had R^2 of approximately 0.6 and while this is acceptable for biological systems it suggests significant explanatory variables may not be present within the models. A second limitation relates to the periodicity analysis. While there was some evidence of a second period in the time series of the ratio of empyema to pneumonia admissions, none of the models evaluated contained significant terms for that period. As the available data only allow up to three cycles of a maximum of 36 months, it was not possible to establish whether the true period was greater than this or whether this represents an epiphenomenon. In addition, the use of months as the unit of the time series introduces inaccuracy to the analysis as all months are not the same length and the finding of an 11.75 month period may reflect this or may be as a consequence of year to year variations in the timing of seasons.

There were several limitations to the analysis of the microbiological data that require more detailed discussion. All the microbiological data were obtained through the HPA national surveillance system. Changes in surveillance data may not result solely from changes in the incidence of individual pathogens but instead from changes in case ascertainment. Factors such as changes in clinical practice e.g. reduced levels of sampling of blood cultures, or increased reporting can confound surveillance data (Flasche *et al.*, 2011). Flasche and colleagues (2011) analysed trends in pneumococcal bacteraemia and two control conditions before and after the introduction of universal pneumococcal vaccination. They found evidence of such changes influencing the estimated incidence of non-pyogenic streptococcal infections in the under 5's in the UK at the rate of approximately 9% increase per year. It would reasonable to consider this as a possibility in the surveillance data presented here. However, this is more of a concern when quantifying absolute changes in incidence rather than when looking at temporal associations.

The presence of autocorrelation in the models P0 and S0 leads to uncertainty as to the accuracy of the estimates of the significance of the co-efficients (Zuur *et al.*, 2009). While the attempts to fit a suitable

autocorrelative function were ultimately unsuccessful in modelling all the autocorrelation present, this approach should have ameliorated the impact as much as possible. The serial dependence may also have inflated the estimate of the model R^2.

The presence of significant autocorrelation is to be expected when modelling counts of infectious disease. The driver of serial dependence is infectious transmission within the population, which is largely unobserved as it occurs through asymptomatic carriers (Cooper and Lipsitch, 2004; Paul *et al.*, 2008). All four organisms can be are spread in this manner. The dependency within the residuals is therefore to some degree determined by the rate of transmission in the asymptomatic population. More complex modelling approaches are being developed which better account for these processes such as the use of hidden Markov models (Cooper and Lipsitch, 2004).

A further limitation of the approach used is the assumption that there is no interaction between the four pathogens modelled. For example, *S. pneumoniae* and *S. aureus* occupy a similar ecological niche within the nasopharynx and are recognised to compete actively with each other. How this interaction manifests on a population level is uncertain. Future analyses should account for and investigate interactions between pathogens.

Finally, the use of isolations data in preference to pathogen specific coding data is both a strength and a limitation of this analysis. The strength comes from avoiding the limitations of pathogen recording in coding data which have been described previously. A major limitation comes from the fact that isolations represent purely culture positive disease. There are recognised differences in the rate of culture positivity between pneumococcal serotypes and changes in pneumococcal serotypes have been implicated in the observed increase in empyema in a number of settings. An increase in pneumococcal disease due to serotypes that have lower rates of culture positivity may result in increases in empyema but not in isolations because these cases would not be recognised. This would impact on the accuracy of the relationship observed between isolations of pneumococcal disease and empyema.

2.4.2 Findings related to other studies

The findings of studies that have explored the hypothesis that an increase in admissions for paediatric pneumonia is responsible for a proportionate increase in admissions for empyema have been contradictory. One Scottish study examined national admission rates for paediatric pneumonia and empyema from 1980 to 2005. These authors observed an increase in the incidence of paediatric pneumonia and empyema over the study period, but concluded that the increase in empyema was a more recent phenomenon, predominantly occurring from 1998 onwards, uncorrelated with the gradual rise in cases of pneumonia (Roxburgh *et al.*, 2008). A more recent study from Spain examined the incidence of probable bacterial pneumonia and empyema from 1995 to 2005 and concluded that the incidence of probable bacterial paediatric pneumonia had also increased but, that the complication rate (including empyema) had remained constant at around 16% (Deiros Bronte *et al.*, 2006). An Australian study reported a trend towards a rise in incidence of pneumonia that was not statistically significant ($p=0.07$, $R^2=0.29$) and a statistically significant rise in empyema from 4 to 9.6 admissions per million children (0-16 yrs) ($p<0.05$, $R^2=0.51$) over the time period 1993/4 – 2004/5. These authors also reported an overall increase in the percentage of empyema as a proportion of pneumonia of 0·27% to 0·7% (Strachan and Jaffé, 2009). Following the introduction of the 7 valent pneumococcal vaccine (PCV-7) in the USA pneumonia hospitalisations complicated by empyema increased 2.01-fold from 3.5 cases per 100,000 children in 1996–1998 to 7.0 cases per 100,000 children in 2005–2007, while total hospitalisations for pneumonia fell (Grijalva *et al.*, 2010). Two further large scale hospital coding studies have further confirmed this trend (Lee *et al.*, 2010; Li and Tancredi, 2010).

In this analysis, variation in pneumonia explained 85% of the variation in empyema admissions, yet the proportion of empyema to pneumonia admissions per month doubled over the study period from 2.04 % to 4.13 %. These results suggest a more complex process than simply an increase in pneumonia being solely responsible for the increase in empyema. Direct comparisons with historical studies are challenging because of the variation in both case definition and methodology in previous studies. Strachan and Jaffé (2009) observed a change in the proportion of pneumonia to empyema of similar magnitude to this analysis, as did the coding studies from the USA (Grijalva *et al.*, 2010; Lee *et al.*, 2010; Li and Tancredi, 2010), although there was significant variation in the baseline rate in their study.

In this analysis pneumococcal and group A streptococcal infections were temporally associated with

empyema hospitalisations and increases in the rate of progression of empyema to pneumonia. While no previous studies have investigated the cause of the increase in incidence of empyema by establishing temporal associations between isolations of different pathogens and empyema on a national scale, one study has reported changes in the incidence of empyema due to different pathogens using coding data. Grijvala et al (2011) examined changes in the incidence of empyema across all age groups in the USA using a nationally representative sample of hospital coding data. They observed an increase in empyema in both adults and children between 1996 and 2008. In children the magnitude of the rise was 1.9 fold (test for trend, $p<0.001$). In contrast to the findings above, pneumococcal empyema rates remained stable among children (rate ratio 1.1, 95% CI 0.7 to 1.7, p value for test for trend $p=0.093$) whereas the rate of streptococcal (non-pneumococcal) empyema increased 2.7 fold (95% CI 1.5 – 5), staphylococcal empyema 2.4 fold (1.4 – 4.1) and empyema without a detected organism 2.5 fold (1.7 – 3.5). The overall organism detection rate within the analysis was 38%.

Culture-negative empyema (empyema initially without a detected organism) in children in the UK has been shown to be predominantly pneumococcal on molecular testing (Eltringham *et al.*, 2003). It is impossible to be certain of the extent of microbiological testing used in the cohort described by Grijvala *et al* as methods may vary significantly between institutions. However, molecular detection techniques were not routine during that time and it is therefore a fair assumption that the majority used only standard culture. It is quite possible that the increase seen in empyema without a detected organism represented changes in pneumococcal empyema, concurring with the temporal correlation between empyema and pneumococcal isolations seen in this analysis of UK data.

The increase in incidence of streptococcal empyema reported by Grijvala *et al* supports the finding of an association with isolations of *S. pyogenes*. However, Grijvala *et al* did not differentiate between different streptococci, as these are not distinguished in coding data based on ICD-9. In adults, *Streptococcus mitis spp* is the commonest bacterial cause of empyema in the UK (Maskell *et al.*, 2005). It appears to be rare in children but it is not unrecognised. It is therefore unlikely but possible that the changes seen in the USA may be a reflection of changes in this group of organisms. This further highlights the difficulties in using coding data to measure longitudinal trends in pathogen specific diseases.

The finding of coherence between the rate of progression between pneumonia and empyema and isolations of *S. pyogenes* at non-seasonal frequencies could provide an explanation for the epidemic like rises in the incidence of empyema observed in 1999 and 2005 in the USA (Grijalva *et al.*, 2010). Examination of the organism specific data presented by Grijalva *et al* in 2011 would support this theory

for 1999 but not for 2005. *S. pyogenes* frequently displays epidemic spread and may have a significantly greater potential for complicating pneumonia in children than pneumococcal infection (Musser *et al.*, 1995).

2.4.3 Implications and future research

Eighty five per cent of the variation in empyema can be explained by changes in pneumonia which, in conjunction with the significant level of correlation of 0.86 ($p<0.001$) between the two conditions, strongly suggests that the increase in empyema admissions has been driven by an increase in underlying bacterial pneumonia. However, given the significant regional variation in the increase in the rate of empyema; the significant increase in the ratio of empyema to pneumonia admissions per month and the significance of a secular increasing trend separate from that predicted by pneumonia suggest other processes are likely to be relevant to the observed increase in empyema. Hypotheses proposed elsewhere for the increase in incidence include a change in circulating pneumococcal serotypes or increasing pathogen virulence (Gupta and Crowley, 2006; Wexler *et al.*, 2006); a change in causative organism (Schultz *et al.*, 2004); changes in community antibiotic prescribing (Strachan and Jaffé, 2009) and increasing usage of Ibuprofen (François *et al.*, 2010). All are possible explanations and such processes are very likely to be multi-factorial. Changes in antibiotic prescribing almost certainly reflect regional variation and it is plausible that reduction in antibiotic prescribing for respiratory tract infections in primary care may increase the risk of progression from pneumonia to empyema. Furthermore antibiotic prescribing for lower respiratory tract infections in children in the UK fell by 56% from 1996 to 2006 (Thompson *et al.*, 2009). Similarly Ibuprofen may modify the rate of progression from pneumonia to empyema and authors in France have demonstrated correlation between an increase in total sales of Ibuprofen and the increase in incidence of empyema (François *et al.*, 2010). However, changes in pathogen are the most likely explanation. Prescribing habits are highly variable between different countries and increases in empyema have been seen in countries described as "low antibiotic users" e.g. the UK and "heavy antibiotic users" e.g. Spain and the USA (Gupta and Crowley, 2006; Obando *et al.*, 2008; Reinert *et al.*, 2010). Furthermore, the contemporaneity with which the increase in incidence of empyema occurred across different countries with 90% of reports occurring within the same 5 year band coupled to the variations in prescribing between countries make this less plausible. In contrast, the contemporaneity favours the hypothesis of pathogen related changes with global spread. Finally, the presence of two separate cycles within the rate of progression between empyema and pneumonia is suggestive of a pathogen related explanation as respiratory pathogens show cyclical epidemic like behaviour e.g. *Streptococcus pyogenes* or *Mycoplasma pneumoniae* (Colman *et al.*, 1993; Rasmussen JN, 2010).

The increase in incidence of empyema seen between 1997 and 2006 in England was associated with pneumococcal and streptococcal (*S. pyogenes)* disease with the biggest association seen with pneumococcal disease. Only *S. pyogenes* was associated with an increase in the rate of progression between pneumonia and empyema, whereas *S. pneumoniae* and *M. pneumoniae* were negatively associated with the rate of progression. This combination suggests that the bulk of empyema was caused by pneumococcal infection. A smaller proportion was associated with *S. pyogenes* but *S. pyogenes* had a greater tendency towards causing pneumonia complicated by empyema.

S. pyogenes has not been previously implicated in the change in incidence of empyema elsewhere despite historically being a significant pathogen in empyema. Different strains of *S. pyogenes* are recognised as characterised by M-typing and these demonstrate somewhat distinct pathogenic characteristics,. Streptococcal M types interact in a similar manner to pneumococcal capsular serotypes, and epidemic spread of individual types is recognised. Future research should attempt to define the M-types associated with childhood empyema and chart longitudinal changes in the incidence of different M-types in childhood empyema to better understand the processes that drive the observed fluctuations in incidence. Recognition of the role of *S. pyogenes* also has significant clinical implications. The clinical profile of streptococcal empyema is different to that of pneumococcal disease and may be more severe. For example, Grijalva *et al* (2011) reported a higher in-hospital case fatality ratio in streptococcal empyema than pneumococcal empyema (0.6 (95% CI 0.2 – 1) vs. 0.8 (0 – 1.5).

S. pneumoniae was temporally associated with empyema hospitalisations in a predictable seasonal pattern strongly suggesting that seasonality in empyema is related to seasonality in pneumococcal disease. This relationship also extends to changes in the relative rate of progression of pneumonia to empyema. *S. pyogenes* is an important driver of empyema and behaves in both a seasonal and epidemic manner in the short time series reported here. Future work should seek to better define the possible relationships between the two organisms and seek to understand the underlying mechanisms that drive both seasonal and epidemic behaviour.

There was no observable relationship between empyema hospitalisations and *S. aureus* isolations, despite the fact that *S. aureus* has been highlighted as a frequent cause of empyema in a number of recent cases series. The reasons for this are likely to be multi-factorial and may reflect differences in methods of detection e.g. blood culture rate, antibiotic prescribing patterns or antibiotic resistance all of which may impact on rates of detection.

Future research into the epidemiology of paediatric empyema at a population scale should take advantage of advances in the science of infectious disease modelling, incorporate methods of adjusting for the unobserved transmission of the relevant pathogens and analyse changes both in time and space (Paul *et al.*, 2008).

A major implication of the increase in incidence is the impact on limited health resources. Several estimates of the cost of treatment of empyema have been published. The most representative for the UK is that of Sonnappa et al (2006) which reported costs associated with medical treatment with fibrinolytics and VATS from a UK centre. Using the increase in annual incidence estimated by the national HES data between 1997 and 2005 which predicts an additional 253 cases per annum, the associated costs would be an extra £1,103,437 per year (fibrinolysis) and £1,494,544 (VATS). To put this in context, the annual budget from the Department of Health for the management of all non-HIV infectious disease in both adults and children in England was £800 million in 2006 (the first year figures were produced) ('Programme budgeting tools and data,' 2011). Assuming this resource is divided equally according to population demographics (proportion of children in population in 2006 17.7%) this provides an annual budget for the care of children of approximately £136 million. The increase in empyema would account for between 0.75 and 1.1% of this depending on treatment methodology which is clearly a substantial sum.

These finding have significant implications for patients. Longer term outcomes from empyema are largely unknown, but assumed to be good following treatment (Balfour-Lynn *et al.*, 2005). The long-term outcome data that is available is limited both in scale and by methodological issues. The commonest outcome reported is lung spirometry in the year following the acute episode, however the majority of empyema cases occur in the under – five age group and children below the age of six years often lack the co-ordination required to perform spirometry measurements in a repeatable manner. The consequence of this is that the majority of children contributing follow-up data were older than average when they contracted the illness and may not be representative of the whole population of interest. The majority of published data show no reduction in lung function compared to that predicted by age, sex and height in children who have had empyema and no differences between different treatment groups (McLaughlin *et al.*, 1984; Redding *et al.*, 1990; Sarihan *et al.*, 1998; Kohn *et al.*, 2002; Mew *et al.*, 2009). The total number of patients reported and the associated methodological problems, such as the absence of pre-morbid data or population level comparisons mean that this should be interpreted with caution.

Eastham et al (2008) suggested that children who have had serious lung infections may be at risk of chronic respiratory symptoms, such as chronic cough, however the longer term implications of this remain unclear. There has been increasing interest in childhood respiratory infections as the antecedents of chronic adult respiratory diseases, in particular chronic obstructive pulmonary disease (COPD). While there is no specific data on a linkage with childhood empyema, this may well be of importance given the dramatic rise in incidence of empyema.

2.5 Conclusions

The incidence of childhood empyema has increased locally, nationally and regionally. In England, cases increased by 287% from 1997 to 2006. Evidence from national data suggests that the rise in empyema was driven to a significant degree by a contemporaneous increase in pneumonia hospitalisations. The proportion of empyema to pneumonia hospitalisations doubled from a median of 2.04 % to 4.13% over the same time period. There is evidence that the relationship between pneumonia and empyema is complex as the rate of increase in both conditions varied markedly at a regional level and because of the presence of periods of strong concordance between pneumonia and empyema. This suggests that other factors such as variation in certain pathogens require further investigation to understand the mechanisms underpinning the increase in empyema. The relevance of variation in pathogens was supported by the positive association of isolations of *S. pneumoniae* and *S. pyogenes* with numbers of admissions to hospital with empyema. Seasonal variation in empyema was associated with both *S. pneumoniae* and *S. pyogenes*. *S. pyogenes* was associated with periodic changes in empyema, suggestive of epidemics of the condition. Further research is required to establish whether these associations are seen in other populations and over different time periods.

3 Spatial epidemiology of paediatric pneumonia and empyema

Declaration

A Blain implemented the Bayesian spatial models used in this chapter and produced the observed and predicted maps. All study design, analysis and interpretation were my own work.

3.1 Introduction

3.1.1 Introduction to spatial epidemiology

Spatial epidemiology is the description and analysis of geographic variations in disease with respect to demographic, environmental, behavioural, socioeconomic, genetic, and infectious risk factors (Elliott and Wartenberg, 2004). In particular, investigating spatial epidemiology can provide insights into the pathogenesis of disease (Mayer, 1983). This relies on the general principle that: "The demonstration of association of a disease with place implies either that the inhabitants of the particular place possess characteristics of etiologic importance in the disease and different from those of the inhabitants of other places, or that etiologic factors are present in the biologic, chemical and physical, or social environments of the people inhabiting the affected places, or that both these types of explanations apply."(Mayer, 1983)

The origins of spatial epidemiology stem from the disease mapping studies of the 1800s (Elliott and Wartenberg, 2004). In these early studies direct mapping of cases elucidated the likely causes and mechanisms of transmission of a range of infectious diseases (Elliott and Wartenberg, 2004). Of these, perhaps the most famous is John Snow's mapping of cholera cases in Soho, London in 1854 and subsequent identification of the Broad Street pump as the epicentre of the outbreak (Snow, 2008). Through the 20th century, spatial epidemiology has led to a wide range of insights into pathogenesis which have been of major importance to public health. Keys study of seven countries leading to the recognition of the importance of dietary fat in coronary heart disease (Epstein, 1996), whilst Cliff and Smallman-Raynor's identification of an association between AIDS cases and trucking routes in Uganda helped to elucidate the patterns of transmission of HIV (Cliff and Smallman-Raynor, 1992). Spatial epidemiology has also provided clues to infectious aetiology in conditions not thought to be infectious e.g. the recognition of the association between Hepatitis B infection and hepatocellular carcinoma (Elliott and Wartenberg, 2004). As well as identifying factors relevant to pathogenesis, spatial analyses have also been used to identify areas of increased risk which may benefit from targeted public health intervention e.g. Law *et al*'s spatial analysis of sexually transmitted infections (STI) in North Carolina which identified

heterogeneity in risk of STI but overlap between areas of risk of different types of infection, thereby identifying areas that may benefit most from targeted public health intervention (Law *et al.*, 2004).

3.1.2 Rationale for analysis of spatial epidemiology of pneumonia and empyema

Pneumonia is the leading cause of childhood mortality in children under 5 years, with greater than 2 million deaths a year occurring worldwide (Wardlaw *et al.*, 2006). Although the greatest burden of the disease occurs in developing and newly industrialised countries (151 million of an estimated total of 156 million cases) it remains a serious cause of child morbidity and mortality in the developed world (Rudan *et al.*, 2008b). Empyema is the presence of infection within the pleural membranes that overlie the lungs and most often arises as a complication of pneumonia in the underlying lung. It appears to have changed over the last 20 years, with a sharp rise in incidence reported across Western Europe, North America and the Far East (Gupta and Crowley, 2006; Grijalva *et al.*, 2010). In the UK, admissions increased from <10 per million in 1998 to 37 per million in 2005 (Roxburgh *et al*, 2008).

As described in more detail below, there is a paucity of information relating to the spatial epidemiology of either disease in childhood. The epidemiology of empyema appears to have changed and is not yet well-characterised. Characterising the spatial epidemiology of empyema is therefore an important step in understanding empyema and identifying factors that have led to the change in epidemiology. Furthermore, in order to understand the epidemiology of empyema it is necessary to explore and understand the epidemiology of pneumonia given the interconnectedness of the two conditions.

There is evidence from the limited previous spatial studies available and from individual level studies that deprivation, ethnicity, migration and healthcare factors may be relevant factors in the spatial epidemiology of both diseases (Crighton *et al.*, 2007; Kazembe and Namangale, 2007; Charland *et al.*, 2011). By extending the analysis to beyond the establishment of the presence of spatial variation in both conditions and testing the influence of these as spatial co-variates, their relevance to the spatial epidemiology of both conditions can be evaluated. Their relevance in either condition will then provide insights into the potential mechanisms underpinning the spatial epidemiology of both conditions.

3.1.3 Current evidence in spatial epidemiology of pneumonia and empyema

Two studies have reported significant spatial variation in pneumonia (Crighton *et al.*, 2007; Kazembe and Namangale, 2007). Crighton *et al* (2007) examined risk of hospitalisation from influenza and pneumonia across the Canadian state of Ontario, using spatial and non-spatial regression methods and identified significant spatial variation. Furthermore within the child population, low education score for district and proportion of population of Aboriginal race were significantly associated with increased rates of hospitalisation. Kazembe *et al* (2007) examined risk of pneumonia, fever and diarrhoea using data from a self-reported health questionnaire in children in Malawi and identified significant spatial variation in pneumonia with increased risk in rural areas. They also identified some clustering in the South of the country where deprivation rates are highest, but the authors were unable to include HIV data or use other specific spatial variables which might have further identified putative causative factors.

Published data on spatial variation in empyema in children is more limited, with only one available study which found no evidence of spatial clustering of cases between 1988 and 1994 in Ohio, USA (Hardie *et al.*, 1996).

3.1.4 Possible risk factors in the spatial epidemiology of pneumonia and empyema

A link between socio-economic deprivation and risk of pneumonia in children has been previously reported in a number of individual level studies, (Hawker *et al.*, 2003; Clark *et al.*, 2007a; Flory *et al.*, 2009) but the results of spatial studies have been mixed (Crighton *et al.*, 2007; Charland *et al.*, 2011). For example, Charland *et al* (2011) in a spatial study of clinic attendance with influenza and pneumonia observed significantly higher rates in areas of greatest material deprivation but could not establish a linear relationship between risk and the level of deprivation. They found no significant link with social deprivation. The influence on socio-economic deprivation on empyema is less certain. It has been associated with increasing severity in pneumonia but a specific link to empyema has not been reported (Clark *et al.*, 2007a). Given the close relationship of the two conditions, it could be hypothesised that similar associations in empyema may exist.

Ethnicity was a predictor of spatial variation in risk of pneumonia and influenza hospitalisations in children in Ontario and a significant determinant of childhood asthma admissions in the UK with

increased in risk in ethnic minority groups in both settings (Gilthorpe *et al.*, 1998; Crighton *et al.*, 2007). The relevance in empyema is unknown. Increased population mixing is well known to potentiate the transmission of infectious diseases, therefore it could be hypothesised that migration levels may be relevant in explaining any spatial variation in either pneumonia or empyema (Wilson, 1995).

Health-care related factors may also be risk factors for spatial variation in pneumonia and empyema. For example, the ratio of population to physicians has been found to be predictive of mortality in childhood lower respiratory tract infections in Bangladesh, although a similar measure was not significantly associated with pneumonia hospitalisations in Canada.(Ali *et al.*, 2001; Crighton *et al.*, 2007) The distance to treatment centre was also a significant predictor of mortality in Bangladesh, highlighting that beyond availability, accessibility may also be relevant in determining spatial variation (Ali *et al.*, 2001).

3.1.5 Challenges of area based analysis in spatial epidemiology

Beale *et al.* (2008) argue that "because of data limitations, most spatial epidemiological studies use data aggregated at the area level … and therefore require that spatial epidemiology analyses be carried out at an ecological level". However, this approach can be associated with significant challenges. Distribution and classification issues in particular can lead to significant bias and error in estimates of association but are intrinsic to area level data. For example, boundaries used to divide areas such as postcodes or SHAs are often primarily administrative and may change over time, potentially causing significant errors of attribution within spatial data. Furthermore, it is likely disease and disease risk varies continuously throughout space but by using an area level approach this variation is subdivided. This will lead to distortion of the estimates of disease risk to some degree according to the boundary structure used. These will also be influenced by the relative homogeneity and heterogeneity of the areas used.

Looking at healthcare related data further complicates matters due to the hierarchy through which care is delivered. For example, if hospitalisations in the UK are used as the outcome measure of interest then several spatially varying factors are likely to be relevant. Access to general practitioners, individual practitioner behaviour and local healthcare networks are likely to be factors prior to referral to hospital. Practitioner behaviour, resource priorities and local clinical guidelines are likely to also have an impact at the hospital institutional level.

Finally, there is also the need to avoid the ecological fallacy when interpreting spatial epidemiological

analyses whereby, associations determined at an ecological level are translated into associations at the individual level, as these may not follow. Any area or spatial based analysis is therefore going to be to involve significant assumptions and be potentially affected by these factors. These can be mollified to some degree by employing a robust modelling strategy and by ensuring that any conclusions drawn are at the community and ecological level.

3.1.6 Aims & hypotheses

The aims of this analysis were twofold. One to establish whether there was spatial variation in pneumonia and empyema in the North East of England. Secondly, to move beyond this and establish what community level risk factors may contribute to the risk of both conditions. Possible risk factors identified previously include socio-economic deprivation, ethnicity and health-care factors.

Both pneumonia and empyema are infectious diseases and therefore factors likely to explain spatial variation in risk of either may be related to transmission or susceptibility within a population or community. For example, socio-economic deprivation may predispose a population to pneumonia in different ways. Socio-economic deprivation is linked to poor quality housing which may increase individual's susceptibility to respiratory infection. It also linked to overcrowding which increases transmission of respiratory pathogens, thereby increasing risk of pneumonia. Both of these factors are often aggregated at a community level. For example, deprived communities are likely to have higher levels of poor quality housing and overcrowding and hence the effects may be manifest at a community level. Transmission may also be potentiated by migration and communities with high levels of migration may be at greater risk than those with lower levels of migration. Ethnicity is linked to variations in an individual's susceptibility to disease, for example Native American populations have been found to have higher rates of both carriage of pneumococci and infection with invasive pneumococcal disease than other populations (Singleton *et al.*, 2007). As communities are often arranged along ethnic lines, these effects are likely to be seen at an area or community level. Similarly health care factors, such as immunisation rates will also influence susceptibility at a community level. Health care factors may also influence risk of pneumonia or empyema by mechanisms other than transmission or susceptibility. For example, the proximity of a community to hospital may influence attendance and hence diagnosis rate. In addition, hospitals and doctors are not homogenous and have different diagnostic thresholds, all of which may alter perceived community level risk.

In order to achieve these aims two hypotheses were investigated:

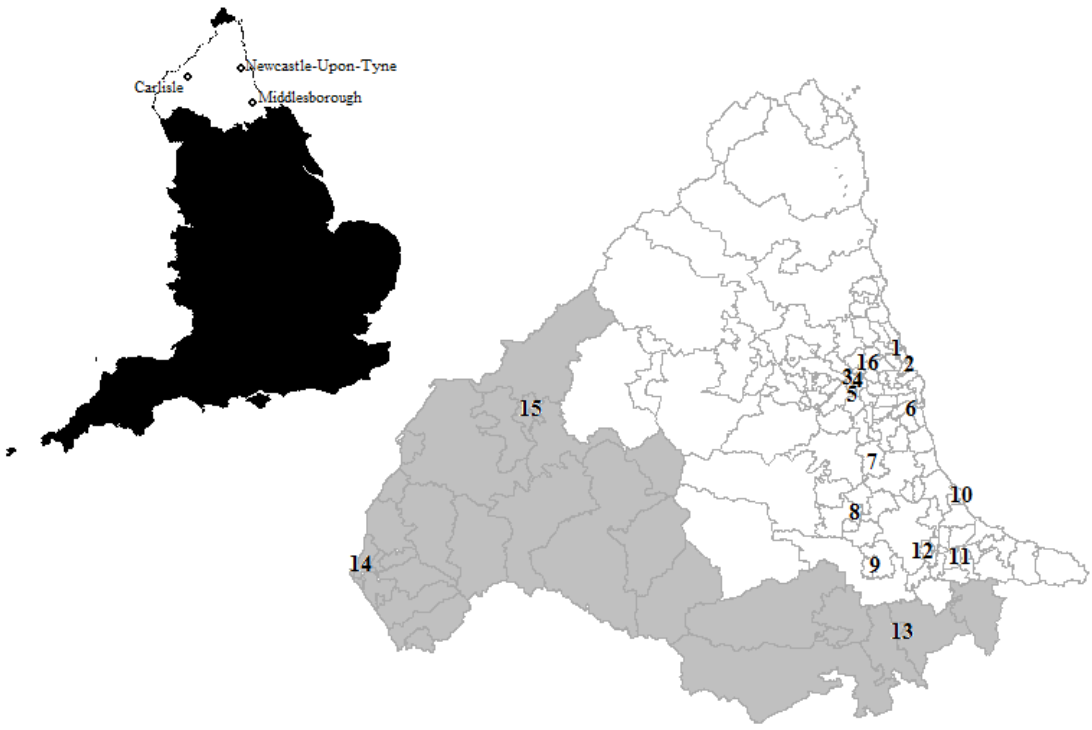
- *There is spatial variation in the risk of pneumonia and empyema.*
- *Spatial variation in risk of pneumonia and empyema will be associated with community level deprivation, ethnicity and proximity to hospital.*

3.2 Methods

3.2.1 Pneumonia data

Data on childhood hospitalisations in the North East of England SHA area (shown in **Figure 3.1**) from 1st April 1997– 30th March 2007 with a diagnosis of bacterial or lobar pneumonia ICD-10 codes J13, J14 and J150–9, J181) in children under 15 years of age were extracted from HES (*HESonline - Hospital Episode Statistics*, 2010). There is a more detailed description of HES data in Chapter 2.2.2.

All data with each disease code were counted and all those with the same HES identification number were amalgamated to convert from episodes to hospital admissions. Duplicate entries were removed and readmissions were excluded if they occurred within a month of the initial admission. Each HES entry contains partial postcode data, through which all hospitalisations over the time period were linked to postcode districts. Residence data were missing or unusable in 15 of 3889 (0.4%) hospitalisations and these were therefore excluded.



1.	North Tyneside General Infirmary, Whitley Bay	9.	Darlington Memorial Hospital, Darlington
2.	South Tyneside District Hospital, South Shields	10.	University Hospital of Hartlepool, Hartlepool
3.	Newcastle General Hospital, Newcastle-upon-Tyne	11.	University Hospital of North Tees, Stockton-on-Tees
4.	Royal Victoria Infirmary, Newcastle-Upon-Tyne	12.	James Cook University Hospital, Middlesbrough
5.	Queen Elizabeth Hospital, Gateshead	13.	Friarage Hospital, Northallerton
6.	Sunderland Royal Infirmary, Sunderland	14.	West Cumberland Hospital, Whitehaven
7.	University Hospital of North Durham	15.	Cumberland Infirmary, Carlisle
8.	Bishop Auckland Hospital, Bishop Auckland	16.	Freeman Hospital, Newcastle-upon-Tyne

Figure 3.1 North East of England with postcode district markings included. ***White area indicates pneumonia study population, grey shading indicates additional districts included in empyema study***

3.2.2 North East empyema data

Records of cases of childhood empyema managed at the tertiary referral centre (Freeman Hospital, Newcastle-upon-Tyne) between May 1995 and 2010 were used, as described in Chapter 2.2.1. The collection of empyema data was part of a wider study of paediatric empyema, the UK Enhanced Surveillance of Paediatric Pneumococcal Empyema project (UK-ESPE). All cases included had full postcode data.

3.2.3 Spatial units and study area boundaries

HES pneumonia data and empyema case data represent individual level or point data i.e. events which have occurred in or to individuals. These individuals clearly have a precise geographical location, however they have been grouped into areal data. This is as a consequence of the nature of the data. HES data contain only partial postcode data to maintain the confidentiality and to prevent identification of individuals. Furthermore, individual level spatial covariate data were not available for HES data and not collected for empyema cases.

The areal unit used in the analysis was the postcode district. Postcode districts are part of postcode classification system in the UK. The postcode classification system is widely used in research, as they are widely known and recognised within the population which makes their collection as a geo-locator straightforward. However, they have several significant drawbacks; one that they change over time, with creation of new codes and recycling of old ones causes difficulty in longitudinal studies. Secondly, they are not always contiguous with other geographical classifications which can lead to errors of attribution during conversion between different classifications. Finally, there is substantial variation in population size between different postcodes (*Mid-year population estimates*). The ONS created and used the output area classification for the 2001 census, in order to counter the problems of the postcode classification system (*HESonline - Hospital Episode Statistics*). A comparison between the two classification systems is shown in **Figure 3.2**.

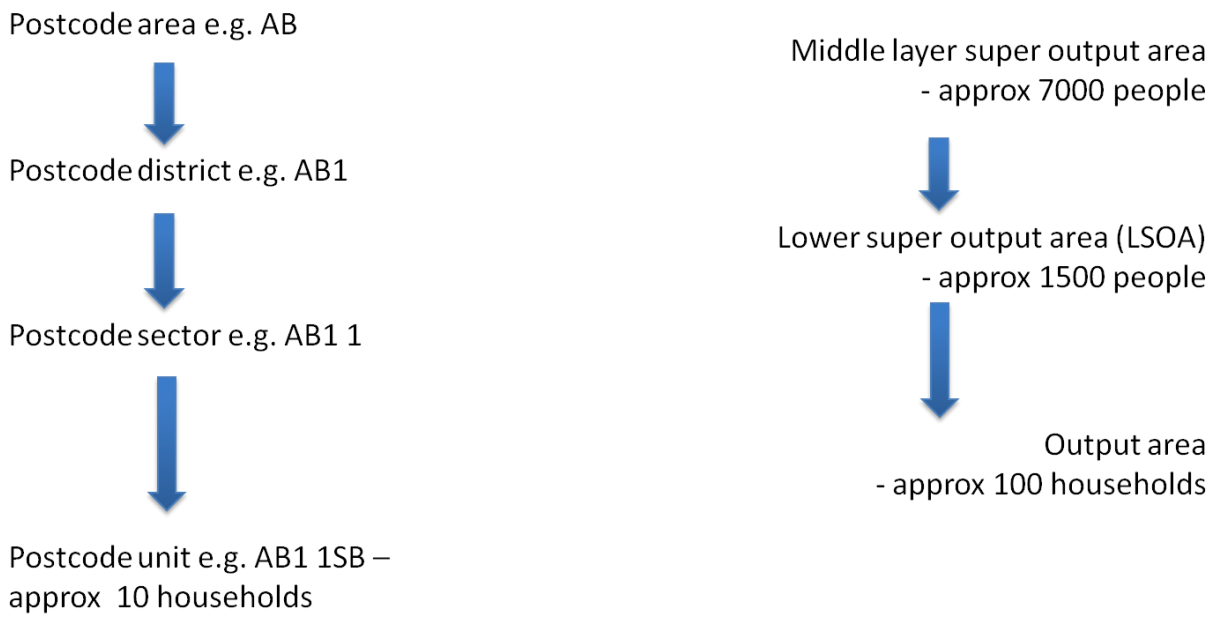


Figure 3.2 Comparing the postcode and output area classification.

Postcode district was used to protect data accuracy and because it provided the most sensible choice in the balance between the need to avoid zero-inflation (from having large numbers of areas without cases, such as would have been the case if postcode sectors were used) and the need to have sufficient contrasting areas to reflect the true spatial heterogeneity.

The two study areas were different reflecting the varying data sources. The pneumonia data represented the areas contained within the North East (SHA) area, a total of 116 postcode districts. Childhood pneumonia in the North East SHA has been previously characterised and reported (Clark *et al.*, 2007a). The empyema data came from the single supra-regional centre for the management of the condition and the study area therefore represents the catchment population for this centre, a total of 150 postcode districts comprising the whole of the North East SHA and Cumbria. This was a conscious decision to maximise the available data given the relative rarity of empyema.

3.2.4 Spatial covariate data

Spatial covariate data were obtained from several sources. Data were obtained from the Local Index of Child Well-being (CWI), a child specific small area deprivation dataset published in 2009 by the department of Communities and Local Government of the UK government (Bradshaw *et al.*, 2009). It includes a measurement for each of the 32,482 lower super output areas in England and comprises seven separate domains covering different aspects of child well-being. These, in turn are comprised of combinations of different indicator measurements and are as follows:

i) Material deprivation - children living in households in receipt of both in work and out of work means tested benefits.

ii) Health - all emergency hospital admissions, all outpatient attendances for children living in the area and children receiving disability living allowance as a proportion of all children in each area.

iii) Education - school absence rates for children living in that area, standardised test and exam scores and entry into higher education rate.

iv) Crime - police recorded crime data for burglary, theft, criminal damage and violence.

v) Housing - access to housing (as measured by overcrowding rate, proportion of shared housing and homelessness) and quality of housing (proportion of children living in accommodation lacking central heating).

vi) Children in need - expected rate of children requiring local authority social support

vii) Environment - environmental quality (air quality, percentage of green space and woodland, number of bird species and child road accident rate) and environmental access (average number of leisure and sports facilities within area and average road distance to school).

The correlation between co-variates is shown **Table 3.1**:

	Material	Education	Health	Environment	Crime	Housing	Children in need	Overall
Material	1.00	0.80	0.56	0.07	0.55	0.63	0.96	0.90
Education		1.00	0.57	0.03	0.53	0.48	0.86	0.83
Health			1.00	0.05	0.36	0.31	0.59	0.68
Environment				1.00	-0.02	0.16	0.07	0.30
Crime					1.00	0.35	0.54	0.63
Housing						1.00	0.59	0.69
Children in need							1.00	0.91
Overall well-being								1.00

Table 3.1 Spearman rank correlation matrix between the domains and overall well-being (all coefficients are statistically significant at the <0.01 level). Adapted from Bradshaw J, Bloor K, Huby M, et al. Local index of child well-being: Summary report: Department of Communities and Local Government, 2009.

Measurements of ethnicity (proportion of non-Caucasian individuals within the district) and migration (all inward migrants, outward migrants and those moving within the district in the year prior to the census) were derived from the UK Census 2001 key statistics dataset at the level of super output area (*Mid-year population estimates*). All spatial covariate data were then matched to counts of cases in each study postcode district. Finally, the distance to each of the admitting hospitals from each postcode district was estimated using the centroids of each district and the centroid of the district containing each hospital. The shortest possible distance for each district was then used and for cases residing in the same district as a hospital this was zero.

3.2.5 Statistical methodology

Relative risks were calculated by using the conditional autoregressive models developed by Besag, York & Mollie (Besag *et al.*, 1991). These were Poisson spatial models with the observed number of cases as the dependent variable, expected number of cases as offset and random effects terms that take the following into account: (a) effects that vary in a structured manner in space (postcode district contiguity) and (b) a component that models the effects that vary among census tracts in an unstructured manner (postcode district heterogeneity). The first of these represents an attempt to model unmeasured spatial dependency associated with the proximity of residential postcodes, whilst the second allows for significant differences between postcodes over the study area.

It was assumed that the numbers of observed cases (O_i) for each postcode district ($i=1,\ldots,n$) follow a Poisson distribution with mean $\mu_i=E_iO_i$, where E_i are the expected cases for each postcode district obtained by indirect standardisation, and θ_i is the relative risk (RR) for each specific area. The expected numbers of cases were calculated with the age-specific population (0-14 years old ONS census 2001 data) of the study region within the study period.

The null model took the following form:

$$O_i \sim Po(E_i\ \mu_i)$$

$$\log(\mu_i) <- \log(E_i) + \alpha + b_i + h_i$$

Where O_i is the observed number of cases in area i; E_i are the expected cases; μi is the relative risk in area i, α is the intercept; h_i is the postcode district heterogeneity term; and b_i is the spatial term. Initially, the null model included no explanatory variable. Subsequent iterations were fitted to include a covariate effect, using the following general model:

$$O_i \sim Po(E_i \mu_i)$$

$$\log(\mu_i) <- \log(E_i) + \alpha + \beta cov_i + b_i + h_i$$

Where cov_i is the covariate value in area i. Models were fitted with a variety of covariates using this general model structure. The null model was compared with models from three groups containing covariates for a) Overall CWI, ethnicity, migration and distance from admitting unit separately b) The seven individual CWI domains that make up the CWI Index, c) all possible combinations of those covariates from a) and b) that were shown to be significant at 95% Bayesian Credibility Intervals (BCI).

The models were fitted using Bayesian Markov chain Monte Carlo (MC) simulation methods within WinBUGS 1.14.(Spiegelhalter DJ *et al.*, 2001) A 10,000 iteration 'burn-in' was followed by a 50,000 iteration sample. In all cases, the MC error for each area was less than 5% of the standard deviation, indicating sufficient iterations of the model had been run after convergence. Model assessment involved the comparison of DIC (Deviance Information Criterion) scores for each model, with the model with the lower DIC score preferred. A difference in DIC greater than 3 indicates a significant difference in model performance (Spiegelhalter *et al.*, 2002). Following the principle of parsimony, the best performing model was judged to be the one with the lowest DIC which contained only significant variables.

Areas were considered significant areas where the 95% BCI around the estimated relative risk excluded 1. A covariate was considered significant if 97.5% of its posterior distribution lay either above or below zero. To depict the spatial distribution of relative risk across the study area postcode district maps obtained from UK Borders (EDINA) were used alongside R 2.12.2 (Team, 2010) and WinBUGS (Spiegelhalter DJ *et al.*, 2001).

3.3 Results

3.3.1 Spatial variation in the risk of pneumonia

There were 3874 pneumonia admissions with postcode data during the study period. The spatial distribution of these admissions is presented at the postcode district level in **Figure 3.3**. The most admissions were found in TS3 (Brambles Farm, Thorntree, Park End, North Ormesby, Berwick Hills, 128 cases) and NE4 (Fenham, Westgate & Wingrove, 106 cases).

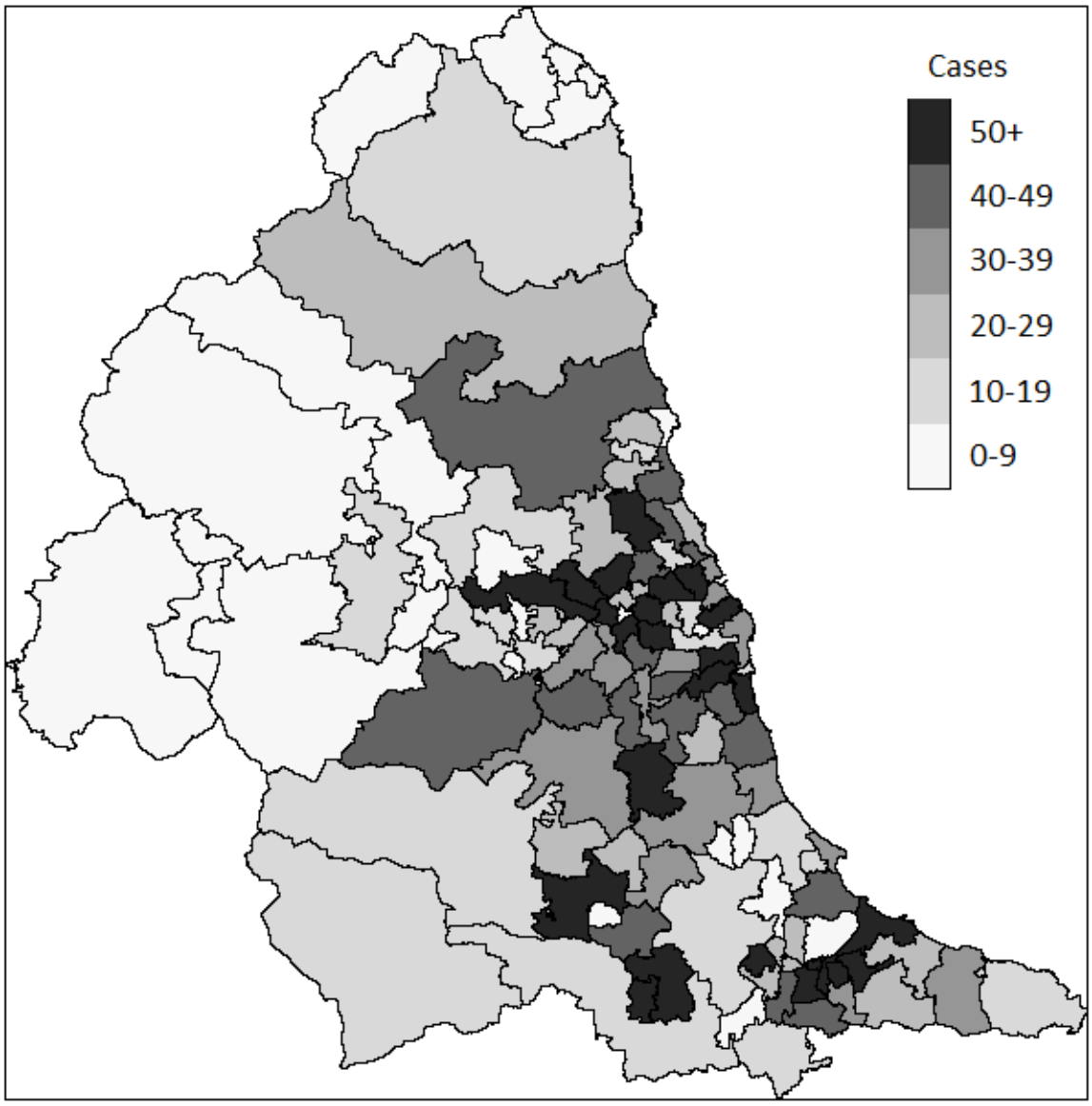


Figure 3.3 The number of observed cases of pneumonia across the postcode districts of North East England.

The expected incidence of pneumonia for each postcode district is presented in **Figure 3.4**.

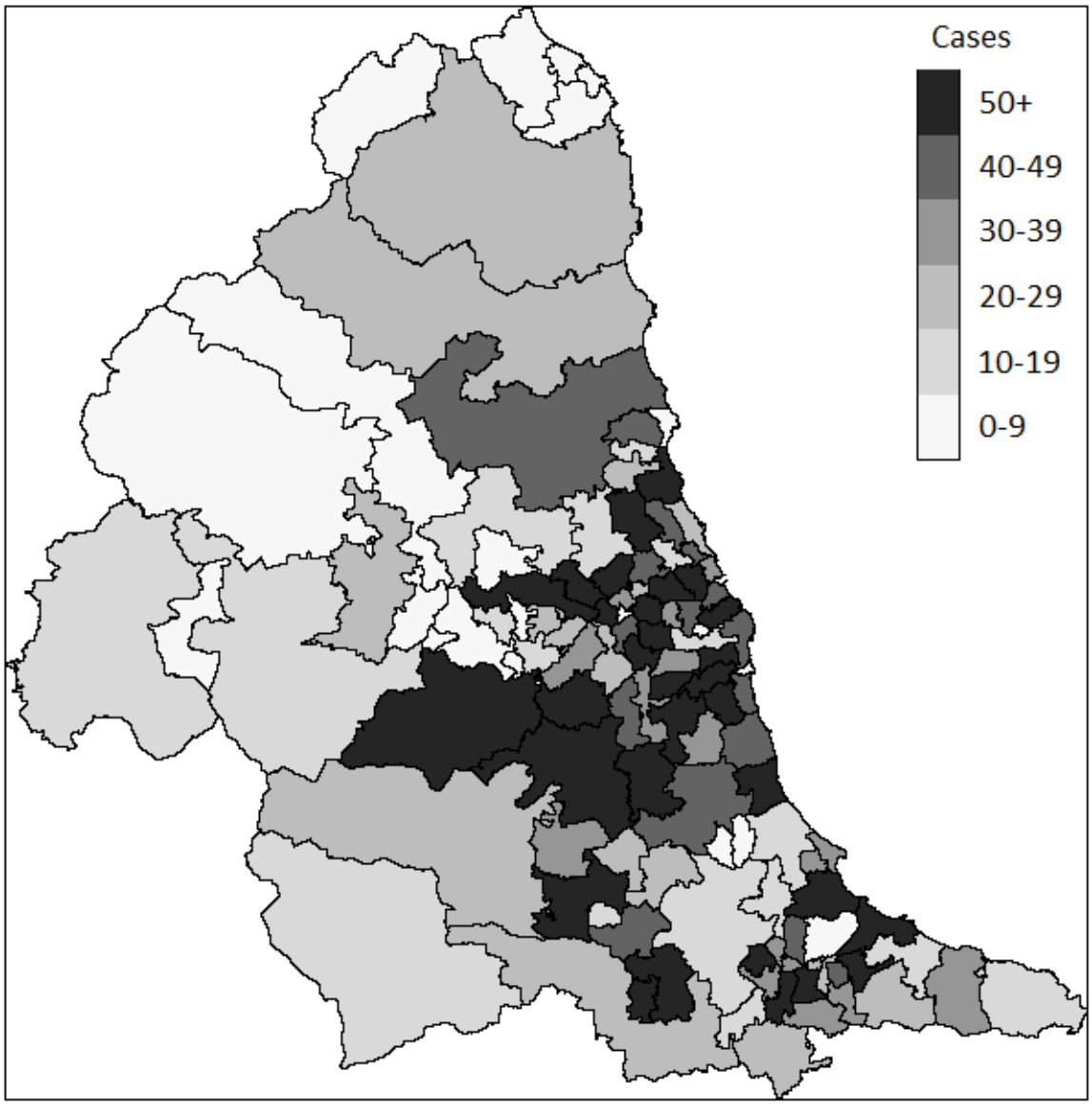


Figure 3.4 The number of expected cases of pneumonia across the postcode districts of North East England.

Figure 3.5 presents the spatial distribution of those postcode districts that, according to the null model, were shown to have either a significantly low (shaded green) or high (shaded red) relative risk of childhood pneumonia.

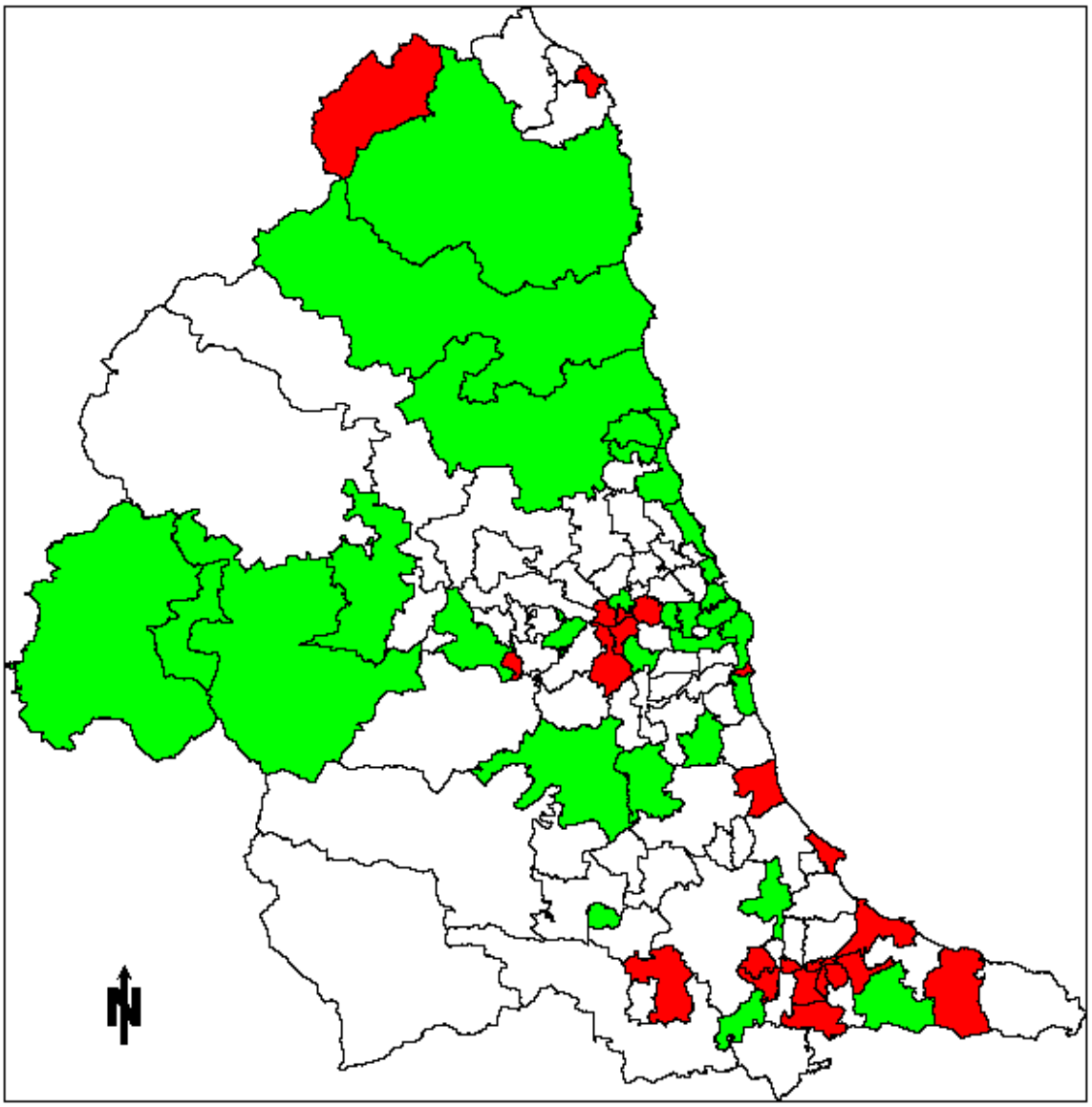


Figure 3.5 The location of postcode districts with significantly high (red) or low (green) relative risk of pneumonia, according to the null model.

Significant spatial variation in risk of admission to hospital with pneumonia was found, with 53 (47%) of districts having either a significantly higher or lower risk compared to that predicted by child population alone (Range of RR: 0.32-2.34). Clusters of areas of highest risk were in the urban centers of major cities of Newcastle and Middlesbrough. Large rural districts to the North and West of the region had the lowest risk.

Model Group	No. Covariates	Covariates	Minimum RR	Maximum RR	No. significant low RR	No. significant high RR	Model DIC
Null	0		0.319	2.338	31	22	749.68
A	1	CWI	0.459	3.017	32	14	746.01
B	7	CWI Individual Scores	0.500	3.086	31	13	742.28
C	3	CWI Significant Scores (Material deprivation, Health, Children in Need)	0.437	2.915	37	13	741.50
A	1	Material deprivation	0.413	2.993	27	13	749.35
A	1	Children in Need	0.404	2.990	28	12	750.88
A	1	Health	0.406	2.898	36	16	743.93
A	1	Distance to hospital	0.279	2.938	29	12	746.34
C	4	Distance to hospital, Health, Material deprivation and Children in Need	0.379	2.917	37	13	740.94
Best performing model							
C	3	Material deprivation, Health, Children in Need	0.437	2.915	37	13	741.50

Abbreviations – CWI = Child Well-being Index, RR = Relative Risk, DIC = Deviance Information Criterion

Table 3.2 Summary statistics for models of relative risk of hospitalisation with pneumonia.

Variable	Median (Range)	Estimate of mean of posterior distribution	SD of posterior distribution
CWI	158.93 (47.82 ; 393.21)	0.0027	0.00045
Distance to hospital (km)	6.83 (0 ; 70.89)	-0.019	0.0057
Material deprivation	0.21 (0.04 ; 0.63)	3.92	1.14
Children in Need	0.03 (0.01 ; 0.09)	-26.74	9.1
Health	0.47 (-1.103 ; 1.62)	0.31	0.092

Table 3.3 Significant spatial covariates for pneumonia ***Abbreviations – CWI = Child Well-being Index, SD = Standard deviation***

Table 3.2 presents summaries of models of the spatial distribution of pneumonia cases. **Table 3.3** contains details of significant spatial covariates. Relative childhood deprivation, as measured by the Local Index of Child Well-being, was associated with a significant increased risk of admission in a district. Greater material deprivation, higher health domain score (indicating higher rates of admissions, clinic attendances and disability in an area) and higher Children in Need score (indicating a greater number of children requiring local authority support) components of the CWI were significant predictors of increased risk.

Distance to hospital was a significant covariate when added to the null model, with increased admissions in districts closest to hospitals. This was a non-significant factor when added to a model containing the significant CWI covariates and did not improve the DIC sufficiently to warrant retaining (DIC 740.94 vs. 741.50). Ethnicity, migration levels, education, housing, crime and the environment domain were all non-significant.

The best performing model contained the three significant CWI covariates alone (DIC 741.50) although the three models containing i) all seven separate components of the CWI, ii) health alone and iii) distance to hospital, health, material deprivation and Children in Need, were not significantly inferior by comparison of DIC (DIC 742.48, 743.93 and 740.94). In the best performing model 50 (44%) of the districts had a significantly increased (n=13) or decreased (n=37) risk (range of RR was

0.44 to 2.92) and are shown in **Figure 3.6**.

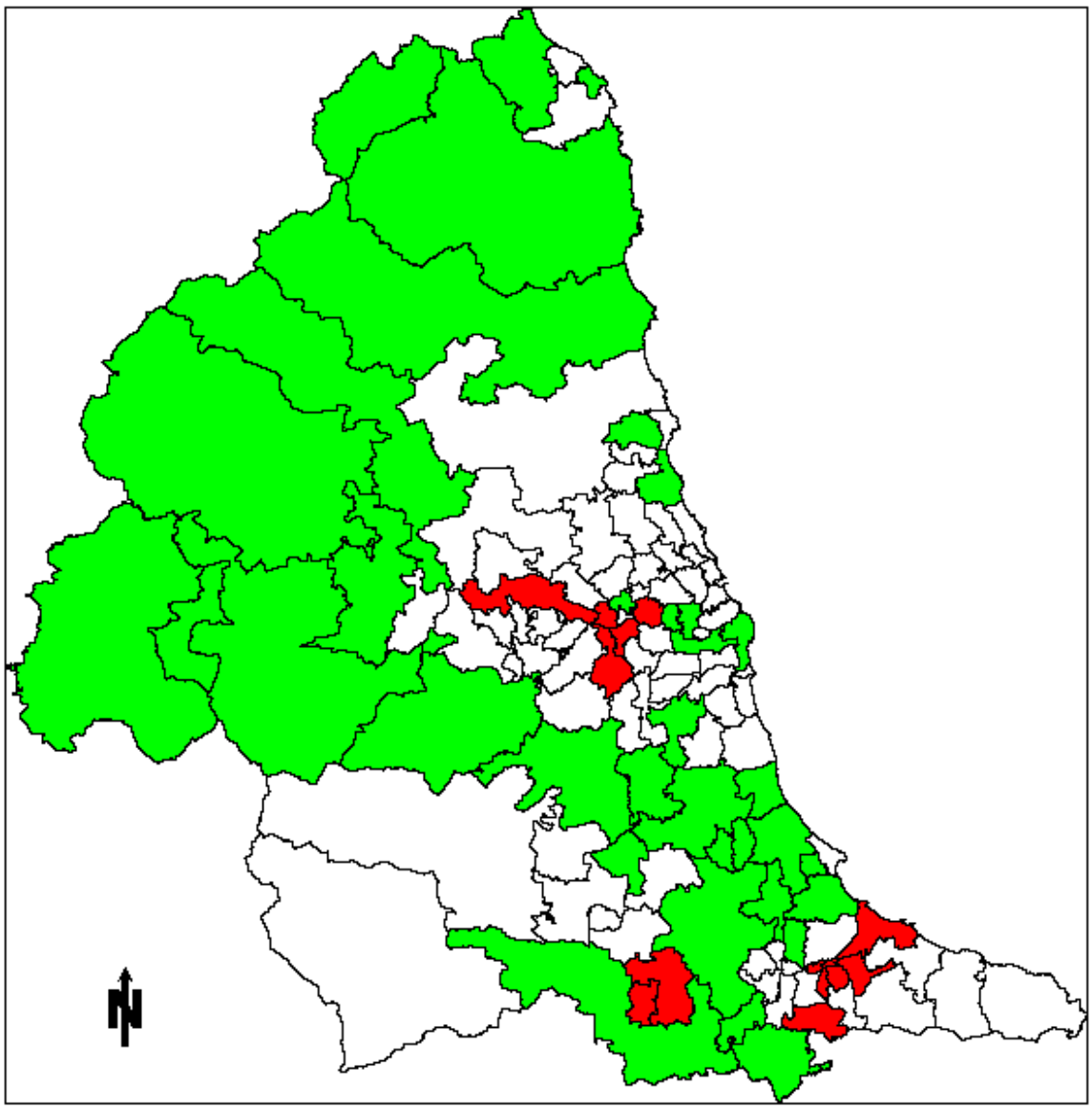


Figure 3.6 The location of postcode districts with significantly high (red) or low (green) relative risk of pneumonia, according to the best performing model.

3.3.2 Spatial variation in the risk of empyema

Figure 3.7 shows the distribution of cases of empyema across the study region. In total, there were 293 cases spread across 150 postcode districts. The most cases in any district were found in DH8 (Consett) and NE6 (Walker, Byker and Heaton), where eight instances were recorded in each. Forty postcode districts contained no reported cases of empyema.

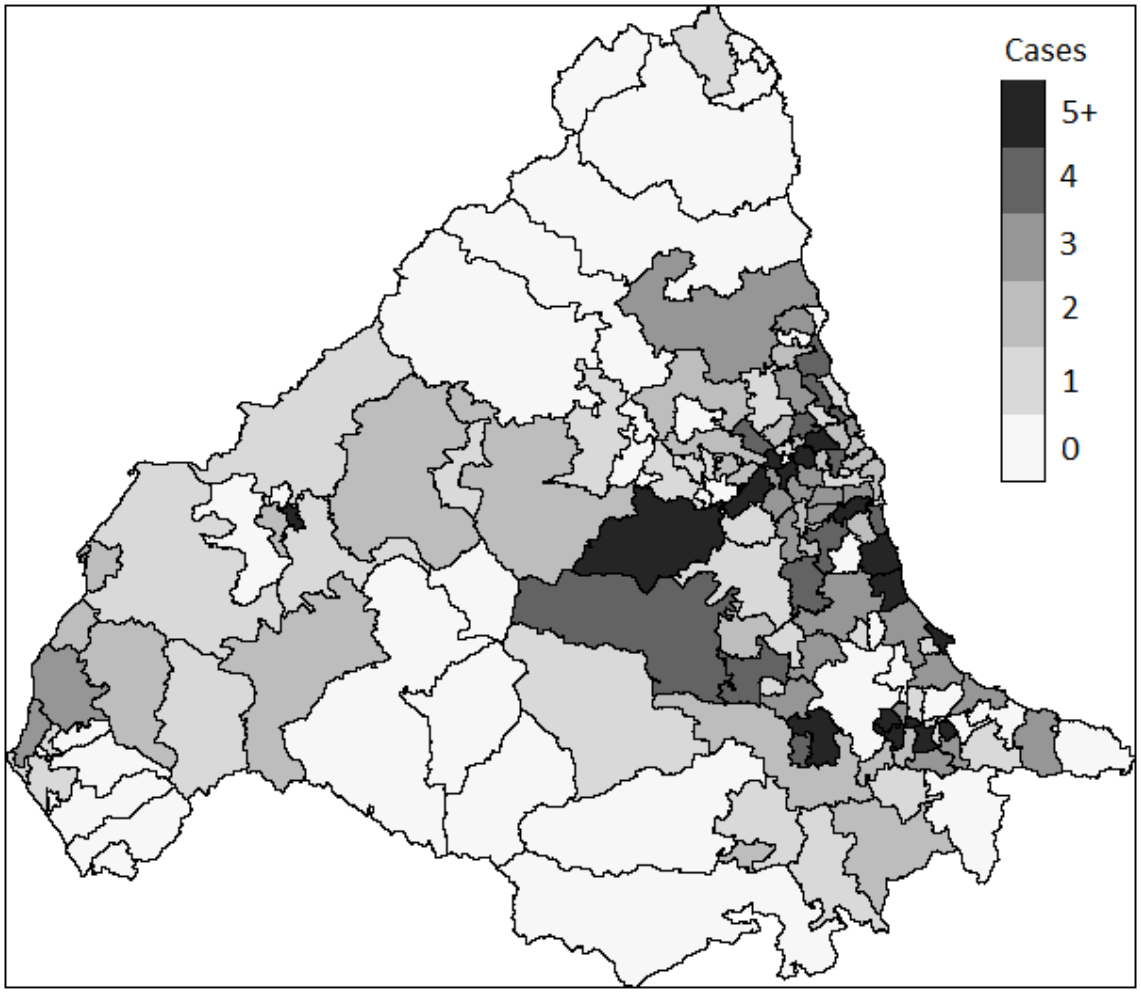


Figure 3.7 The number of observed cases of empyema across the postcode districts of Northern England.

The expected incidence of empyema for each postcode district is presented in **Figure 3.8.**

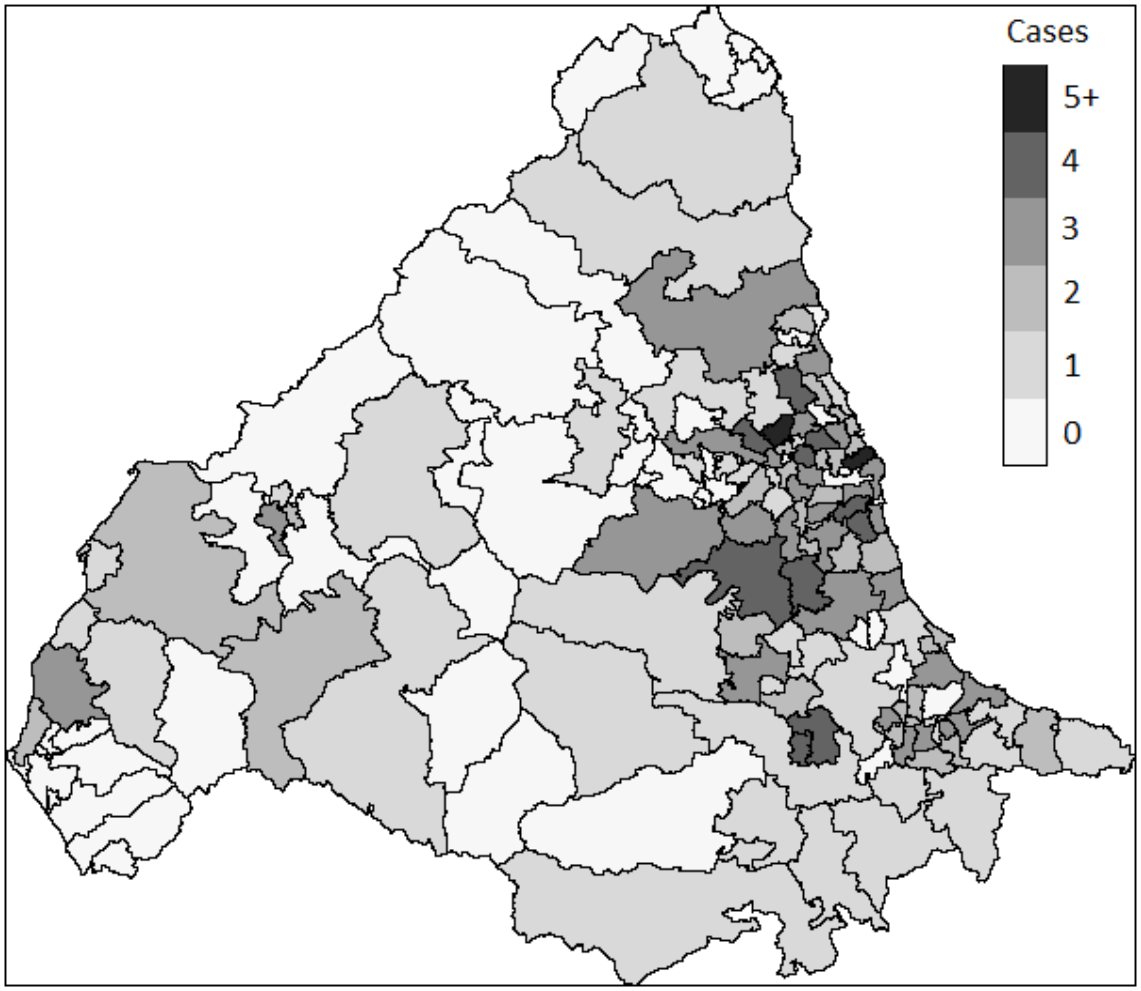


Figure 3.8 The number of expected cases of empyema, according to relative childhood populations across postcode districts in Northern England.

The low numbers of observed and expected cases resulted in no postcode districts showing significantly high or low relative risks. This remained the case when the credible interval was lowered to 90% and also when the spatial boundaries were increased to the coarser scale of postcode area. An analysis using expected cases calculated from the rate of bacterial or lobar pneumonia in each district was also carried out. This identified four areas of significantly lower risk. However, there was no evidence of clustering and the expected counts of disease in each were low (<5 cases). Consequently, this finding does not represent reliable evidence of spatial variation in risk of empyema.

3.4 Discussion

3.4.1 Strengths and limitations

The strengths of these analyses include the use of child specific deprivation measurements not been previously available for this type of study in the population in question, the use of a well-validated and reported statistical methodology robust in the handling of spatial auto-correlation, and a comprehensive geographical dataset.

As this is an ecological study there are limitations regarding causation, due to the use of aggregated rather than individual level data. The use of hospital coding data to determine cases is a further limitation. Coding data has been used to examine childhood pneumonia in a number of other studies, but concerns have been raised about its specificity.(Elemraid *et al.*, 2010; Koshy *et al.*, 2010).Although the study relates to a large and varied region, it only relates to one area of the United Kingdom and factors unique to the study area will inevitably influence the generalisability of the findings.

3.4.2 Findings related to other studies

There was significant spatial variation in the risk of hospitalisation of children with pneumonia between postcode districts in the North of England. The range of relative risks was 0.44 to 2.92 indicating an approximately six fold variation between districts. Significant covariates in explaining the variation were material deprivation, CWI health domain score, number of children requiring local authority support and distance to hospital. Ethnicity, migration levels, housing, education, crime and environment were not significant covariates. There was no evidence of spatial variation in the risk of hospitalisation with empyema.

The existence of significant spatial variation in risk of hospital admission for pneumonia is in keeping with the findings of other spatial studies of risk of pneumonia (Crighton *et al.*, 2007; Kazembe and Namangale, 2007). Direct comparisons of the magnitude of the effect are difficult to make, because of differences in statistical methodology. The variation in pneumonia and influenza admission rates between counties described by Crighton *et al* (2007) was three-fold, which is lower than this finding of a six-fold variation (0.44 to 2.92). This may be due to the smaller geographical units used in our study, which are likely to provide heterogeneity between individual areas.

A number of studies have reported an association between deprivation and increased risk of hospitalisation for respiratory tract infections in children and this study supports these previous findings (Spencer *et al.*, 1996; Hawker *et al.*, 2003; Crighton *et al.*, 2007). Defining the mechanisms underpinning this increased risk has been controversial (Crighton *et al.*, 2007). Proposed hypotheses have included, increased pathogen transmission from overcrowding and increased host susceptibility resulting from poor nutrition (Cohen, 1999; Hawker *et al.*, 2003). Absolute material deprivation is rare in the UK child population and primary health care access is free at the point of use, suggesting that other factors linked to relative deprivation are likely to be relevant (Wilkinson, 1997). The most obvious is exposure to exhaled tobacco smoke through living with adult smokers. Smoking rates are correlated with socio-economic status and have been linked to increasing hospitalisation for childhood respiratory infections (Colley *et al.*, 1974; Spencer *et al.*, 1996; Cook and Strachan, 1999). It is not possible to separate the contribution of smoking from that of material deprivation within this dataset. Given the magnitude of disparity in risk between different areas, public health interventions emphasising the benefits of reducing children's exposure to second hand smoke should be targeted at these areas to reduce these inequalities.

The health domain was also a significant explanatory variable for the variation in risk between districts. It is comprised of three indicators, total child hospital admissions, total child hospital out-patient attendances and the proportion of children receiving disability living allowance. Factors within a community that modify overall admissions to hospital will therefore contribute to both the number of pneumonia admissions in a district and the health domain score. Differences in access to healthcare resources have been reported as a significant factor in the variation in childhood hospital admissions in other settings (Erzen *et al.*, 1997; Crighton *et al.*, 2007). The significance of distance to hospital as a covariate would suggest this in our population. Admission policies for children with pneumonia at individual hospitals vary, contributing to the variation in risk. These disparities are insufficient to fully explain the magnitude of variation between areas, as national guidelines for the management of childhood pneumonia were widely implemented during the study period. Children with co-morbidities such as cerebral palsy predisposing to pneumonia are more likely to be in receipt of disability living allowance, and this will also contribute to the association with the health domain (Owayed *et al.*, 2000).

The Children in Need domain was also a significant predictor, but the association was counter-intuitive with higher levels of risk in districts with lower numbers of children requiring local authority

social support. It is recognised that children in local authority care suffer from worse long-term health outcomes (Polnay and Ward, 2000). However, there were difficulties with missing data and this domain was partially estimated using a combination of income and education scores which are highly correlated with material deprivation (correlation co-efficient 0.96) (Bradshaw *et al.*, 2009). It is therefore not easy to explain the difference in polarity between the two observed associations. One possibility is that the effect of material deprivation is to some degree ameliorated by children being in local authority care. Most likely however is that the problem of data derivation means this is an epiphenomenon.

Housing characteristics were not significant determinants of risk at a community level, as measured by either overcrowding or housing quality. This was surprising as individual level studies have identified both as risk factors for pneumonia (McCarthy *et al.*, 1985; Mann *et al.*, 1992). This finding may be a consequence of the community level aggregation, although it is also possible that these levels of inadequate housing were simply not of significant magnitude to have adverse consequences in the population studied.

No variation in the relative risk of hospitalisation was found for empyema, even when the spatial scale was increased. This could reflect the true absence of spatial variation in risk. It is also possible that it is a consequence of the low number of cases per district. Conditional autoregressive models have been shown to be conservative, and they may not identify significant variation in risk if the variation between areas is less than two-fold and the expected number of cases less than 50 (Richardson S *et al.*, 2004).

3.4.3 Implications and future research

There was a six-fold variation in risk of admission to hospital for children with pneumonia in different postcode districts in the North East of England between 1997 and 2007. These variations appear to be associated with deprivation, specifically levels of material deprivation, childhood illness and disability and proximity to hospital. The relationships between hospital admissions determined by coding data and the true incidence of disease are complex and many factors such as coding accuracy, physician behaviour, and access to secondary care may be relevant (Campbell *et al.*, 2001). The magnitude of these variations suggests that these findings indicate true differences in the incidence of this disease. This has important public health implications, particularly given the increasing evidence of significant

associations between childhood pneumonia and morbidity in later life (Sethi and Murphy, 2001).

There was no evidence of spatial variation in risk of empyema. This is unexpected given the close relationship between pneumonia and empyema. It may represent a true negative but given the relatively low number of cases per postcode district and the conservatism of CAR models in assessing risk in rarer conditions, it is more likely to represent a failure to detect variation rather than true absence of variation in risk.

3.5 Conclusions

There is significant spatial variation in the risk of hospital admission with pneumonia but not empyema, in the North of England above that determined by population alone. Levels of material deprivation, childhood illness and disability help explain these variations but significant unexplained variation remains. Further studies are required to quantify to what extent these spatial differences in risk represent differences in health behaviour as opposed to a genuine increase in the risk of disease. Research directed at understanding the role of public health measures at a community level, such as smoking cessation campaigns in reducing hospital admissions for childhood pneumonia could also help to elucidate which factors are truly relevant in determining variations in risk of pneumonia.

4 The impact of the seven valent pneumococcal conjugate vaccine on paediatric empyema in the North East of England

Note: This chapter is a manuscript that was submitted to the European Respiratory Journal in 2011.

4.1 Introduction

The incidence of empyema thoracis in children has risen in Europe, Australia and North America over the last 15 years (Rees *et al.*, 1997; Byington *et al.*, 2002b; Strachan and Jaffé, 2009; Van Ackere *et al.*, 2009). In England, admissions for empyema rose from 14 per million children (0–14 years) in 1995 to 26 per million in 2003 (Gupta and Crowley, 2006). Similarly, in Scotland admissions increased from <10 per million in 1998 to 37 per million in 2005 (Roxburgh *et al.*, 2008).

While a range of pathogens are known to cause empyema, paediatric empyema in the United Kingdom (UK) is predominantly a pneumococcal disease (Eastham *et al.*, 2004). This is in contrast to adult disease where the most common cause of culture positive disease is streptococci of the anginosus group (Maskell *et al.*, 2005). Pneumococci display antigenic diversity in their polysaccharide capsules and over 90 different serotypes or serogroups are recognised (Lipsitch, 1999). Pneumococcal conjugate vaccines comprise multiple pneumococcal capsular polysaccharides covalently coupled to molecules of carrier protein, but these vaccines only contain antigen derived from a limited number of serotypes (Lipsitch, 1999).

PCV-7 was added to the UK routine immunisation schedule in September 2006 and a replacement 13-valent pneumococcal conjugate vaccine was introduced in April 2010. The current schedule includes doses at two, four and 12 months. A catch-up programme for children under two years of age was undertaken when PCV-7 was introduced.

PCV-7 contains antigen from serotypes 4, 6B, 9V, 14, 18C, 19F and 23F accounting for approximately 65% of culture-positive paediatric invasive pneumococcal disease according to national surveillance prior to the introduction of PCV-7 (Miller *et al.*, 2000). These serotypes previously accounted for only 20% of culture-positive and negative paediatric empyema cases, suggesting that any impact of the immunisation on empyema may be modest (Eastham *et al.*, 2004; Fletcher *et al.*, 2006).

Studies investigating the incidence of paediatric empyema after the introduction of PCV-7 have produced conflicting results. One study, using hospital coding data, reported a drop in admissions in England in 2008 from a peak in 2006 which was attributed to PCV-7 (Koshy *et al.*, 2010). This finding was in marked contrast to three large US studies of hospital coding data using robust methodologies reporting an *increas*e in paediatric empyema following the introduction of PCV-7 (Grijalva *et al.*, 2010; Lee *et al.*, 2010; Li and Tancredi, 2010).

4.1.1 Aims and objectives

The objective of this study was to investigate whether the introduction of PCV-7 was associated with a change in the incidence of paediatric empyema. An interrupted time-series analysis approach was used, which has been successfully used in the past to assess the impact of PCV-7 on childhood pneumonia, as well as other health outcomes (Wagner *et al.*, 2002; Grijalva *et al.*, 2007). A secondary objective was to test the hypothesis that any vaccine impact will be primarily focused towards the age group who are direct vaccine recipients as opposed to those benefitting from the consequences of enhanced herd immunity.

4.2 Methods

The case data used were those from the North East dataset described in chapter 2.2.1. PCV-7 immunisation rates were obtained from the HPA Immunisation department and covered each quarter from July 2007 until June 2010 (Immunisation Department). The transition period for the introduction of the vaccine was considered to be from September 2006 until July 2007 when uptake of two vaccine doses by 12 months of age was greater than 90%.

Data on the individual vaccination status of cases were not available. All children born after 5 September 2004 were possible vaccine recipients, either through the primary or catch-up vaccine programmes. The pattern of vaccine uptake was progressive, thus in any one year after introduction in 2006 an increasing proportion of the at risk population will have been vaccinated. The proportion of empyema cases in this cohort of possible vaccine recipients was calculated for each year after July 2007 and compared to the numbers of cases in the equivalent cohort for all years from 1995–2006.

4.2.1 Seasonality

Empyema incidence is strongly seasonal, with most cases occurring in winter. As monthly count data were used to increase the robustness of the time series analysis, a variable to account for the seasonal variation was required and air temperature was used in same manner as in Chapter 2.2.1. Data were obtained from MIDAS as previously described.

4.2.2 Statistical methods

Rates of cases per month per head of population were calculated using the UK ONS mid-year population estimates for the relevant health authority areas (North East Strategic Health Authority area (NE-SHA) and Cumbria Teaching Primary Care Trust (CT-PCT)).The rate of empyema cases per month per million individuals aged 0-14 years was used as the outcome measure. Interrupted time-series regression analysis was used to investigate the relationships between this rate and the introduction of PCV-7. All models were fitted using a least squares regression approach.

The outcome data were log-transformed to stabilise the variance. Analysis of the residuals of the final models did not demonstrate any violation of distributional assumptions. Autocorrelation was quantified against all lags up to 12 months and no evidence of primary or secondary autocorrelation was observed. Since the variance in case rate increased with time from study start, a weighting function was used to counterbalance the effect of heteroscadascity and prevent violation of the model assumptions (Faraway, 2005). The power of the study to detect a 20% change in the rate of cases based on a significance level of 0.05, with 174 months of data and in the absence of significant autocorrelation was 0.996 (Cohen, 1998).

Pearson's chi-squared test was used to assess the differences between age distributions. The R statistical package version 2.9.1 (The R Foundation for Statistical Computing, 2009) was used for all analyses.

4.3 Results

A total of 298 cases were included in the analysis, over a time period of 179 months from May 1995 to April 2010. Fifty-six percent were males and the median age at diagnosis was 4.6 years (range: 0–14). Separate analyses were performed to either include or exclude the transition period. As no significant differences were found between the two estimates the reported results exclude the transition period.

The incidence of empyema increased significantly from a mean monthly rate of 1.1 cases per million children in 1996 to 5.2 per million in 2009 ($p<0.0001$). The introduction of PCV-7 was not associated with a decrease in cases; instead both the incidence and the rate of rise in incidence appeared to *increase* following the introduction of PCV-7 (Figure 1). By March 2010, predicted cases of empyema per month per million children were 17 % higher than the level predicted by extrapolation of the linear trend prior to introduction of the PCV-7 (**Figure 4.1**). This was equivalent to an additional 2.07 cases per million children per year after the introduction of PCV-7, although this effect was not significant (95 % confidence interval -3.38, 10.98, $p=0.53$). Furthermore, the variable of time from vaccine introduction was not significant suggesting that despite the increasing number of individuals in the population who had received the vaccine there was no significant change in incidence ($p=0.25$).

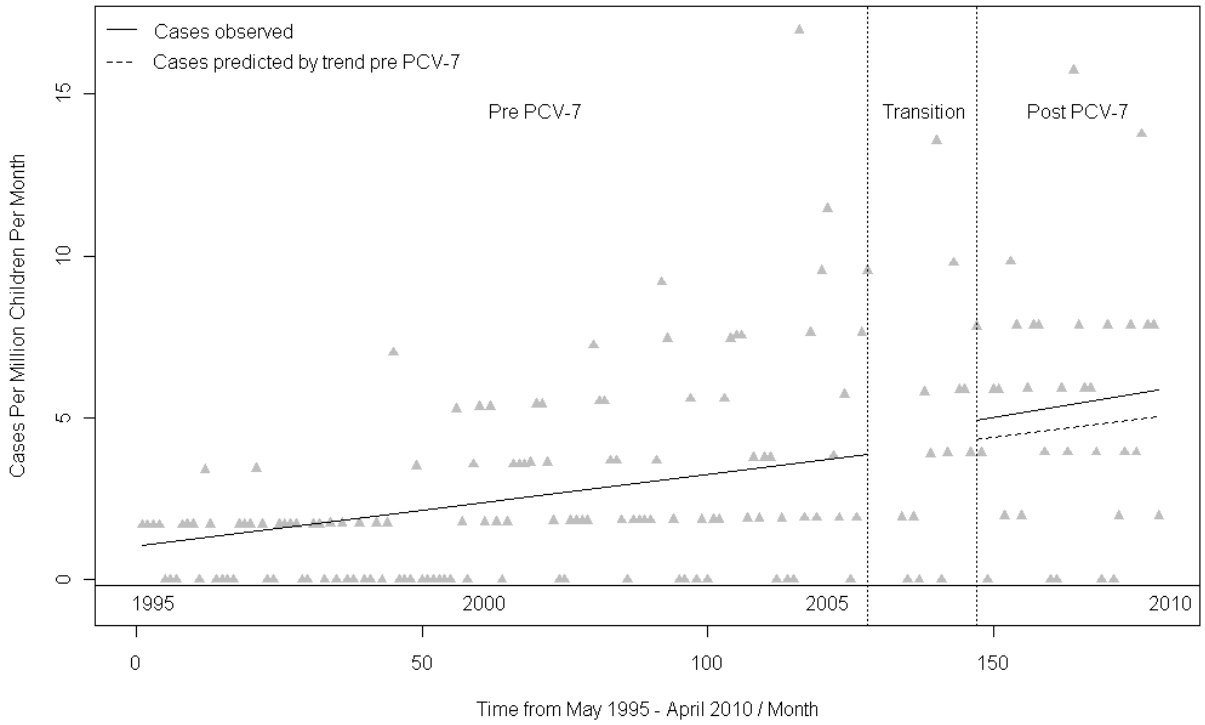


Figure 4.1 Trends in the incidence of paediatric empyema in the North of England from 1995 to 2010, pre and post the introduction of PCV-7.

The results of the regression models are summarised in **Table 4.1** and **Table 4.2**.

	Co-efficient†	**95 % Confidence Intervals**	**p-value**
Mean Monthly Minimum Temperature / ^{o}C	-0.0042	-0.086 ; 0.0052	0.083
Month From Study Start	0.010	0.0061 ; 0.014	<0.0001
Vaccine programme present	0.17	-0.28 ; 0.91	0.53
Month From Vaccine Programme Start	-0.010	-0.027 ; 0.0071	0.25
Interaction Term (Month From Study Start * Mean Monthly Minimum Temperature)	-0.00049	-0.00098 ; 0.00001	0.046
Akaike Information Criteria (AIC)	390.73		

Table 4.1 Regression model predicting monthly cases of empyema using time, temperature and introduction of the PCV-7 vaccine (n= 174).

† - All co-efficients represent change in cases per million individuals per month for a unit change in the co-variate.

	Co-efficient[†]	**95 % Confidence Intervals**	**p-value**
Interaction Term (Month From Start * Mean Monthly Minimum Temperature)	-0.00083	-0.0011 ; -0.00056	<0.0001
Month From Study Start	0.011	0.0088 ; 0.0013	<0.0001
Akaike Information Criteria (AIC)	374.66		

Table 4.2 Final parsimonious regression model containing only significant variables in predicting monthly cases of empyema (n=174, R^2 0.402).

† - All co-efficients represent change in cases per million individuals per month for a unit change in the co-variate

There was strong evidence of an interaction between temperature and time on the rate of cases, suggesting the need for an interaction term between the two variables. The included term was significant ($p<0.0001$) which equates to the observation that the number of cases increased more in association with low temperatures throughout the sampling period than in summer. Also, on inclusion of the interaction term, measures of the ability of the model to explain the observed data, e.g. the Akaike Information Criteria (AIC) significantly improved (AIC -84.3 without vs. -87.83 with).

The results of Pearson's chi-squared testing for difference in age distribution are shown in Table 3. There were no significant differences in the age distribution of cases in the three years after vaccine that were tested.

Year	Cases in vaccine cohort†	Cases outside vaccine cohort	X^2	Degrees of freedom	p – value
1995-05	67	105	0.013	1	0.91
Jul 2007- Jun 08	13	21			
1995-05	88	84	2.1	1	0.15
Jul 2008 - Jun 09	13	18			
1995-05	105	67	0.70	1	0.40
Jul 2009 - Apr 10	18	7			

Table 4.3 Chi-squared testing for differences in age distribution before and after the introduction of PCV-7.

† Corresponds to:

Jul 2007- Jun 08: Cases aged 0 – 3 years 10 months

Jul 2008 - Jun 09: Cases aged 0 – 4 years 10 months

Jul 2009 - Apr 10: Cases aged 0 – 5 years 10 months

4.4 Discussion

Cases of empyema increased following the introduction of the PCV-7 in September 2006, however this change was non-significant and no change in age cohort of cases was seen.

4.4.1 Strengths and limitations

There are three main strengths of this study. Firstly, the length of the time period reported is sufficient to allow accurate extrapolation of secular trends and judgement of an intervention whose efficacy might change with time. Secondly, the study has sufficient power to detect an effect of the magnitude previously described (Koshy *et al.*, 2010). Finally, the analysis is based on a stable population with a high rate of vaccine uptake, by July 2007 91.3% of children had received two doses of PCV-7 by their first birthday (Immunisation Department).

This was an ecological study addressing changes in total cases of empyema, therefore specific causation of the observed changes cannot be attributed directly or solely to the effects of PCV-7. This is a limitation of all studies using this methodology and while it is important to recognise this, the observation that the introduction of PCV-7 was not associated with a decrease in incidence of empyema remains valid.

The method of data collection switched from retrospective to prospective identification in 2008. This is significant in a time series analysis as case ascertainment may change as a result resulting in false associations. However, when included as a variable within the time series analysis, this factor was not associated with a change in incidence (co-efficient -0.322, 95% CI -1.16; 5.21, p=0.45) and the use of a consistent case definition should have limited the impact of this change.

A further potential limitation was that the second period of data collection required informed consent from participants, but only one case over the whole study period refused this and was therefore excluded. A final limitation stems from the use of a procedure based case definition and a study period of 15 years. There have been no overt changes in empirical therapy, referral threshold, imaging usage and threshold for invasive management at the study centre over the study period. However, we recognise that subtle changes in clinical practice may remain unrecognised and could potentially influence time series data over such a period of time. On balance, as the increase in incidence was of such magnitude and was sustained, it is unlikely that these factors significantly influenced the results.

4.4.2 Findings related to other studies

The incidence of empyema in children in the North East of England increased from 13.6 to 62.9 cases per million from 1996 to 2009. This finding of a large increase in cases from 1995 onwards is in line with other UK based studies (Gupta and Crowley, 2006; Roxburgh *et al.*, 2008; Koshy *et al.*, 2010). This equates to a change in the mean monthly rate from 1.1 to 5.2 cases per million children over the same time period. These data are similar to previously reported figures from Scotland (Roxburgh *et al.*, 2008) and England, (Gupta and Crowley, 2006) but this estimate is somewhat larger than that of Koshy *et al* (2010) who reported a change from 4.7 admissions per million children in England to a peak of 17.5 per million and a subsequent fall to 13.7 per million by 2008. Koshy *et al* acknowledged that their estimates of empyema incidence were lower than those of other studies and cited their usage of more specific diagnostic codes as a potential explanation for the difference. They also estimated a 22% reduction in empyema between the calendar years 2006 and 2008 which they relate to the introduction of PCV-7, whereas we find no such change despite our study having sufficient power to detect such a magnitude of effect. One possible explanation for these differences is that estimates based on single paired calendar years may be distorted by the distribution of empyema incidence during individual winter seasons, and by random fluctuations in the annual totals for the paired years. An alternative explanation would be that local factors might have been operating to substantially increase the incidence of empyema in our region, although there is no evidence to suggest this.

The observation of an increase in empyema despite the introduction of PCV-7 is in keeping with a number of studies in the USA where it was introduced into the routine immunisation schedule in 2000-01. For example, in Utah, the overall incidence of cases continued to rise substantially, from a mean of 38 to 71.5 cases per year in the periods of 1996-2000 and 2001-2004 respectively following the introduction of PCV-7. Although the percentage of cases due to serotypes covered by the vaccine did decrease from 37% to 14% (Byington *et al.*, 2006). This same trend of increasing empyema incidence after introduction of PCV-7, particularly in children between the ages of 1 and 4 years, has been recently corroborated by three large scale studies based on hospital coding data (Grijalva *et al.*, 2010; Lee *et al.*, 2010; Li and Tancredi, 2010).

Perhaps the most likely explanation for the rise in empyema despite the introduction of PCV-7 is that non-PCV-7-vaccine serotypes are responsible for the greatest burden of disease and therefore disease due to these serotypes is not reduced by the introduction of the vaccine. An alternative explanation that has been proposed in several settings is serotype replacement disease (Byington *et al.*, 2006; Hendrickson *et al.*, 2008). It is difficult to be certain that this phenomenon is truly responsible for the increase in empyema given the relatively low rate of vaccine serotypes causing the disease and the fact that many centres reported a synchronous upward trend in incidence *before* the vaccine was introduced (Fletcher *et al.*, 2006; Roxburgh *et al.*, 2008; Grijalva *et al.*, 2010). Furthermore, pneumococcal serotypes exhibit natural variation over time and in any given population. This may account for the fact that an increase in serotype 1, which is a non-PCV-7 serotype was seen *before* the introduction of the conjugate vaccine (Byington *et al.*, 2002b).

The changing incidence of other infections causing empyema in children also needs to be considered, along with their possible relationships to changes in pneumococcal ecology. For example in one study from Texas *Staphylococcus aureus* infection increased substantially, from 18% to 60% over ten years, and several other series have reported similar findings (Buckingham *et al.*, 2003; Schultz *et al.*, 2004). Analysis of the incidence of other causative organisms was beyond the scope of this analysis, so we cannot determine whether this was a significant factor within the North East population

The time–temperature interaction term

There was a significant interaction between temperature and time on the number of incident cases. In effect, winter case numbers increased through time, to a greater degree than did summer case numbers. This may be as a result of the overall distribution of cases or as a result of different mechanisms involved in the aetiology of empyema. We are unable to delineate this further.

Changes in age distribution

It was hypothesised that any preventative effect of the vaccine would be seen initially in those children who were direct vaccine recipients. There were no differences in the age distribution of cases year on year following the introduction of vaccine. This is further evidence that introduction of PCV-7 was not responsible for any reduction in the incidence of paediatric empyema in the North East of England.

4.4.3 Implications and future research

Our results suggest that the introduction of the PCV-7 immunisation was not associated with a reduction in the incidence of empyema in the North East of England. It is perhaps surprising that we are unable to report any effect as previous estimates have suggested that up to 20% of UK cases may be due to serotypes covered by the vaccine (Eltringham *et al.*, 2003; Fletcher *et al.*, 2006). PCV-7 has been shown to be very effective in the prevention of invasive pneumococcal disease due to vaccine serotypes and uptake in our region was high (O'Brien *et al.*, 2003; Cutts *et al.*, 2005). Serotype replacement may have contributed to this lack of impact. It is more likely that PCV-7 serotypes did not contribute significantly to these cases and other factors such as the circulation of epidemic strains of serotype 1, changes in organism virulence or an increase in the underlying cases of pneumonia may be responsible. It is not possible from the data available to address these possibilities. However, the ongoing UK-ESPE study, which is a multi-centre

prospective observational study of paediatric empyema will have the scope and the power to address these factors.

Further detailed epidemiological studies are required to establish the processes that are driving the change in empyema incidence, and to monitor the impact of the next generation of pneumococcal conjugate vaccines.

4.5 Conclusions

The introduction of PCV-7 did not alter the incidence of empyema within the North East of England.

5 The influence of pathogen and pneumococcal serotype on illness length and the impact of primary pleural drainage methodology on outcomes in empyema

5.1 Introduction

5.1.1 Pathogen, pneumococcal serotype and outcomes in empyema

Historically empyema has been considered in terms of its pathophysiology related to the accumulation of pus with loculation within the pleural cavity. Considerably less attention has been given to the underlying pathogenic cause. There has however been an awareness of differing risks associated with infection by different pathogens from the earliest scientific literature onwards. The Hippocratic Corpus described different constitutions of infected fluid drained in empyema and attributed them both to different predicted outcomes and different causes (Christopoulou-Aletra and Papavramidou, 2008). Heuer (1932) commented that streptococcal (implying *Streptococcus pyogenes*) empyema carried a greater risk of death than pneumococcal empyema. Since 1997, when the first modern reports of an increase in the incidence of paediatric empyema emerged (Playfor *et al.*, 1997; Rees *et al.*, 1997), only one study has directly examined the clinical profile and outcomes of empyema caused by different pathogens. This study found significant increases in hospital stay associated with staphylococcal empyema (Blaschke *et al.*, 2011).

Attention has been paid to empyema associated with different pneumococcal serotypes and certain pneumococcal serotypes appear to be associated with particular complications of empyema. In one London case series from 2002 to 2009, 91% empyema cases complicated by formation of a bronchopleural fistula were associated with Serotype 3 (equivalent to 10/16 cases, $p<0.0001$). This was accompanied by a significant increase in median hospital stay (7 vs. 10 days, $p<0.001$) and surgical intervention rate (2% vs. 14%, $p=0.001$) (McKee *et al.*, 2011). Serotype 19A has been associated with an increased risk of other complications including necrotising pneumonia and haemolytic uraemic syndrome (Reinert *et al.*, 2010; Andrade *et al.*, 2011). Mortality in adults with pneumococcal pneumonia differs between pneumococcal serotypes with serotypes 3, 6A, 6B, 9N, and 19F being associated with an increased risk of death. Not of all of the patients in this series suffered from, or died from empyema (Weinberger *et al.*, 2010).

Understanding the varying patterns of empyema between different pathogens is important, as the increases in the incidence of the condition which occurred after 1997 appear to be related to changes in the nature of the individual organisms and may have been responsible for the significant changes in the clinical profile of the disease.

5.1.2 Impact of primary pleural drainage procedure in empyema

As the incidence of empyema has increased since 1997, another important question is what is the best way to manage empyema? Clinical management has routinely been based on draining the fluid in the pleural cavity. This is required to relieve the pressure and the pain associated with empyema and to allow healing to occur. The optimal method for achieving adequate pleural drainage has not been agreed between different specialists. The UK national guidelines explicitly acknowledge the paucity of firm evidence underpinning the management of empyema in children (Balfour-Lynn *et al.*, 2005). Areas of particular controversy include the choice of primary pleural drainage procedure, the role of fibrinolysis in treatment and the usage of computerised tomography scanning in diagnosis of empyema (Balfour-Lynn *et al.*, 2005). The UK guidelines recommend that initial management may be either thoracocentesis with the installation of intrapleural fibrinolytics, such as Urokinase or surgical drainage using either a mini-thoracotomy or VATS approach. The evidence underpinning this is limited. Intrapleural drainage with fibrinolysis has been shown to shorten hospital stay in children when compared to drainage with saline placebo in one small (n=60 children in total) randomised controlled trial (Thomson *et al.*, 2002). Interestingly, the use of fibrinolytic agents in adults is associated with increased morbidity and mortality (Tokuda *et al.*, 2006).

The data comparing surgical and medical primary pleural drainage procedure is contradictory. Two small RCT's (n=60 and 30 respectively) have demonstrated no difference between medical and surgical management in terms of length of stay, days of pleural drainage, days of fever and procedure number (Sonnappa *et al.*, 2006; St. Peter *et al.*, 2009). In contrast one small RCT (18 patients), two large cohort studies and a meta-analysis have shown some benefit from a primary surgical approach (Avansino *et al.*, 2005; Kurt *et al.*, 2006; Li and Gates, 2008; Shah *et al.*, 2011). Further large scale studies comparing outcomes between the different treatment

modalities would be a significant contribution to the clinical management of paediatric patients with empyema.

5.1.3 Measuring outcomes in empyema

One factor which has contributed to the lack of concensus regarding treatment in empyema is the lack of agreed outcomes with which to compare treatments. Of those that have been used, none have included patient related outcome measures – outcome measures derived from direct questioning of patients and/or their parents.

A number of different outcome measures have been used previously in studies of paediatric empyema. These include mortality, length of hospital stay, time to defervescence, treatment costs, complication rate and treatment failure rate (Thomson *et al.*, 2002; Sonnappa *et al.*, 2006; Li and Gates, 2008). Other comparators are required because specialists do not always concur on the relevance of each of these parameters. Death rates are not a useful outcome measure as the mortality from empyema is extremely low in children in industrialised nations (Balfour-Lynn *et al.*, 2005).

Length of hospital stay (LOS) has been consistently used as the predominant outcome measurement. LOS is a surrogate measure of the need for relatively intensive clinical care. The more severe the condition the longer the need for in-hospital care. There is a strong argument for the usage of LOS metrics to compare treatments because it is simple and easy to measure. LOS is objective but has several limitations. However, this outcome is confounded by treatment which may prevent early discharge and by other factors determining discharge. It can also be dependent on organisational factors such as the availability of hospital beds, and the ability and willingness of parents to take their child home to recuperate. LOS does not relate to any measure of patient or parental satisfaction.

When comparing pathogen and serotype differences, there is some evidence of significant differences in the duration of pre-hospital symptoms between pathogens, suggesting that a measure of pre-hospital symptom length should also be included in any comparator (Blaschke *et al.*, 2011).

5.1.4 Aims & hypotheses

The aims of this analysis were twofold. Firstly, to investigate whether the clinical profile of empyema differed between pathogens and pneumococcal serotypes. Secondly, to evaluate the effectiveness of different methods of treating empyema to better inform management of the condition. It was assumed that LOS was an accurate surrogate of illness severity and that hospital discharge was the final event in the patient's experience of the illness. By using discharge as an event or outcome in this manner, survival analysis can be used to model factors impacting on this outcome. Differences in the clinical profile of paediatric empyema between pathogens and pneumococcal serotypes were then investigated by survival analysis with discharge from hospital as the primary outcome. Similarly, LOS stay was modelled to compare primary pleural drainage methodologies in the treatment of empyema using Cox's proportional hazards models to assess factors impacting on risk of discharge. Readmission and complication rates were also compared between different drainage methodologies to investigate failure and safety of the methods. The key hypotheses were therefore:

Key hypotheses:

- *Time to disease resolution in empyema is dependent on pathogen and pneumococcal serotype.*
- *Surgical methods of primary pleural drainage are associated with shorter hospital stay in paediatric empyema.*

5.2 Methods

5.2.1 Ethical approval

The UK-ESPE study received a favourable opinion from the Sunderland Regional Ethics Committee in October 2007 (**Appendix Q).** The National Information Governance Board approved the study, with informed consent required for the prospective collection of patient identifiable information.

5.2.2 Clinical data

Demographic, clinical and microbiological data on children with empyema undergoing pleural drainage were obtained from 19 paediatric respiratory centres from September 2006 until March 2011 as part of the UK-ESPE study (Study protocol is included as separate Appendix P). Data were collected through the use of a web-access form stored on a secure database. Forms were completed by a nominated clinical team in each respiratory centre for each patient.

The inclusion criteria were:

- Age up to 16 years
- Clinical diagnosis of empyema
- Undergoing some form of invasive pleural drainage

The clinical data were linked by surname and date of birth to the separate database of detected pneumococcal serotypes from the HPA. In September 2006, the HPA established enhanced surveillance of paediatric culture negative empyema for England in collaboration with members of the British Paediatric Respiratory Society (BPRS). Samples were forwarded from admitting hospitals managing patients with empyema. Those that were pneumococcal PCR positive underwent non-culture serotyping using a multiplex antigen detection assay capable of detecting 14 serotypes/groups (1, 3, 4, 5, 6A/C, 6B, 7F/A, 8, 9V, 14, 18, 19A, 19F and 23F) (Sheppard *et al.*, 2011).

Primary pleural drainage management strategy was defined as that used in the first 48 hours of pleural drainage and all cases were divided into two categories – primary medical management and primary surgical management and four sub-categories – drainage alone, drainage and fibrinolysis (primary medical management), mini-thoracotomy and VATS (primary surgical management). The cut off of 48 hours was used to distinguish between patients who had primary surgical management from those that had failure of primary medical management requiring later surgical treatment.

Two LOS measurements were used – total hospital stay and tertiary hospital stay. Tertiary hospital stay was defined as the stay in the hospital where definitive drainage was carried out and in those cases where children were admitted directly to the tertiary centre was equivalent to the total hospital stay. For comparisons between empyema caused by different organisms and different pneumococcal serotypes, a third time metric named illness length was used. This was defined as the total hospital stay in days added to the length of pre-hospital symptoms. The relationship between all three measurements is shown in **Figure 5.1**.

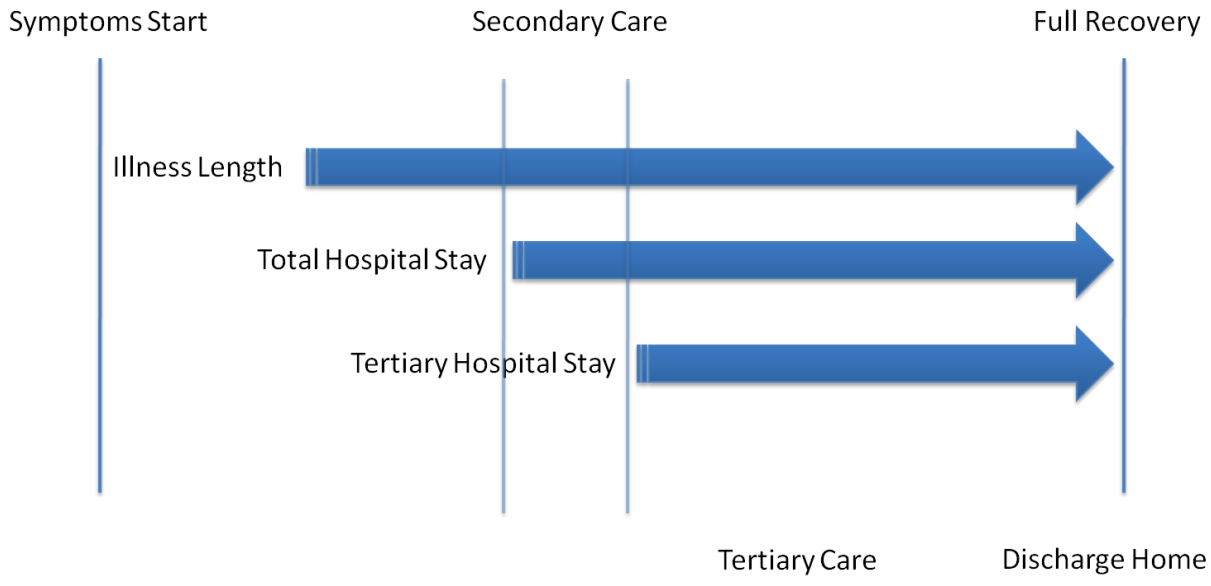


Figure 5.1 Relationship between illness length and LOS measurements.

Three further outcome measures were used for comparisons between treatment groups. These were the readmission rate, the complication rate and the pneumothorax rate. The readmission rate was proposed as an indicator of treatment failure. The complication and pneumothorax rates were included as important safety indicators for different treatments. Complications were defined by the clinician collecting data in each centre and were recorded in a free-text box in the web-access proforma. Pneumothoraces were all listed as complications in this manner, rather than as a separate question.

5.2.3 Statistical methods

Survival analysis is a broad term referring to a range of techniques employed whereby the main outcome variable is time to observe a particular events (Kirkwood and Sterne, 2003). Data recorded during this analysis were generally not normally distributed and were often positively skewed. Secondly, not all individuals in the study reached the endpoint of interest during the observation period leading to the phenomenon of right censoring (Everitt and Rabe-Hesketh, 2001). Several statistical techniques are available that allow for the analysis of these data and are known collectively as survival methods (Kirkwood and Sterne, 2003).

Three primary outcome measures were examined. These were illness length, total hospital stay and tertiary centre stay. Each was assessed for normality by plotting of the frequency distribution and using the Shapiro-Wilks test for normality (Royston, 1982) (Appendix - **Figure 9.57, Figure 9.58, Figure 9.59**). All three measures were found to not conform to the underlying statistical assumptions of normality insofar as they were positively skewed. These findings support a survival analysis approach. Furthermore, age, sex, hospital or centre and admission to intensive care were hypothesised to be likely to relevant to determining illness length and hospital stay length, hence a multivariate approach, as opposed to a simple univariate survival approach using survivor functions was required. The most widely applied method of analysing multivariate regression of survival data is the Cox proportional hazards model (Kirkwood and Sterne, 2003). The Cox proportional hazards model can be expressed as:

$$Log\big(h(t)\big) = \log\big(h_0(t)\big) + \beta_1 x_1 + \beta_2 x_2 + \beta_p x_p$$

where *h(t)* is the hazard at time t, $h_0(t)$ is the baseline hazard (the hazard for an individual in whom all exposure variables = 0) at time *t*, and x_1 and x_p are the $_p$ exposure variables (Kirkwood and Sterne, 2003). An assumption of the Cox model is that the hazards remain proportional over the time period.

Cox proportional hazards models were built using the Survival package in R 2.14.0 (Therneau, 2011). Ties were handled using the Efron approximation. (Therneau, 2011).

Two models were proposed to investigate differences in illness length between pathogens and pneumococcal serotypes.

Model Org:

$$\text{Illness length} \sim \log\big(h_0(t)\big) + \beta_1\text{*Sex} + \beta_2\text{*Age} + \beta_3\text{*Organism detected} + \epsilon$$

Model Sero:

$$\text{Illness length} \sim \log\big(h_0(t)\big) + \beta_1\text{*Sex} + \beta_2\text{*Age} + \beta_3\text{*Serotype detected} + \epsilon$$

Age, sex, pre-hospital illness length, paediatric intensive care unit (PICU) admission, chest drain type, co-morbidity, antibiotic choice and treatment group were included as co-variates in the initial models comparing treatment. Four initial models were proposed, two comparing primary medical and primary surgical methods and two comparing the four specific treatment methodologies such as mini-thoracotomy (**Table 5.1**). The principles of parsimony and backward selection were followed so non-significant parameters were removed stepwise and significant parameters retained in the models.

There were significant differences in both LOS measures between different centres. Stratification by centre was initially considered in order to remove the influence of centre. Stratification has often been used to adjust for the unwanted effect of centre within multi-centre trials of treatments (Everitt and Rabe-Hesketh, 2001). However, while there was intra-centre variation in primary pleural drainage procedure, not all centres carried out all methods of pleural drainage, meaning that stratification by centre would not be possible. Therefore, in order to adjust for the influence of centre when comparing the LOS between treatment methods, a frailty term was used. Frailty terms represent random effects within survival models and in the context of a multivariate regression model represent common unmeasured covariates between the groupings specified as the frailty term (Andersen *et al.*, 1999; Therneau and Grambsch, 2000). For example, centres may differ in approaches to pain management in children with empyema, but this is likely to be consistent within the hospital. Pain management may in turn influence LOS but would usually not be specifically measured. It is likely there are multiple unmeasured variables of potential influence that may vary considerably between centres. This unmeasured variation was included in the models as a frailty term.

Model HSa	Total hospital stay ~ $\mathbf{log}(\boldsymbol{h_0(t)})$ + β_1*Sex + β_2*Age + β_3*PICU Admission + β_4*Chest drain type + β_5*Co-morbidity + β_6*Antibiotic choice + β_7*Treatment method (Medical vs. Surgical) + frailty (Centre) + ϵ
Model TCa	Tertiary centre stay ~ $\mathbf{log}(\boldsymbol{h_0(t)})$ + β_1*Sex + β_2*Age + β_3*PICU Admission + β_4*Chest drain type + β_5*Co-morbidity + β_6*Antibiotic choice + β_7*Treatment method (Medical vs. Surgical) + frailty (Centre) + ϵ
Model HSb	Total hospital stay ~ $\mathbf{log}(\boldsymbol{h_0(t)})$ + β_1*Sex + β_2*Age + β_3*PICU Admission + β_4*Chest drain type + β_5*Co-morbidity + β_6*Antibiotic choice + β_7*Treatment method (Specific methodology) + frailty (Centre) + ϵ
Model TCb	Tertiary centre stay ~ $\mathbf{log}(\boldsymbol{h_0(t)})$ + β_1*Sex + β_2*Age + β_3*PICU Admission + β_4*Chest drain type + β_5*Co-morbidity + β_6*Antibiotic choice + β_7*Treatment method (Specific methodology) + frailty (Centre) + ϵ

Table 5.1 Modelling framework for analysis of primary pleural drainage method on length of stay in empyema

Testing of non-proportionality of hazards was carried out by plotting the Schoenfeld residuals for the models and by regression of the weighted residuals, as outlined by Grambsch and Therneau (1994).

Readmission, complication and pneumothorax rates represented binomial outcomes. The effect of different treatment groups on each rate was interrogated using binomial generalised linear models where odds ratios for each treatment group were calculated. Not all outcomes occurred in all centres and as such centre was omitted from the model. A binomial generalised linear model was also used to compare baseline differences in age, sex, co-morbidity and PICU admission between the medical and surgical groups.

LOS data are prone to the influence of outliers (Sills *et al.*, 2000). For example, children with serious underlying co-morbidity or complex psycho-social needs are at risk of staying significantly longer in hospital (Mahon and Kibirige, 2004). The removal of the 2% highest and lowest LOS values has been used elsewhere to counterbalance this (Markowitz *et al.*, 1996; Sills *et al.*, 2000). The range the three outcome measures before and after this adjustment are shown in

Table 5.2**:**

Measure	Pre-adjustment	Post-adjustment
Illness length / days	6 – 134	9 – 88
Total hospital stay / days	3 – 134	5 – 43
Tertiary hospital stay / days	2 – 103	3 – 33

Table 5.2 Range of outcome measures pre and post adjustment.

5.3 Results

5.3.1 Summary of clinical data

Data on a total of 637 children were obtained (56% male, median age 4.3 years). Their characteristics are summarised in

Table 5.3. The spread of cases by year of the study is shown in

Figure 5.2.

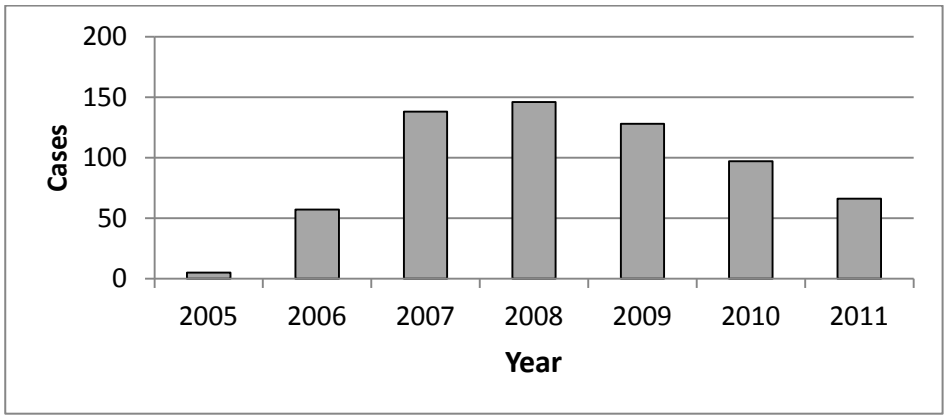


Figure 5.2 Cases recruited in each year of the UK-ESPE study

Where a microbiological cause was identified (n=287), *Streptococcus pneumoniae* accounted for 68%. The full details of microbiological data are included in

Table 5.4. A pneumococcal serotype was detected in a 123 patients with serotype 1 accounting for 46% of these. Other prominent serotypes were 3(22%), 7(12%) and 19A(9%) (

Table 5.5).

The median tertiary LOS was 8 days (range 3-33) and median total hospital stay (THS) 11 days (range 5-43). One in five children was admitted to PICU. There was an even split between medical (51%) and surgical (49%) primary pleural drainage procedures, with mini-thoracotomy being the commonest approach used (46%).

Variable	**% (Count) / Median (Range)**	**Respondents (n=637)**
Age *(years)*	4.3 (0-16)	637
Sex	56% (359)	637
Pre-hospital illness length *(days)*	6 (0-60)	613
Pre-hospital antibiotics given?	34% (163)	477
Any co-morbidity?	10% (65)	637
Antibiotics in tertiary centre		637
- Penicillins	5% (31)	
- Clindamycin	22% (141)	
- Cephalosporins	18% (120)	
Drain type		610
- Large bore	62% (378)	
- Pigtail	38% (232)	
PICU admission	21% (125)	609
Any listed complication?	9% (60)	637
Illness length *(days)*	18 (9-88)	603
Total hospital stay *(days)*	11 (5-43)	637
Stay at tertiary centre *(days)*	8 (3-33)	637
Primary pleural drainage procedure		637
- Drainage & fibrinolysis	45% (288)	
- Drainage alone	6% (36)	
- Mini-thoracotomy	46% (295)	
- VATS	3% (18)	
- Primary medical (Drainage alone & drainage and fibrinolysis)	51% (324)	
- Primary surgical (Mini-thoracotomy & VATS)	49% (313)	

Table 5.3 Demographics of cases of paediatric empyema included in the analysis (n=637).

Organism recovered	% (Count)
S. aureus	10 (16)
S. pyogenes	6.3 (40)
S. mitis	1.1 (7)
S. pneumoniae	28.7 (183)
Other organism	6.8 (43)
No organism recovered	48.7 (306)
Unknown	6.1 (39)

Table 5.4 Organisms recovered from the 637 children included in the analysis.

Pneumococcal serotype recovered	% (Count)
1	43.4 (59)
14	2.2 (3)
19A	10.3 (14)
3	20.6 (28)
5	2.2 (3)
7	11 (15)
8	0.7 (1)
Non assay serotype	9.6 (13)

Table 5.5 Distribution of pneumococcal serotypes detected in 136 children matched to the HPA pneumococcal serotype database.

5.3.2 Variation between centres

A total of 19 centres contributed data. The number of patients included from each centre is shown in **Figure 5.3**. One centre (Newcastle) provided 23% of study patients. There were no significant differences in the age of patients (Kruskal-Wallis, p= 0.1255) and the sex distribution (Fisher's test, p= 0.1496) between centres. However, pre-hospital antibiotic usage (Fisher's test, p<0.001) and pre-hospital illness length (Kruskal-Wallis, p<0.001) did vary significantly. There was also significant variation in management between centres. Primary pleural drainage approach by centre is shown in

Figure 5.4 confirming both inter-centre and intra-centre variation in approach. Chest drain type (Fisher's test, p<0.001) and PICU admission (Fisher's test, p<0.001) varied significantly between centres. Perhaps most importantly, length of stay outcomes were also significantly different between centres (

Figure 5.5**,**

Figure 5.6 **and Figure 5.7**). These differences confirmed the need to model the centre effect explicitly when analysing the effects of primary pleural drainage procedure on the outcome measures.

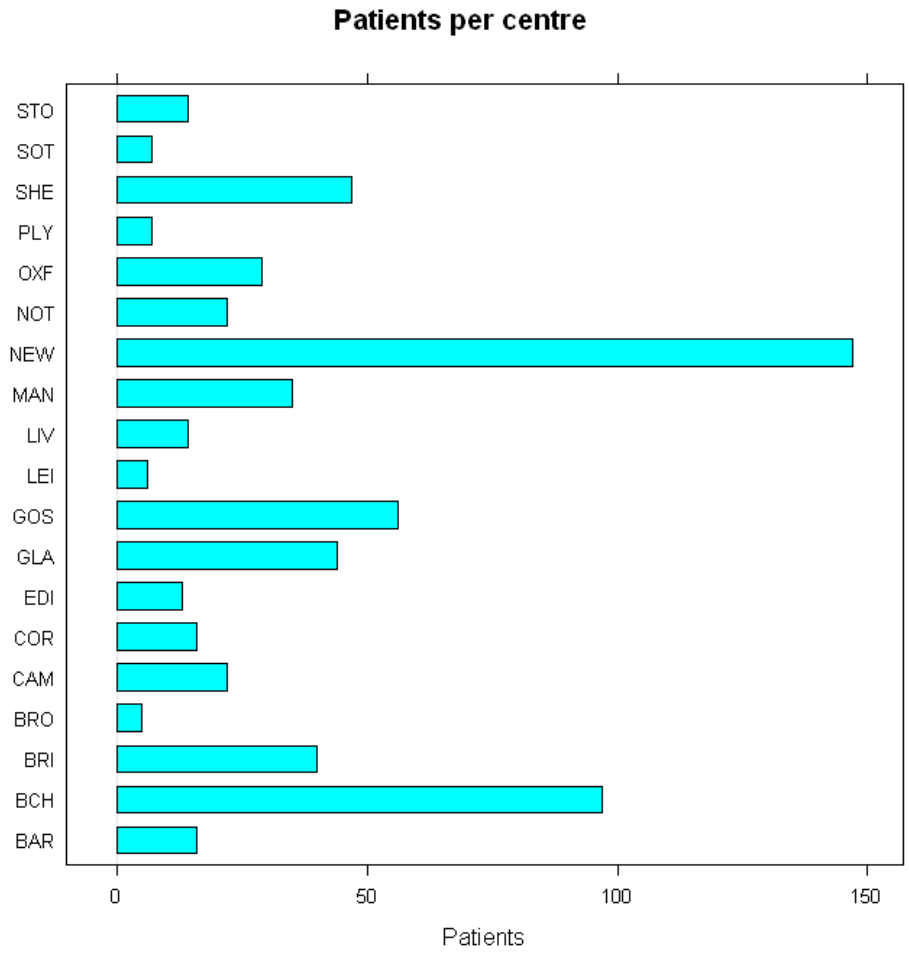


Figure 5.3 Numbers of patients included from each centre ***(STO – North Staffordshire Hospitals, SOT – Southampton General Hospital, SHE – Sheffield Children's Hospital, PLY – Derriford Hospital, Plymouth, OXF – John Radcliffe Infirmary, Oxford, NOT – Nottingham Royal Infirmary, NEW – Freeman Hospital, Newcastle, MAN – Manchester Children's Hospital, LIV – Alder Hey Hospital for Sick Children, LEI – Leicester Royal Infirmary, GOS – Great Ormond Street Hospital for Sick Children, GLA – Royal Hospital for Sick Children, Glasgow, EDI – Royal Hospital for Sick Children, Edinburgh, COR – Royal Cornwall Hospital, Truro, CAM – Addenbrookes Hospital, Cambridge, BRO – Royal Brompton Hospital, London, BRI – Bristol Royal Infirmary, BCH – Birmingham Children's Hospital, BAR – Royal London Hospital, London)***

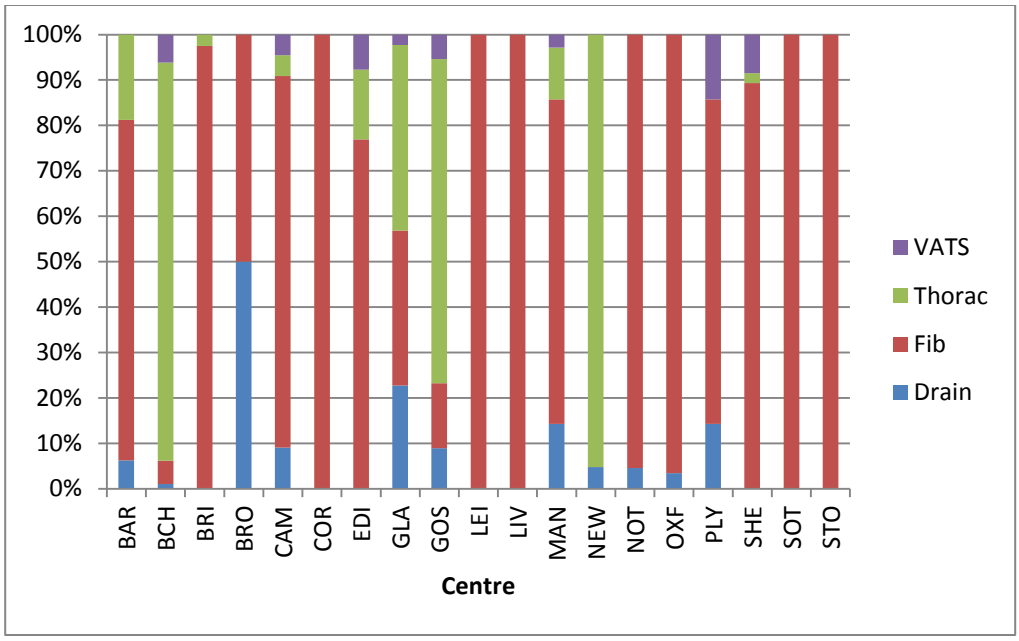


Figure 5.4 Proportion of different primary pleural drainage procedure in each study centre (*VATS – video assisted thoracoscopic surgery, Thorac – thoracotomy, Fib – drainage and intrapleural fibrinolysis, Drain – pleural drainage only; STO – North Staffordshire Hospitals, SOT – Southampton General Hospital, SHE – Sheffield Children's Hospital, PLY – Derriford Hospital, Plymouth, OXF – John Radcliffe Infirmary, Oxford, NOT – Nottingham Royal Infirmary, NEW – Freeman Hospital, Newcastle, MAN – Manchester Children's Hospital, LIV – Alder Hey Hospital for Sick Children, LEI – Leicester Royal Infirmary, GOS – Great Ormond Street Hospital for Sick Children, GLA – Royal Hospital for Sick Children, Glasgow, EDI – Royal Hospital for Sick Children, Edinburgh, COR – Royal Cornwall Hospital, Truro, CAM – Addenbrookes Hospital, Cambridge, BRO – Royal Brompton Hospital, London, BRI – Bristol Royal Infirmary, BCH – Birmingham Children's Hospital, BAR – Royal London Hospital, London).*

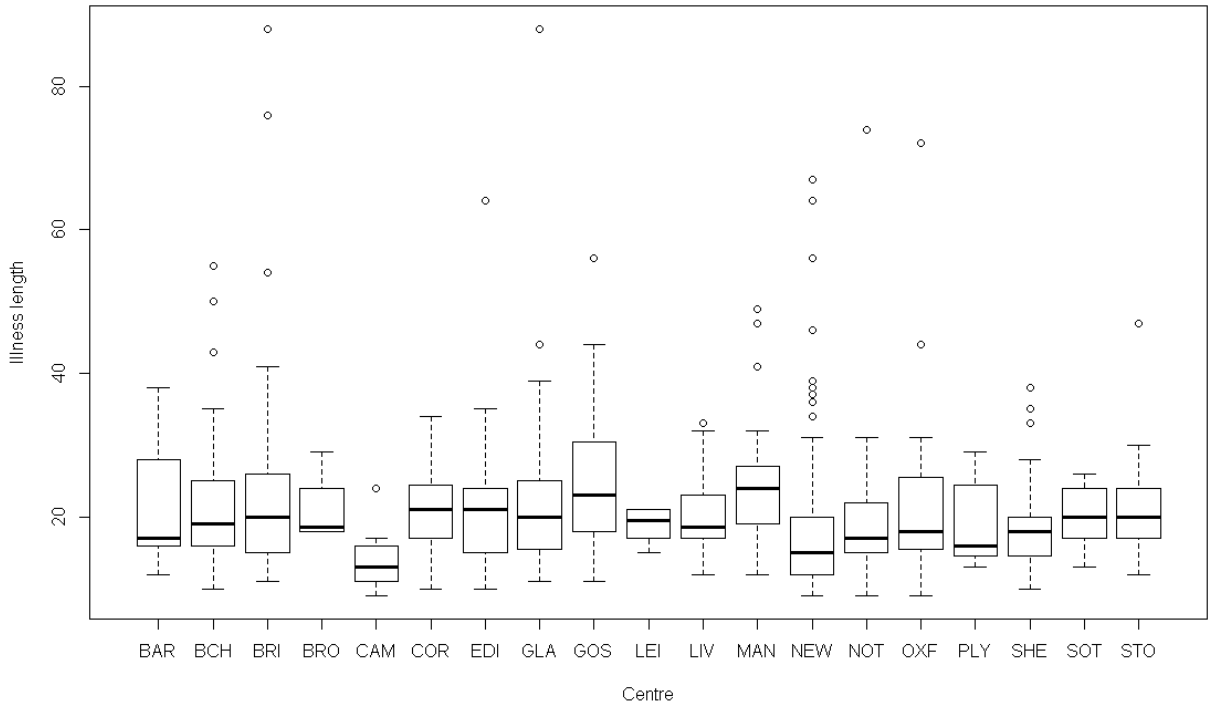


Figure 5.5 Variation in total illness length between study centres (Test for difference in illness length, Kruskal-Wallis, p<0.001).

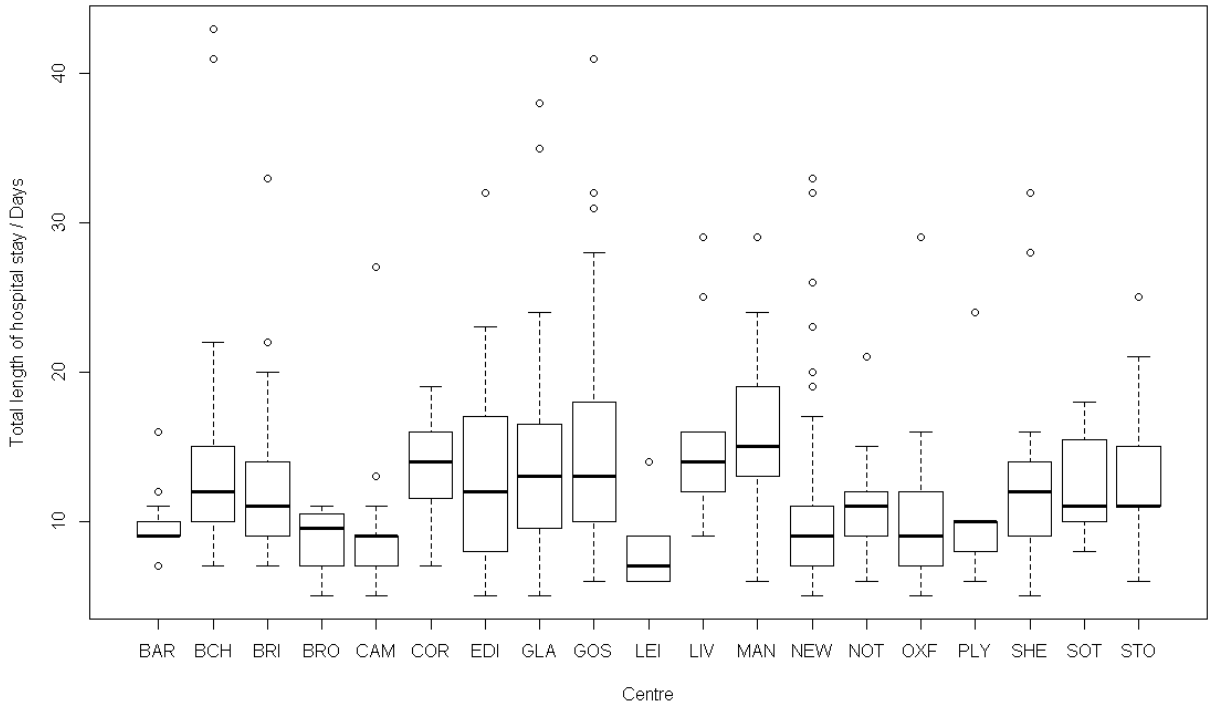


Figure 5.6 Variation in total length of hospital stay between study centres (Test for difference in total length of hospital stay, Kruskal Wallis, p<0.001).

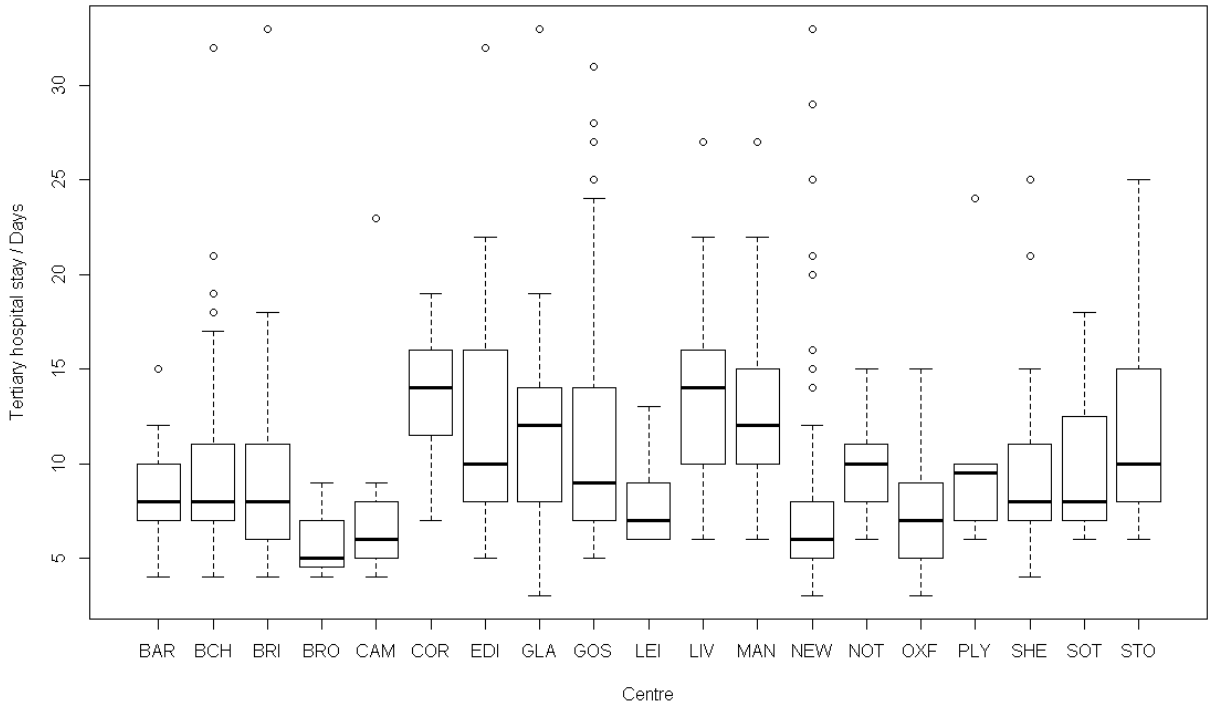


Figure 5.7 Variation in tertiary hospital stay between study centres (Test for difference in tertiary hospital stay, Kruskal-Wallis, p<0.001).

5.3.3 Influence of pathogen on illness length

In order to test whether changes in the epidemiology of empyema had led to changes in the severity of clinical disease, the relationship between illness length and the pathogen causing empyema was investigated.

While there was no relationship between sex and pathogen (Fisher's test, p=0.5246), there was a significant interaction between age and pathogen, with those known to have been infected with *S. aureus* and *S. pyogenes* being the youngest (Figure 5.8, Kruskal-Wallis, p<0.001). As a consequence, age was included as a covariate in the initial model Org. However, this was shown to be non-significant and was therefore subsequently excluded from the analysis (Age – Hazard 0.9885 (95% CI 0.969 – 1.008), p=0.2487).

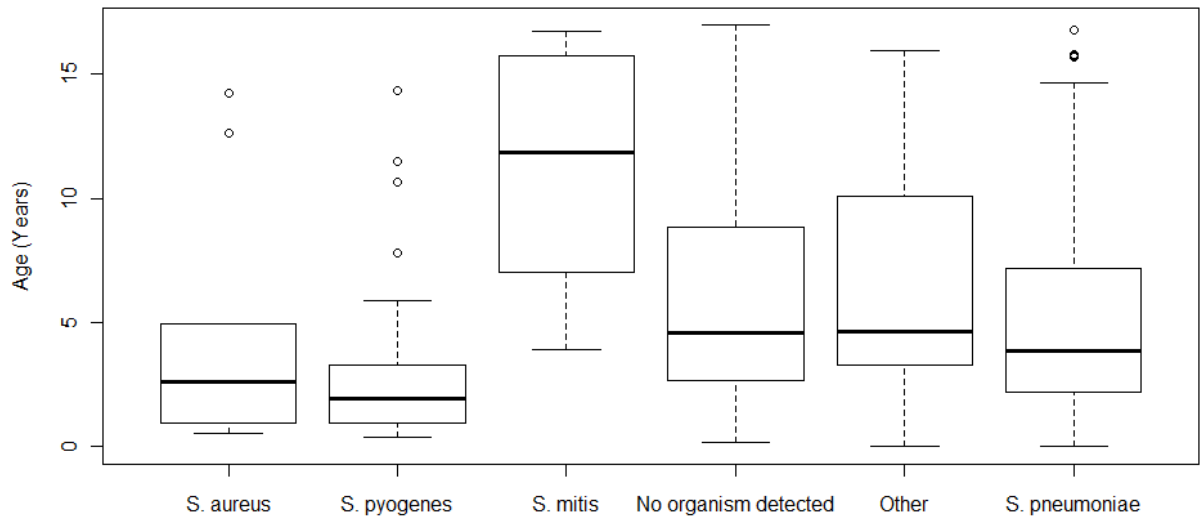


Figure 5.8 Significant differences in the age of children with empyema caused by different pathogens (Kruskal-Wallis chi-squared = 43.1226, df = 5, p-value = <0.001)

There was a significant relationship between illness length and pathogen. The event curves for each pathogen are shown in **Figure 5.9**. **Table 5.6** contains the hazard estimates and model diagnostic information. There was no violation of the proportional hazards assumption (Plots of Schoenfeld residuals and results of regression of residuals are shown in Appendix - **Table 9.4**

and **Figure 9.60**). Children who isolated *S. pneumoniae* had a 110% reduction in their illness length when compared to children who isolated *S. aureus* (Hazard 2.121, (95% CI 1.117 – 4.028), p=0.021) suggesting that the predominance of pneumococcus in causing empyema may have led to less severe disease.

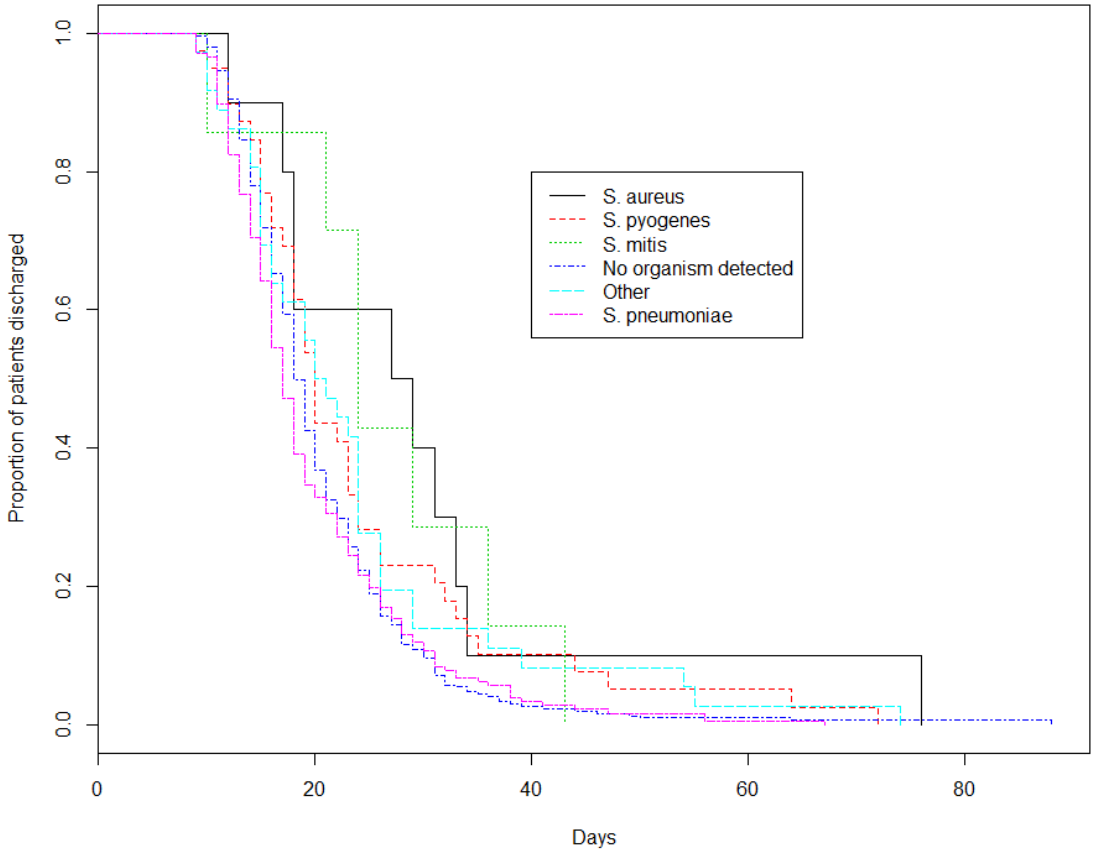


Figure 5.9 Event curves comparing illness length between pathogens in children with empyema in the UK.

Illness length					
Variable	**Co-efficient**	**Hazard**	**Standard Error**	**95% CI**	**P-value**
S. aureus	REFERENCE				
S. pyogenes	0.367	1.444	0.356	0.719 ; 2.899	0.302
S. mitis	0.096	1.101	0.494	0.418 ; 2.901	0.846
No organism detected	0.670	1.954	0.324	1.036 ; 3.685	0.039
Other	0.408	1.504	0.359	0.745 ; 3.039	0.255
S. pneumoniae	0.752	2.121	0.327	1.117 ; 4.028	0.021
Concordance = 0.548 (se = 0.014)					
R squared = 0.026					
Likelihood ratio test = 14.82 on 5 df, p=0.011					
Wald test = 13.22 on 5 df, p=0.021					
Score (logrank) test = 13.5 on 5 df, p=0.019					

Table 5.6 Results of parsimonious model Org investigating the relationship between illness length and pathogen in UK children with empyema.

5.3.4 Influence of pneumococcal serotype on illness length

The relationship between pneumococcal serotype and illness length was investigated using the same approach as for the relationship between pathogen and illness length. There was no difference in the sex distribution between pneumococcal serotypes (Fisher's test, p-value = 0.8572). There was a significant interaction between age and serotype. Serotype 19A and serotype 3 had the lowest median age and whereas serotype 1 showed the greatest median age (**Figure 5.10** (Kruskal Wallis, p-value <0.001). Age was therefore included as a covariate in the initial model Sero. This was found to be non-significant and was therefore excluded from subsequent analysis (Age – Hazard -0.0171 (95% CI 0.9341 - 1.0345), p=0.5104).

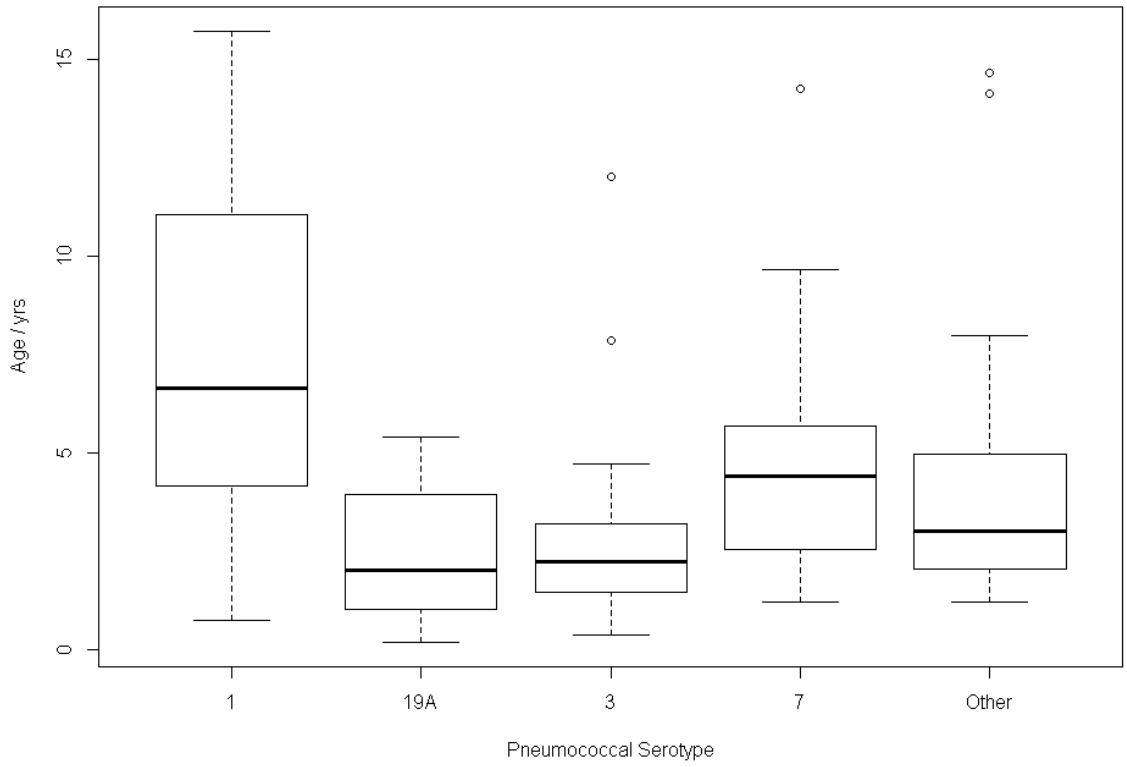


Figure 5.10 Significant differences in age between cases of pneumococcal empyema caused by individual pneumococcal serotypes (Kruskal Wallis - chi-squared = 44.9934, df = 4, p-value <0.001).

Illness length varied significantly between pneumococcal serotypes. The event analysis curves are shown in **Figure 5.11** and the relationship between illness length and serotype is displayed

in **Table 5.7**. Children infected with pneumococcal serotype 19A had on average a 60% longer illness (95% CI 25% - 78%) compared to those infected with serotype 1. The hazards remained constant over the whole time period (Plots of Schoenfeld residuals and results of regression of residuals are shown in Appendix **Figure 9.61** and **Table 9.5**).

Illness length					
Variable	**Co-efficient**	**Hazard**	**Standard Error**	**95% CI**	**P-value**
Serotype 1	REFERENCE				
Serotype 19A	-0.8965	0.4080	0.3189	0.2184 ; 0.7623	0.0049
Serotype 3	-0.2915	0.7472	0.2331	0.4731 ; 1.1799	0.2112
Serotype 7	-0.6159	0.5402	0.3151	0.2913 ; 1.0018	0.0507
Other serotype	-0.2210	0.8017	0.2618	0.48 ; 1.3392	0.3986
Concordance = 0.568 (standard error = 0.031)					
R-squared = 0.077					
Likelihood ratio test = 10.66 on 4 df, p=0.03067					
Wald test = 9.73 on 4 df, p=0.04519					
Score (logrank) test = 10.06 on 4 df, p=0.03946					

Table 5.7 The results of the parsimonious model Sero investigating the relationship between illness length and pneumococcal serotype in UK children with empyema.

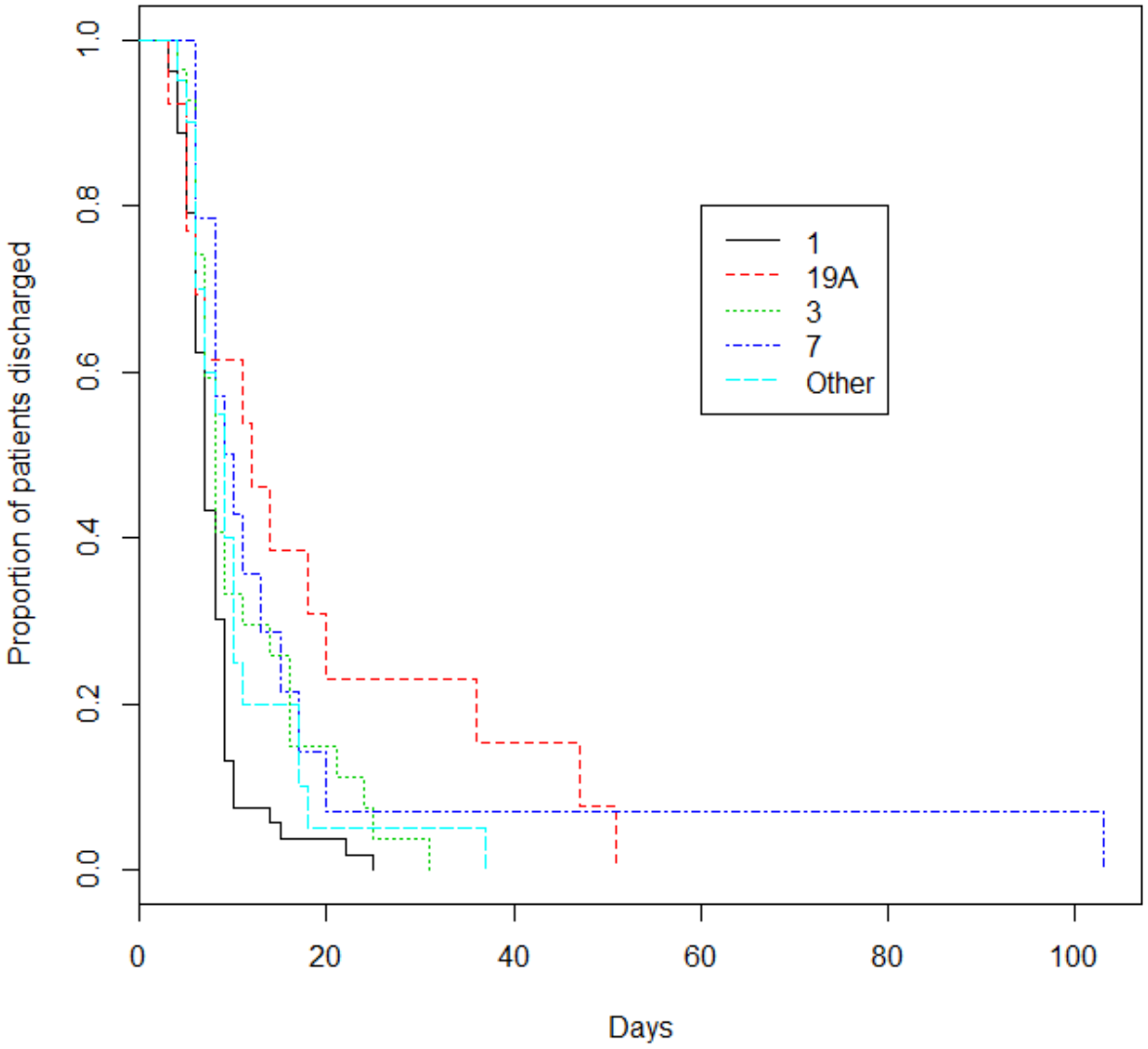


Figure 5.11 Event curves comparing illness length between pneumococcal serotypes in children with empyema in the UK.

5.3.5 Impact of primary pleural drainage methodology on outcomes in empyema

The impact of primary pleural drainage method on total hospital stay and tertiary centre stay was investigated by survival modelling including centre as a frailty term. The results of the parsimonious models investigating total hospital stay (HSa and HSb) and tertiary centre stay (TCa and TCb) are shown in **Table 5.8** and **Table 5.9**. Surgical methods (Thoracotomy or VATS) were associated with a decrease in total hospital stay of 40% (95% CI 5-87%) when compared to non-surgical methods (thoracocentesis +/- fibrinolysis). A similar relationship was seen in tertiary centre stay (30% (95% CI -0.03-71%) but was of borderline significance (p=0.056).

When comparing the four subgroups of treatments and using thoracocentesis and fibrinolysis as the reference treatment, thoracocentesis alone was associated with a significant increase in tertiary centre stay. A similar direction of effect was seen with total hospital stay but was non-significant (p=0.061). Thoracotomy was associated with a reduction in both length of stay measures but neither association reached significance.

PICU admission was associated with a significant increase in LOS in all models. Older children had a slightly longer total hospital stay (2% (95% CI 0-4%) HSa and 2% (0-3.9% (HSb)). The centre frailty term was significant in all models.

Total hospital stay (HSa)										
Variable	**Co-efficient**	**Hazard**	**SE**	**95% CI**	**P-value**	**Concordance**	**R-squared**	**LRT**	**Wald test**	**Variance of frailty**
PICU Admission	-0.816	0.442	0.121	0.349; 0.56	<0.001	0.695 (se: 0.016)	0.234	156 on 17.7 df, p<0.001	52.6 on 17.7 df, p<0.001	0.299 I-L: -3087.7
Age (Years)	-0.020	0.980	0.010	0.961; 1	0.049					
Medical Treatment	REFERENCE									
Surgical Treatment	0.338	1.402	0.146	1.053; 1.87	0.021					
Frailty (Centre)					<0.001					
Tertiary centre stay (TCa)										
Variable	**Co-efficient**	**Hazard**	**SE**	**95% CI**	**P-value**	**Concordance**	**R-squared**	**LRT**	**Wald test**	**Variance of frailty**
PICU Admission	-0.985	0.374	0.121	0.295; 0.473	<0.001	0.720 (se: 0.016)	0.284	195 on 17 df, <0.001	70.5 on 17 df, p<0.001	0.323 I-L: -3062.5
Medical Treatment	REFERENCE									
Surgical Treatment	0.265	1.303	0.139	0.997; 1.711	0.056					
Frailty (Centre)					<0.001					

Table 5.8 Analysis of primary medical vs. surgical pleural drainage method in UK children with empyema *(LRT = likelihood ratio test, I-L = I likelihood).*

Total hospital stay (HSb)										
Variable	**Co-efficient**	**Hazard**	**SE**	**95% CI**	**P-value**	**Concordance**	**R-squared**	**LRT**	**Wald test**	**Variance of frailty**
PICU Admission	-0.793	0.452	0.120	0.357 ; 0.573	<0.001	0.694 (se:0.016)	0.244	163 on 19.4 df, p<0.001	57.6 on 19.4 df, p<0.001	0.310 I-L: -3084.5
Age (Years)	-0.020	0.980	0.010	0.961 ; 1	0.046					
Drainage & fibrinolysis	REFERENCE									
Drainage alone	-0.417	0.659	0.222	0.426 ; 1.019	0.061					
Mini-thoracotomy	0.280	1.323	0.199	0.896 ; 1.954	0.160					
VATS	-0.091	0.913	0.293	0.514 ; 1.622	0.760					
Frailty(Centre)	<0.001									
Tertiary centre stay (TCb)										
Variable	**Co-efficient**	**Hazard**	**SE**	**95% CI**	**P-value**	**Concordance**	**R-squared**	**LRT**	**Wald test**	**Variance of frailty**
PICU Admission	-0.972	0.378	0.12	0.299 ; 0.479	<0.001	0.726 (se: 0.016)	0.297	206 on 18.9 df, P<0.001	78.6 on 18.9 df, p<0.001	0.343 I-L: -3057.5
Drainage & fibrinolysis	REFERENCE									
Drainage alone	-0.546	0.579	0.221	0.375 ; 0.894	0.014					
Mini-thoracotomy	0.155	1.168	0.182	0.817 ; 1.67	0.390					
VATS	-0.248	0.78	0.289	0.442 ; 1.375	0.390					
Frailty(Centre)	<0.001									

Table 5.9 Comparison of the four primary pleural drainage procedure in UK children with empyema (*LRT = likelihood ratio test, I-L = I likelihood*).

Assumptions governing the use of proportional hazards was tested for all models by visual examination of the plotted scaled Schoenfeld residuals and by regression of the weighted residuals (Appendix **Table 9.6, Figure 9.62, Figure 9.63, Figure 9.64 and Figure 9.65**).

Regression of the weighted residuals indicated there was significant non-proportionality in all four models. Examination of the plotted Schoenfeld residuals indicated that the non-proportionality was moderate and unlikely to significantly affect the accuracy of the estimates of the co-variate effects (Therneau and Grambsch, 2000).

Higher values of non-proportionality were seen in the surgical methods and thoracotomy groups in all four models. The estimate of the effect of surgery was positive (i.e. a reduction in length of stay) across all the models until approximately 14 days when the effect changes polarity and the probability of discharge associated with surgery declines (example shown in **Figure 5.12**).

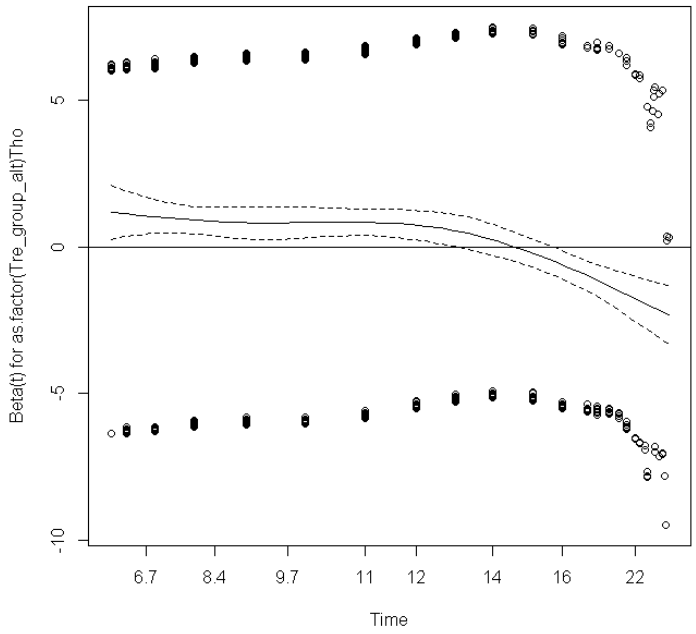


Figure 5.12 Non-proportional hazards in surgical treatment group in total hospital stay model (HSa).

Early surgery appears to confer a shorter stay when compared to medical treatment in the majority of patients, given the median stay was 11 days. However a minority seem to experience the opposite effect.

Comparison of the baseline characteristics of patients in the medical and surgical treatment groups is shown in **Table 5.10**. Overall patients who had primary surgery were significantly older and more likely to have co-morbidity. Comparison of the characteristics of just those who stayed longer than 14 days indicated that surgical patients were significantly more likely to have a co-morbidity than those given medical treatment (OR 2.76, 95% CI 1.076 to 7.34, p=0.0365). There were no significant differences in age, sex and PICU admission between the two treatment groups in those that stayed more than 14 days. Co-morbidity alone was associated with an increase in hospital stay across the whole cohort of patients (co-efficient 2.47, 95% CI 0.94 to 4.00, p=0.0016).

The absence of a significant difference in the rate of PICU admission between the two groups suggests that the observed variation in effect of surgery may not be related to disease severity, although PICU admission is at best only a proxy measurement of severity and may be influenced by other uncontrolled variables such as bed availability.

Variable	Odds Ratio	95% CI	P-value	Deviance residuals	Null Deviance	Residual Deviance	AIC
Age (Years)	1.041	1.000 ; 1.083	0.049	-1.782 to 1.377	844.25 on 608 DF	824.85 on 604 DF	834.85
Sex	1.031	0.744 ; 1.430	0.853				
PICU Admission	0.844	0.600 ; 1.269	0.416				
Co-morbidity	2.750	1.567 ; 5.016	<0.0001				

Table 5.10 Comparison of baseline characteristics between medical and surgical treatment groups (medical treatment group reference).

Treatment	Odds Ratio	95% CI	P-value	Range of Deviance Residuals	Null Deviance	Residual Deviance	AIC
Readmissions							
Medical management	REFERENCE			-0.478 to 2.675	248.98 on 591 DF	243.00 on 588 DF	251
Surgical management	0.410	0.182 ; 0.863	0.0231				
Drainage & fibrinolysis	REFERENCE			-0.485 to 2.697	248.98 on 591 DF	242.45 on 586 DF	254.45
Drainage alone	0.792	0.122 ; 2.916	0.7614				
Mini-thoracotomy	0.378	0.161 ; 0.826	0.0185				
VATS	0.846	0.0455; 4.566	0.875				
Any complications							
Medical management	REFERENCE			-0.518 to 2.564	350.44 on 636 DF	338.00 on 633 DF	346
Surgical management	0.338	0.169 ; 0.634	0.00116				
Drainage & fibrinolysis	REFERENCE			-0.756 to 2.588	350.44 on 636 DF	333.87 on 631 DF	345.87
Drainage alone	2.602	1.026 ; 6.061	0.033				
Mini-thoracotomy	0.383	0.184; 0.750	0.00684				
VATS	0.535	0.0291; 2.764	0.551				

Treatment	Odds Ratio	95% CI	P-value	Deviance residuals	Null Deviance	Residual Deviance	AIC
Pneumothorax							
Medical management	REFERENCE			-0.424 to 2.986	191.32 on 636 DF	183.15 on 633 DF	191.15
Surgical management	0.377	0.133 ; 0.934	0.0454				
Drainage & fibrinolysis	REFERENCE			-0.630 to 2.965	191.32 on 636 DF	179.66 on 631 DF	191.66
Drainage alone	3.054	0.807 ; 9.562	0.033				
Mini-thoracotomy	0.482	0.165 ; 1.268	0.007				
VATS	0	∞	0.551				

Table 5.11 Analysis of readmission rates, complications and pneumothorax rates by treatment group in UK children with paediatric empyema.

A total of 32 patients were readmitted to hospital following their initial discharge. Fifty patients experienced complications (Full list of complications is shown in Appendix **Table 9.7**). Of these, 22 had a pneumothorax.

The results of the analysis of readmission, complication and pneumothorax rates between the different treatment classifications are shown in **Table 5.11**. Early surgery (VATS or mini-thoracotomy) was associated with a reduced risk of readmission (OR 0.41, 95% CI: 0.18-0.86), any complications (0.34, 0.17-0.63) and pneumothorax (0.38, 0.13-0.93). When comparing the four treatment groups and using drainage and fibrinolysis as the reference group, drainage alone was associated with increased risk of complications (2.60, 1.03-6.06). Mini-thoracotomy was associated with a reduced risk of readmission (0.38, 0.16-0.83) and complications (0.38, 0.18-0.75).

5.4 Discussion

A typical patient with paediatric empyema in the UK between 2006 and 2011 was a four year old boy who stayed in hospital for a total of eleven days. If a cause was found for the empyema it was likely to be pneumococcal and of pneumococcal cases, pneumococcal serotype 1 was the most likely to be isolated. Staphylococcal disease was associated with significantly increased length of illness. Children with disease due to serotype 19A had a significantly prolonged illness length relative to other pneumococcal serotypes.

There were significant differences in the in-patient stay between different centres. Independent of centre effects, surgery as the primary pleural drainage procedure was associated with an approximate 40% reduction in LOS but there was significant heterogeneity in response between patients, possibly related to co-morbidity. Furthermore surgery was associated with a reduced risk of readmission, complications and pneumothoraces. There were no significant differences in length of stay when comparing the four individual treatment groups.

There were baseline differences in the four treatment groups. It therefore seems likely that the non-proportionality of the hazards associated with surgical treatment reflect in part a patient selection issue. Patients with co-morbidity are likely to have a longer stay independent of treatment method and are more likely to be selected to undergo surgery.

5.4.1 Strengths and limitations

Strengths of this analysis include the size of the cohort and the use of a robust modelling framework. The size of the study means that it will significantly add to the literature examining the impact of the change in epidemiology of empyema on the natural history of the disease as experienced by patients. Furthermore 637 patients represent a cohort that is substantially larger that have been based on directly collected clinical data than any previously used to evaluate empyema management in children. Only studies based on hospital coding data and therefore

with limited accuracy have been larger in scope (Li and Gates, 2008; Shah *et al.*, 2011). Other studies that have used directly collected clinical data have either been based at a single centre or have not used methods for assessing the influence of centre related effects (Khalil *et al.*, 2007; Shomaker *et al.*, 2011). This data will therefore provide an important contribution to the evidence base supporting clinicians making decisions about the treatment of paediatric empyema.

The modelling framework used allows estimation and adjustment for centre effects thereby allowing accurate comparisons of different treatment methodologies. This is particularly important in the context of the widespread variation in approach to the management of paediatric empyema in the UK and given the observational nature of the study. While national guidance for the management of empyema in the UK exists, this study highlights the fact that not only approaches but outcomes vary significantly between centres. The presence of this heterogeneity therefore required explicit modelling of a centre effect.

Despite these strengths there remain significant limitations in this analysis. One limitation is the microbiological data. Pathogens were identified in several ways. Standard microbial culture of blood or pleural fluid was applied to all patients entered into the study. However, a proportion of patients were identified as having pneumococcal infection by use of non-culture molecular detection through the use of pneumococcal antigen specific PCR on pleural fluid. This was carried out by the HPA in England and the Scottish Haemophilus Legionella Meningococcus & Pneumococcus Reference Laboratory. The protocol was for all enrolled patients whose pleural fluid was culture negative to be forwarded for molecular testing. Due to a technical error by the third party collaborator however no record of which samples were tested were kept until March 2010. Furthermore not all laboratories consistently forwarded samples, leaving significant uncertainty about the status of patients recorded as having no organism identified.

To compound matters the web-based proforma did not distinguish between identifications made via culture or non-culture methods. Individual centres did not generally keep records of which samples were forwarded for non-culture testing. It is therefore possible that samples identified as pneumococcal were either detected through culture and non-culture methods. Those where no

pathogen was detected may be as a consequence two different chains of events. Standard cultures may have been negative and no further testing was carried out or standard cultures were negative and non-culture pneumococcal testing was negative. Therefore those samples that were listed as no organism detected may in fact represent both undetected pneumococcal disease and culture negative non-pneumococcal disease, significantly compromising the interpretability of the microbiology results.

There are also some limitations from the analytical approach. Readmission, complication and rates of pneumothoraces were insufficiently common to allow adjustment for a centre effect. Yet it is likely that in particular readmission rates are subject to centre related factors such as bed availability, admission criteria and perhaps more importantly discharge criteria. The presence of a significant centre effect would therefore have acted as a potential source of bias.

The non-proportionality of the hazards associated with surgical treatment methods potentially limits the generalisability of the results. The overall non-proportionality was moderate. It was not judged to have significantly affected the results and had a plausible causal explanation in the form of selection bias towards those patients with co-morbidity.

5.4.2 Findings related to other studies

Within the UK-ESPE cohort, children who had pneumococcal empyema had an illness length of approximately half of that experienced by those who had a staphylococcal empyema. Blaschke and colleagues (2011) compared outcomes of children with *S. pneumoniae*, *S. pyogenes* and *S. aureus*. Children with staphylococcal empyema had hospital stays approximately twice as long as those with pneumococcal and streptococcal empyema (21 days vs. 10 and 11 days respectively) consistent with the UK-ESPE cohort, although the overall numbers of children were small and no statistically significant differences were found between the groups.

The difference in illness length between pathogens is likely to be a consequence of several factors. Firstly, culture positivity is associated with shorter pre-hospital symptom length but significantly prolonged hospital stay. Secondly, differences in culture positivity rates between different pathogens are well recognised in paediatric empyema (Blaschke *et al.*, 2011). Culture positivity may also be affected by a number of factors other than the pathogen involved including previous antibiotic usage, sampling frequency and laboratory technique. These caveats are supported by significant geographical variation in blood culture positivity rates between different countries, notably the USA and countries in Western Europe (Jefferson *et al.*, 2006). In addition, both pathogen and host specific factors are considered to be relevant in determining illness length. These factors are also likely to interact and impact on pathogen acquisition, probability of invasiveness and subsequent response to treatment. All of which in turn will have a direct impact on illness length. In a similar manner, pathogen related factors such as presence of virulence factors, invasiveness potential and antibiotic resistance will also impact on illness length. Finally treatment factors may be also be relevant. Discordant initial antibiotic use was the only significant risk factor found for death in empyema in adults in one study of empyema in Asia (Tsang *et al.*, 2007).

Children with pneumococcal infection had a much reduced illness length compared to those infected with other pathogens, differences between pneumococcal serotypes were observed with serotype 19A associated with a significantly prolonged illness length when compared to the most common serotype, serotype 1. The pneumococcal capsule is an important virulence factor, and hence differences in disease severity between serotypes are well described, although illness length has not previously been specifically reported (Reinert *et al.*, 2010). Picazo and colleagues found high rates of PICU admission (63.6%) and complications (27.3%) in Spanish children with empyema due to serotype 19A (Picazo *et al.*, 2011). Similar high rates of PICU admission associated with empyema due to Serotype 19A have been described elsewhere (Blaschke *et al.*, 2011).

As outlined previously, the literature comparing medical and surgical primary pleural drainage approaches is heterogeneous. Since 2005 a total of six high quality studies have been published, including three randomised controlled trials. The findings of these are re-summarised in Table 5.12. A finding in favour of primary surgery as primary pleural drainage procedure is consistent

with the findings of previous retrospective cohort studies and the previous meta-analysis (Avansino *et al.*, 2005; Li and Gates, 2008; Shah *et al.*, 2011). Interestingly of the three separate RCT's addressing this question, two have reported no difference between primary surgery and medical management and one found in favour of VATS. It is worth commenting that the total number of patients enrolled in RCT's was small when compared to the other studies (114 vs. 8728 including those from the UK-ESPE study). It is therefore possible that the RCT's carried out thus far have not had sufficient power to discriminate between the different approaches.

It is also of interest that consistent with the findings of Shah et al (2011) there were no significant differences between the four individual treatment sub-categories. This may indicate that there is no true difference in LOS between different surgical approaches.

Study	Type of study	Comparison	Results
Avansino (2005)	Meta-analysis	Any surgery vs. CTD or IFT	Primary operative approach (VATS/OT) lowered mortality, need for re-intervention, LOS, duration of chest tube and antibiotics compared to primary medical therapy
Kurt (2006)	RCT	VATS vs. CTD	VATS lowered LOS and cost and risk of transfer
Sonnappa (2006)	RCT	VATS vs. IFT	No difference between VATS and IFT in terms of LOS, days of CTD and procedure number. IFT less expensive
Li and Gates (2008)	Retrospective cohort review	Any surgery vs. CTD or IFT	Primary operative approach lowered LOS and cost and risk of transfer
St Peter (2009)	RCT	VATS vs. IFT	No difference between VATS and IFT in terms of LOS, febrile days and oxygen requirement
Shah (2011)	Retrospective cohort review	VATS vs. IFT	VATS lowered need for re-intervention; No difference in LOS between VATS and IFT
UK-ESPE	*Prospective (& retrospective) cohort review*	*Any surgery vs. CTD or IFT*	*Reduced LOS, readmission and complication rate in favour of surgery*

CTD = Continuous tube drainage without fibrinolysis, IFT = Drainage and intrapleural fibrinolysis, OT = open thoracotomy

Adapted from Anselmo M *Can Resp J* (2010)

Table 5.12 Update of Table 1.2 Summary of studies comparing primary pleural drainage procedure in empyema.

5.4.3 Implications for further research

The above findings raise a number of intriguing questions. If the observed change in the epidemiology of empyema is as consequence of predominantly pneumococcal disease then does this imply an era of more frequent but much less severe disease in contrast to that historically associated with staphylococcal infection? Furthermore do changes in the incidence of different pneumococcal serotypes have important clinical implications? Detailed above is further evidence of the association of serotype 19A infection with increasing disease severity. Continuing surveillance will be required to monitor changes in the pneumococcal serotype distribution, particularly following the introduction of the latest generation of pneumococcal vaccines as these changes are likely to have clinical impacts.

While this study provides further evidence to support surgery as the most efficacious primary pleural drainage procedure, two large cohorts have now not found any difference between open and VATS approaches suggesting that expending effort investigating this further may not be fruitful. Instead, future work should focus on establishing a multi-centre randomised controlled trial of sufficient power to evaluate the effectiveness of early surgery in the treatment of empyema in children when compared to optimal medical management. Similar trials have been carried out very successfully in adult patients with empyema in the UK and would probably be achievable in children (Rahman *et al.*, 2011).

A significant weakness of this work and that previously published in the field is that none of the outcome measures are patient based. Future trials need to address this and include adequate and appropriate measures of patient and parent satisfaction, as well as health economic factors and more traditional outcome measures such as LOS. A further weakness is the lack of longer term outcome data such as outcome at 1 year after discharge. Future studies must address this gap if truly informed decisions regarding the most appropriate methods of managing empyema are to be made.

5.5 Conclusions

Staphylococcal disease was associated with significantly increased illness length in contrast to the more prevalent pneumococcal disease. Children with disease due to pneumococcal serotype 19A had a significantly prolonged illness length. Further work both in epidemiological surveillance and in specific surveillance of pneumococcal serotypes are needed to monitor future changes, as such changes are likely to have clinical impacts.

Surgery as the primary pleural drainage procedure was associated with an approximate 40% reduction in length of stay but there was significant heterogeneity in response between patients possibly reflective of co-morbidity. Surgery was associated with a reduced risk of readmission, complications and pneumothoraces.

6 Conclusions

6.1 Summary of findings

This thesis set out to explore and define the epidemiology of paediatric empyema in response to the increase in the incidence of the disease observed across different continents from 1995 onwards.

The first step towards this aim was to confirm as robustly as possible that paediatric empyema had increased in the populations available to study. Analyses confirmed a consistent and substantial rise in the incidence of empyema in children. In England, the incidence of empyema as measured by hospital admissions increased from 9.68 (95% CI 7.79; 11.89) admissions per million children (0-14 years) in 1997/8 to 37.47 (95% CI 33.6; 41.66) in 2004/5. This was equivalent to an extra 253 cases per annum. Similarly, cases increased in the North East of England from a baseline of 13.5 (95% CI 5.83; 26.6) cases per million children in 1995/6 to 51.43 (95% CI 33.59; 75.35) in 2010/11.

The increase in the incidence of empyema in England between 1997 and 2006 was associated with a concurrent increase in bacterial pneumonia. The relationship between the two conditions varied over time with a doubling of the proportion of admissions to hospital with pneumonia complicated by empyema over the same period. Similarly, there were significant variations in the rate of increase of both conditions between different SHAs in England.

Isolations of *S. pneumoniae* and *S. pyogenes* were positive predictors of empyema admissions, suggesting that both pathogens were responsible for the increase in incidence of empyema, with the bigger effect size seen with *S. pneumoniae*. *S. pyogenes* was also a positive predictor of the monthly rate of progression from empyema to pneumonia. Hospital admissions for empyema and the rate of progression between empyema and pneumonia both displayed significant seasonal variation. This seasonal variation correlated with activity of both *S. pneumoniae* and *S. pyogenes*. There was some evidence of cyclicity within the rate of progression of pneumonia to empyema operating at an approximate 3 year cycle. This correlated with activity of *S. pyogenes* indicating that streptococcal pneumonia has a higher rate of progression to empyema than pneumococcal pneumonia.

There was no spatial variation in the risk of hospital admission with empyema in children in the North East of England at postcode district or area level. There was, however, a six-fold variation in the risk of hospital admission with pneumonia in the North East between different postcode districts. These differences were partially explained by variation in levels of family poverty in each district, total numbers of hospital admissions per district, levels of disability living allowance and hospital out-patient attendances. The introduction of PCV-7 to the routine UK immunisation schedule in September 2006 was not associated with a reduction in the monthly incidence of empyema in the North East (co-efficient of the vaccine programme 0.17, 95% CI -0.28; 0.91), nor a change in the age profile of cases.

Children with empyema in the UK between 2006 and 2011 were more likely to be male (56%), under the age of five (median age 4.3 years) and had a median total hospital stay of eleven days. If a cause was found for the empyema it was likely to be pneumococcal and of pneumococcal cases, serotype 1 was the most frequently isolated. Staphylococcal empyema was associated with double the illness length when compared to pneumococcal empyema (Hazard 2.121, 95% CI 1.117; 4.028). Approximately half of all cases received primary medical treatment (46%) and half primary surgical treatment (49%). Those who had primary surgery were older (OR 1.041 95% CI 1.000; 1.083) and more likely to have co-morbidity (OR 2.750 95% CI 1.567; 5.016). Those in the primary surgical treatment group had a significantly shorter total hospital stay (Hazard ratio 1.402 95% CI 1.053; 1.870), lower risk of readmission and any complication.

6.2 Understanding the increase in the incidence of childhood empyema thoracis

One major conclusion from this thesis is that the predominant mechanism underpinning the increase in the incidence of childhood empyema thoracis in England between 1997 and 2006 was an increase in childhood bacterial pneumonia caused by *S. pneumoniae* and to a lesser degree *S. pyogenes*.

The predominance of pneumococcal empyema over staphylococcal empyema is a return to the pattern of causation of empyema seen in the pre-penicillin era (Heuer, 1932; Lionakis *et al.*, 1958). This raises the intriguing question of what factors have led to the resurgence of pneumococcal empyema.

One explanation for the resurgence of pneumococcal infection as the predominant cause of childhood empyema is the spread of a virulent pneumococcal serotype or clone (Byington *et al.*, 2002a). This thesis reports that pneumococcal serotype 1 was the most common cause of paediatric empyema, as have a number of others (Byington *et al.*, 2002b; Eltringham *et al.*, 2003; Byington *et al.*, 2006; Fletcher *et al.*, 2006) raising the possibility that the increase in empyema was driven by expansion of this serotype. Serotype 1 has been observed to cause empyema within an outbreak situation and is recognised to spread epidemically (Kalin M, 1998; Gupta A *et al.*, 2008). In the USA between 1993 and 2002 89% of isolates of serotype 1 causing invasive pneumococcal disease in were multi-locus sequence type (MLST) 227 suggesting a close common lineage (Gonzalez *et al.*, 2004). Similarly all serotype 1 isolates from children with invasive pneumococcal disease in Utah between 1996 and 2002 were sequence type ST227 (Byington *et al.*, 2005).

Obando et al (2008) found that all serotype 1 isolates causing parapneumonic empyema in Spain between 2005 and 2007 were from a major lineage of ST types which included ST227 found across Europe and North America. Byington *et al* (2010) also reported that serotype 1 pneumococci causing paediatric empyema in Utah were related to the same lineage. In addition, clonal expansion of similar MLST types have been associated with an epidemic of meningitis in Ghana (Leimkugel *et al.*, 2005). Some commentators have gone as far as to refer to these sequences types of serotype 1 as behaving near pandemically (Marimon *et al.*, 2009). It is therefore tempting to conclude that the increase in the incidence of paediatric empyema thoracis was driven by clonal expansion and near pandemic spread of certain sequence types of serotype 1 pneumococcus.

What remains uncertain is the mechanisms that determined the expansion of these sequence types. The pneumococcus is recognised to be a highly recombinogenic pathogen with a very active phylogeny and associated high spontaneous mutation rate (Croucher *et al.*, 2011). Different factors can exert selection pressures on pathogens like the pneumococcus. These include anthropological changes such as increasing hygiene and decreasing overcrowding altering pneumococcal transmission, host factors such as immunodeficiency related to HIV infection and clinical factors such as antibiotic usage and pneumococcal vaccination (Croucher *et al.*, 2011). The recognised selection pressures of antibiotic usage and limited valency immunisation appear unlikely to be directly relevant in this case, as the rate of antibiotic resistance remains very low in these sequence types and their spread and the associated increase in incidence of empyema predates the introduction of the pneumococcal conjugate vaccine (Byington *et al.*, 2010). Furthermore, serotype 1 is recognised to cause invasive pneumococcal disease in otherwise healthy individuals, suggesting that host factors are less likely to be relevant (Grau *et al.*, 2012). Perhaps the simplest explanation for the increase in incidence was a

random mutation giving rise to highly virulent clone whose particular expression of virulence factors predisposes to infection of the pleural space. This gave it a significant selection advantage relative to other clones and predisposed it to pandemic spread. At present, our understanding of the biology of serotype 1 and in particular its recent evolution is incomplete and this hypothesis cannot be directly proven but on the basis of the available evidence appears most likely.

Serotype 1 is not the only serotype reported in paediatric empyema. For example, Hsieh *et al* (2004) reported a significant increase in the incidence of paediatric empyema in Taiwan but found no cases of serotype 1. It is also difficult to explain the observed increase in streptococcal empyema in the context of the spread of a virulent pneumococcal clone. Other potential contributory factors that may explain the rise of pneumococcal and streptococcal empyema include reductions in community antibiotic prescribing for respiratory tract infections in children, increasing pneumococcal and streptococcal antibiotic resistance and a decline in invasive staphylococcal infections.

6.3 Projected impact of PCV-13 on paediatric empyema in the UK

In April 2010 PCV-7 was replaced by PCV-13 in the routine immunisation schedule of the UK. Inferred from the limited data available about the pneumococcal serotype distribution in empyema, the projected reduction from PCV-7 was an approximately 20% reduction in cases (Fletcher *et al.*, 2006). The potential impact of PCV-13 is much greater with PCV-13 serotypes accounting for 96% of cases of culture-negative paediatric empyema in England between 2006 and 2009 (Sheppard *et al.*, 2011). However, the experience in the USA following the introduction of PCV-7 demonstrates a need for caution and continued surveillance. There was a significant expansion in non-PCV-7 serotypes (notable serotype 1, 3 and 19A) causing empyema in children in the years following the introduction of PCV-7 (Byington *et al.*, 2006). Analysis of the MLST types of these pneumococci highlighted that this process was driven by clones that were circulating prior to the introduction of PCV-7 (Byington *et al.*, 2010). The expansion of non-vaccine serotypes into the ecological niche left by vaccine serotypes is but one mechanism of vaccine related serotype replacement. Another potential mechanism is serotype switching or vaccine escape where individual pneumococcal clones change their expression of capsular serotype in response to the selective pressures of limited valency vaccines. This process has already been demonstrated to have occurred in response to PCV-7 in the USA and has been associated with the rise of serotype 19A as the predominant cause of invasive pneumococcal disease in that country (Pelton *et al.*, 2007). Furthermore, these shifts in pneumococcal serotype distribution have serious clinical consequences, as shown by the significant increase in length of

hospital stay seen in serotype 19A empyema cases in this cohort compared to the commonest serotype, serotype 1. The propensity of limited valency conjugate vaccines to act as a selective pressure driving the evolution of pneumococci in empyema has already been demonstrated and careful surveillance of the molecular epidemiology of paediatric empyema will be required for the foreseeable future.

6.4 Moving forward the management of empyema in children

The management of empyema in children remains controversial, largely due to a dearth of high quality evidence. In this cohort, patients managed with primary surgical drainage had significantly reduced hospital stay, complication and readmission rates. Furthermore, there were significant differences in length of stay between UK centres. These centre based differences need further investigation to ensure that all patients receive the highest quality care and best outcomes.

This analysis has demonstrated with the finding of heterogeneity in patients' response to surgery and a significant centre effect. Comparisons of management in empyema are complex and can be challenging to interpret. Outcomes measures such as length of stay are not the only factors relevant to patients and their families, and there is a need to develop patient related outcome measures in this condition. Data regarding the acceptability of surgery to patients and their families are also needed.

Two large scale multi-centre RCT's of treatment of pleural infection in adults in the UK have been carried out in recent years. Assessment of the optimal management of paediatric empyema needs to be investigated in a similar fashion with national collaboration (Maskell *et al.*, 2005; Rahman *et al.*, 2011). The UK-ESPE study has shown the potential for multi-centre co-operation between centres in the UK managing this condition, and will hopefully be the beginning of further fruitful collaborations in this area.

Future work needs to be broad based, employ a range of methodologies and address a series of interlinked problems. Firstly, there is a need to define what outcomes in empyema matter most to patients and their families. This question is likely to require mixed approach with both quantitative and qualitative approaches needed. Qualitative interview methods can better identify outcomes that may not have been previously considered relevant in empyema. Quantitative methods could then be used to develop and standardise tools for the measurement of these outcomes. Secondly, appropriately

controlled randomised trials are needed to evaluate current management using the identified outcomes, in particular comparisons of primary surgery versus medical management, minimally invasive surgery versus open surgery and early surgery versus later surgery are required. Newer therapies used in adults, such as combination therapy with intrapleural DNAse and fibrinolytics also need investigation in children. Single centre retrospective reviews need to be avoided given the significant bias from centre effects. Finally, longer term outcomes in empyema need careful investigation. As discussed previously, data on the long term effects of empyema on lung function and respiratory health are currently very limited. While the assumption is that the long term outlook is excellent for the respiratory health of children who develop empyema, there is increasing evidence that the antecedents of adult respiratory disease begin in childhood and the outcomes in empyema need investigating longitudinally in cohorts of sufficient size to detect potential effects. The UK-ESPE study has identified a large prospective cohort of children with empyema and can provide the basis for both the identification of patient related and clinically relevant outcomes in paediatric empyema and the longitudinal study of the long term outcomes of the condition.

There is also the possibility of further changes in the clinical profile of empyema which need consideration in the context of the management of the condition. In this study, serotype 19A was associated prolonged hospital stay. Other studies have found other non-PCV-7 serotypes associated with increased risk of necrotic complications of empyema, such as bronchopleural fistula formation (McKee *et al.*, 2011). If disease due to non-vaccine serotypes continues to rise and the experience remains similar then arguments for earlier definitive treatment become stronger. Future developments in the management of empyema need to account for changes in the clinical profile of the disease.

In conclusion, further research in this area should comprise on-going careful clinical epidemiological studies longitudinally evaluating clinical changes in pleural infection in children. An extension, perhaps concentrated in several sentinel centres of the current UK-ESPE study could fulfil this role. In conjunction with this a detailed study of the molecular epidemiology of the pneumococci causing paediatric empyema using non-culture methods e.g. non-culture MLST typing to map further evolution of this group of pathogens in particular in response to the introduction of pneumococcal conjugate vaccines. With regard to improving the management of the condition, a small scale study using quantitative and qualitative methods is required to identify patient-related outcome measures (a small pilot study is underway in Newcastle) which will then feed into a series of UK wide RCT's aimed at establishing the role of surgery in the management of empyema.

6.5 **Conclusions**

The increase in the incidence of paediatric empyema in England was driven predominantly by an increase in pneumococcal and streptococcal pneumonia re-establishing the epidemiological picture seen in the pre-antibiotic era. The increase in pneumococcal pneumonia appears likely to have been driven by the near pandemic spread of virulent pneumococcal serotype 1 clones. Although no reduction in the incidence of paediatric empyema was seen in North East England in response to PCV-7, the potential impact of broader valency vaccines on paediatric empyema, in particular PCV-13 remains very promising. Careful surveillance will be required however to monitor changes both in the pneumococcal serotypes causing paediatric empyema but also any change in the clinical profile of the disease associated with the introduction of this vaccine. Primary surgery in empyema allowed earlier discharge in this cohort, but further research is needed to establish which outcomes are most acceptable to patients and their families, what the longer term outcomes of empyema are and how best to manage the condition.

7 List of Publications, Presentations and Conference Abstracts

Publications

Thomas MF, Sheppard CL, Guiver M, Slack MP, George RC, Gorton R, Paton JY, Simmister C, Cliff D, Elemraid MA, Clark JE, Rushton SP, Spencer DA. Emergence of pneumococcal 19A empyema in UK children. Arch Dis Child. 2012 Oct 16. [Epub ahead of print] PubMed PMID: 23076341.

Thomas MF, Spencer DA. Necrotising pneumonia, pneumatoceles and the pneumococcus. Thorax. 2012 Oct;67(10):925. Epub 2012 Feb 16. PubMed PMID: 22343709.

Thomas MF, Spencer DA. Management and complications of pneumonia. Paediatr and Child Health. 2011; 21: 207-12

Elemraid MA, Pollard K, **Thomas MF**, Gennery AR, Eastham KM, Rushton SP, Hampton F, Singleton P, Gorton R, Spencer DA, Clark JE; North East of England Paediatric Respiratory Infection Study Group. Validity of using Hospital Episode Statistics data on monitoring disease trends. Thorax. 2011 Sep;66(9):827; author reply 827-8. Epub 2010 Dec 2. PubMed PMID: 21131299.

Presentations and conference abstracts

MF Thomas, C Simmister, D Cliff, MA Elemraid, JE Clark, SP Rushton, R Gorton, JY Paton, DA Spencer. The UK-ESPE study: Paediatric empyema in the UK. *Oral presentation, Royal College of Paediatrics and Child Health Annual Meeting, July 2012*

M F Thomas, C Simmister, D Cliff, M A Elemraid, J E Clark, S P Rushton, R Gorton, J Y Paton, D A Spencer. Comparison of primary pleural drainage strategies in paediatric empyema. *Extended poster presentation, British Thoracic Society Winter Meeting, December 2011*

M F Thomas, C L Sheppard, M Guiver, R C George, C Simmister, D Cliff, R Gorton, M A Elemraid, J E Clark, S P Rushton, J Y Paton, D A Spencer. Changes in pneumococcal serotype distribution of paediatric empyema in the age of pneumococcal conjugate vaccines. *Extended poster presentation, British Thoracic Society Winter Meeting, December 2011*

D A Spencer, **M F Thomas**, C L Sheppard, M Guiver, R C George, R Gorton, J Y Paton, C Simmister, D Cliff, M A Elemraid, J E Clark, S P Rushton. Emergence of pneumococcal serotype 19A as a cause of severe complicated pneumonia with empyema in children in England. *Extended poster presentation, British Thoracic Society Winter Meeting, December 2011*

Thomas MF, Rushton SP, Shirley MDF, Elemraid MA, Clark JE, Gorton R, Spencer DA. Understanding the changing epidemiology of paediatric empyema: the relationship with pneumonia. *Poster presentation, Health Protection 2011, May 2011*

Thomas MF, Blain AP, Shirley MDF, Simmister C, Elemraid MA, Gorton R, Pearce MS, Clark JE, Rushton SP, Spencer DA. Geographical variation in the risk of childhood pneumonia and relationships to socio-economic and health deprivation. *Poster presentation, European Respiratory Society Annual Congress, September 2011*

Spencer DA, Close AJ, Simmister C, Cliff D, Elemraid MA, Clark JE, Rushton SP, **Thomas MF**. Limited impact of the 7-valent pneumococcal vaccine on paediatric empyema in the North of England. *Elite poster presentation European Society of Paediatric Infectious Diseases Congress, June 2011*

Spencer DA, Close AJ, Simmister C, Cliff D, Elemraid MA, Clark JE, Rushton SP, **Thomas MF**. Impact of the 13-valent pneumococcal vaccine on the incidence of paediatric empyema in the North

of England. *Elite poster presentation European Society of Paediatric Infectious Diseases Congress, June 2011*

Thomas MF, Simmister C, Rushton SP, Spencer DA. The changing incidence of paediatric empyema in NE England 2006 – 2010. *Poster presentation, British Thoracic Society Winter Meeting, December 2010*

Thomas MF, Cliff D, Rushton SP, Gorton R, Shirley MDF, Clark J, Spencer DA. Trends in paediatric pneumonia and empyema in England 1997-2006. *Oral presentation, European Respiratory Society Annual Congress, September 2010*

8 References

Ali, M., Emch, M., Tofail, F. and Baqui, A.H. (2001) 'Implications of health care provision on acute lower respiratory infection mortality in Bangladeshi children', *Social Science & Medicine*, 52(2), pp. 267-277.

Ali, N.J., Sillis, M., Andrews, B.E., Jenkins, P.F. and Harrison, B.D.W. (1986) 'The Clinical Spectrum and Diagnosis of Mycoplasma pneumoniae Infection', *QJM*, 58(3-4), pp. 241-251.

Allen, F. (2010) 'The first draft of DSM-V', *BMJ*, 340.

Andersen, P.K., Klein, J.P. and Zhang, M.-J. (1999) 'Testing for centre effects in multi-centre survival studies: a Monte Carlo comparison of fixed and random effects tests', *Statistics in Medicine*, 18(12), pp. 1489-1500.

Andrade, A.L., Toscano, C.M., Minamisava, R., Costa, P.S. and Andrade, J.G. (2011) 'Pneumococcal disease manifestation in children before and after vaccination: What's new?', *Vaccine*, 29, Supplement 3(0), pp. C2-C14.

Antony, V.B. (2003) 'Immunological mechanisms in pleural disease', *ERJ*, 21(3), pp. 539-544.

Aragon, T. (2010) *epitools: Epidemiology Tools* (Version R package version 0.5-6.) [Computer program]. Available at: http://CRAN.R-project.org/package=epitools.

Austrian R, Douglas RM, Schiffman G, Coetzee AM, Koornhof HJ, Hayden-Smith S and Reid RD (1976) 'Prevention of pneumococcal pneumonia by vaccination', *Trans Assoc Am Physicians*, 89, pp. 184-94.

Avansino, J.R., Goldman, B., Sawin, R.S. and Flum, D.R. (2005) 'Primary Operative Versus Nonoperative Therapy for Pediatric Empyema: A Meta-analysis', *Pediatrics*, 115(6), pp. 1652-1659.

Balfour-Lynn, I.M., Abrahamson, E., Cohen, G., Hartley, J., King, S., Parikh, D., Spencer, D., Thomson, A.H. and Urquhart, D. (2005) 'BTS guidelines for the management of pleural infection in children', *Thorax*, 60(SUPPL. 1), pp. i1-i21.

Barnes, N.P., Hull, J. and Thomson, A.H. (2005) 'Medical management of parapneumonic pleural disease', *Pediatr Pulmonol*, 39(2), pp. 127-134.

Beale, L., Abellan, J.J., Hodgson, S. and Jarup, L. (2008) 'Methodologic Issues and Approaches to Spatial Epidemiology', *Environ Health Perspect*, 116(8).

Bekri, H., Cohen, R., Varon, E., Madhi, F., Gire, R., Guillot, F. and Delacourt, C. (2007) '[Streptococcus pneumoniae serotypes involved in children with pleural empyemas in France]', *Arch Pediatr*, 14(3), pp. 239-43.

Besag, J., York, J. and Mollié, A. (1991) 'Bayesian image restoration, with two applications in spatial statistics', *Annals of the Institute of Statistical Mathematics*, 43(1), pp. 1-20.

Bessen, D.E. (2009) 'Population biology of the human restricted pathogen, Streptococcus pyogenes', *Infection, Genetics and Evolution*, 9(4), pp. 581-593.

Blaschke, A.J., Heyrend, C., Byington, C.L., Obando, I., Vazquez-Barba, I., Doby, E.H., Korgenski, E.K., Sheng, X., Poritz, M.A., Daly, J.A., Mason, E.O., Pavia, A.T. and Ampofo, K. (2011) 'Molecular Analysis Improves Pathogen Identification and Epidemiologic Study of Pediatric Parapneumonic Empyema', *Pediatr Infect Dis J* , 30(4), pp. 289-294 10.1097/INF.0b013e3182002d14.

Boersma, W.G., Löwenberg, A., Holloway, Y., Kuttschrütter, H., Snijder, J.A. and Koëter, G.H. (1993) 'Rapid detection of pneumococcal antigen in pleural fluid of patients with community acquired pneumonia', *Thorax*, 48(2), pp. 160-162.

Box, G., Jenkins, G. and Reinsel, G. (2008) *Time Series Analysis: Forecasting and control*. Hoboken: Wiley.

Box, G.E.P. and Cox, D.R. (1964) 'An Analysis of Transformations', *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2), pp. 211-252.

Bradshaw, J., Bloor, K., Huby, M., Rhodes, D., Sinclair, I., Gibbs, I., Noble, M., McLennan, D. and Wilkinson, K. (2009) *Local index of child well-being: Summary report*. Government, D.o.C.a.L. [Online]. Available at: http://www.communities.gov.uk/publications/communities/childwellbeing2009 (Accessed: 18 February 2011).

Brims, F.J.H., Lansley, S.M., Waterer, G.W. and Lee, Y.C.G. (2010) 'Empyema thoracis: new insights into an old disease', *European Respiratory Review*, 19(117), pp. 220-228.

Brook, I. (1990) 'Microbiology of empyema in children and adolescents', *Pediatrics*, 85(5), pp. 722 - 726.

Buckingham, S.C., King, M.D. and Miller, M.L. (2003) 'Incidence and etiologies of complicated parapneumonic effusions in children, 1996 to 2001', *Pediatr Infect Dis J*, 22(6), pp. 499-504.

Byington, C.L., Castillo, H., Gerber, K., Daly, J.A., Brimley, L.A., Adams, S., Christenson, J.C. and Pavia, A.T. (2002a) 'The Effect of Rapid Respiratory Viral Diagnostic Testing on Antibiotic Use in a Children's Hospital', *Arch Pediatr Adolesc Med*, 156(12), pp. 1230-1234.

Byington, C.L., Hulten, K.G., Ampofo, K., Sheng, X., Pavia, A.T., Blaschke, A.J., Pettigrew, M., Korgenski, K., Daly, J. and Mason, E.O. (2010) 'Molecular Epidemiology of Pediatric Pneumococcal Empyema from 2001 to 2007 in Utah', *J. Clin. Microbiol.*, 48(2), pp. 520-525.

Byington, C.L., Korgenski, K., Daly, J., Ampofo, K., Pavia, A. and Mason, E.O. (2006) 'Impact of the pneumococcal conjugate vaccine on pneumococcal parapneumonic empyema', *Pediatr Infect Dis J*, 25(3), pp. 250 - 254.

Byington, C.L., Samore, M.H., Stoddard, G.J., Barlow, S., Daly, J., Korgenski, K., Firth, S., Glover, D., Jensen, J., Mason, E.O., Shutt, C.K. and Pavia, A.T. (2005) 'Temporal Trends of Invasive Disease Due to Streptococcus pneumoniae among Children in the Intermountain West: Emergence of Nonvaccine Serogroups', *Clin Infect Dise*, 41(1), pp. 21-29.

Byington, C.L., Spencer, L.Y., Johnson, T.A., Pavia, A.T., Allen, D., Mason, E.O., Kaplan, S., Carroll, K.C., Daly, J.A., Christenson, J.C. and Samore, M.H. (2002b) 'An epidemiological investigation of a sustained high rate of pediatric parapneumonic empyema: Risk factors and microbiological associations', *Clin Infect Dis*, 34(4), pp. 434-440.

Calder, A. and Owens, C.M. (2009) 'Imaging of parapneumonic pleural effusions and empyema in children', *Pediatric Radiology*, 39(6), pp. 527-537.

Callery, P., Kyle, R.G., Campbell, M., Banks, M., Kirk, S. and Powell, P. (2010) 'Readmission in children's emergency care: an analysis of hospital episode statistics', *Arch Dis Child*, 95(5), pp. 341-346.

Campbell, S.E., Campbell, M.K., Grimshaw, J.M. and Walker, A.E. (2001) 'A systematic review of discharge coding accuracy', *J Public Health*, 23(3), pp. 205-211.

Cazelles, B., Chavez, M., Berteaux, D., Ménard, F., Vik, J., Jenouvrier, S. and Stenseth, N. (2008) 'Wavelet analysis of ecological time series', *Oecologia*, 156(2), pp. 287-304.

Cazelles, B., Chavez, M., Magny, G.C.d., Guégan, J.-F. and Hales, S. (2007) 'Time-dependent spectral analysis of epidemiological time-series with wavelets', *Journal of The Royal Society Interface*, 4(15), pp. 625-636.

Cazelles, B., Chavez, M., McMichael, A.J. and Hales, S. (2005) 'Nonstationary Influence of El Niño on the Synchronous Dengue Epidemics in Thailand', *PLoS Med*, 2(4), p. e106.

Chan, P.W.K., Chew, F.T., Tan, T.N., Chua, K.B. and Hooi, P.S. (2002) 'Seasonal variation in respiratory syncytial virus chest infection in the tropics', *Pediatr Pulmonol*, 34(1), pp. 47-51.

Charland, K.M., Brownstein, J.S., Verma, A., Brien, S. and Buckeridge, D.L. (2011) 'Socio-economic disparities in the burden of seasonal influenza: the effect of cocial and material deprivation on rates of influenza infection', *PLoS ONE*, 6(2), p. e17207.

Chernick, V., Boat, T., Wilmott, R. and Bush, A. (eds.) (2006) *Kendig's Disorders of the Respiratory Tract in Children*. Seventh Edition edn. Philadelphia: Saunders Elsevier.

Chonmaitree, T. and Powell, K.R. (1983) 'Parapneumonic pleural effusion and empyema in children. Review of a 19-year experience, 1962-1980', *Clinical Pediatrics*, 22(6), pp. 414-9.

Christopoulou-Aletra, H. and Papavramidou, N. (2008) '"Empyemas" of the Thoracic Cavity in the Hippocratic Corpus', *Ann Thorac Surg*, 85, pp. 1132-1134.

Clagett, O.T. (1943) 'Chronic empyema', *Journal of Thoracic Surgery*, 12, pp. 464-83.

Clagett, O.T. (1973) 'Changing aspect of the etiology and treatment of pleural empyema', *Surgical Clinics of North America*, 53, pp. 863-66.

Clark, J.E., Hammal, D., Hampton, F., Spencer, D. and Parker, L. (2007) 'Epidemiology of community-acquired pneumonia in children seen in hospital', *Epidemiol and Infect*, 135(2), pp. 262-269.

Cliff, A.D. and Smallman-Raynor, M.R. (1992) 'The Aids Pandemic: Global Geographical Patterns and Local Spatial Processes', *The Geographical Journal*, 158(2), pp. 182-198.

Cohen, J. (1998) *Statistical power analysis for the behavioural sciences*. 2nd edn. New York: Psychology Press.

Cohen, S. (1999) 'Social Status and Susceptibility to Respiratory Infections', *Annals of the New York Academy of Sciences*, 896(1), pp. 246-253.

Cole, D. (1937) 'Management of Empyema', *Chest*, 3, pp. 40-42.

Colley, J.R.T., Holland, W.W. and Corkhill, R.T. (1974) 'Influence of passive smoking and parental phlegm on pneumonia and bronchitis in early childhood', *Lancet*, 304(7888), pp. 1031-1034.

Colman, G., Tanna, A., Efstratiou, A. and Gaworzewska, E.T. (1993) 'The serotypes of Streptococcus pyogenes present in Britain during 1980–1990 and their association with disease', *Journal of Medical Microbiology*, 39(3), pp. 165-178.

Cook, D.G. and Strachan, D.P. (1999) 'Summary of effects of parental smoking on the respiratory health of children and implications for research', *Thorax*, 54(4), pp. 357-366.

Cooper, B. and Lipsitch, M. (2004) 'The analysis of hospital infection data using hidden Markov models', *Biostatistics*, 5(2), pp. 223-237.

Coote, N. and Kay, E. (2005) 'Surgical versus non-surgical management of pleural empyema', *Cochrane Data Sys Rev*, (4).

Crighton, E.J., Elliott, S.J., Moineddin, R., Kanaroglou, P. and Upshur, R. (2007) 'A spatial analysis of the determinants of pneumonia and influenza hospitalisations in Ontario (1992-2001)', *Social Science & Medicine*, 64(8), pp. 1636-1650.

Cutts, F.T., Zaman, S.M.A., Enwere, G., Jaffar, S., Levine, O.S., Okoko, J.B., Oluwalana, C., Vaughan, A., Obaro, S.K., Leach, A., McAdam, K.P., Biney, E., Saaka, M., Onwuchekwa, U., Yallop, F., Pierce, N.F., Greenwood, B.M. and Adegbola, R.A. (2005) 'Efficacy of nine-valent pneumococcal conjugate vaccine against pneumonia and invasive pneumococcal disease in The Gambia: randomised, double-blind, placebo-controlled trial', *Lancet*, 365(9465), pp. 1139-1146.

Davila, R. and Crouch, E. (1995) 'Anatomic organisation and function of the human pleura', *Semin Respir Crit Care Med*, 16(4), pp. 261-68.

Deiros Bronte, L., Baquero-Artigao, F., Garcia-Miguel, M.J., Hernandez Gonzalez, N., Pena Garcia, P. and del Castillo Martin, F. (2006) '[Parapneumonic pleural effusion: an 11-year review]', *An Pediatr (Barc)*, 64(1), pp. 40-5.

Dethlefsen, C., Lundbye-Christensen, S. and Christensen, A. (2009) *sspir: State Space Models in R* (Version R package version 0.2.8) [Computer program]. Available at: http://CRAN.R-project.org/package=sspir.

Diggle, P. (1990) *Time series: A biostatistical introduction*. Oxford: Clarendon Press.

Dowell, S.F. and Ho, M.S. (2004) 'Seasonality of infectious diseases and severe acute respiratory syndrome–what we don't know can hurt us', *Lancet Infect Dis*, 4(11), pp. 704-708.

Dowell, S.F., Whitney, C.G., Wright, C., Rose, C.E., Jr. and Schuchat, A. (2003) 'Seasonal patterns of invasive pneumococcal disease', *Emerg Infect Dis*, 9(5), pp. 573-9.

Dunham, E.K., Graham, E.A., Mitchell, J.F., Moschcowitz, A.V., Kinsella, R.A., Bell, R.D., Stevens, F.A., Tower, W.L., Hartman, C.C., Rivers, T.M., Zeman, F.D., Cohen, M.B., Hays, M.H., Stocking, B.E. and Jacobs, E.P. (1918) 'Cases of empyema at Camp Lee, VA. Preliminary Report', *JAMA*, 71(5), pp. 366-373.

Eastham, K.M., Freeman, R., Kearns, A.M., Eltringham, G., Clark, J., Leeming, J. and Spencer, D.A. (2004) 'Clinical features, aetiology and outcome of empyema in children in the north east of England', *Thorax*, 59(6), pp. 522-5.

Eastham, K.M., Hammal, D.M., Parker, L. and Spencer, D.A. (2008) 'A follow-up study of children hospitalised with community-acquired pneumonia', *Archives of Disease in Childhood*, 93(9), pp. 755-759.

EDINA '2001 UK Census: Digitised Boundary Data (England and Wales)'. 21 February 2011. Census Geography Data Unit (UKBORDERS), EDINA (University of Edinburgh). Available at: http://edina.ac.uk/ukborders/.

Elemraid, M.A., Pollard, K., Thomas, M.F., Gennery, A.R., Eastham, K.M., Rushton, S.P., Hampton, F., Singleton, P., Gorton, R., Spencer, D.A. and Clark, J.E. (2010) 'Validity of using hospital episode statistics data on monitoring disease trends', *Thorax*, [Online]. Available at: http://thorax.bmj.com/content/early/2010/12/02/thx.2010.153551.short DOI: 10.1136/thx.2010.153551 (Accessed: December 2, 2010).

Elliott, P. and Wartenberg, D. (2004) 'Spatial Epidemiology: Current Approaches and Future Challenges', *Environ Health Perspect*, 112(9).

Eltringham, G., Kearns, A., Freeman, R., Clark, J., Spencer, D., Eastham, K., Harwood, J. and Leeming, J. (2003) 'Culture-negative childhood empyema is usually due to penicillin-sensitive Streptococcus pneumoniae capsular serotype 1', *J Clin Microbiol*, 41(1), pp. 521-2.

Epstein, F.H. (1996) 'Cardiovascular Disease Epidemiology : A Journey From the Past Into the Future', *Circulation*, 93(9), pp. 1755-1764.

Erzen, D., Carriere, K.C., Dik, N., Mustard, C., Roos, L.L., Manfreda, J. and Anthonisen, N.R. (1997) 'Income level and asthma prevalence and care patterns', *Am J Respir Crit Care Med*, 155(3), pp. 1060-1065.

Eskola J and Anittila, M. (1999) 'Pneumococcal conjugate vaccines', *Pediatr Infect Dis J*, 18(6), pp. 543-551.

Everitt, B. and Rabe-Hesketh, S. (2001) *Analyzing Medical Data Using S-PLUS*. New York: Springer-Verlag.

Fajardo JE and Chang MJ (1987) 'Pleural empyema in children: A nationwide retrospective study', *Southern Medical Journal*, 80(5), pp. 593-96.

Fanshawe, T.R., Diggle, P.J., Rushton, S., Sanderson, R., Lurz, P.W.W., Glinianaia, S.V., Pearce, M.S., Parker, L., Charlton, M. and Pless-Mulloli, T. (2008) 'Modelling spatio-temporal variation in exposure to particulate matter: a two-stage approach', *Environmetrics*, 19(6), pp. 549-566.

Faraway, J. (2005) 'Checking Error Assumptions', in *Texts in Statistical Science: Linear Models with R*. Boca Raton: Chapman & Hall/CRC, pp. 53-8.

Finley, C., Clifton, J., Fitzgerald, J.M. and Yee, J. (2008) 'Empyema: an increasing concern in Canada', *Can Respir J*, 15(2), pp. 85-9.

Fisman, D.N. 28 (2007) 'Seasonality of infectious diseases'. pp. 127-143. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-34249742034&partnerID=40&md5=43c00c222607ac52614a44d20808af34.

Flasche, S., Slack, E. and Miller, L. (2011) 'Long term trends introduce a potential bias when evaluating the impact of the pneumococcal conjugate vaccination programme in England and Wales', *Eurosurveillance*, 16(20), pp. 1-5.

Fletcher, M., Leeming, J., Cartwright, K. and Finn, A. (2006) 'Childhood empyema: limited potential impact of 7-valent pneumococcal conjugate vaccine', *Pediatr Infect Dis J*, 25(6), pp. 559-60.

Flory, J.H., Joffe, M., Fishman, N.O., Edelstein, P.H. and Metlay, J.P. (2009) 'Socioeconomic risk factors for bacteraemic pneumococcal pneumonia in adults', *Epidemiology and Infection*, 137(05), pp. 717-726.

François, P., Desrumaux, A., Cans, C., Pin, I., Pavese, P. and Labarère, J. (2010) 'Prevalence and risk factors of suppurative complications in children with pneumonia', *Acta Pædiatrica*, 99(6), pp. 861-866.

Geha AS (1970) 'Pleural empyema: Changing etiologic, bacteriologic, and therapeutic aspects', *Journal of Thoracic and Cardiovascular Surgery*, 61(4), pp. 626-635.

Gilthorpe, M.S., Lay-Yee, R., Wilson, R.C., Walters, S., Griffiths, R.K. and Bedi, R. (1998) 'Variations in hospitalisation rates for asthma among Black and minority ethnic communities', *Respiratory Medicine*, 92(4), pp. 642-648.

Gonzalez, B.E., Hulten, K.G., Kaplan, S.L., Mason, E.O. and the, U.S.P.M.P.S.S.G. (2004) 'Clonality of Streptococcus pneumoniae Serotype 1 Isolates from Pediatric Patients in the United States', *Journal Clin Micro*, 42(6), pp. 2810-2812.

Graham, E.A. and Berck, M. (1933) 'Principles versus details in the treatment of acute empyema', *Ann Surg*, 98, pp. 520-27.

Grambsch, P. and Therneau, T. (1994) 'Proportional hazards tests and diagnostics based on weighed residuals', *Biometrika*, 81(3), pp. 515-26.

Grenfell, B.T., Bjornstad, O.N. and Kappey, J. (2001) 'Travelling waves and spatial hierarchies in measles epidemics', *Nature*, 414(6865), pp. 716-723.

Grijalva, C.A.G., Nuorti, J.A P., Zhu, Y. and Griffin, Marie R. (2010) 'Increasing Incidence of Empyema Complicating Childhood Community Acquired Pneumonia in the United States', *Clin Infect Dis*, 50(6), pp. 805-813.

Grijalva, C.G., Nuorti, J.P., Arbogast, P.G., Martin, S.W., Edwards, K.M. and Griffin, M.R. (2007) 'Decline in pneumonia admissions after routine childhood immunisation with pneumococcal conjugate vaccine in the USA: a time-series analysis', *Lancet*, 369(9568), pp. 1179-1186.

Grijalva, C.G., Zhu, Y., Nuorti, J.P. and Griffin, M.R. (2011) 'Emergence of parapneumonic empyema in the USA', *Thorax, 66,* pp 663-668.

Groff DB, Randolph JG and Blades B (1966) 'Empyema in childhood', *JAMA*, 195(7), pp. 164-66.

Gruttola, V.D. and Tu, X.M. (1994) 'Modelling Progression of CD4-Lymphocyte Count and Its Relationship to Survival Time', *Biometrics*, 50(4), pp. 1003-1014.

Gupta A, Khaw FM, Stokle E, George RC, Pebody R, Stansfield C, Sheppard CL, Slack M, Gorton R and Spencer DA (2008) 'Outbreak of Streptococcus pneumoniae serotype 1 pneumonia in a United Kingdom school', *BMJ*, 337.

Gupta, R. and Crowley, S. (2006) 'Increasing paediatric empyema admissions', *Thorax*, 61(2), pp. 179-180.

Hamm, H. and Light, R.W. (1997) 'Parapneumonic effusion and empyema', *ERJ*, 10(5), pp. 1150-1156.

Hardie, W., Bokulic, R., Garcia, V.F., Reising, S.F. and Christie, C.D. (1996) 'Pneumococcal pleural empyemas in children', *Clin Infect Dis*, 22(6), pp. 1057-63.

Harley, R. (1987) 'Anatomy of the pleura', *Seminars in respiratory medicine*, 9, pp. 1-6.

Harrison, L.M., Morris, J.A., Telford, D.R., Brown, S.M. and Jones, K. (1999) 'The nasopharyngeal bacterial flora in infancy: effects of age, gender, season, viral upper respiratory tract infection and sleeping position', *FEMS Immunology & Medical Microbiology*, 25(1-2), pp. 19-28.

Hawker, J.I., Olowokure, B., Sufi, F., Weinberg, J., Gill, N. and Wilson, R.C. (2003) 'Social deprivation and hospital admission for respiratory infection: an ecological study', *Respiratory Medicine*, 97(11), pp. 1219-1224.

Hendrickson, D.J., Blumberg, D.A., Joad, J.P., Jhawar, S. and McDonald, R.J. (2008) 'Five-fold increase in pediatric parapneumonic empyema since introduction of pneumococcal conjugate vaccine', *Pediatr Infect Dis J*, 27(11), pp. 1030-2.

HESonline - Hospital Episode Statistics (2010). Available at: http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=537 (Accessed: 22/04/2010).

Heuer, G.J. (1932) 'Acute empyema', *Journal of Thoracic Surgery*, 1, pp. 461-84.

Hewett, F.C. (1876) 'Thoracentesis: The Plan of Continuous Aspiration', *Bmj*, 1(793), pp. 317-317.

HPA (2012) *Epidemiological data*. Available at: http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/MeningococcalDisease/EpidemiologicalData/ (Accessed: 21/02/2012).

Hsieh, Y.-C., Hsueh, P.-R., Lu, C.-Y., Lee, P.-I., Lee, C.-Y. and Huang, L.-M. (2004) 'Clinical manifestations and molecular epidemiology of necrotizing pneumonia and empyema caused by Streptococcus pneumoniae in children in Taiwan', *Clin Infect Dis*, 38(6), pp. 830-5.

Immunisation Department, C.F.I. 'Cover of Vaccination Evaluated Rapidly (COVER) '. Health Protection Agency.

Ishiguro, T., Takayanagi, N., Ikeya, T., Yoshioka, H., Yanagisawa, T., Hoshi, E., Hoshi, T., Sugita, Y. and Kawabata, Y. (2010) 'Isolation of Candida Species is an Important Clue for Suspecting Gastrointestinal Tract Perforation as a Cause of Empyema', *Internal Medicine*, 49(18), pp. 1957-1964.

Jaffé, A. and Balfour-Lynn, I.M. (2005) 'Management of empyema in children', *Pediatr Pulmonol*, 40(2), pp. 148-156.

Jefferson, T., Ferroni, E., Curtale, F., Giorgi Rossi, P. and Borgia, P. (2006) 'Streptococcus pneumoniae in western Europe: serotype distribution and incidence in children less than 2 years old', *Lancet Infect Dis*, 6(7), pp. 405-410.

Jensen, E., Lundbye-Christensen, S., Samuelsson, S., Sørensen, H. and Carl Schønheyder, H. (2004) 'A 20-year ecological study of the temporal association between influenza and meningococcal disease', *European Journal of Epidemiology*, 19(2), pp. 181-187.

Kalin M (1998) 'Pneumococcal serotypes and their clinical relevance', *Thorax*, 1998(53), pp. 159-62.
Kazembe, L. and Namangale, J. (2007) 'A Bayesian multinomial model to analyse spatial patterns of childhood co-morbidity in Malawi', *European Journal of Epidemiology*, 22(8), pp. 545-556.

Kern, J.A. and Rodgers, B.M. (1993) 'Thoracoscopy in the management of empyema in children', *Journal of Pediatric Surgery*, 28(9), pp. 1128-1132.

Khalil, B., Corbett, P., Jones, M., Baillie, C., Southern, K., Losty, P. and Kenny, S. (2007) 'Less is best? The impact of urokinase as the first line management of empyema thoracis', *Pediatric Surgery International*, 23(2), pp. 129-133.

Kiesewetter WB, Rusnock JR and Girdany BR (1959) 'Pediatric empyema: a second look at its incidence and importance', *Journal of Pediatrics*, 54(81), pp. 81-6.

Kirkwood, B. and Sterne, J. (2003) *Essential Medical Statistics*. Oxford: Blackwells Publishing.
Kitagawa, G. (2010) *Introduction to Time Series Modelling*. Boca Raton: Chapman & Hall.

Ko, S.-C., Chen, K.-Y., Hsueh, P.-R., Luh, K.-T. and Yang, P.-C. (2000) 'Fungal Empyema Thoracis', *Chest*, 117(6), pp. 1672-1678.

Kohn, G., Walston, C., Feldstein, J., Warner, B., Succop, P. and Hardie, W. (2002) 'Persistent abnormal lung function after childhood empyema', *Am J Respir Med*, 1(6), pp. 441-5.

Kokoska, E.R. and Chen, M.K. (2009) 'Position paper on video-assisted thoracoscopic surgery as treatment of pediatric empyema', *Journal of Pediatric Surgery*, 44(1), pp. 289-293.

Koshy, E., Murray, J., Bottle, A., Sharland, M. and Saxena, S. (2010) 'Impact of the seven-valent pneumococcal conjugate vaccination (PCV7) programme on childhood hospital admissions for

bacterial pneumonia and empyema in England: national time-trends study, 1997–2008', *Thorax*, 65(9), pp. 770-774.

Kurt, B.A., Winterhalter, K.M., Connors, R.H., Betz, B.W. and Winters, J.W. (2006) 'Therapy of Parapneumonic Effusions in Children: Video-Assisted Thoracoscopic Surgery Versus Conventional Thoracostomy Drainage', *Pediatrics*, 118(3), pp. e547-e553.

Lahti, E., Mertsola, J., Kontiokari, T., Eerola, E., Ruuskanen, O. and Jalava, J. (2006) 'Pneumolysin polymerase chain reaction for diagnosis of pneumococcal pneumonia and empyema in children', *European Journal of Clinical Microbiology & Infectious Diseases*, 25(12), pp. 783-789.

Lamagni TL, Neal S, Keshishian C, Alhaddad N, George R, Duckworth G, Vuopio-Varkila J and Efstratiou, A. (2008) 'Severe Streptococcus pyogenes infections, United Kingdom, 2003–2004', *Emerg Infect Dis*, 14(2), pp. 201-209.

Law, D.C.G., Serre, M.L., Christakos, G., Leone, P.A. and Miller, W.C. (2004) 'Spatial analysis and mapping of sexually transmitted diseases to optimise intervention and prevention strategies', *Sexually Transmitted Infections*, 80(4), pp. 294-299.

Le Monnier, A., Carbonnelle, E., Zahar, J.R., Le Bourgeois, M., Abachin, E., Quesne, G., Varon, E., Descamps, P., De Blic, J., Scheinmann, P., Berche, P. and Ferroni, A. (2006) 'Microbiological diagnosis of empyema in children: comparative evaluations by culture, polymerase chain reaction, and pneumococcal antigen detection in pleural fluids', *Clin Infect Dis*, 42(8), pp. 1135 - 1140.

Lee, G.E., Lorch, S.A., Sheffler-Collins, S., Kronman, M.P. and Shah, S.S. (2010) 'National Hospitalisation Trends for Pediatric Pneumonia and Associated Complications', *Pediatrics*, 126(2), pp. 204-213.

Leimkugel, J., Adams Forgor, A., Gagneux, S., Pflüger, V., Flierl, C., Awine, E., Naegeli, M., Dangy, J.-P., Smith, T., Hodgson, A. and Pluschke, G. (2005) 'An Outbreak of Serotype 1 Streptococcus pneumoniae Meningitis in Northern Ghana with Features That Are Characteristic of Neisseria meningitidis Meningitis Epidemics', *J of Infect Dis*, 192(2), pp. 192-199.

Lewis, R.A. and Feigin, R.D. (2002) 'Current issues in the diagnosis and management of pediatric empyema', *Seminars in Pediatric Infectious Diseases*, 13(4), pp. 280-288.

Li, S.-T.T. and Gates, R.L. (2008) 'Primary Operative Management for Pediatric Empyema: Decreases in Hospital Length of Stay and Charges in a National Sample', *Arch Pediatr Adolesc Med*, 162(1), pp. 44-48.

Li, S.-T.T. and Tancredi, D.J. (2010) 'Empyema Hospitalisations Increased in US Children Despite Pneumococcal Conjugate Vaccine', *Pediatrics*, 125(1), pp. 26-33.

Light, R.W., Macgregor, M.I., Luchsinger, P.C. and Ball, W.C. (1972) 'Pleural Effusions: The Diagnostic Separation of Transudates and Exudates', *Annals of Internal Medicine*, 77(4), pp. 507-513.

Lionakis, B., Gray, S.W., Skandalakis, J.E. and Hopkins, W.A. (1958) 'Empyema in children: A twenty-five-year study', *The Journal of Pediatrics*, 53(6), pp. 719-725.

Lipsitch, M. (1999) 'Bacterial Vaccines and Serotype Replacement: Lessons from Haemophilus influenzae and Prospects for Streptococcus pneumoniae', *Emerg Infect Dis*, 5(3), pp. 336-345.

Lundbye-Christensen, S., Dethlefsen, C., Gorst-Rasmussen, A., Fischer, T., Schønheyder, H., Rothman, K. and Sørensen, H. (2009) 'Examining secular trends and seasonality in count data using dynamic generalized linear modelling: a new methodological approach illustrated with hospital discharge data on myocardial infarction', *European Journal of Epidemiology*, 24(5), pp. 225-230.

MacLeod, C.M., Hodges, R.G., Heidelberger, M. and Bernhard, W.G. (1945) 'PREVENTION OF PNEUMOCOCCAL PNEUMONIA BY IMMUNIZATION WITH SPECIFIC CAPSULAR POLYSACCHARIDES', *The Journal of Experimental Medicine*, 82(6), pp. 445-465.

Mahon, M. and Kibirige, M.S. (2004) 'Patterns of admissions for children with special needs to the paediatric assessment unit', *Arch Dis Child*, 89(2), pp. 165-169.

Mann, S.L., Wadsworth, M.E. and Colley, J.R. (1992) 'Accumulation of factors influencing respiratory illness in members of a national birth cohort and their offspring', *Journal of Epidemiology and Community Health*, 46(3), pp. 286-292.

Markowitz, J., Pashko, S., Gutterman, E., Linde-Zwirble, W. and Newbold, R. (1996) 'Death rates among patients hospitalized with community-acquired pneumonia: a reexamination with data from three states', *Am J Public Health.*, 86(8), pp. 1152-4.

Maskell, N.A., Davies, C.W.H., Nunn, A.J., Hedley, E.L., Gleeson, F.V., Miller, R., Gabe, R., Rees, G.L., Peto, T.E.A., Woodhead, M.A., Lane, D.J., Darbyshire, J.H. and Davies, R.J.O. (2005) 'U.K. Controlled Trial of Intrapleural Streptokinase for Pleural Infection', *N Eng J Med*, 352(9), pp. 865-874.

Mayer, J. (1983) 'The role of spatial analysis and geographic data in the detection of disease causation', *Social Science & Medicine*, 17(16), pp. 1213-1221.

McCarthy, P., Byrne, D., Harrisson, S. and Keithley, J. (1985) 'Respiratory conditions: effect of housing and other factors', *Journal of Epidemiology and Community Health*, 39(1), pp. 15-19.

McKee, A.J., Ives, A. and Balfour-Lynn, I.M. (2011) 'Increased incidence of bronchopulmonary fistulas complicating pediatric pneumonia', *Pediatr Pulmonol*, 46(7), pp. 717-721.

McLaughlin, F., Goldmann, D., Rosenbaum, D., Harris, G., Schuster, S. and Strieder, D. (1984) 'Empyema in children: clinical course and long-term follow-up', *Pediatrics.*, 73(5), pp. 587-93.

Melegaro, A., Edmunds, W.J., Pebody, R., Miller, E. and George, R. (2006) 'The current burden of pneumococcal disease in England and Wales', *Journal of Infection*, 52(1), pp. 37-48.

Menezes-Martins, L.F., Menezes-Martins, J.J., Michaelsen, V.S., Aguiar, B.B., Ermel, T. and Machado, D.C. (2005) 'Diagnosis of parapneumonic pleural effusion by polymerase chain reaction in children', *Journal of Pediatric Surgery*, 40(7), pp. 1106-1110.

Merino, J.M., Carpintero, I., Alvarez, T., Rodrigo, J., Sánchez, J. and Coello, J.M. (1999) 'Tuberculous Pleural Effusion in Children', *Chest*, 115(1), pp. 26-30.

Mew, R., Jaffe, A., Biassoni, L. and Sonnappa, S. (2009) 'Ventilation-perfusion scans in children treated for empyema', *Thorax.*, 64(3), p. 273.

Mid-year population estimates (2011). Available at: http://www.ons.gov.uk/ons/publications/all-releases.html?definition=tcm:77-22371 (Accessed: 22/03/2011).

MIDAS 'UK MIDAS Land Surface Stations Data '. 10/01/2012. United Kingdom Met Office. Available at: http://badc.nerc.ac.uk/data/dataset_index/.

Middelkamp, J.N., Purkerson, M.L. and Burford, T.H. (1964) 'The Changing Pattern of Empyema Thoracis in Pediatrics', *Journal of Thoracic & Cardiovascular Surgery*, 47, pp. 165-73.

Miller, E., Andrews, N.J., Waight, P.A., Slack, M.P.E. and George, R.C. (2011) 'Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: an observational cohort study', *Lancet Infect Dis*, 11(10), pp. 760-768.

Miller, E., Waight, P., Efstratiou, A., Brisson, M., Johnson, A. and George, R. (2000) 'Epidemiology of invasive and other pneumococcal disease in children in England and Wales 1996–1998', *Acta Pædiatr*, 89, pp. 11-16.

Molnar, T.F., Hasse, J., Jeyasingham, K. and Rendeki, S. (2004) 'Changing dogmas: history of development in treatment modalities of traumatic pneumothorax, hemothorax, and posttraumatic empyema thoracis', *Ann Thorac Surg*, 77(1), pp. 372-378.

Morens, D.M., Taubenberger, J.K. and Fauci, A.S. (2008) 'Predominant Role of Bacterial Pneumonia as a Cause of Death in Pandemic Influenza: Implications for Pandemic Influenza Preparedness', *Journal of Infectious Diseases*, 198(7), pp. 962-970.

Musser, J.M., Kapur, V., Szeto, J., Pan, X., Swanson, D.S. and Martin, D.R. (1995) 'Genetic diversity and relationships among Streptococcus pyogenes strains expressing serotype M1 protein: recent intercontinental spread of a subclone causing episodes of invasive disease', *Infection and Immunity*, 63(3), pp. 994-1003.

O'Brien, K.L., Moulton, L.H., Reid, R., Weatherholtz, R., Oski, J., Brown, L., Kumar, G., Parkinson, A., Hu, D., Hackell, J., Chang, I., Kohberger, R., Siber, G. and Santosham, M. (2003) 'Efficacy and safety of seven-valent conjugate pneumococcal vaccine in American Indian children: group randomised trial', *Lancet*, 362(9381), pp. 355-361.

Obando, I., Munoz-Almagro, C., Arroyo, L.A., Tarrago, D., Sanchez-Tatay, D., Moreno-Perez, D., Dhillon, S.S., Esteva, C., Hernandez-Bou, S., Garcia-Garcia, J.J., Hausdorff, W.P. and Brueggemann, A.B. (2008) 'Pediatric parapneumonic empyema, Spain', *Emerg Infect Dis*, 14(9), pp. 1390-7.

Obaro, S.K., Adegbola, R.A., Banya, W.A. and Greenwood, B.M. (1996) 'Carriage of pneumococci after pneumococcal vaccination', *Lancet*, 348(9022), pp. 271-2.

Odell, J. (1994) 'Management of empyema thoracis', *JRSM*, 87, pp. 466-470.

Owayed, A.F., Campbell, D.M. and Wang, E.E.L. (2000) 'Underlying causes of recurrent pneumonia in children', *Arch Pediatr & Adol Med*, 154(2), pp. 190-194.

Padman, R., King, K.A., Iqbal, S. and Wolfson, P.J. (2007) 'Parapneumonic Effusion and Empyema in Children: Retrospective Review of the duPont Experience', *Clinical Pediatrics*, 46(6), pp. 518-522.

Paul, M., Held, L. and Toschke, A.M. (2008) 'Multivariate modelling of infectious disease surveillance data', *Statistics in Medicine*, 27(29), pp. 6250-6267.

Pelton, S.I., Huot, H., Finkelstein, J.A., Bishop, C.J., Hsu, K.K., Kellenberg, J., Huang, S.S., Goldstein, R. and Hanage, W.P. (2007) 'Emergence of 19A as Virulent and Multidrug Resistant Pneumococcus in Massachusetts Following Universal Immunization of Infants With Pneumococcal Conjugate Vaccine', *Pediatr Infect Dis J*, 26(6), pp. 468-472 10.1097/INF.0b013e31803df9ca.

Penberthy, G.C. and Benson, C. (1936) 'A ten year study of empyema in children:1926-1936', *Ann Surg*, 104, pp. 579-84.

Peters, R. (1989) 'Empyema thoracis: Historical perspective', *The Annals of Thoracic Surgery*, 48(2), pp. 306-308.

Picazo, J., Ruiz-Contreras, J., Hernandez, B., Sanz, F., Gutierrez, A., Cercenado, E., Meseguer, M.A., Delgado-Iribarren, A., Rodriguez-Avial, I. and Méndez, C. (2011) 'Clonal and clinical profile of Streptococcus pneumoniae serotype 19A causing pediatric invasive infections: A 2-year (2007–2009) laboratory-based surveillance in Madrid', *Vaccine*, 29(9), pp. 1770-1776.

Pinhero, J.C. and Bates, D.M. (2000) *Mixed-Effects Models in S and S-Plus*. 1st edn. New York: Springer.

Pinhero, J.C., Bates, D.M., DebRoy, S. and Sarkar, D. (2011) *nlme: Linear and Nonlinear Mixed Effects Models* (Version R package version 3.1-102) [Computer program].

Playfor, S.D., Smyth, A.R. and Stewart, R.J. (1997) 'Increase in incidence of childhood empyema', *Thorax*, 52(10), p. 932.

Polnay, L. and Ward, H. (2000) 'Promoting the health of looked after children', *BMJ*, 320(7236), pp. 661-662.

Proesmans, M. and De Boeck, K. (2009) 'Clinical practice: Treatment of childhood empyema', *European Journal of Pediatrics*, 168(6), pp. 639-645.

'Programme budgeting tools and data' (2011). 21/02/2012. UK Government: Department of Health. Available at: http://www.dh.gov.uk/en/Managingyourorganisation/Financeandplanning/Programmebudgeting/DH_075743#_3.

R-Development-Core-Team (2011) *R: A Language and Environment for Statistical Computing* (Version 2.14) [Computer program]. R Foundation for Statistical Computing. Available at: http://www.R-project.org.

Rahman, N.M., Maskell, N.A., West, A., Teoh, R., Arnold, A., Mackinlay, C., Peckham, D., Davies, C.W.H., Ali, N., Kinnear, W., Bentley, A., Kahan, B.C., Wrightson, J.M., Davies, H.E., Hooper, C.E., Lee, Y.C.G., Hedley, E.L., Crosthwaite, N., Choo, L., Helm, E.J., Gleeson, F.V., Nunn, A.J. and Davies, R.J.O. (2011) 'Intrapleural Use of Tissue Plasminogen Activator and DNase in Pleural Infection', *NEJM*, 365(6), pp. 518-526.

Rasmussen JN, V.M., Andersen RL, Ellermann-Eriksen S, Jensen TG, Johansen HK, Kolmos B, Mølvadgaard M, Nielsen SS, Olsen E, Schønning K, Uldum SA (2010) 'Increased incidence of Mycoplasma pneumoniae infections detected by laboratory-based surveillance in Denmark in 2010', *Euro Surveill.*, 15(45).(pii), p. 19708.

Ravitch MM and Fein R (1961) 'The changing picture of pneumonia and empyema in infants and children', *JAMA*, 175(12), pp. 87-91.

Redding, G., Walund, L., Walund, D., Jones, J., Stamey, D. and Gibson, R. (1990) 'Lung function in children following empyema', *Am J Dis Child.*, 144(12), pp. 1337-42.

Rees, J.H., Spencer, D.A., Parikh, D. and Weller, P. (1997) 'Increase in incidence of childhood empyema in West Midlands, UK', *Lancet*, 349(9049), p. 402.

Reinert, R., Jacobs, M.R. and Kaplan, S.L. (2010) 'Pneumococcal disease caused by serotype 19A: Review of the literature and implications for future vaccine development', *Vaccine*, 28(26), pp. 4249-4259.

Richardson S, Thomson A, Best N and Elliott P (2004) 'Interpreting posterior relative risk estimates in disease-mapping studies', *Environmental Health Perspectives.*, 112(9), pp. 1016-25.

Rohani, P., Green, C.J., Mantilla-Beniers, N.B. and Grenfell, B.T. (2003) 'Ecological interference between fatal diseases', *Nature*, 422(6934), pp. 885-888.

Rothman, K., Greenland, S. and Lash, T. (2008) *Modern Epidemiology*. Third edn. Philadelphia: Lippincott Williams & Wilkins.

Roxburgh, C.S., Youngson, G.G., Townend, J.A. and Turner, S.W. (2008) 'Trends in pneumonia and empyema in Scottish children in the past 25 years', *Arch Dis Child*, 93(4), pp. 316 - 318.

Royston, P. (1982) 'An extension of Shapiro and Wilk's W test for normality to large samples', *Applied Statistics*, 31, pp. 115-124.

Rudan, I., Boschi-Pinto, C., Biloglav, Z., Mulholland, K. and Campbell, H. (2008) 'Epidemiology and etiology of childhood pneumonia', *Bull World Health Organ*, 86(5), pp. 408-16.

Ruoff, K.L. (1988) 'Streptococcus anginosus (''Streptococcus milleri''): the unrecognised pathogen', *Clin Microbiol Rev*, 1, pp. 102-08.

Saglani, S., Harris, K.A., Wallis, C. and Hartley, J.C. (2005) 'Empyema: the use of broad range 16S rDNA PCR for pathogen detection', *Arch Dis Child*, 90(1), pp. 70-73.

Samson, P.C. (1971) 'Empyema Thoracis: Essentials of Present-Day Management', *Ann Thorac Surg*, 11(3), pp. 210-221.

Sarihan, H., Cay, A., Aynaci, M., Akyazici, R. and Baki, A. (1998) 'Empyema in children', *J Cardiovasc Surg (Torino)*. 39(1), pp. 113-6.

Sarkar, D. (2008) *Lattice: Multivariate Data Visualisation with R* [Computer program].

Schultz, K.D., Fan, L.L., Pinsky, J., Ochoa, L., Smith, E.O., Kaplan, S.L. and Brandt, M.L. (2004) 'The changing face of pleural empyemas in children: epidemiology and management', *Pediatrics*, 113(6), pp. 1735-40.

Schwartz, L., Flippin, H. and Turnbull, W. (1939) 'Treatment of pneumococci pneumonia: a comparative study of 351 patients treated at the Philadelphia General Hospital', *Ann Intern Med*, 13, pp. 1005-12.

Sethi, S. and Murphy, T.F. (2001) 'Bacterial infection in chronic obstructive pulmonary disease in 2000: a state-of-the-art review', *Clin Microbiol Rev*, 14(2), pp. 336-363.

Shah, S.S., Hall, M., Newland, J.G., Brogan, T.V., Farris, R.W.D., Williams, D.J., Larsen, G., Fine, B.R., Levin, J.E., Wagener, J.S., Conway, P.H. and Myers, A.L. (2011) 'Comparative effectiveness of pleural drainage procedures for the treatment of complicated pneumonia in childhood', *J Hosl Med*, 6(5), pp. 256-263.

Shankar, K.R., Kenny, S.E., Okoye, B.O., Carty, M.L., Lloyd, D.A. and Losty, P.D. (2000) 'Evolving experience in the management of empyema thoracis', *Acta Pædiatrica*, 89(4), pp. 417-420.

Sheppard, C.L., Guiver, M., Hartley, J., Harrison, T.G. and George, R.C. (2011) 'The use of a multiplexed immunoassay for detection of serotype-specific Streptococcus pneumoniae antigen in pleural fluid and CSF specimens', *J Med Microbiol*, 60(12), pp 1879-81.

Shomaker, K.L., Weiner, T. and Esther, C.R. (2011) 'Impact of an evidence-based algorithm on quality of care in pediatric parapneumonic effusion and empyema', *Pediatr Pulmonol*, 46(7), pp. 722-728.

Sills, M.R., Huang, Z.J., Shao, C., Guagliardo, M.F., Chamberlain, J.M. and Joseph, J.G. (2000) 'Pediatric Milliman and Robertson Length-of-Stay Criteria: Are They Realistic?', *Pediatrics*, 105(4), pp. 733-737.

Singh, M., Mathew, J.L., Chandra, S., Katariya, S. and Kumar, L. (2004) 'Randomized controlled trial of intrapleural streptokinase in empyema thoracis in children', *Acta Pædiatrica*, 93(11), pp. 1443-1445.

Singleton, R.J., Hennessy, T.W., Bulkow, L.R., Hammitt, L.L., Zulz, T., Hurlburt, D.A., Butler, J.C., Rudolph, K. and Parkinson, A. (2007) 'Invasive pneumococcal disease caused by nonvaccine serotypes among Alaska Native children with high levels of 7-valent pneumococcal conjugate vaccine coverage', *JAMA*, 297(16), pp. 1784-1792.

Snow, S.J. (2008) 'John Snow: the making of a hero?', *The Lancet*, 372(9632), pp. 22-23.
Sogaard, M., Kornum, J., Schonheyder, H. and Thomsen, R. (2011) 'Positive predictive value of the ICD-10 hospital diagnosis of pleural empyema in the Danish National Registry of Patients', *Clin Epidemiol.*, 3, pp. 85-9.

Sonnappa, S., Cohen, G., Owens, C.M., van Doorn, C., Cairns, J., Stanojevic, S., Elliott, M.J. and Jaffe, A. (2006) 'Comparison of Urokinase and Video-assisted Thoracoscopic Surgery for Treatment of Childhood Empyema', *Am J Respir Crit Care Me.*, 174(2), pp. 221-227.

Spencer, D., Iqbal, S., Hamilton, J. and Hasan, A. (2006a) 'Twin Peaks: The changing epidemiology of complicated pneumonia and empyema in children', *British Thoracic Society Winter Meeting*. London. p. ii96. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2104788/.

Spencer, D.A., Iqbal, S.M., Hasan, A. and Hamilton, L. (2006b) 'Empyema thoracis is still increasing in UK children', *Bmj*, 332(7553), p. 1333.

Spencer, N., Logan, S., Scholey, S. and Gentle, S. (1996) 'Deprivation and bronchiolitis', *Archives of Disease in Childhood*, 74(1), pp. 50-52.

Spiegelhalter DJ, Thomas A, Best NG and D, L. (2001) *WinBUGS, Version 1.4* [Computer program]. MRC Biostatistics Unit. Available at: http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Linde, A.v.d. (2002) 'Bayesian Measures of Model Complexity and Fit', *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 64(4), pp. 583-639.

St. Peter, S.D., Tsao, K., Harrison, C., Jackson, M.A., Spilde, T.L., Keckler, S.J., Sharp, S.W., Andrews, W.S., Holcomb Iii, G.W. and Ostlie, D.J. (2009) 'Thoracoscopic decortication vs tube thoracostomy with fibrinolysis for empyema in children: a prospective, randomized trial', *Journal of Pediatric Surgery*, 44(1), pp. 106-111.

Stovroff, M., Teague, G., Heiss, K.F., Parker, P. and Ricketts, R.R. (1995) 'Thoracoscopy in the management of pediatric empyema', *Journal of Pediatric Surgery*, 30(8), pp. 1211-1215.

Strachan, R. and Jaffe, A. (2009) 'Assessment of the burden of paediatric empyema in Australia', *J Paediatr Child Health*, 45(7-8), pp. 431-6.

Team, R.D.C. (2010) *R: A language and environment for statistical computing*. [Computer program]. R Foundation for Statistical Computing. Available at: http://www.R-project.org/.

Therneau, T. (2011) *Survival analysis, including penalised likelihood* (Version 2.36-10) [Computer program]. Available at: http://r-forge.r-project.org.

Therneau, T. and Grambsch, P. (2000) *Modelling survival data - Extending the Cox model*. New York Springer.

Thomas, M., Cliff, D., Beaton, S., Rushton, S., Paton, J. and Spencer, D. (2009) 'Paediatric empyema management in the UK', *Thorax*, 64 Suppl. 4.

Thomas, M., Simmister, C., Cliff, D., Elemraid, M., Clark, J., Rushton, S., Gorton, R., Paton, J. and Spencer, D. (2011) 'Comparison of primary pleural drainage strategies in paediatric empyema ', *Thorax*, 66 Suppl. 4, pp. A137-A138

Thompson, A., Reid, A., Shields, M., Steen, H. and Taylor, R. (1999) 'Increased incidence in childhood empyema thoracis in Northern Ireland', *Ir Med J*, 92(7), p. 438.

Thompson, L.D., Edwards, J.C. and Hoagland, C.L. (1940) 'Experiences in the treatment of lobar pneumonia', *Ann Intern Med*, 13, pp. 1138-49.

Thompson, P.L., Spyridis, N., Sharland, M., Gilbert, R.E., Saxena, S., Long, P.F., Johnson, A.P. and Wong, I.C.K. (2009) 'Changes in clinical indications for community antibiotic prescribing for children in the UK from 1996 to 2006: will the new NICE prescribing guidance on upper respiratory tract infections just be ignored?', *Arch Dis Child*, 94(5), pp. 337-340.

Thomson, A.H., Hull, J., Kumar, M.R., Wallis, C. and Balfour Lynn, I.M. (2002) 'Randomised trial of intrapleural urokinase in the treatment of childhood empyema', *Thorax*, 57(4), pp. 343-347.

Tian, H. and Cazelles, B. (2011) *WaveletCo: Wavelet Coherence Analysis.* (Version R package version 1.0) [Computer program]. Available at: http://CRAN.R-project.org/package=WaveletCo.

Tillet, W.S., Cambier, M.J. and McCormack, J.E. (1944) 'The treatment of lobar pneumonia and pneumococcal empyema with penicillin', *Bull N Y Acad Sci*, 20, pp. 142-78.

Tokuda, Y., Matsushima, D., Stein, G.H. and Miyagi, S. (2006) 'Intrapleural Fibrinolytic Agents for Empyema and Complicated Parapneumonic Effusions:A Meta-analysis', *Chest*, 129(3), pp. 783-790.

Tsang, K., Leung, W., Chan, V., Lin, A. and Chu, C. (2007) 'Complicated parapneumonic effusion and empyema thoracis: microbiology and predictors of adverse outcomes', *Hong Kong Medical Journal*, 13, pp. 178-86.

Van Ackere, T., Proesmans, M., Vermeulen, F., Van Raemdonck, D. and De Boeck, K. (2009) 'Complicated parapneumonic effusion in Belgian children: increased occurrence before routine pneumococcal vaccine implementation', *Eur J Pediatr*, 168(1), pp. 51-58.

van der Poll, T. and Opal, S.M. (2011) 'Pathogenesis, treatment, and prevention of pneumococcal pneumonia', *The Lancet*, 374(9700), pp. 1543-1556.

Venables, W. and Ripley, B. (2002) *MASS: Modern applied statistics with S* [Computer program].

Virkki, R., Juven, T., Rikalainen, H., Svedström, E., Mertsola, J. and Ruuskanen, O. (2002) 'Differentiation of bacterial and viral pneumonia in children', *Thorax*, 57(5), pp. 438-441.

Wagner, A.K., Soumerai, S.B., Zhang, F. and Ross-Degnan, D. (2002) 'Segmented regression analysis of interrupted time series studies in medication use research', *J Clin Pharm Thera*, 27(4), pp. 299-309.

Waites, K.B. and Talkington, D.F. (2004) 'Mycoplasma pneumoniae and Its Role as a Human Pathogen', *Clin Microbiol Rev*, 17(4), pp. 697-728.

Wang, N. (1985) 'Anatomy and physiology of the pleural space', *Clinics in chest medicine*, 6(1), pp. 3-16.

Wardlaw, T., Salama, P., Johansson, E.W. and Mason, E. (2006) 'Pneumonia: the leading killer of children', *Lancet*, 368(9541), pp. 1048-1050.

Weinberger, D.M., Harboe, Z.B., Sanders, E.A.M., Ndiritu, M., Klugman, K.P., Rückinger, S., Dagan, R., Adegbola, R., Cutts, F., Johnson, H.L., O'Brien, K.L., Scott, J.A. and Lipsitch, M. (2010) 'Association of Serotype with Risk of Death Due to Pneumococcal Pneumonia: A Meta-Analysis', *Clin Infect Dis*, 51(6), pp. 692-699.

Wexler, I.D., Knoll, S., Picard, E., Villa, Y., Shoseyov, D., Engelhard, D. and Kerem, E. (2006) 'Clinical characteristics and outcome of complicated pneumococcal pneumonia in a pediatric population', *Pediatr Pulmonol*, 41(8), pp. 726 - 734.

WHO (1990) *International Classification of Diseases (ICD) 10* Available at: http://www.who.int/classifications/icd/en/ (Accessed: 22/04/2010).

Wilkinson, R.G. (1997) 'Socioeconomic determinants of health: health inequalities: relative or absolute material standards?', *BMJ*, 314(7080), p. 591.

Williams, D.J., Hall, M., Brogan, T.V., Farris, R.W.D., Myers, A.L., Newland, J.G. and Shah, S.S. (2011) 'Influenza Coinfection and Outcomes in Children With Complicated Pneumonia', *Arch Pediatr Adolesc Med*, 165(6), pp. 506-512.

Wilson, M., E. (1995) 'Travel and the emergence of infectious diseases', *Emerg Infect Dis.*, 1(2), pp. 39-46.

Wolfe WG, Spock A and Bradford WD (1968) 'Pleural fluid in infants and children', *American Review of Respiratory Disease*, 98, pp. 1027-32.

Zhang, Z., Cazelles, B., Tian, H., Christian Stige, L., Bräuning, A. and Stenseth, N.C. (2009) 'Periodic temperature-associated drought/flood drives locust plagues in China', *Proceedings of the Royal Society B: Biological Sciences*, 276(1658), pp. 823-831.

Zocchi, L. (2002) 'Physiology and pathophysiology of pleural fluid turnover', *ERJ*, 20(6), pp. 1545-1558.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. and Smith, G.M. (2009) *Mixed Effects Models and Extensions in Ecology with R*. New York: Springer.

9 Appendices

9.1 Appendix A - ICD-10 Codes for bacterial pneumonia

ICD-10 Codes	Bacterial pneumonia
J13	Pneumonia due to Streptococcus pneumoniae
J14	Pneumonia due to Haemophilus influenzae
J15.0	Pneumonia due to Klebsiella pneumoniae
J15.1	Pneumonia due to Pseudomonas
J15.2	Pneumonia due to staphylococcus
J15.3	Pneumonia due to streptococcus, group B
J15.4	Pneumonia due to other streptococci
J15.5	Pneumonia due to Escherichia coli
J15.6	Pneumonia due to other aerobic Gram-negative bacteria
J15.7	Pneumonia due to Mycoplasma pneumoniae
J15.8	Other bacterial pneumonia
J15.9	Bacterial pneumonia, unspecified
J18.1	Lobar pneumonia, unspecified

Table 9.1 ICD-10 codes included in the definition of pneumonia.

9.2 Appendix B - Model diagnostic plots for changes in the incidence of empyema in the NE (Models T0-T1)

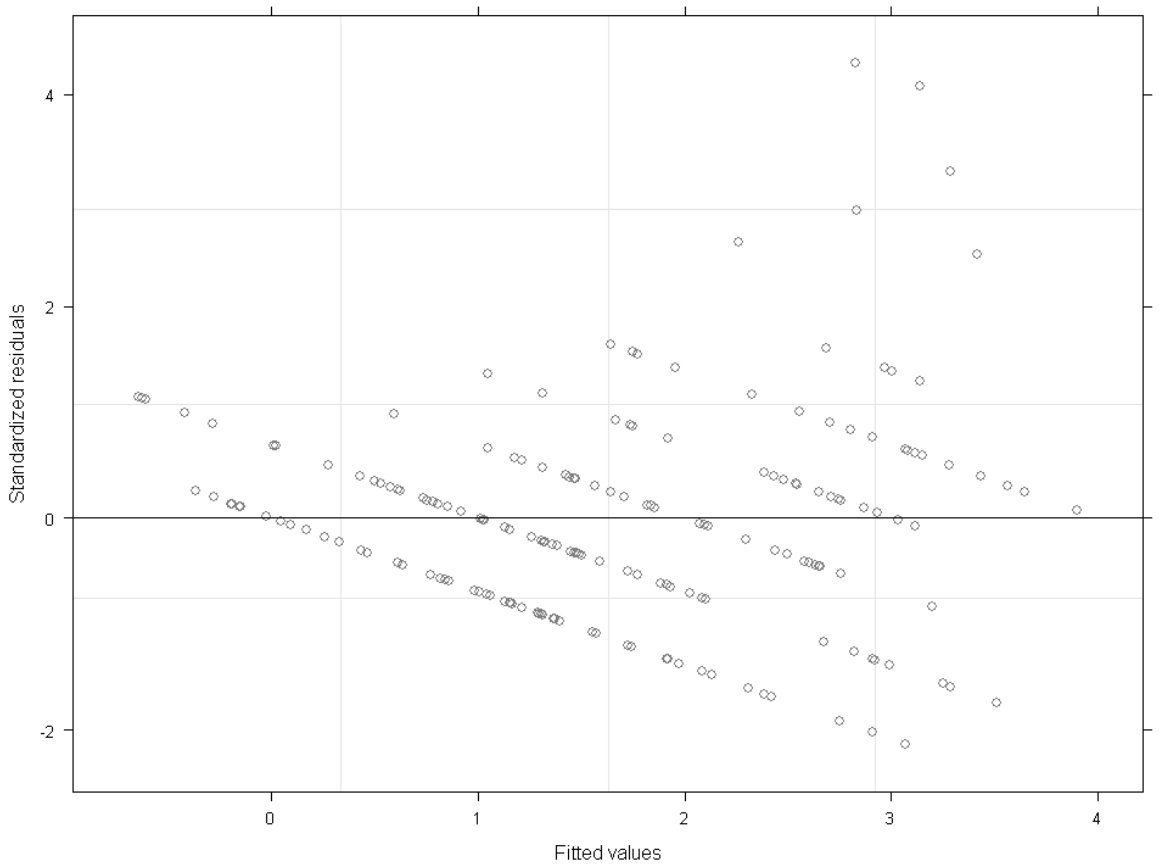


Figure 9.1 Scatter plot of residual and fitted values to assess homoscedascity, independence and variance of the residual distribution for model T0. Distribution of the residuals reflects the constraints and lack of variation in numbers of cases per months (Range: 0 – 9 cases).

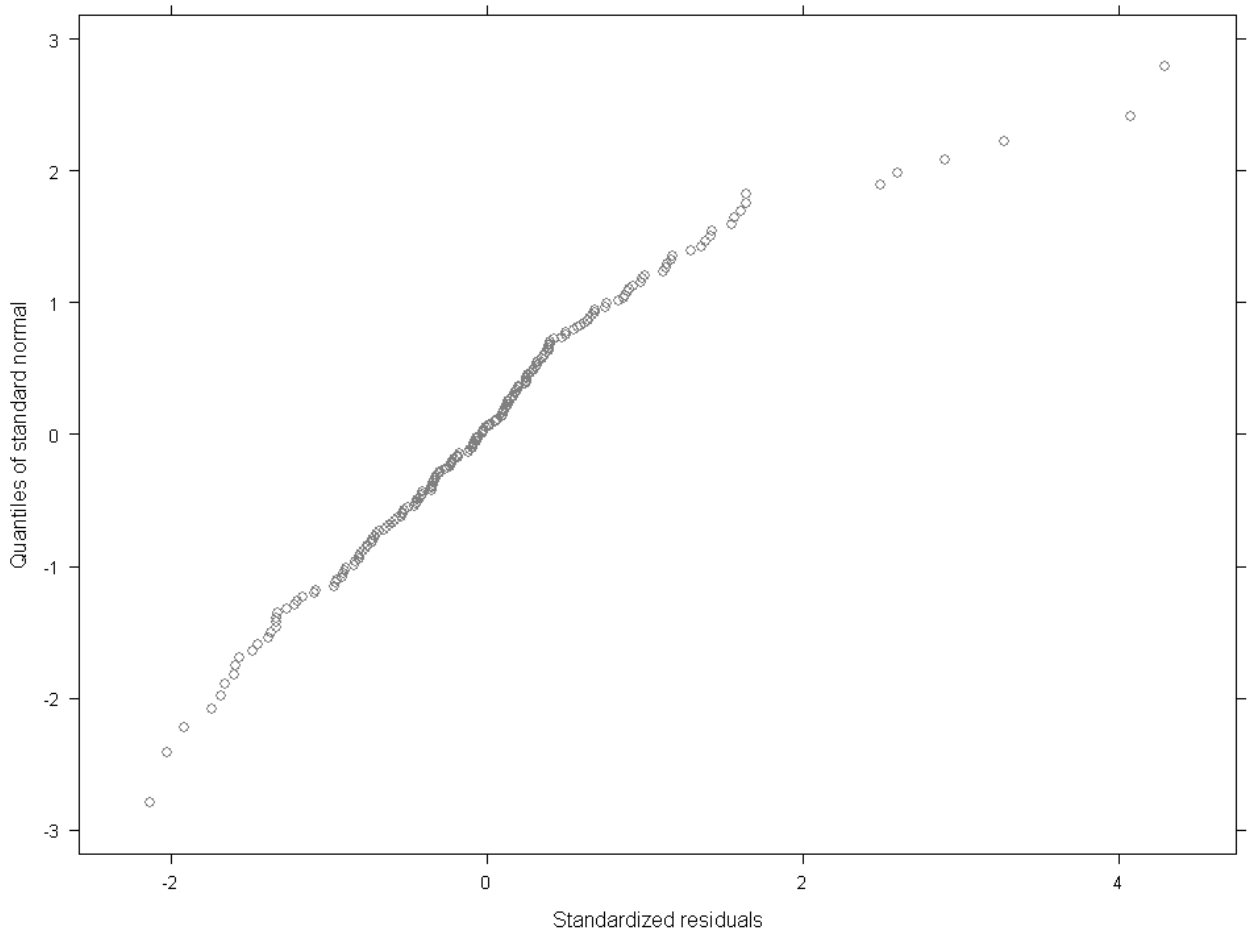


Figure 9.2 Normal probability plot of standardised residuals to assess distribution of errors and detect any evidence of non-normal residual distribution and abnormal variance from model T0.

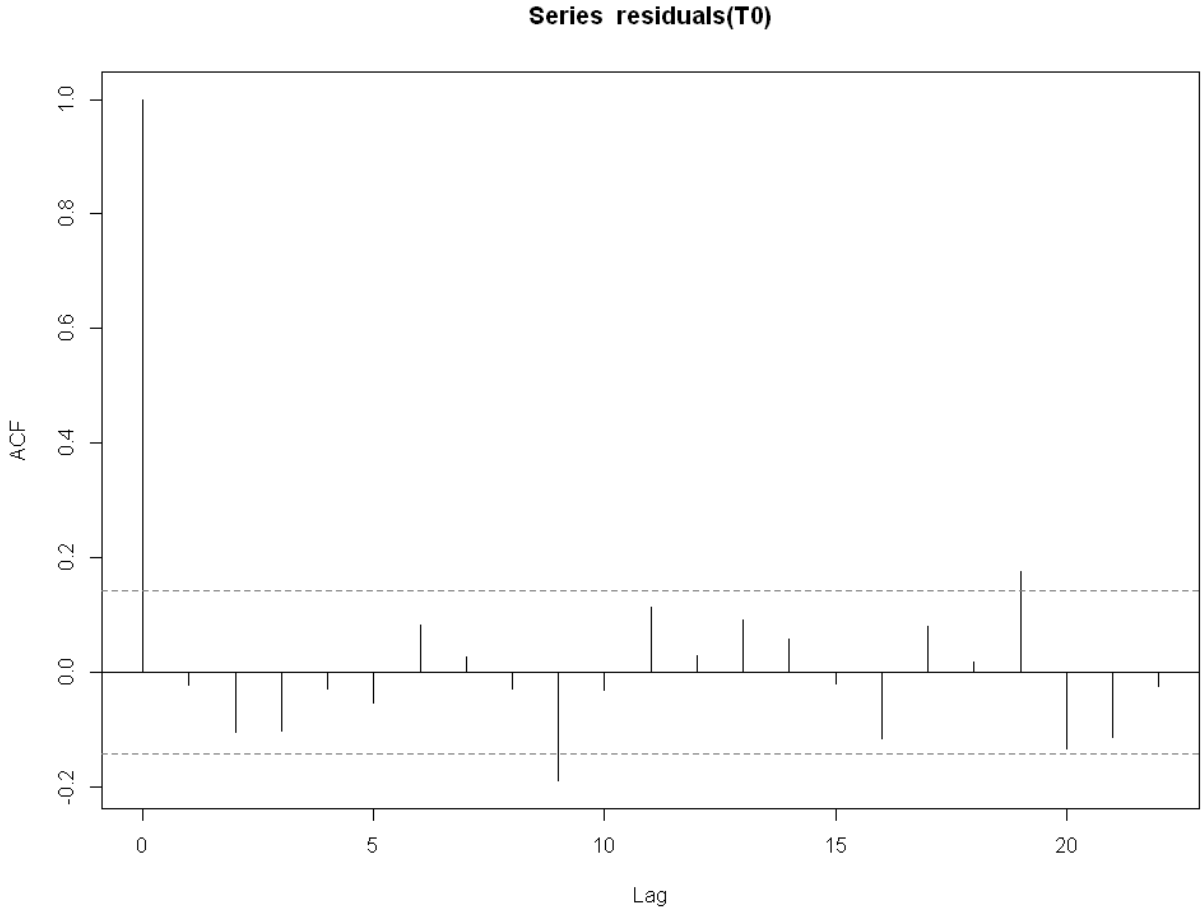


Figure 9.3 Autocorrelation plot of model T0 showing evidence of significant autocorrelation at lags 9 and 19. ***Dotted lines indicate cut-off for presences of significant autocorrelation. ACF – autocorrelation function (size of autocorrelation, lag indicates number of months between observations when correlation occurs)***

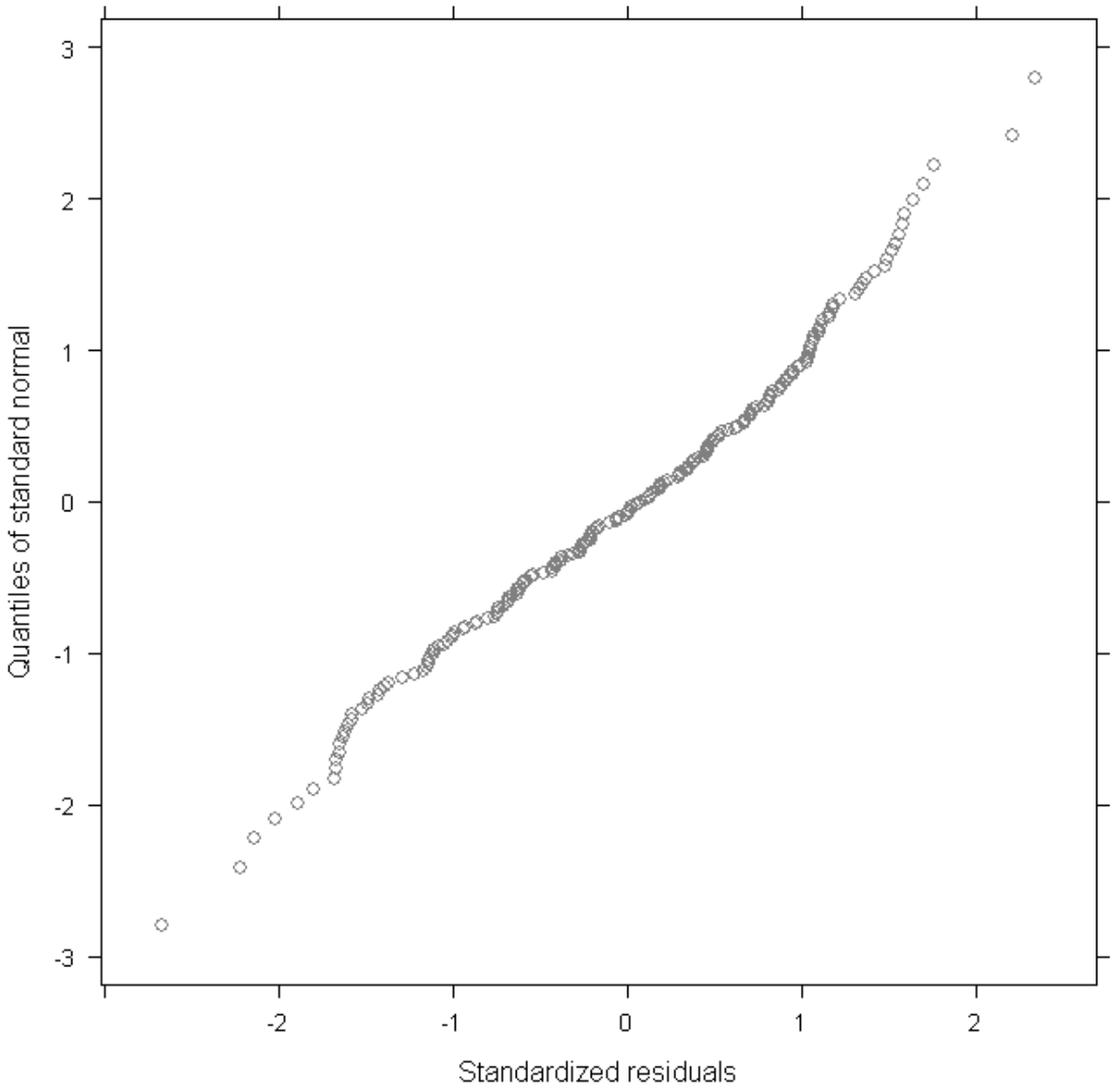


Figure 9.4 Normal probability plot of standardised residuals to assess distribution of errors and detect any evidence of non-normal residual distribution and abnormal variance from model T1.

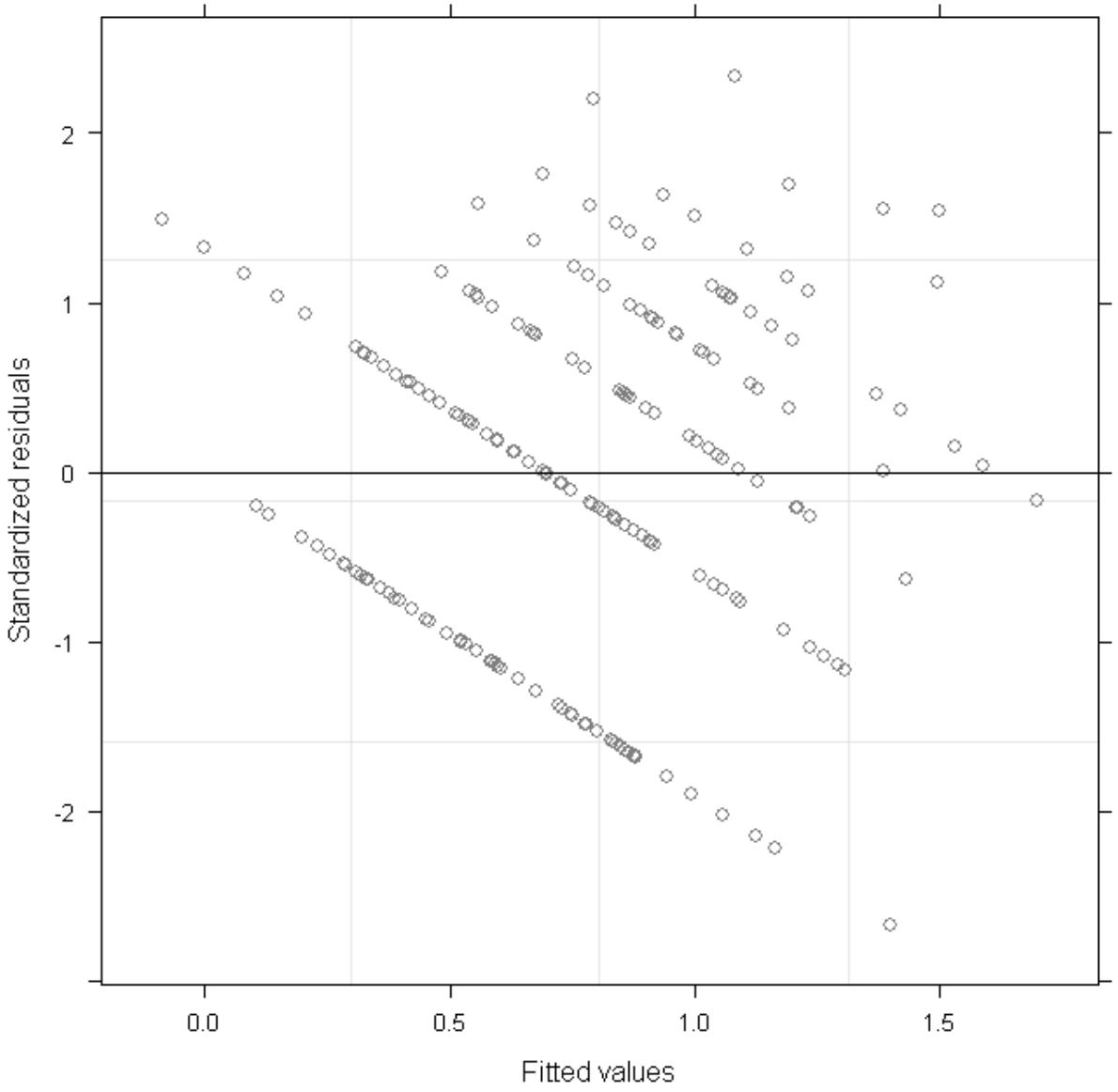


Figure 9.5 Scatter plot of residual and fitted values to assess homoscedascity, independence and variance of the residual distribution for model T1.

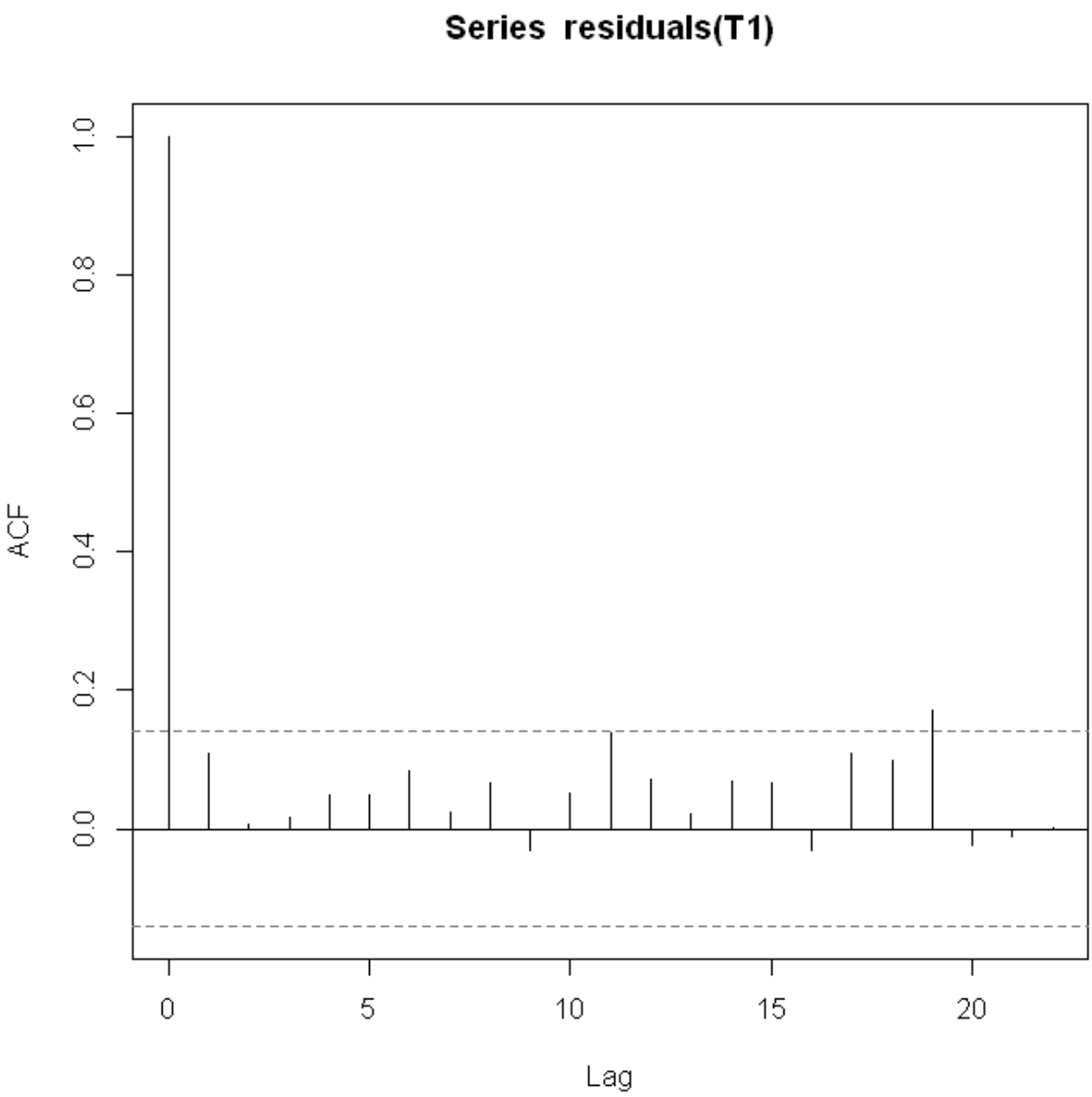


Figure 9.6 Autocorrelation plot of model T1, with lag 19 of borderline significance. ***Dotted lines indicate cut-off for presences of significant autocorrelation. ACF – autocorrelation function (size of autocorrelation, lag indicates number of months between observations when correlation occurs)***

9.3 Appendix C - Model diagnostic plots for national trends in empyema and pneumonia (Models M0a/b – M4)

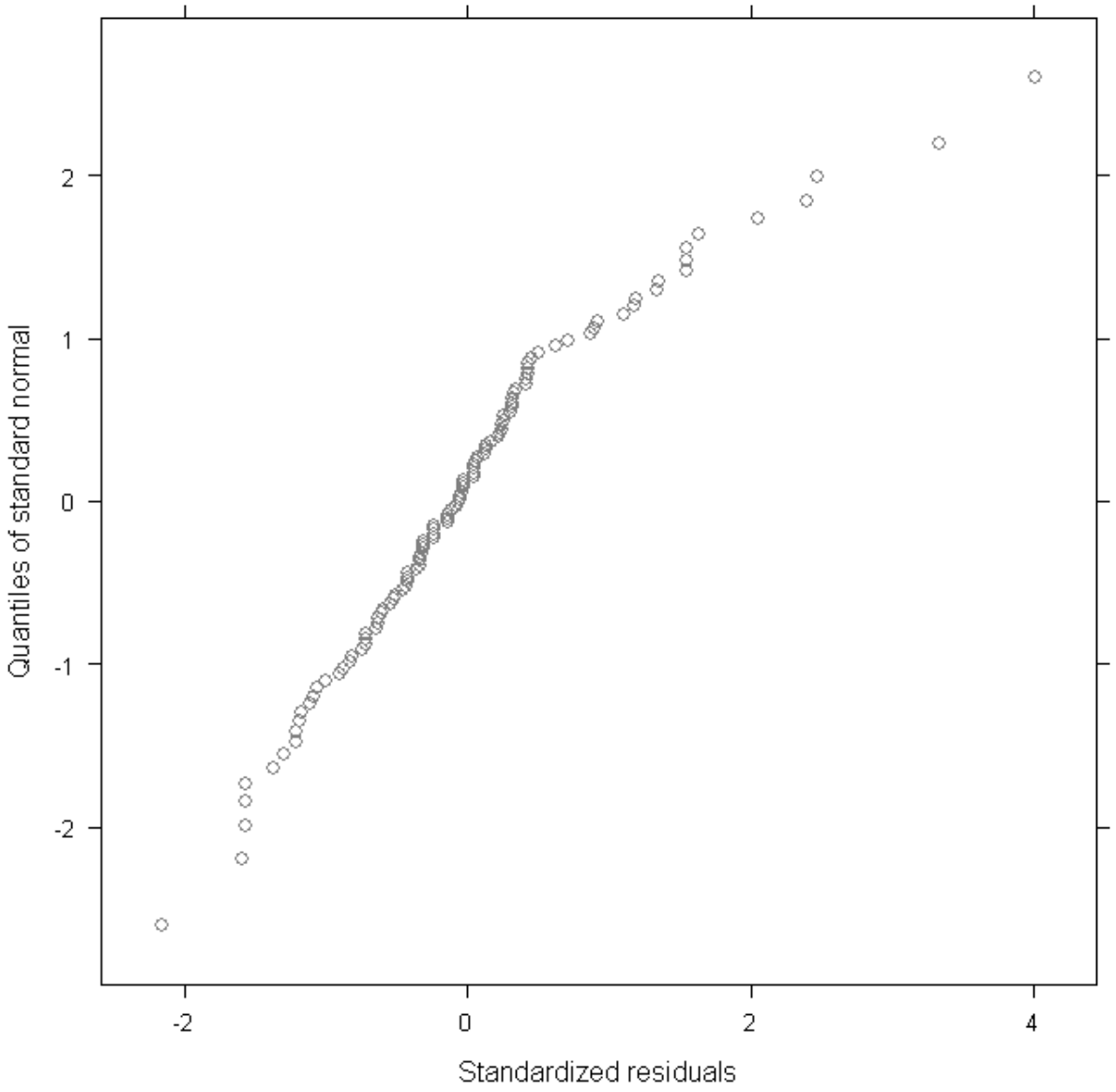


Figure 9.7 Normal probability plot of standardised residuals to assess distribution of errors and detect any evidence of non-normal residual distribution and abnormal variance from model M0a.

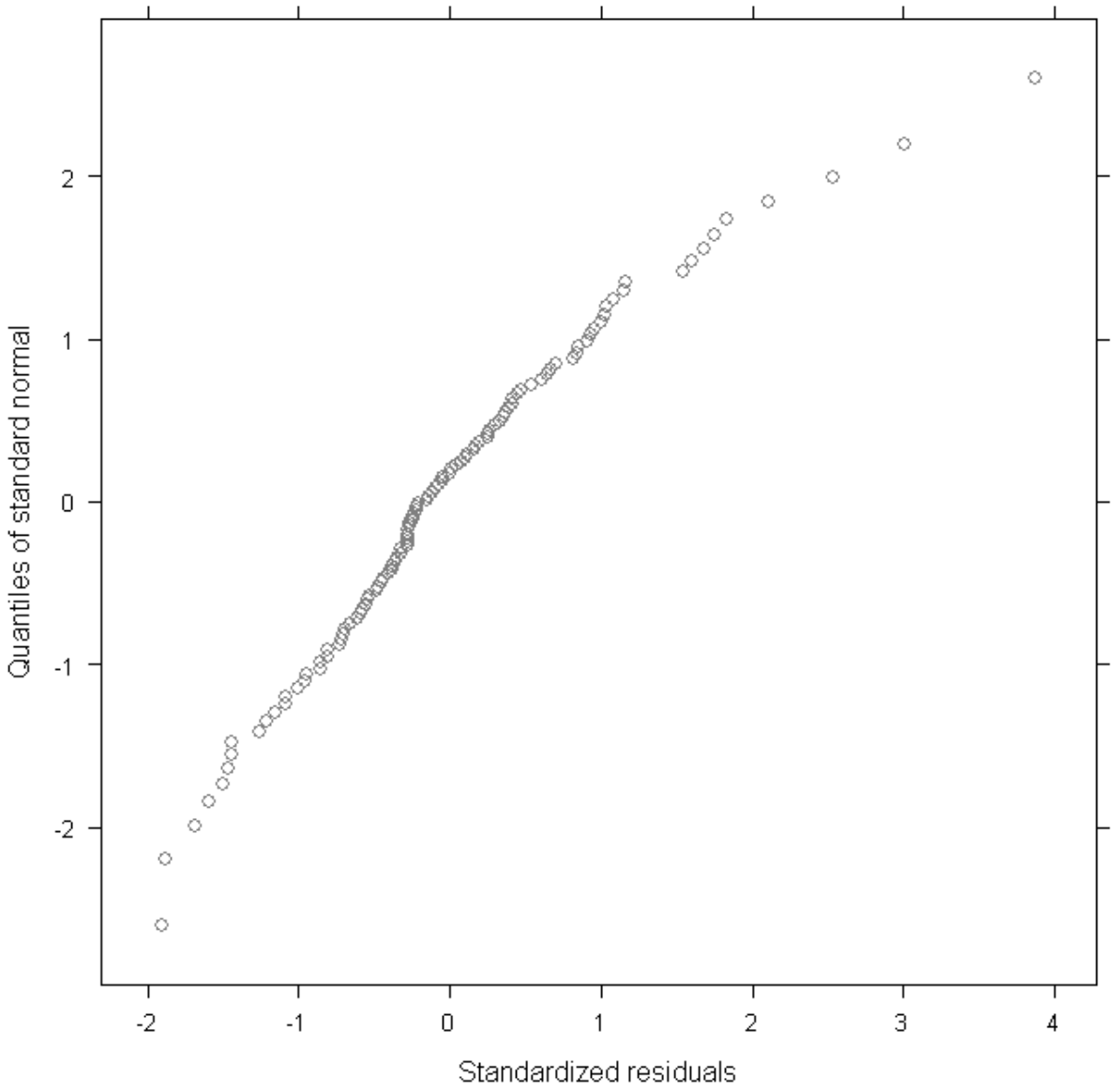


Figure 9.8 Normal probability plot of standardised residuals to assess distribution of errors and detect any evidence of non-normal residual distribution and abnormal variance from model M0b.

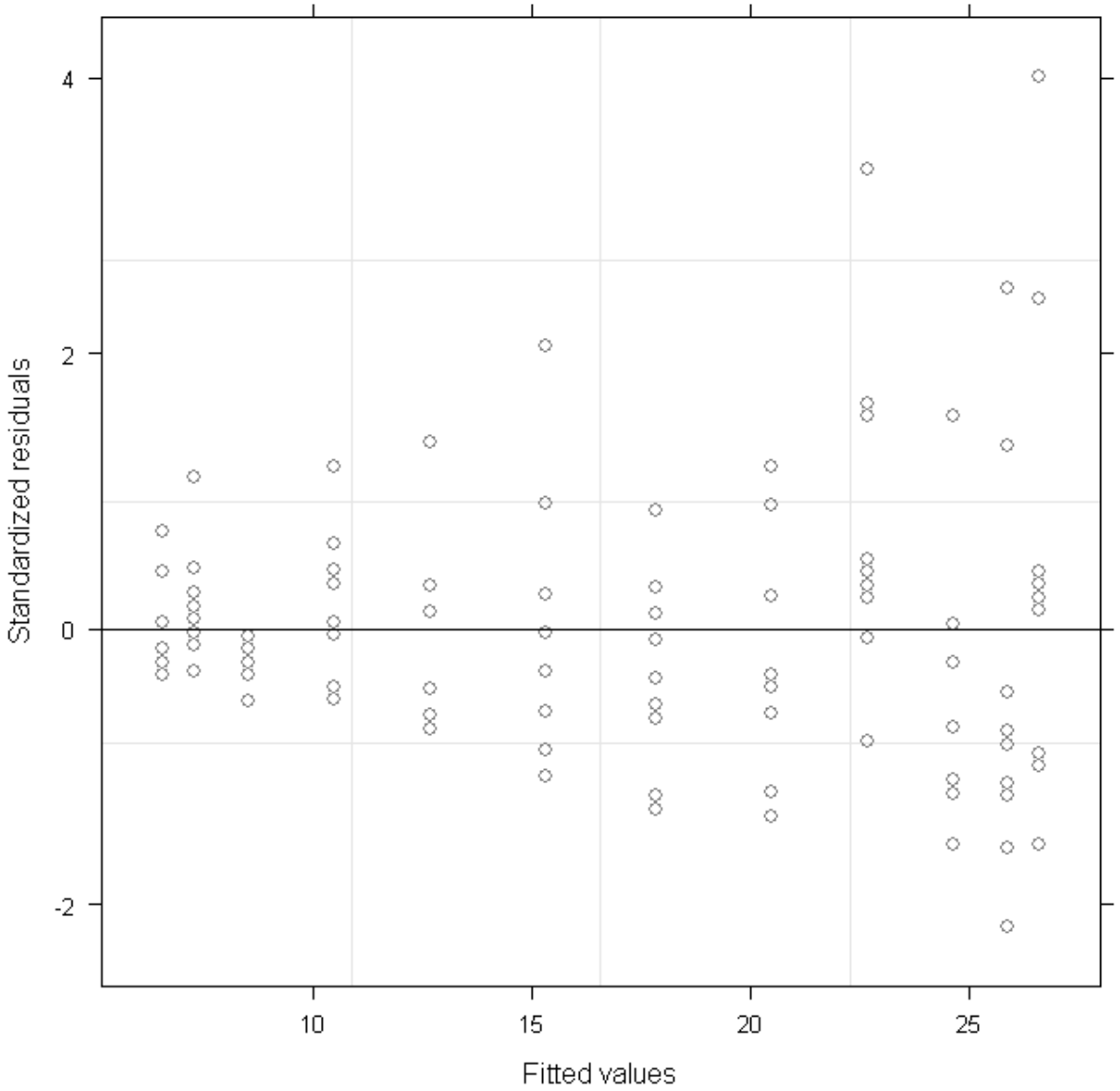


Figure 9.9 Scatter plot of residual and fitted values to assess homoscedascity, independence and variance of the residual distribution for Model M0a.

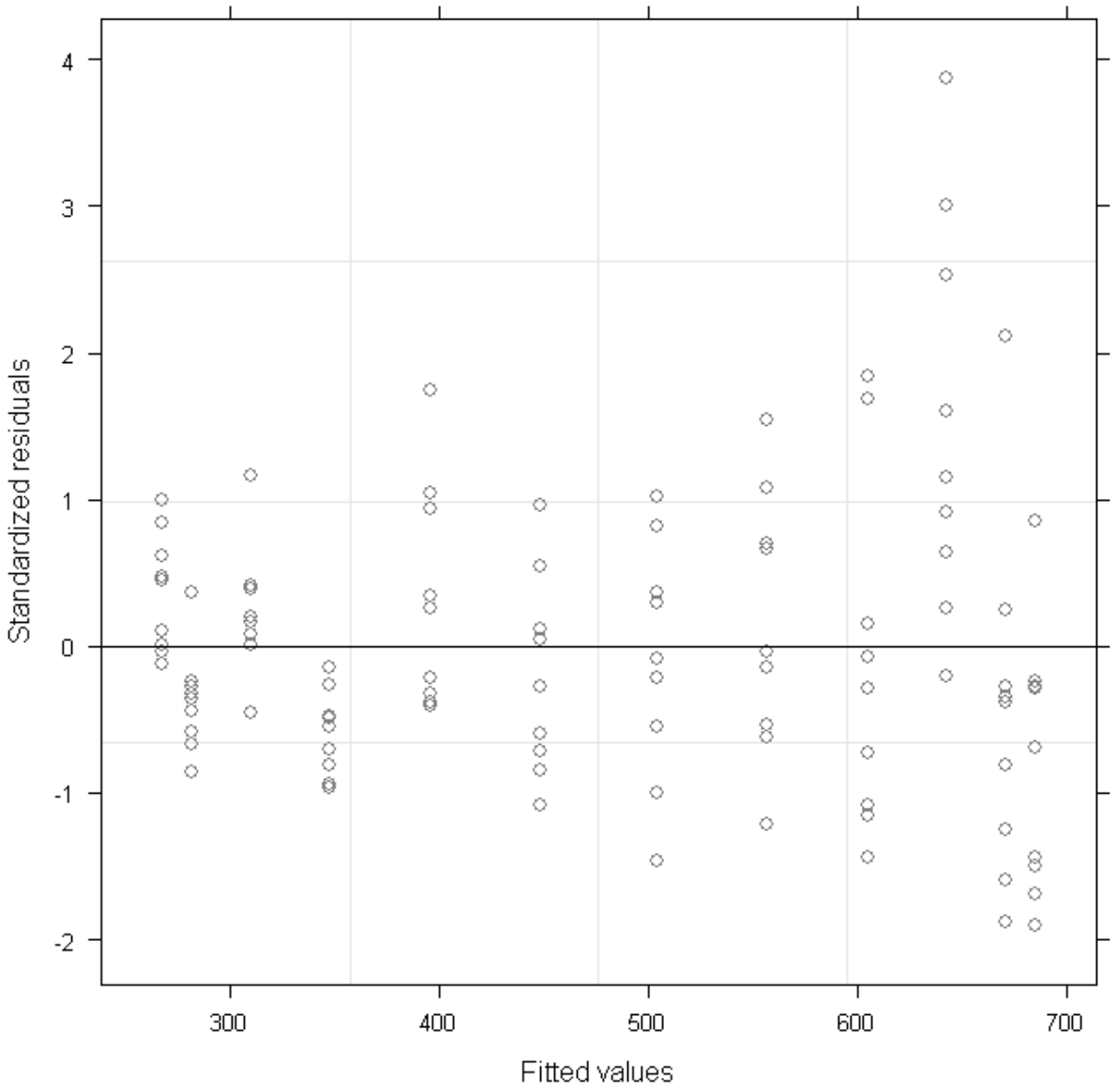


Figure 9.10 Scatter plot of residual and fitted values to assess homoscedascity, independence and variance of the residual distribution for Model M0b. Increasing variance over time is seen.

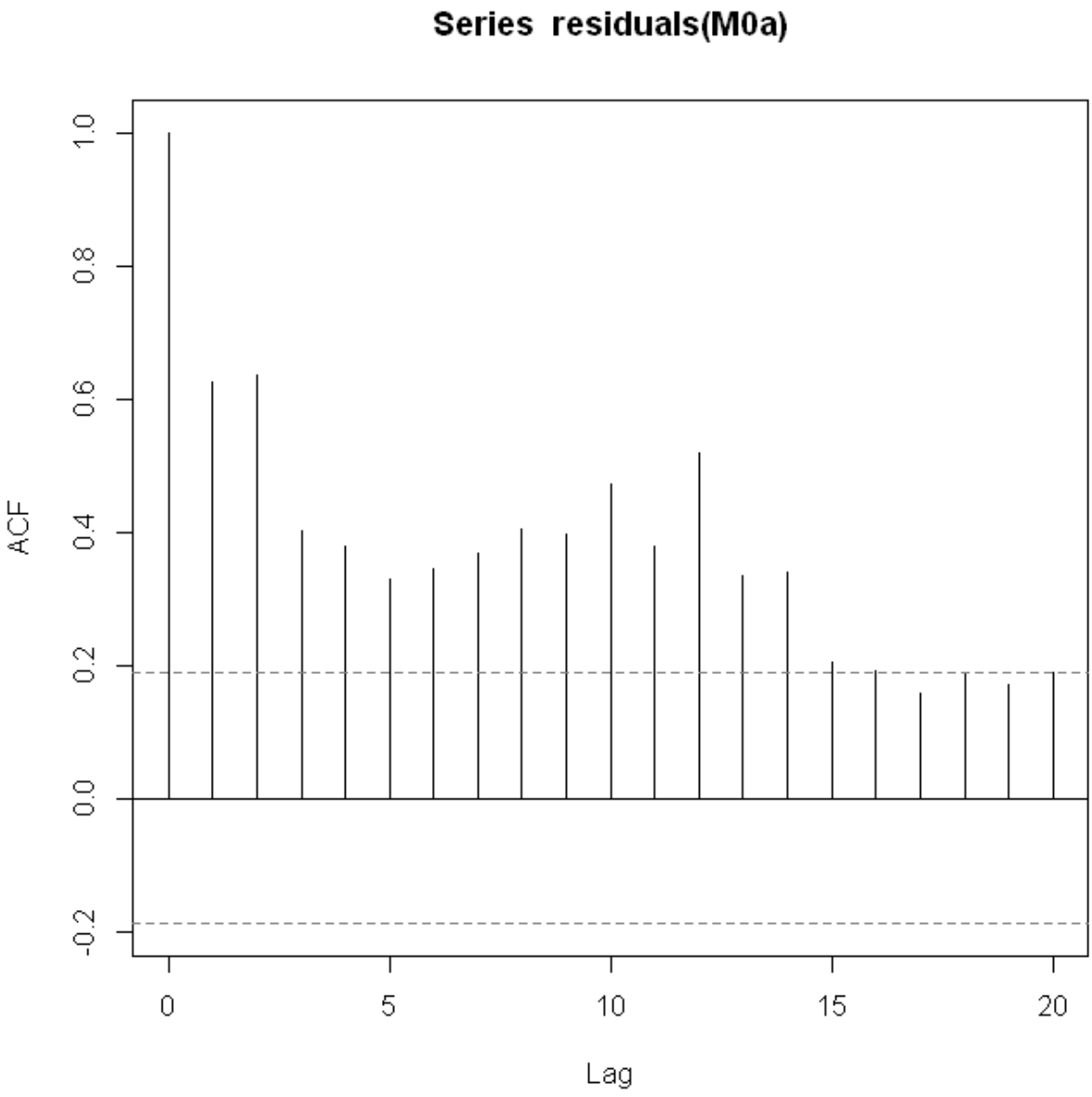


Figure 9.11 Plot illustrating autocorrelation over all lags in model M0a. ***Dotted lines indicate cut-off for presences of significant autocorrelation. ACF – autocorrelation function (size of autocorrelation, lag indicates number of months between observations when correlation occurs).***

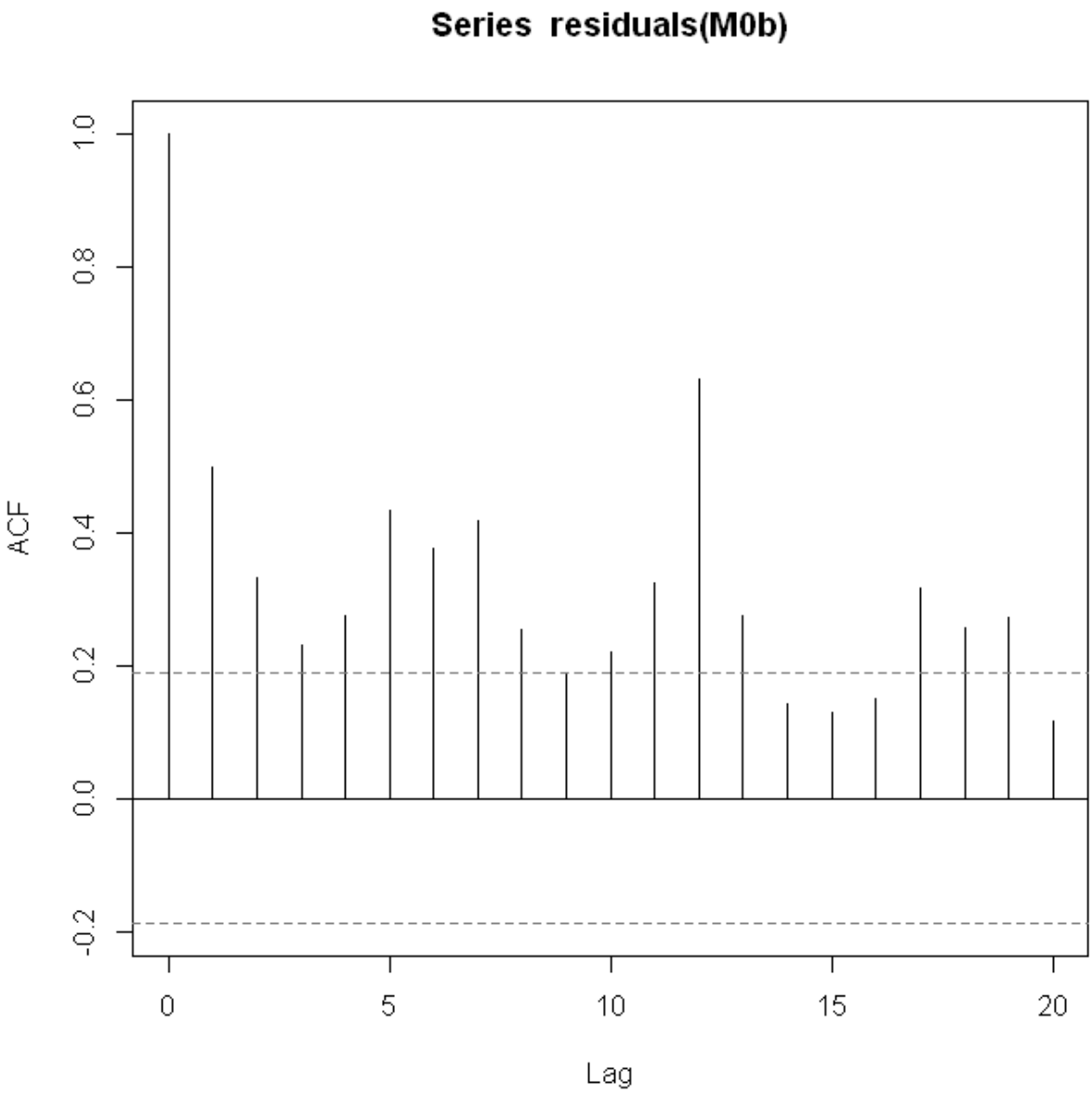


Figure 9.12 Plot illustrating autocorrelation over all lags in model M0b. ***Dotted lines indicate cut-off for presences of significant autocorrelation. ACF – autocorrelation function (size of autocorrelation, lag indicates number of months between observations when correlation occurs)***

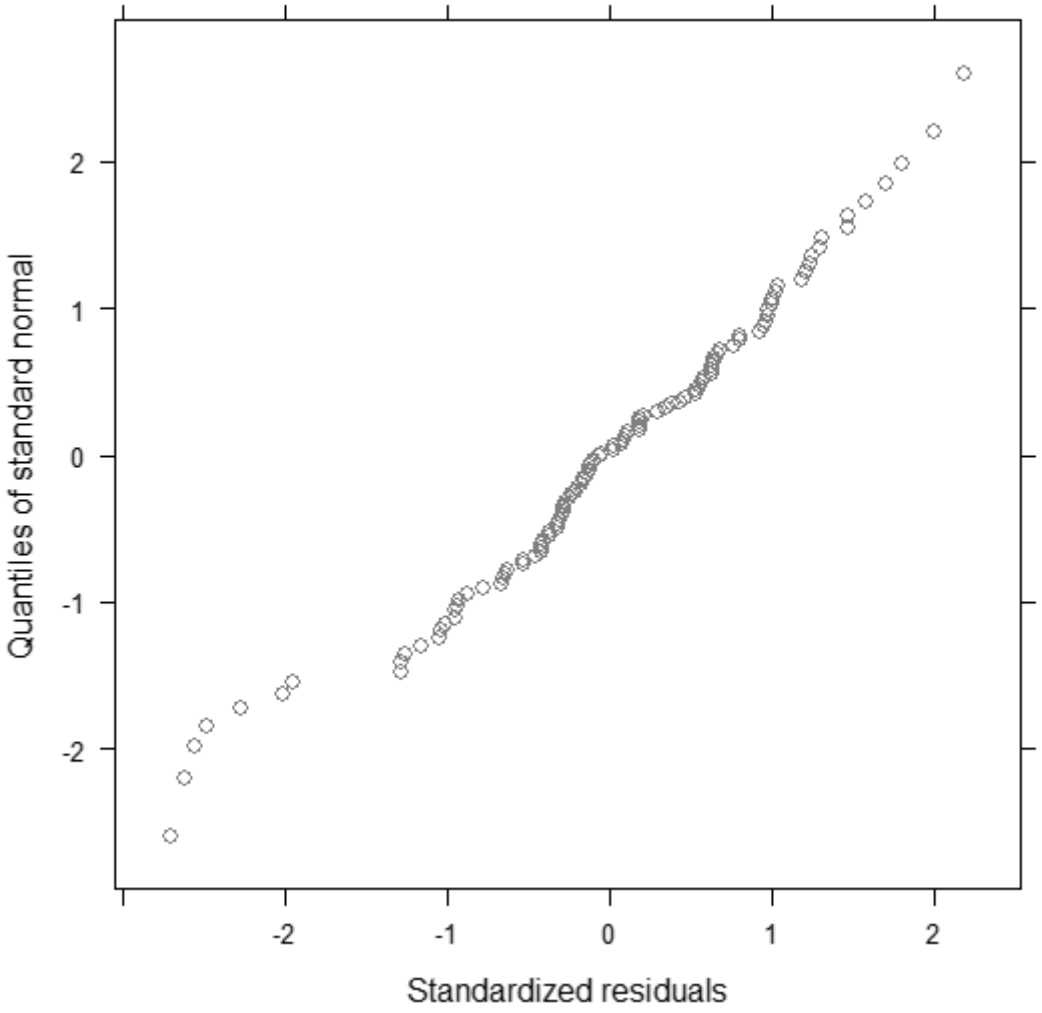


Figure 9.13 Normal probability plot of standardised residuals to assess distribution of errors and detect any evidence of non-normal residual distribution and abnormal variance from model M1.

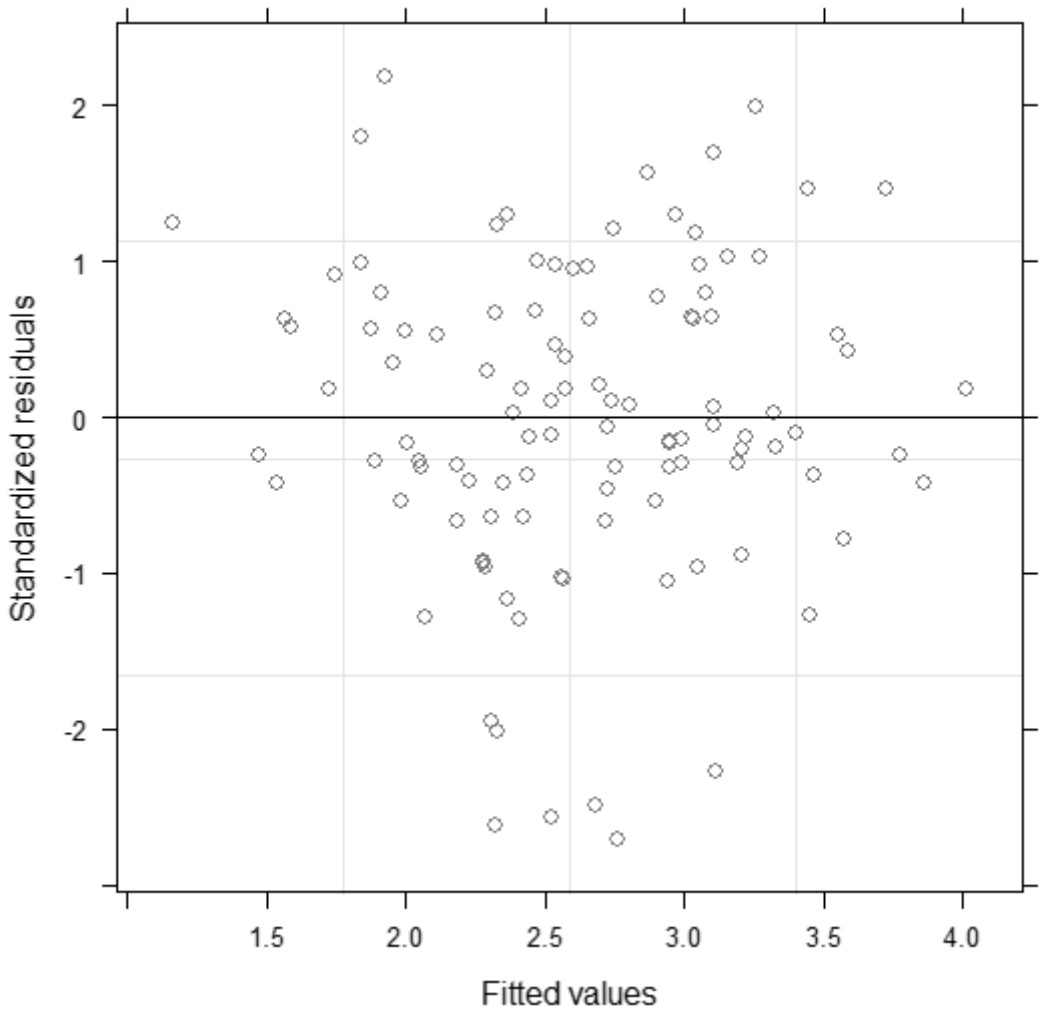


Figure 9.14 Scatter plot of residual and fitted values to assess homoscedascity, independence and variance of the residual distribution for Model M1. A normal distribution of fitted values are seen.

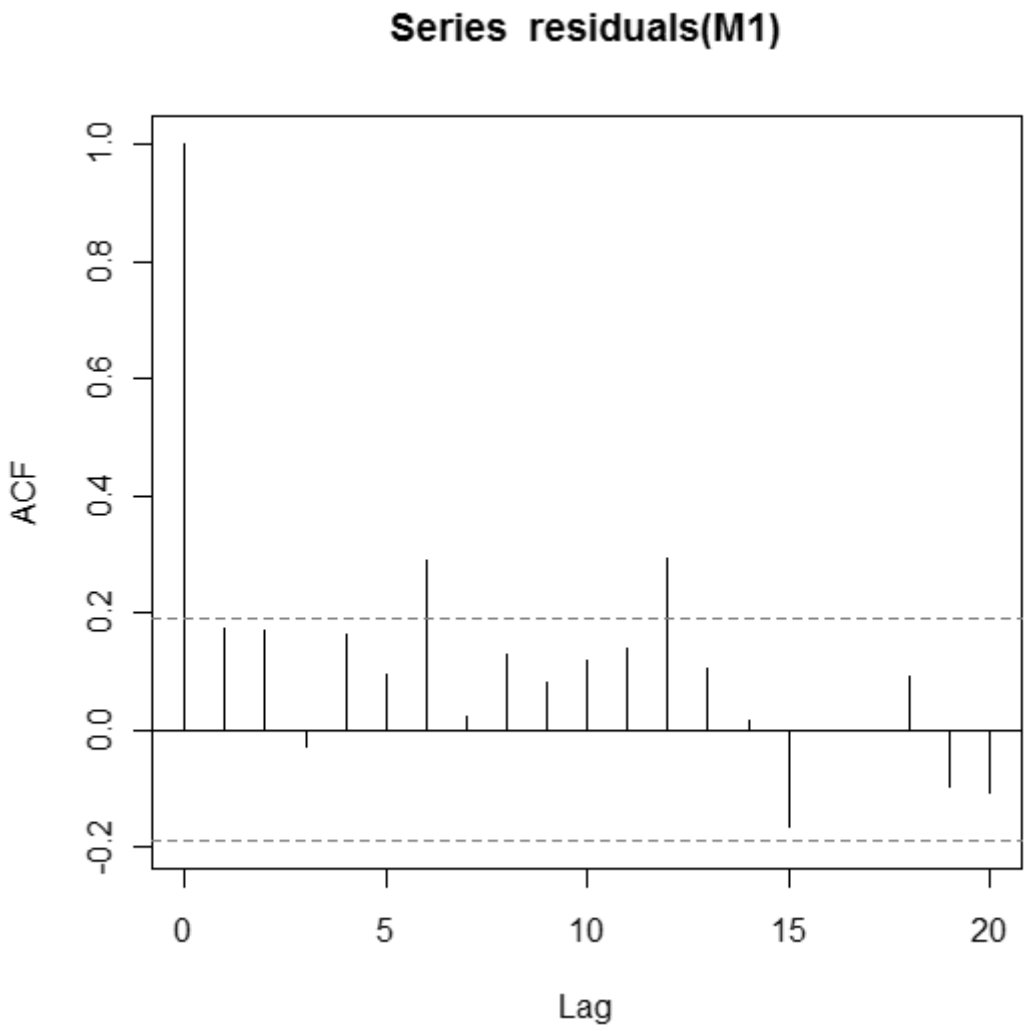


Figure 9.15 Autocorrelation plot demonstrating significant autocorrelation at lags 2, 6 and 12 in model M1. ***Dotted lines indicate cut-off for presences of significant autocorrelation. ACF – autocorrelation function (size of autocorrelation, lag indicates number of months between observations when correlation occurs).***

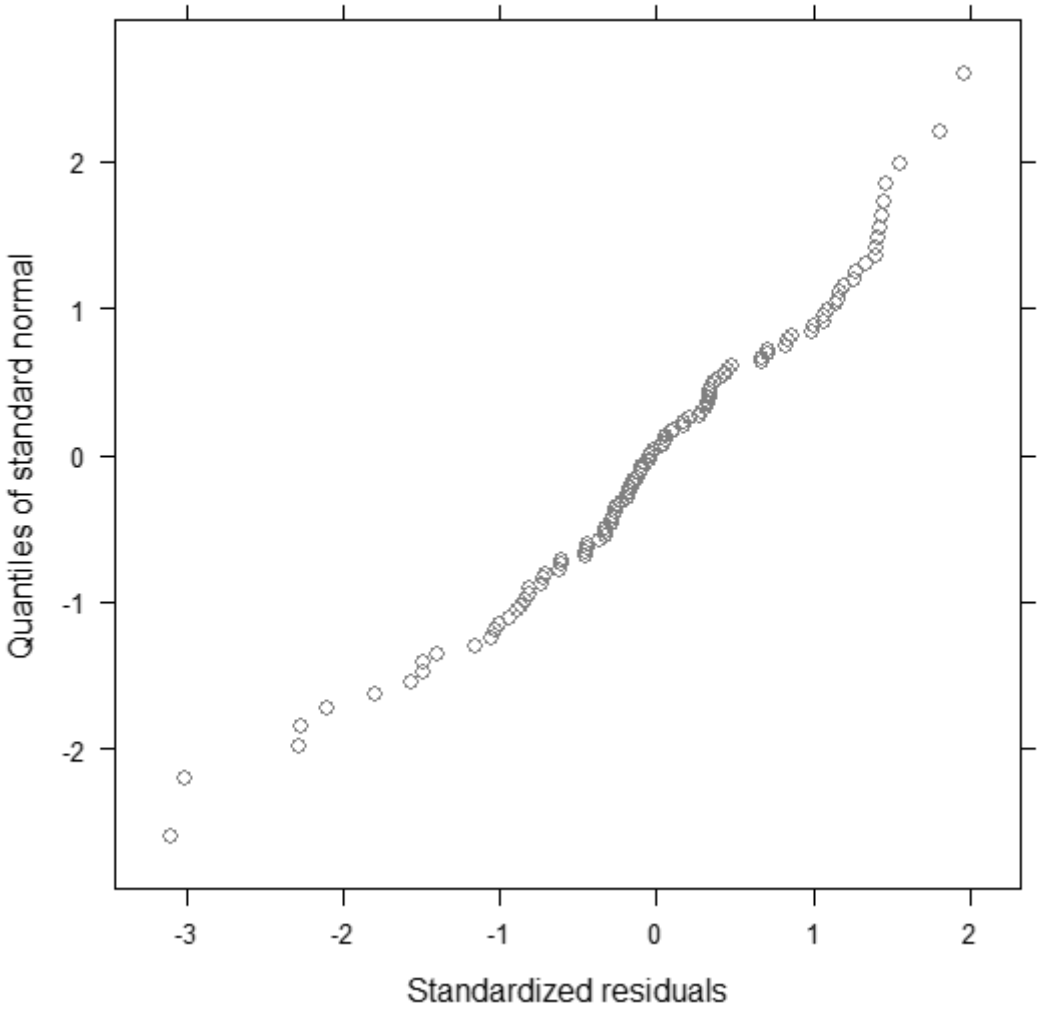


Figure 9.16 Normal probability plot of standardised residuals to assess distribution of errors and detect any evidence of non-normal residual distribution and abnormal variance from model M2 showing potential overdispersion at lower residual values.

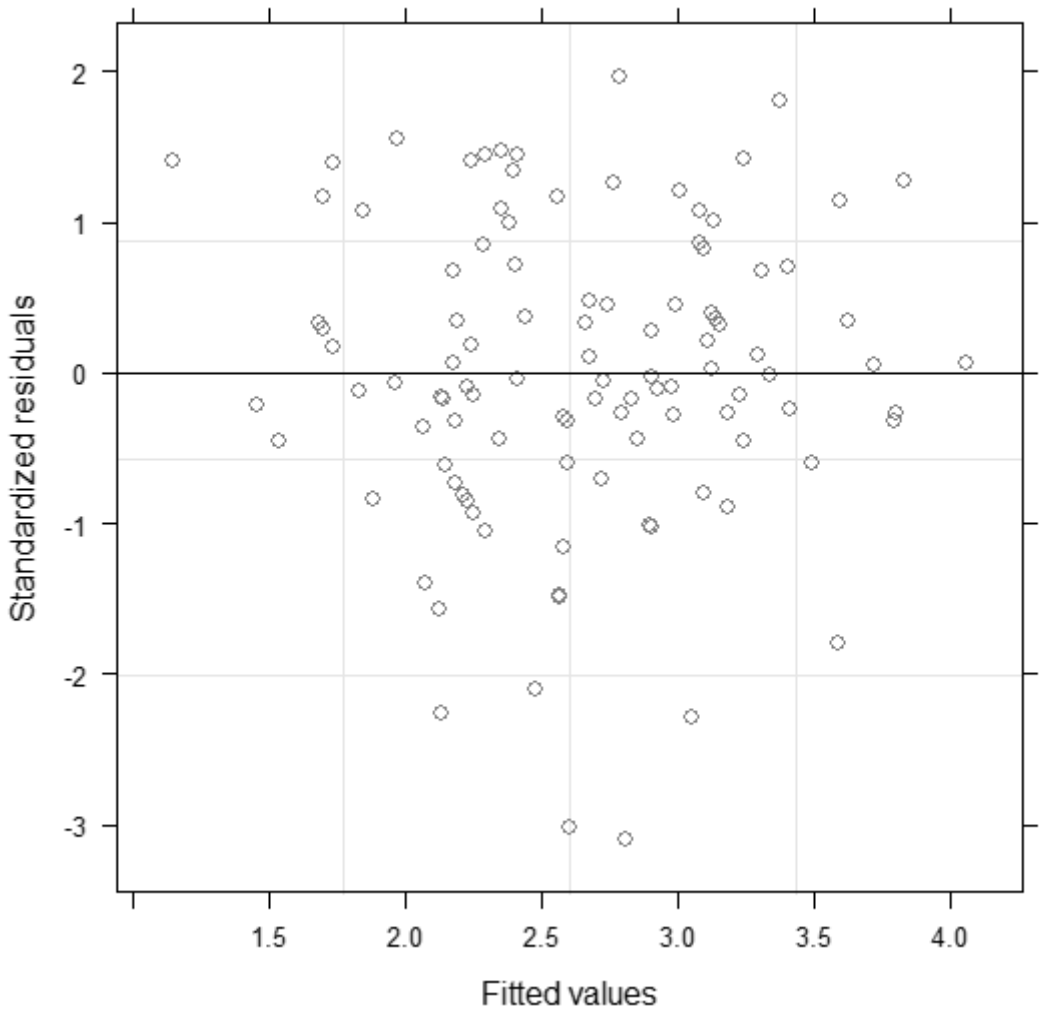


Figure 9.17 Scatter plot of residual and fitted values to assess homoscedascity, independence and variance of the residual distribution for Model M2. Normal distribution of the residuals is seen.

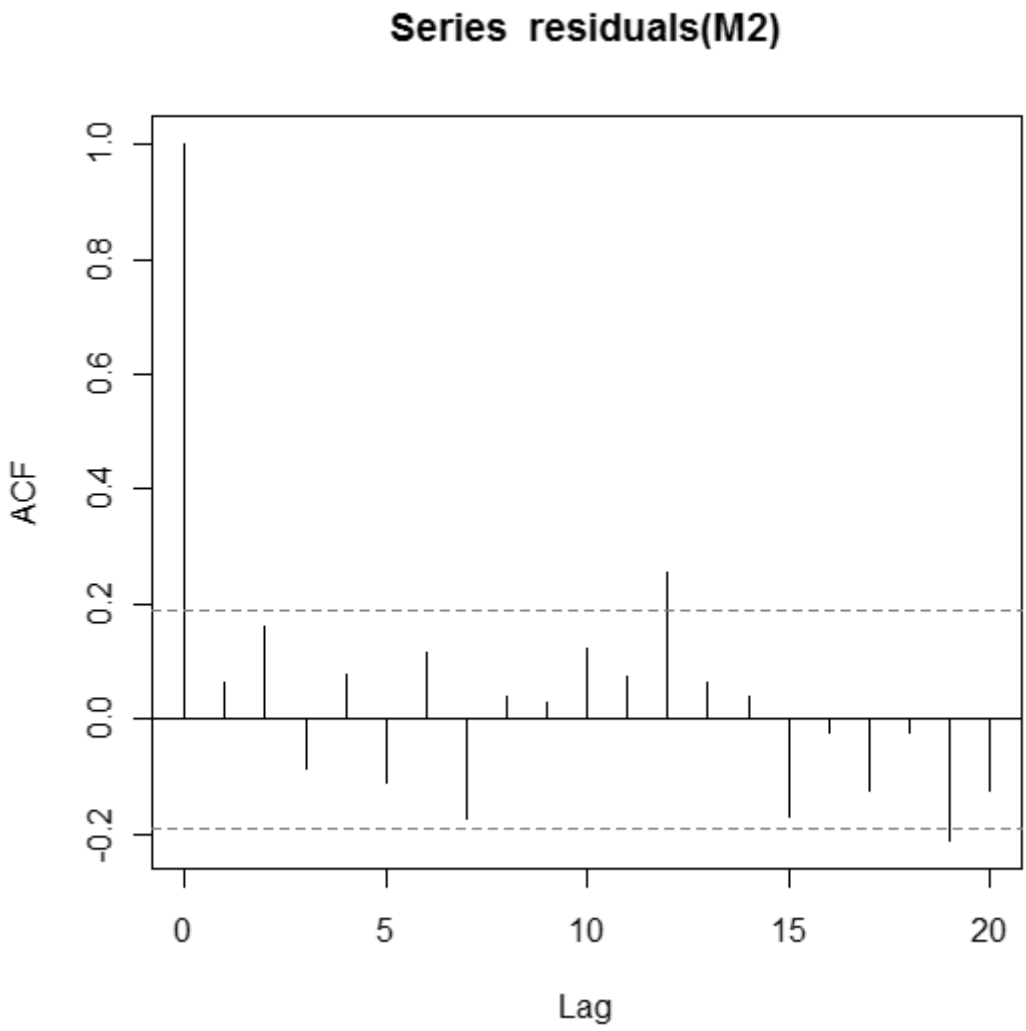


Figure 9.18 Autocorrelation plot of model M2 demonstrating significant autocorrelation at lags 12 and 19. ***Dotted lines indicate cut-off for presences of significant autocorrelation. ACF – autocorrelation function (size of autocorrelation, lag indicates number of months between observations when correlation occurs).***

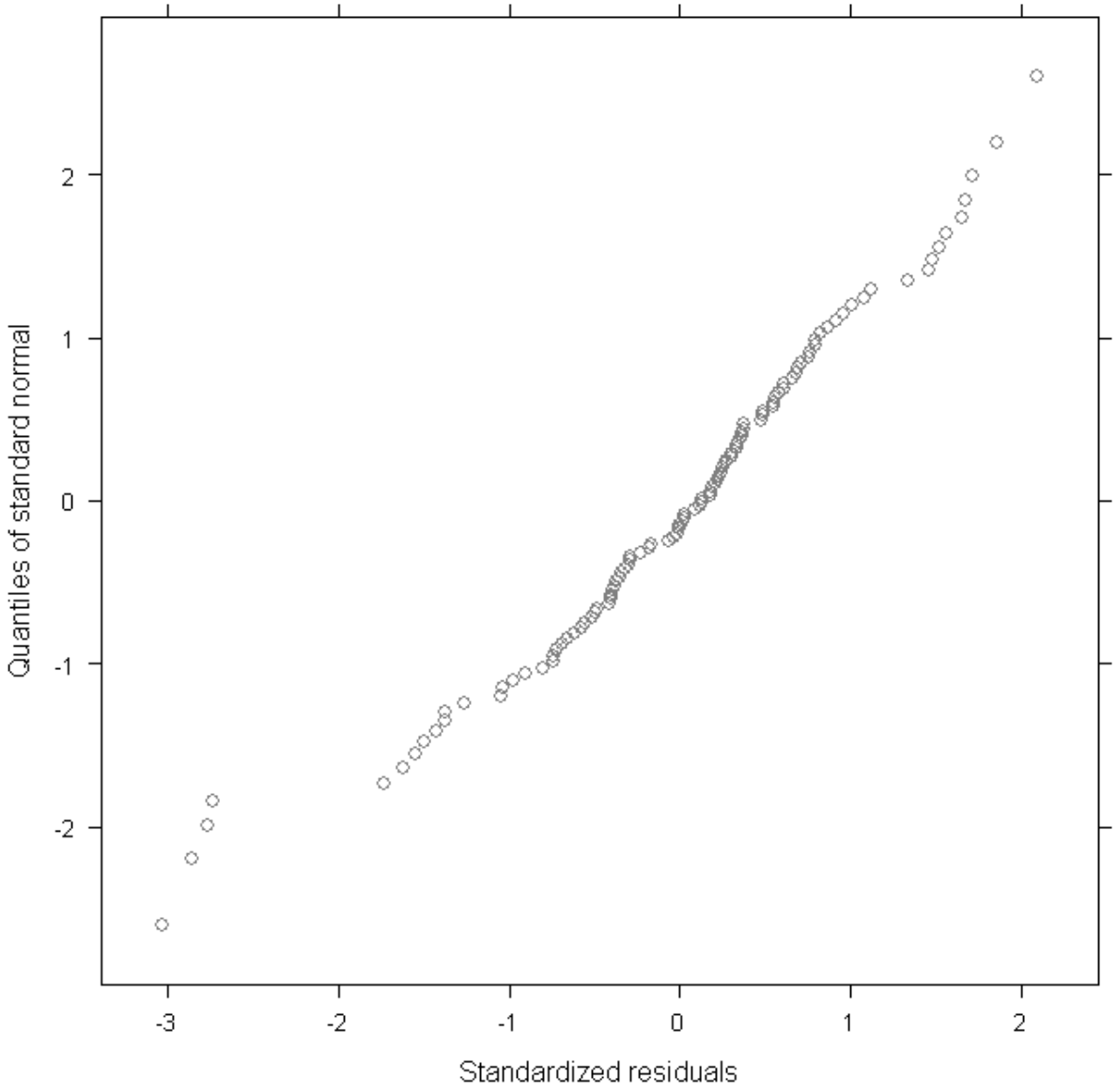


Figure 9.19 Normal probability plot of standardised residuals to assess distribution of errors and detect any evidence of non-normal residual distribution and abnormal variance from model M4 illustrating minor overdispersion of the residuals at the extremes.

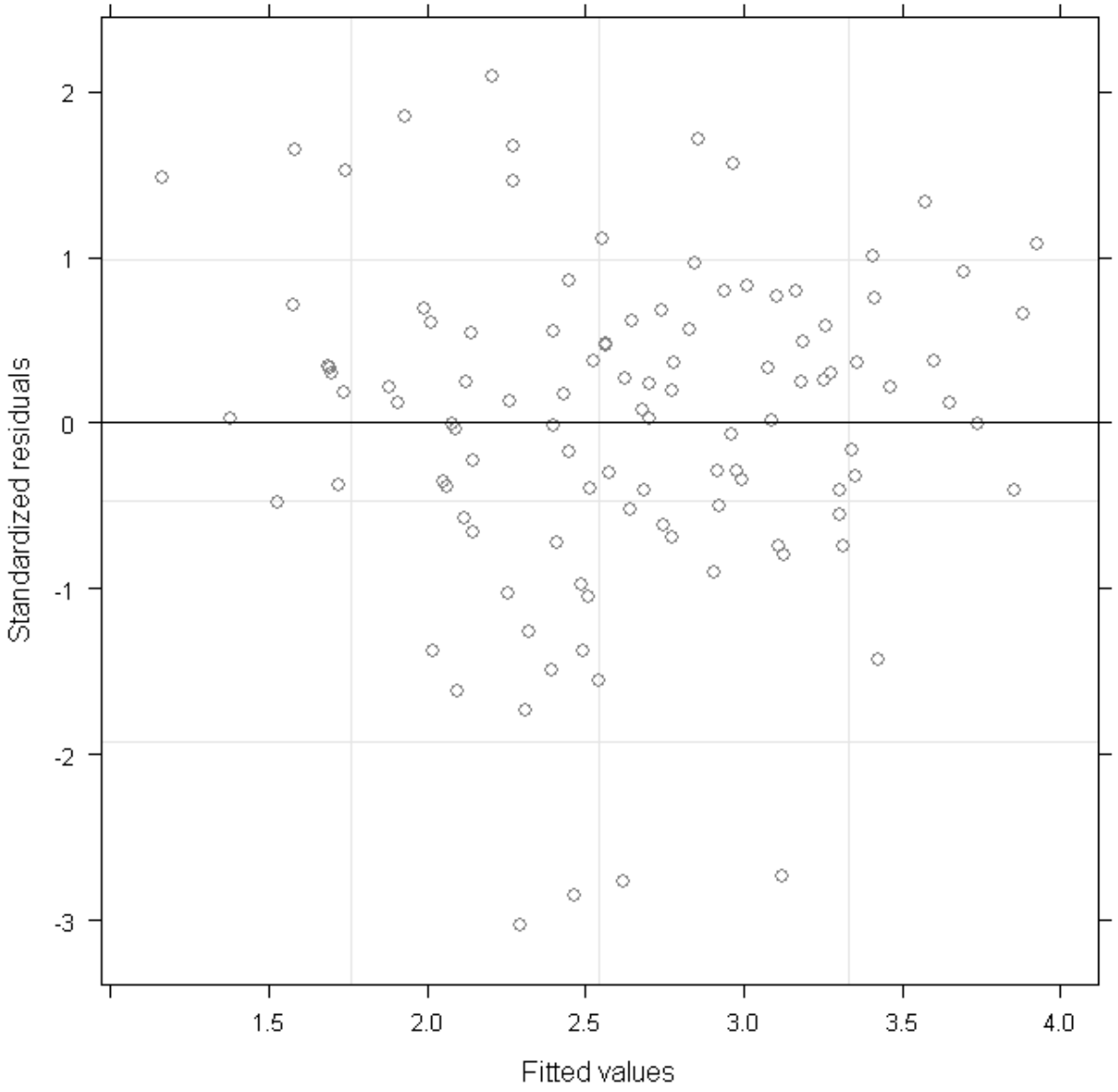


Figure 9.20 Scatter plot of residual and fitted values to assess homoscedascity, independence and variance of the residual distribution for model M4. Normal distribution seen.

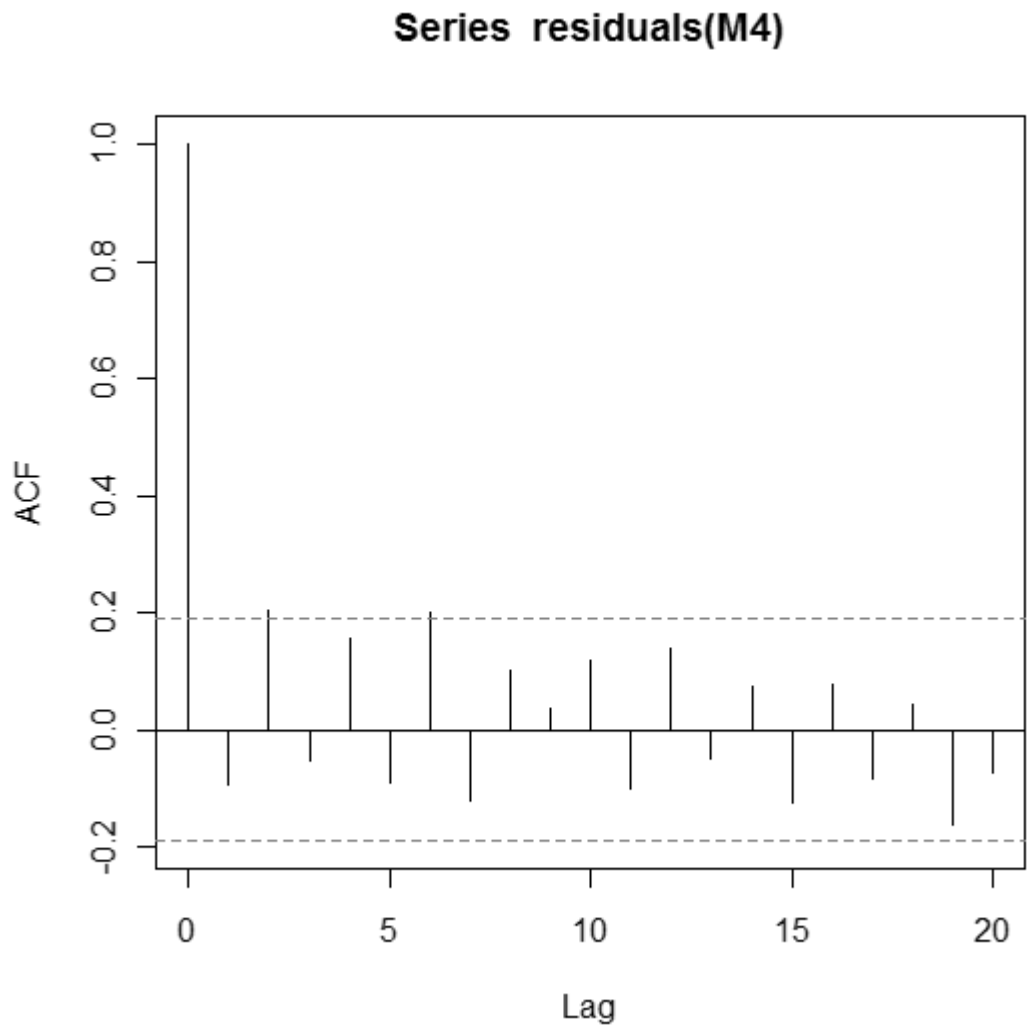


Figure 9.21 Autocorrelation plot of model M4 with autocorrelation of marginal significance at lags 3 and 6. ***Dotted lines indicate cut-off for presences of significant autocorrelation. ACF – autocorrelation function (size of autocorrelation, lag indicates number of months between observations when correlation occurs).***

9.4 Appendix D - Model diagnostic plots for definition of periodicity and cyclicity within national empyema admissions (Model M5)

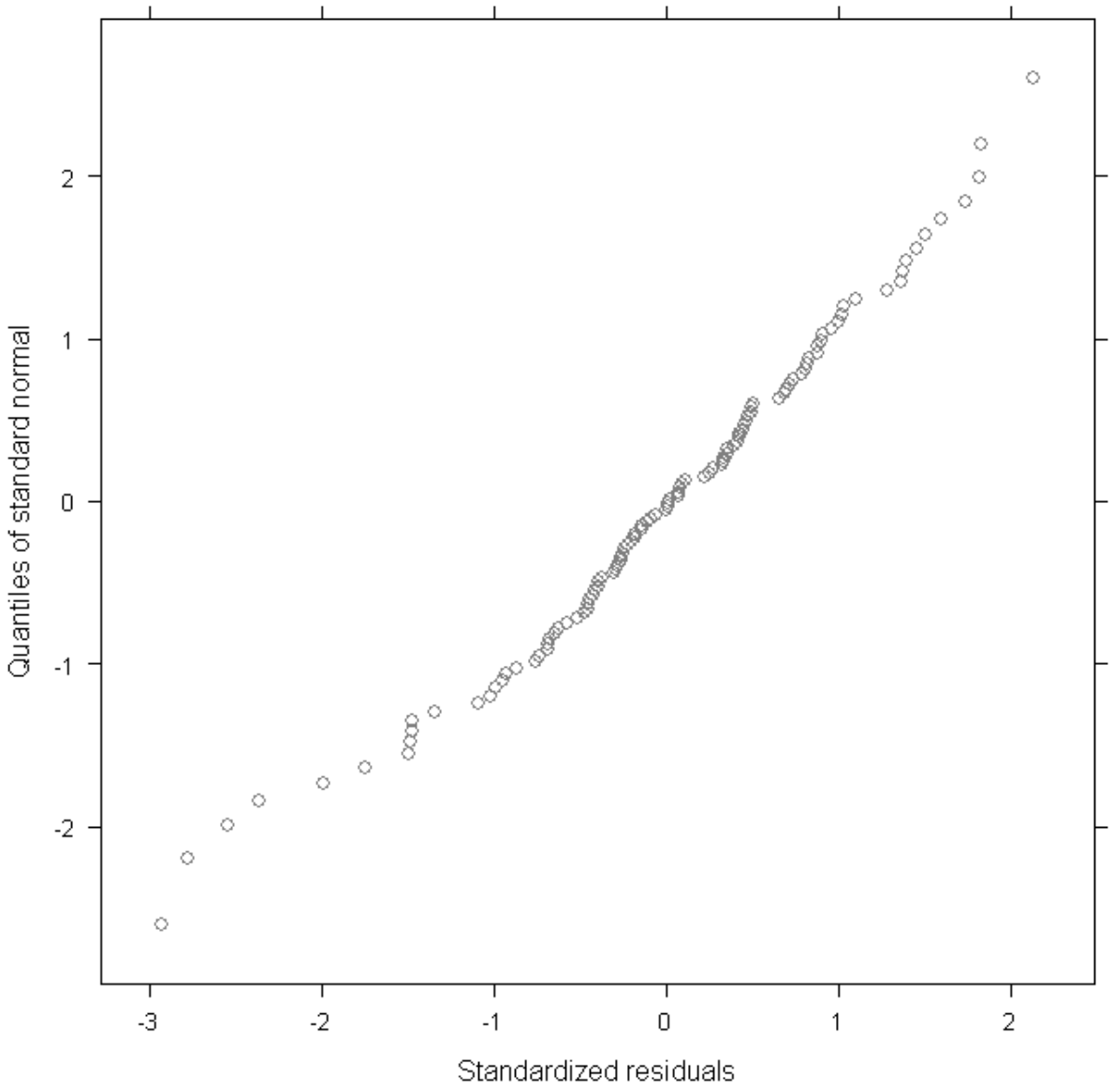


Figure 9.22 Normal probability plot of standardised residuals to assess distribution of errors and detect any evidence of non-normal residual distribution and abnormal variance from model M5 illustrating minor overdispersion of the residuals at the extremes.

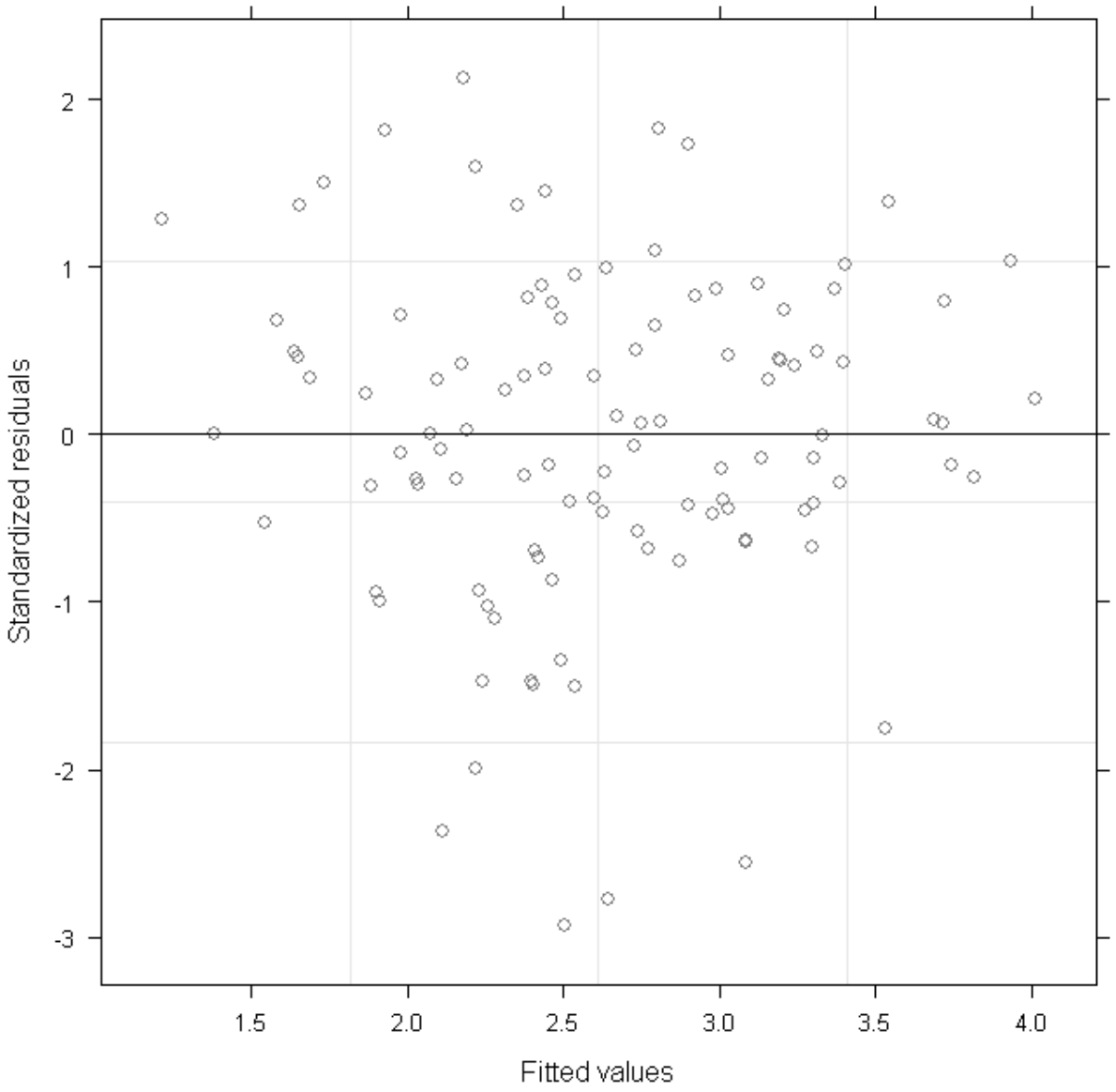


Figure 9.23 Scatter plot of residual and fitted values to assess homoscedascity, independence and variance of the residual distribution for model M5.

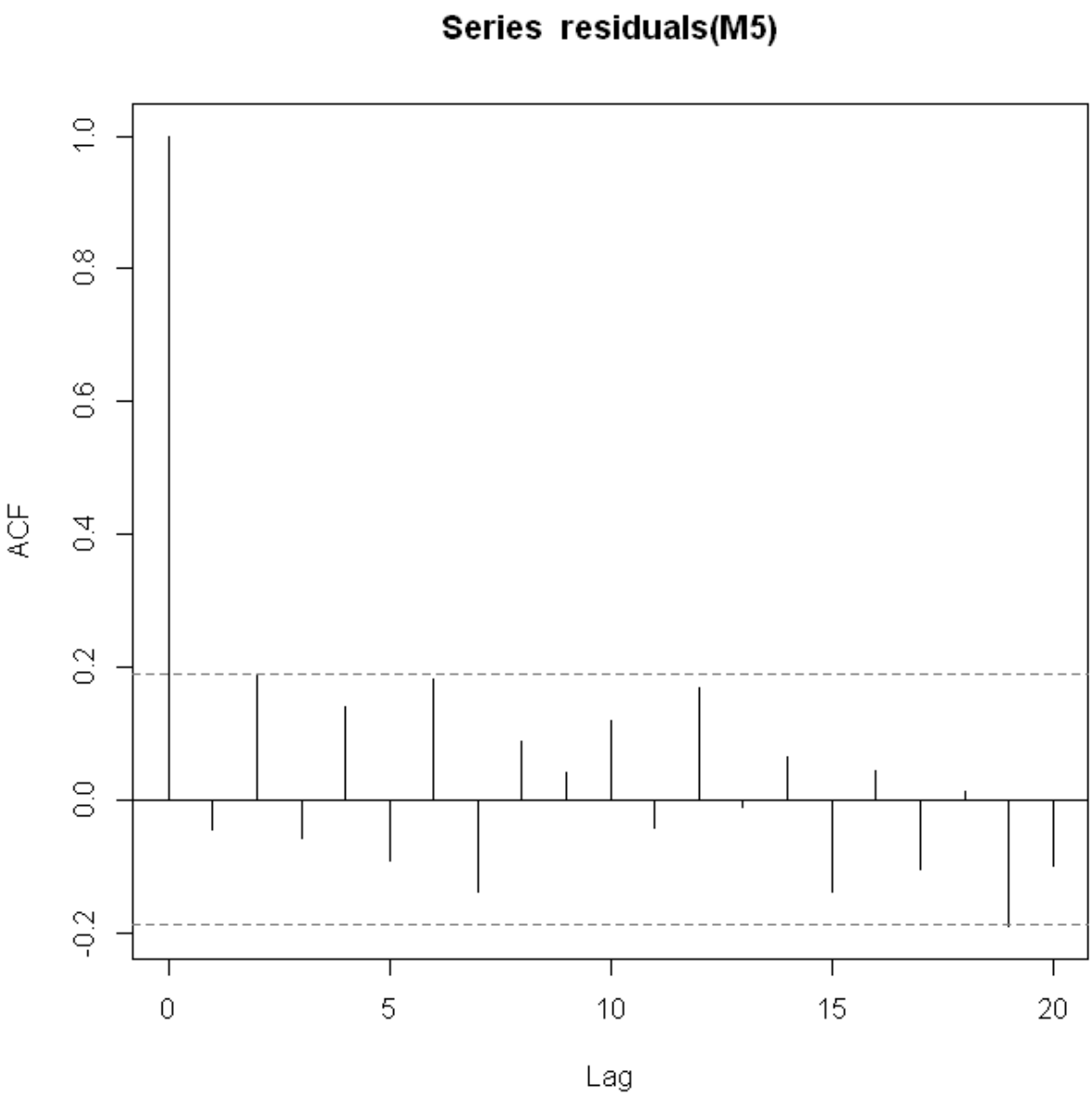


Figure 9.24 Autocorrelation plot showing no significant autocorrelation in model M5. ***Dotted lines indicate cut-off for presences of significant autocorrelation. ACF – autocorrelation function (size of autocorrelation, lag indicates number of months between observations when correlation occurs).***

9.5 Appendix E - Model diagnostic plots for population adjusted pneumonia and empyema admissions (Models S0a/b-S3a/b)

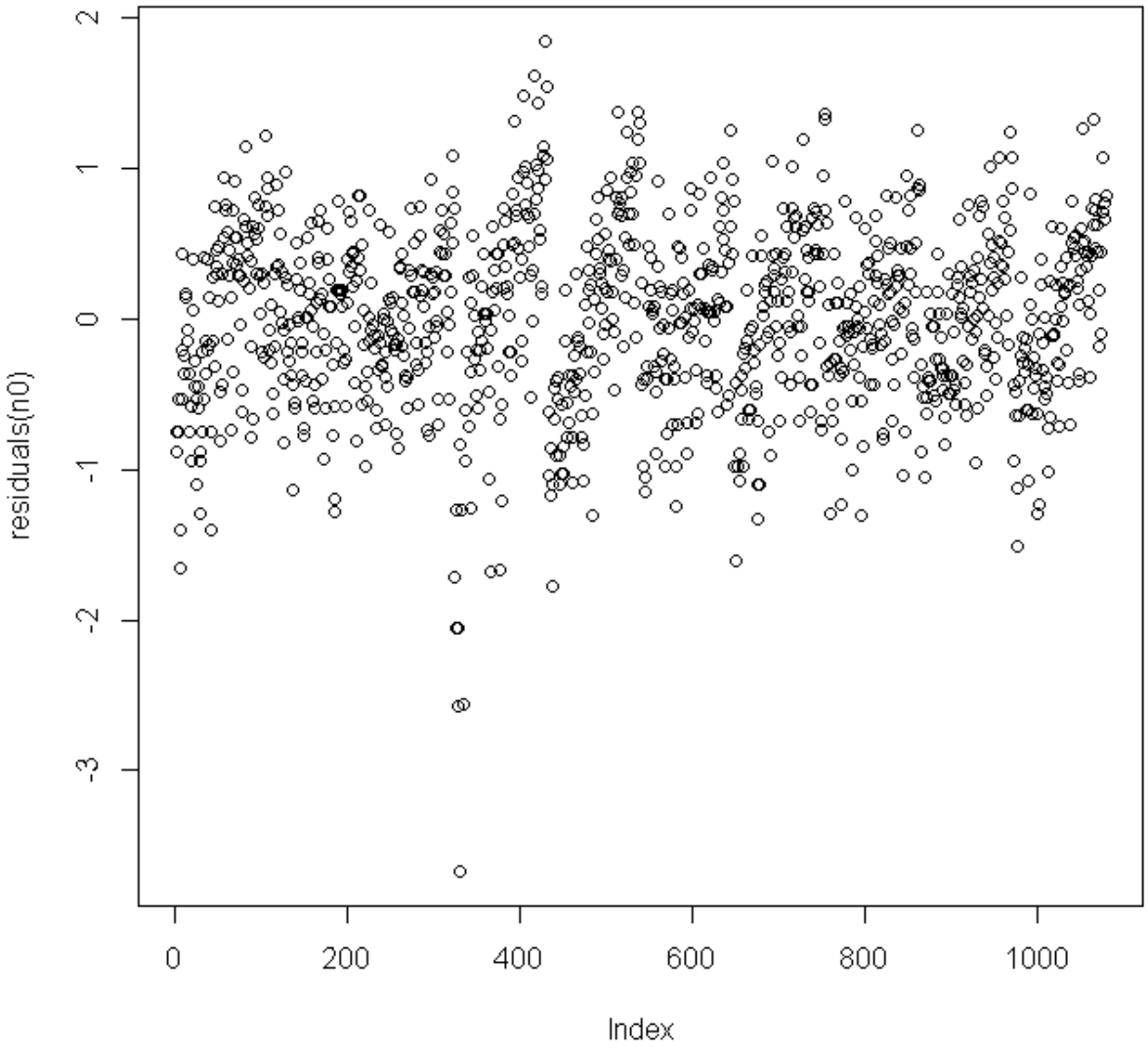


Figure 9.25 Residuals from model S0a – linear model relating population adjusted pneumonia admissions to SHA. Three outlying values identified from the North East SHA corresponding to June, July and August 1997. There were two, two and zero recorded cases of pneumonia in those months respectively. Similarly low values are observed in the same months at different points and admissions for pneumonia are recognised to be low at this time of year suggesting that these values are real rather than artefact.

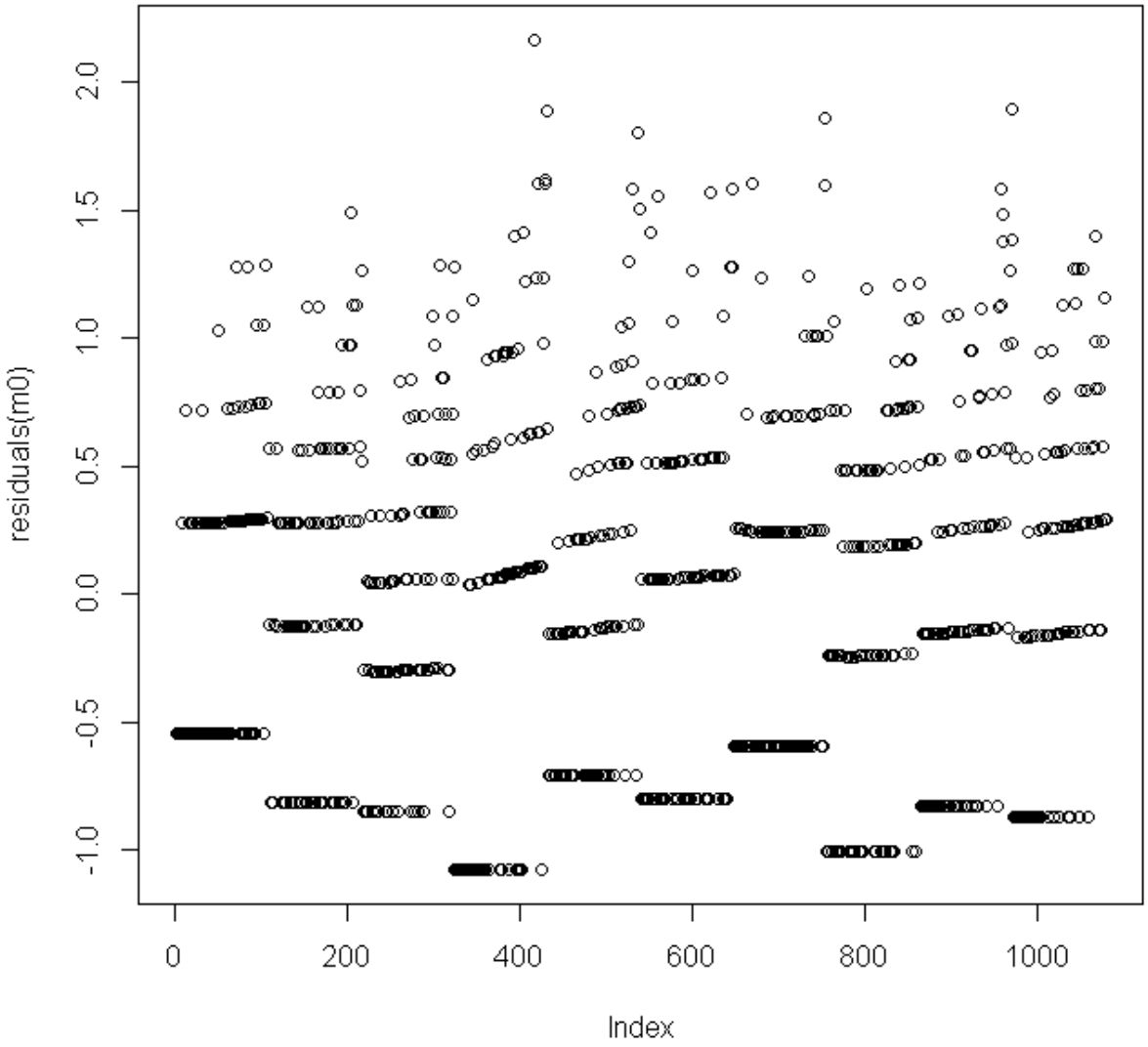


Figure 9.26 Residuals from model S0a – linear model relating population adjusted empyema admissions to SHA. The residuals are clustered due to the limited range of values of population adjusted empyema admission seen within the data.

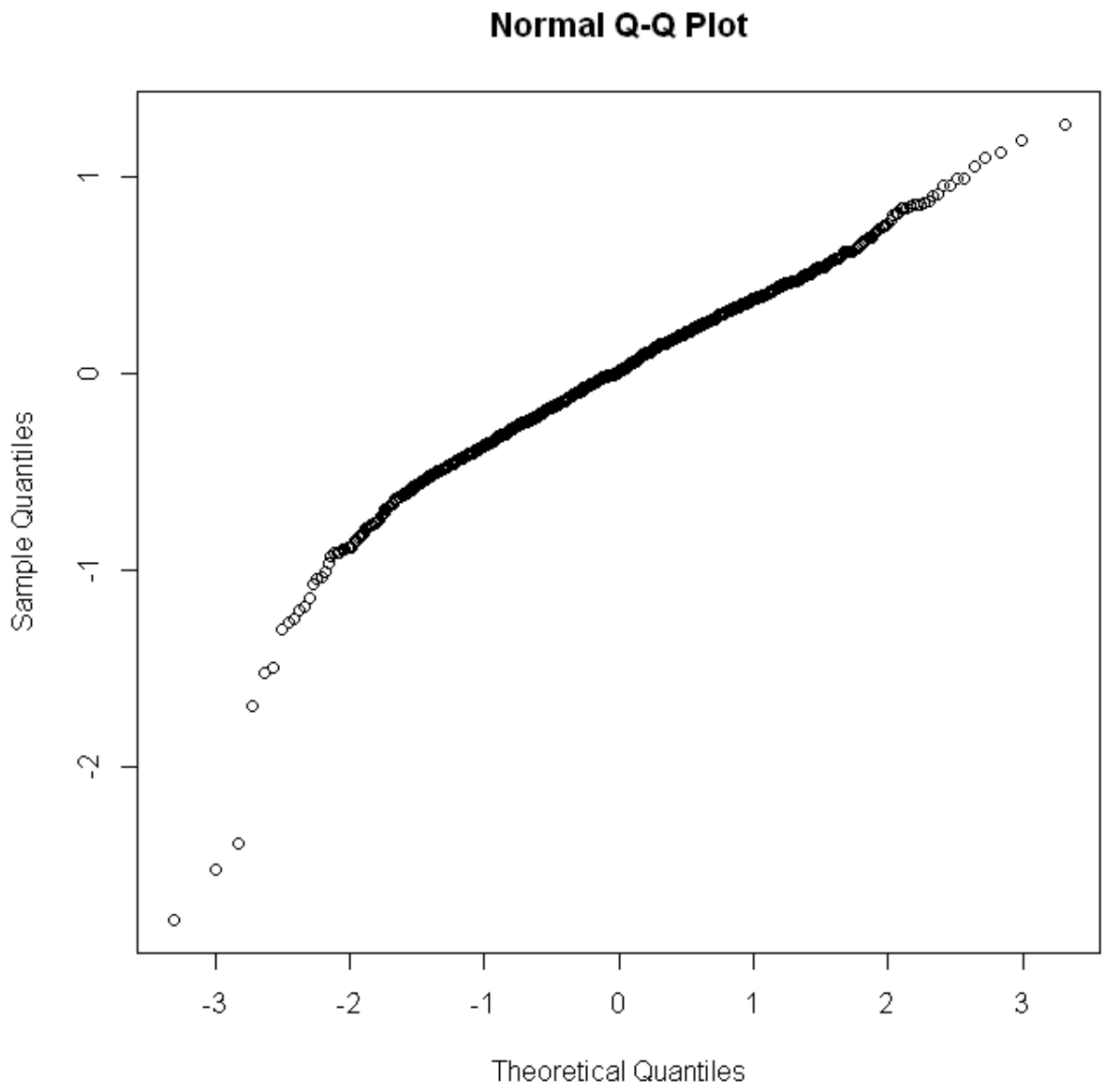


Figure 9.27 Normal probability plot of standardised residuals to assess distribution of errors and detect any evidence of non-normal residual distribution and abnormal variance from model S1a. Three outlying values as in model S0a were identified.

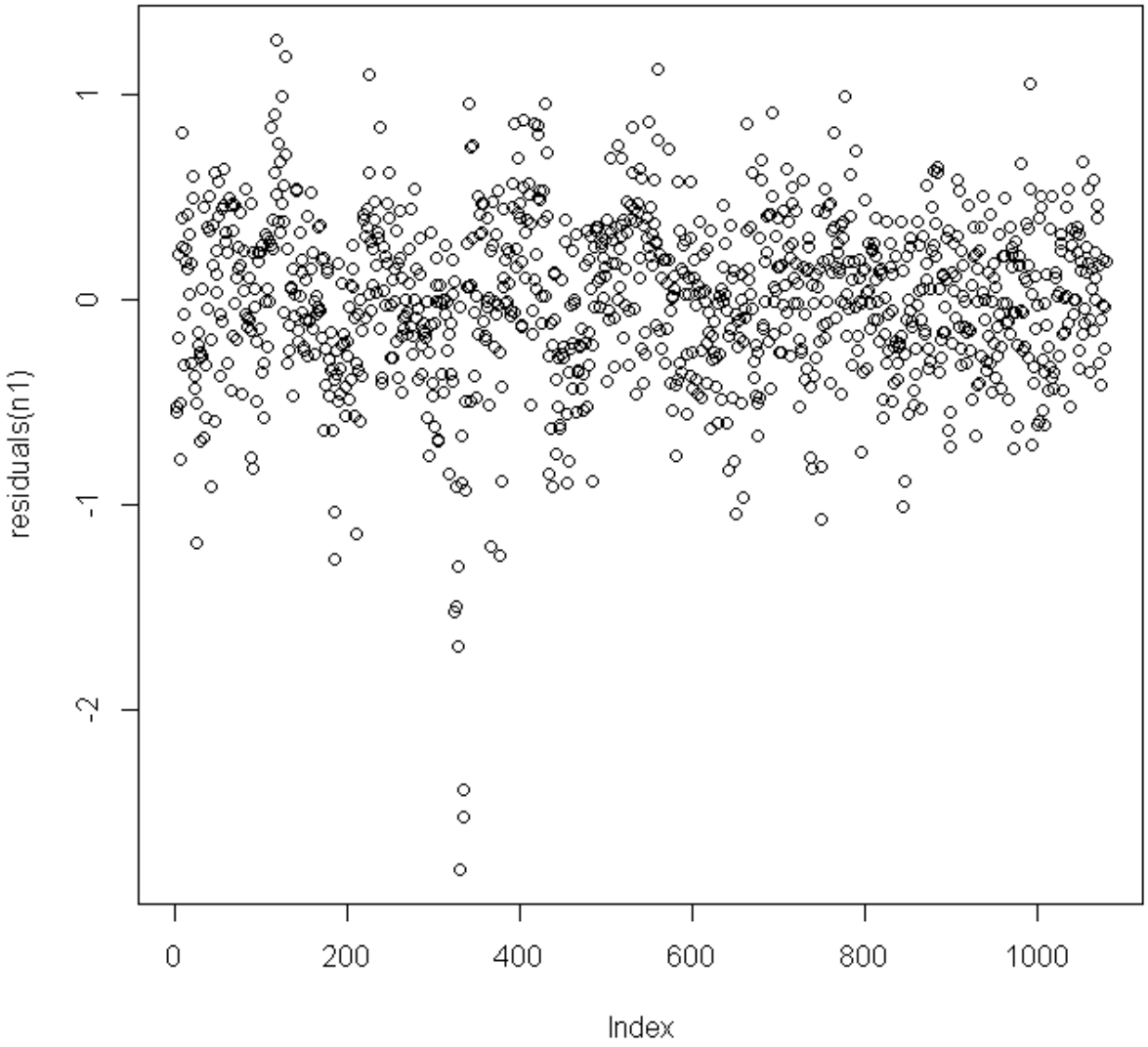


Figure 9.28 Scatter plot of residual to assess homoscedascity, independence and variance of the residual distribution from model S1a. Normal distribution except for three outlying values are seen.

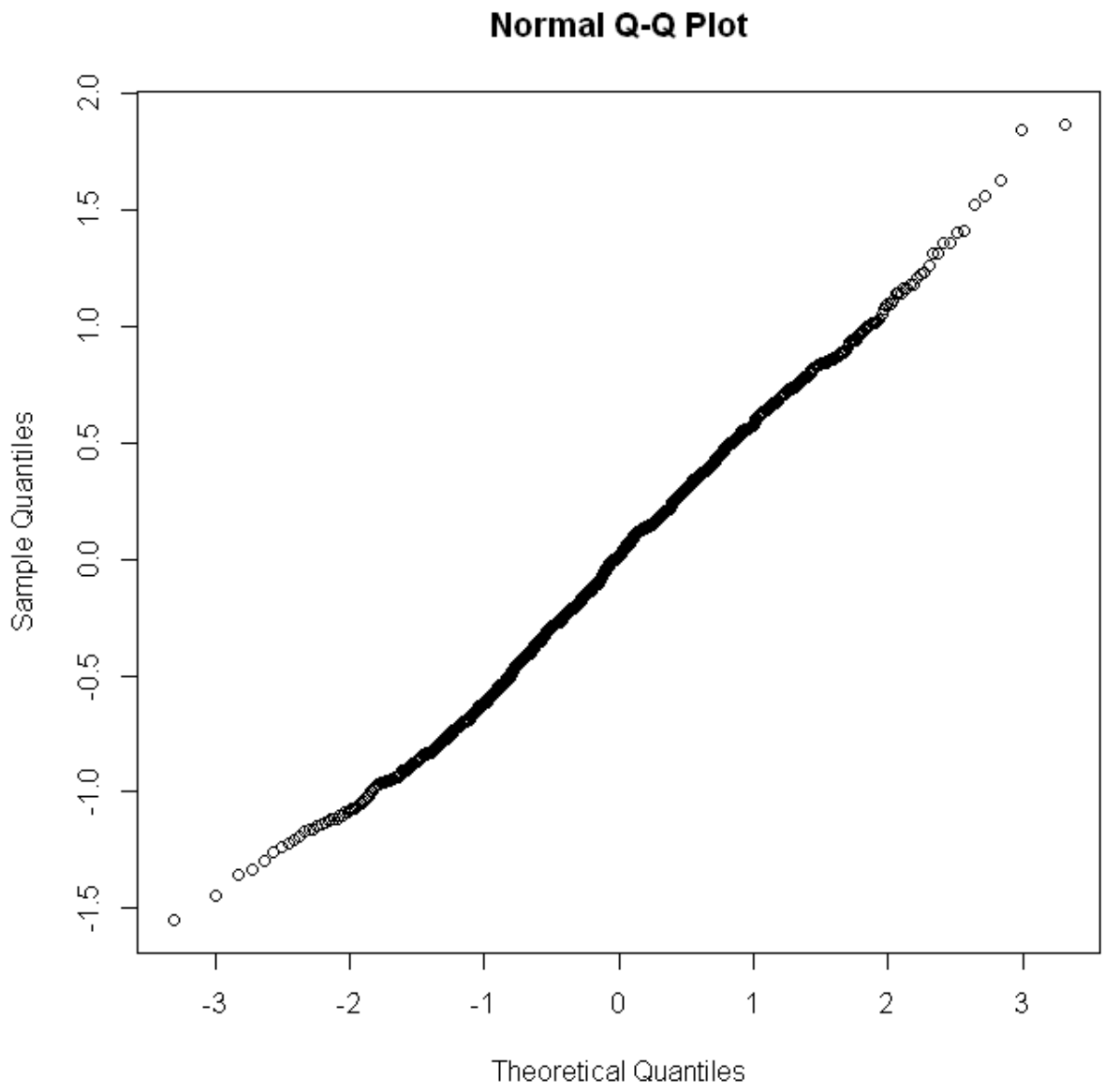


Figure 9.29 Normal probability plot of standardised residuals to assess distribution of errors and detect any evidence of non-normal residual distribution and abnormal variance from model S1b.

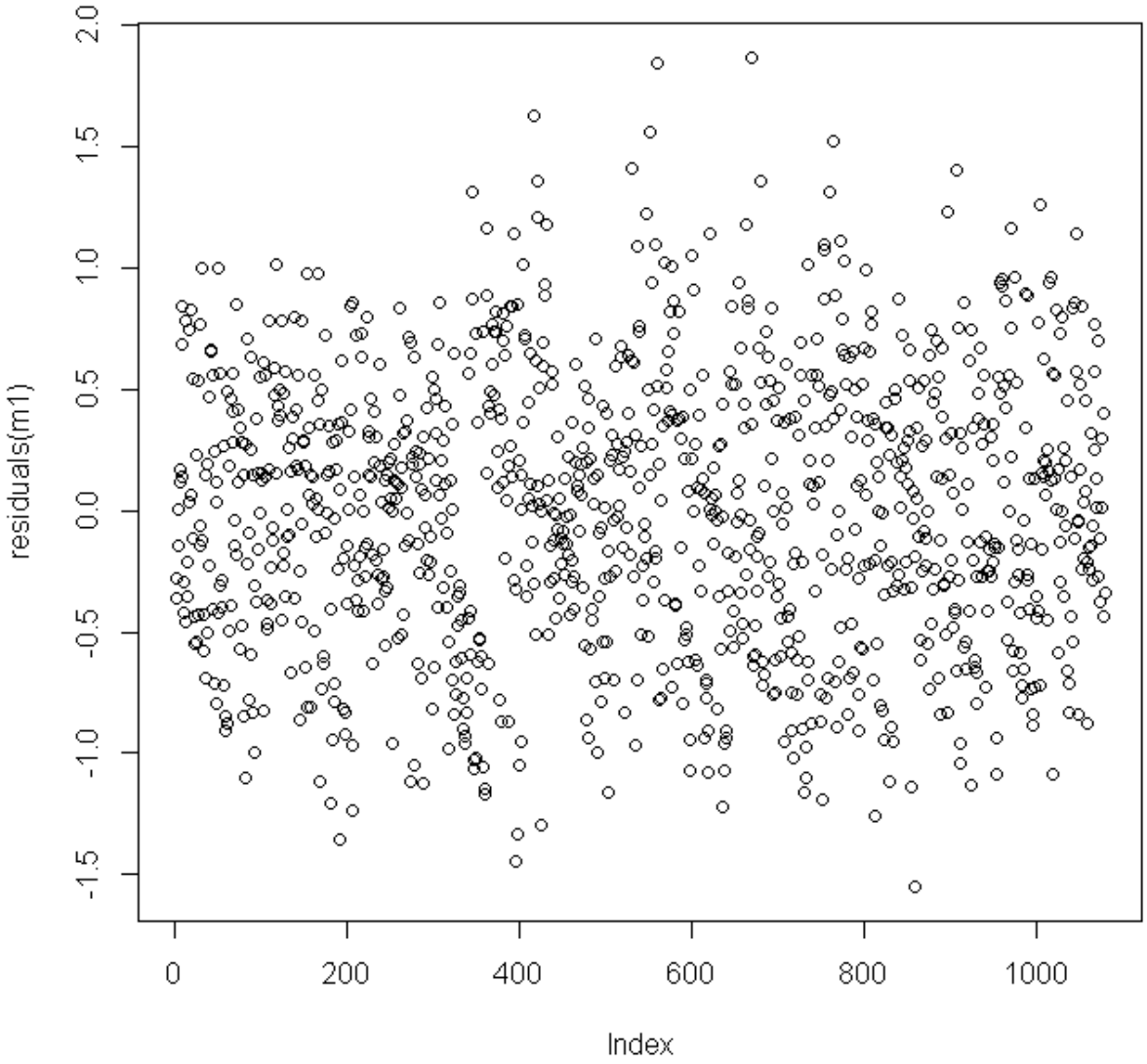


Figure 9.30 Scatter plot of residual to assess homoscedascity, independence and variance of the residual distribution from model S1b.

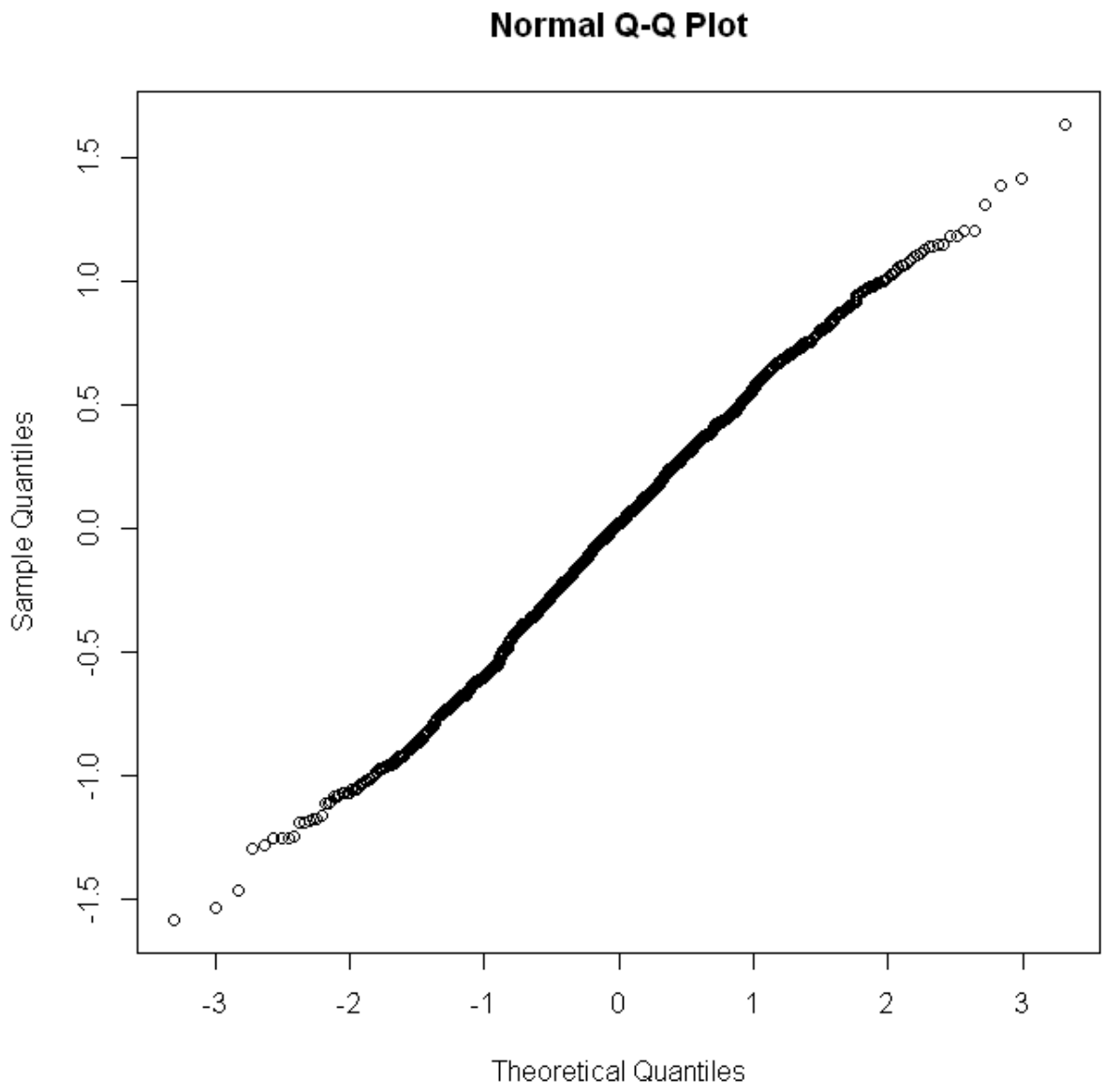


Figure 9.31 Normal probability plot of standardised residuals to assess distribution of errors and detect any evidence of non-normal residual distribution and abnormal variance from model S2.

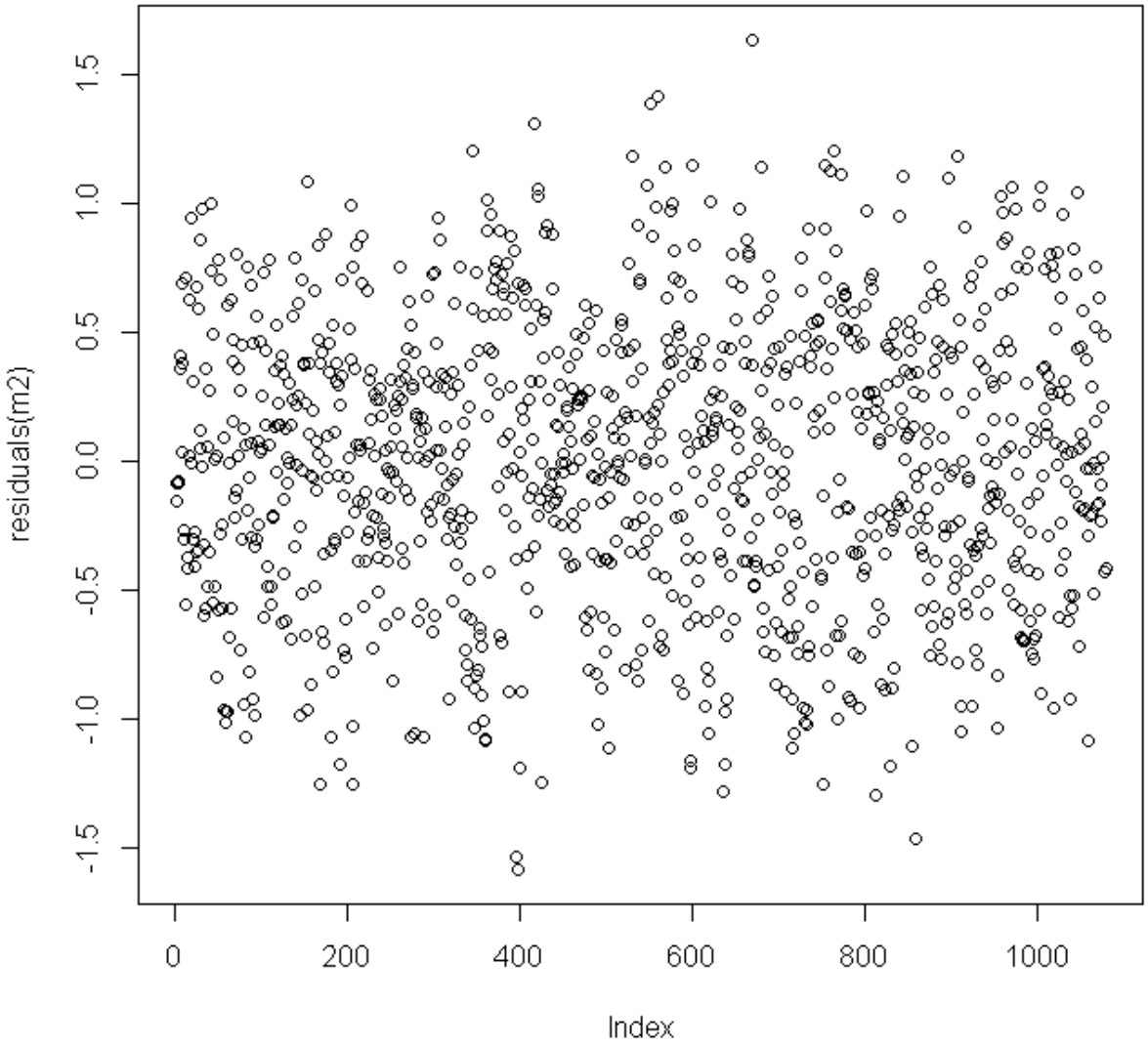


Figure 9.32 Scatter plot of residuals to assess homoscedascity, independence and variance of the residual distribution from model S2.

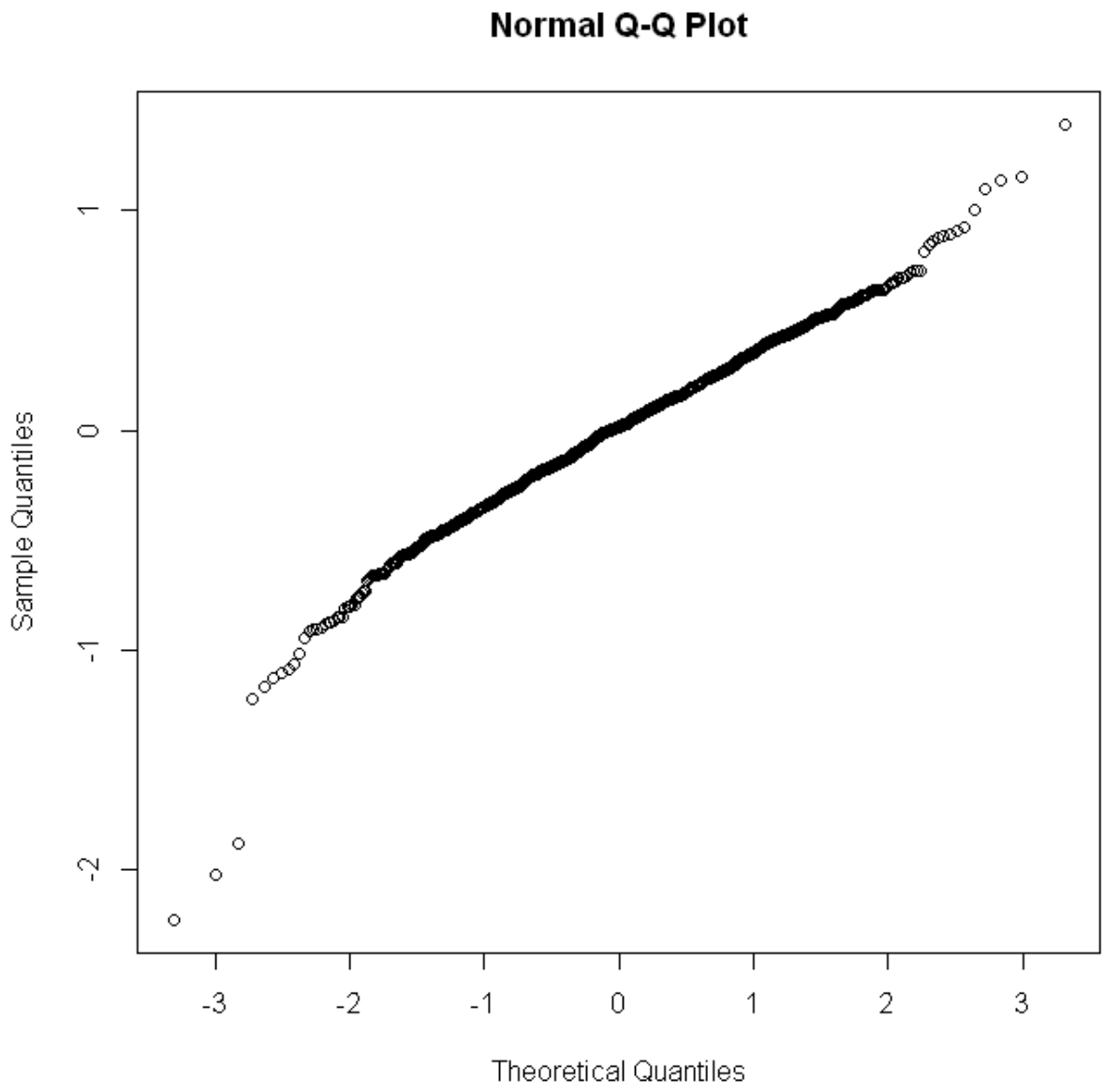


Figure 9.33 Scatter plot of residual and fitted values to assess homoscedascity, independence and variance of the residual distribution from model S3a.

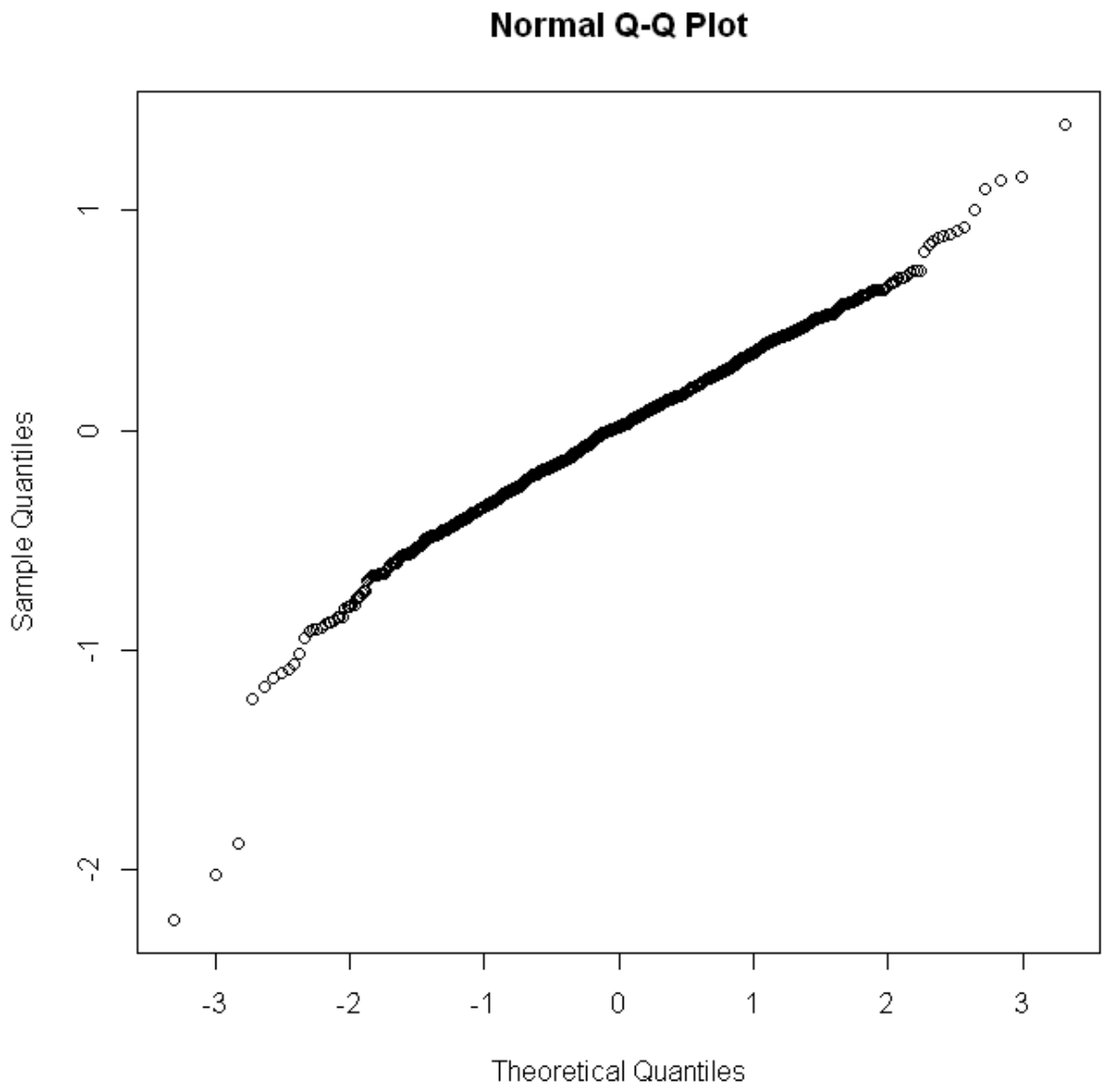


Figure 9.34 Residuals from model S3a.

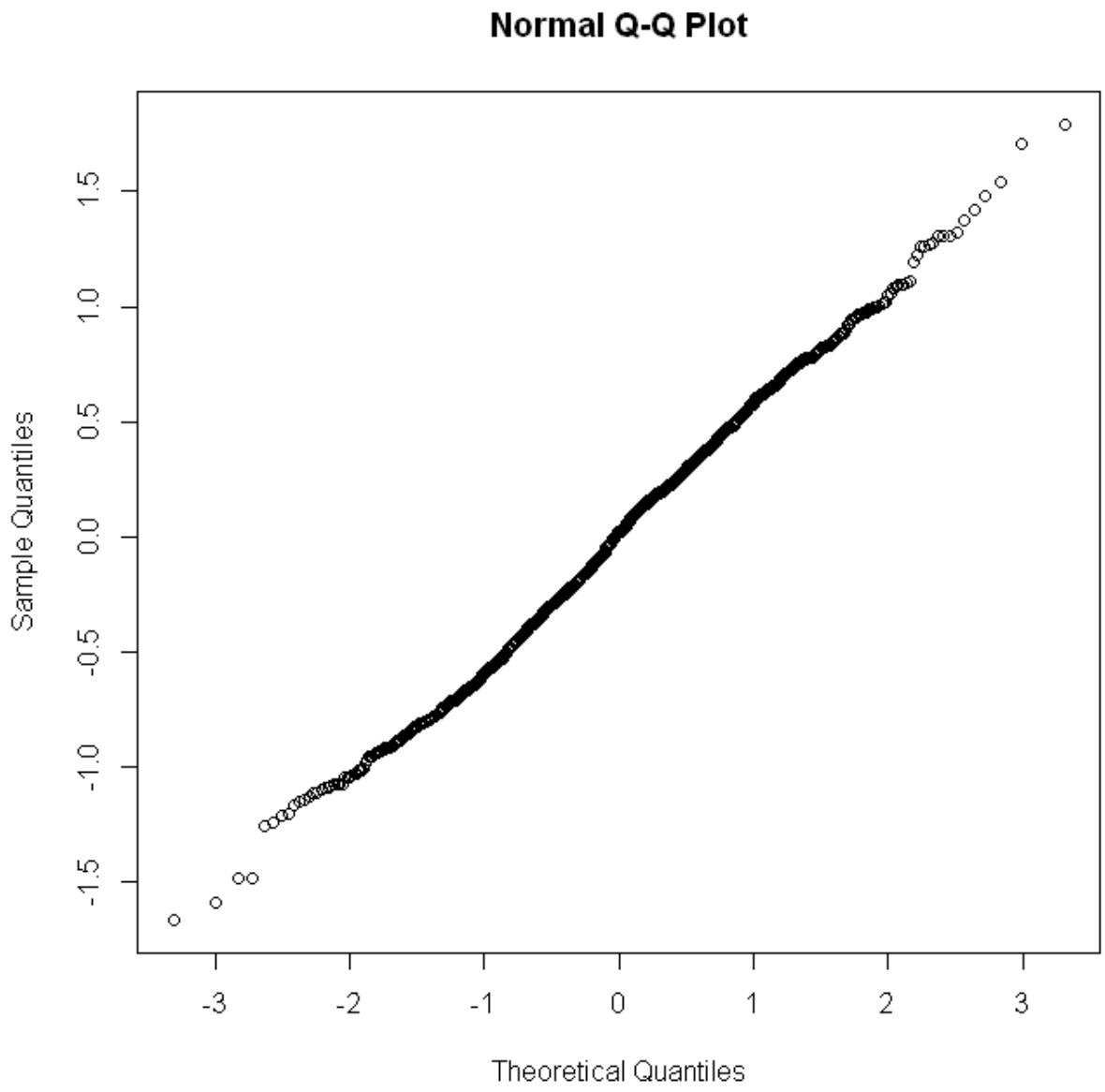


Figure 9.35 Normal probability plot of standardised residuals to assess distribution of errors and detect any evidence of non-normal residual distribution and abnormal variance from model S3b.

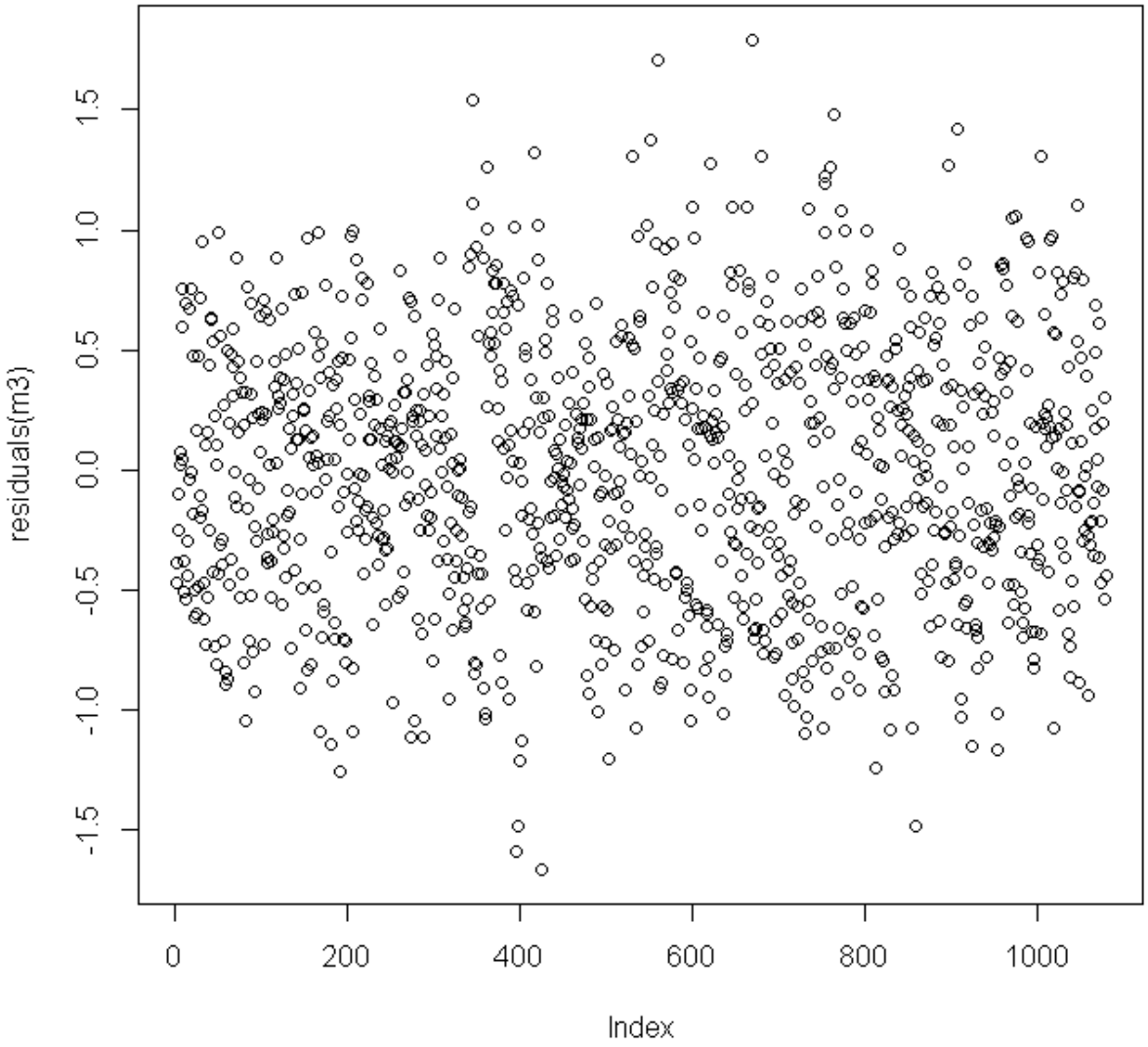


Figure 9.36 Scatter plot of residuals to assess homoscedascity, independence and variance of the residual distribution from model S3b.

9.6 Appendix F - Model diagnostic plots for ratio of empyema admissions to pneumonia admissions (Models R0-2)

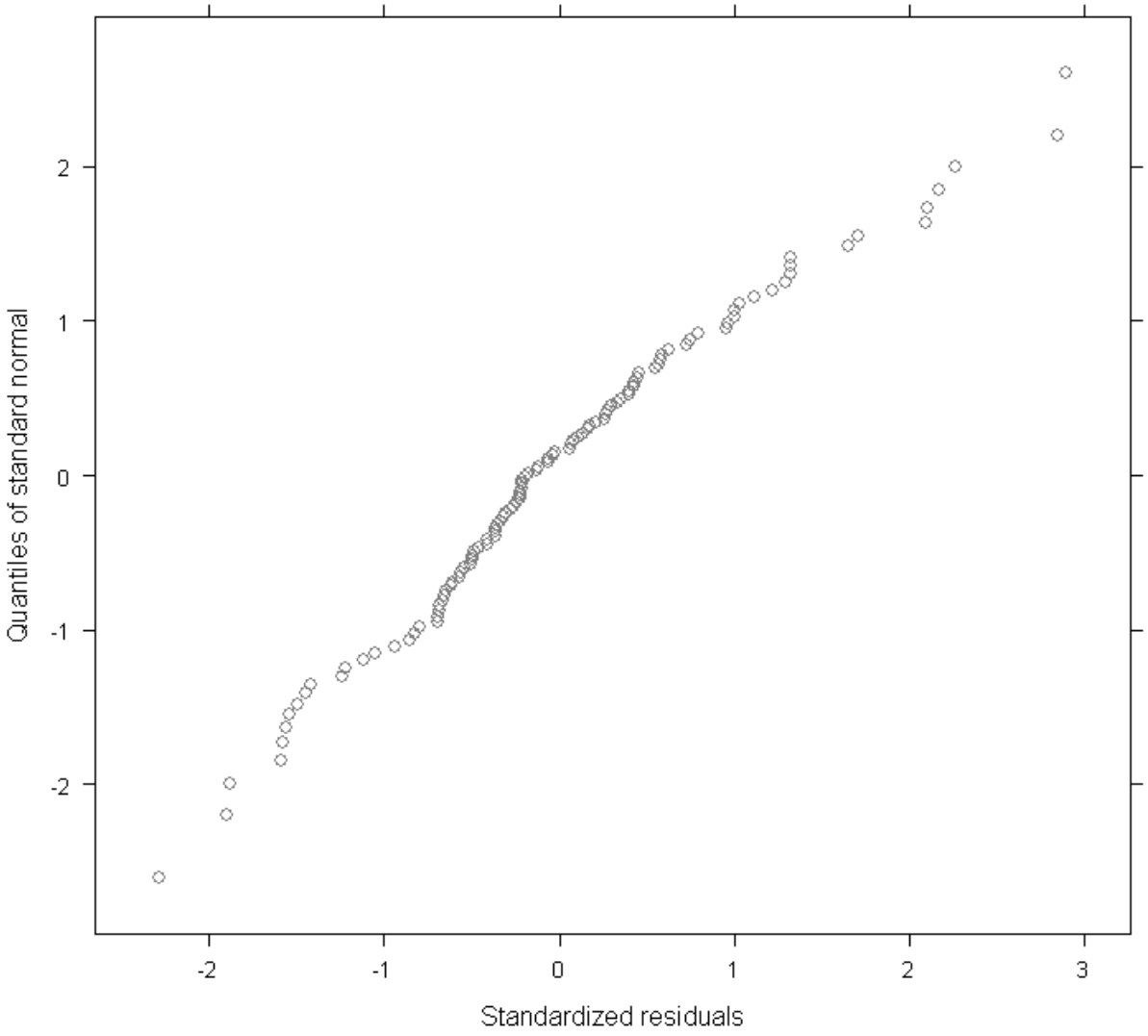


Figure 9.37 Normal probability plot of standardised residuals to assess distribution of errors and detect any evidence of non-normal residual distribution and abnormal variance from model R0 illustrating some overdispersion of the residuals at the extremes.

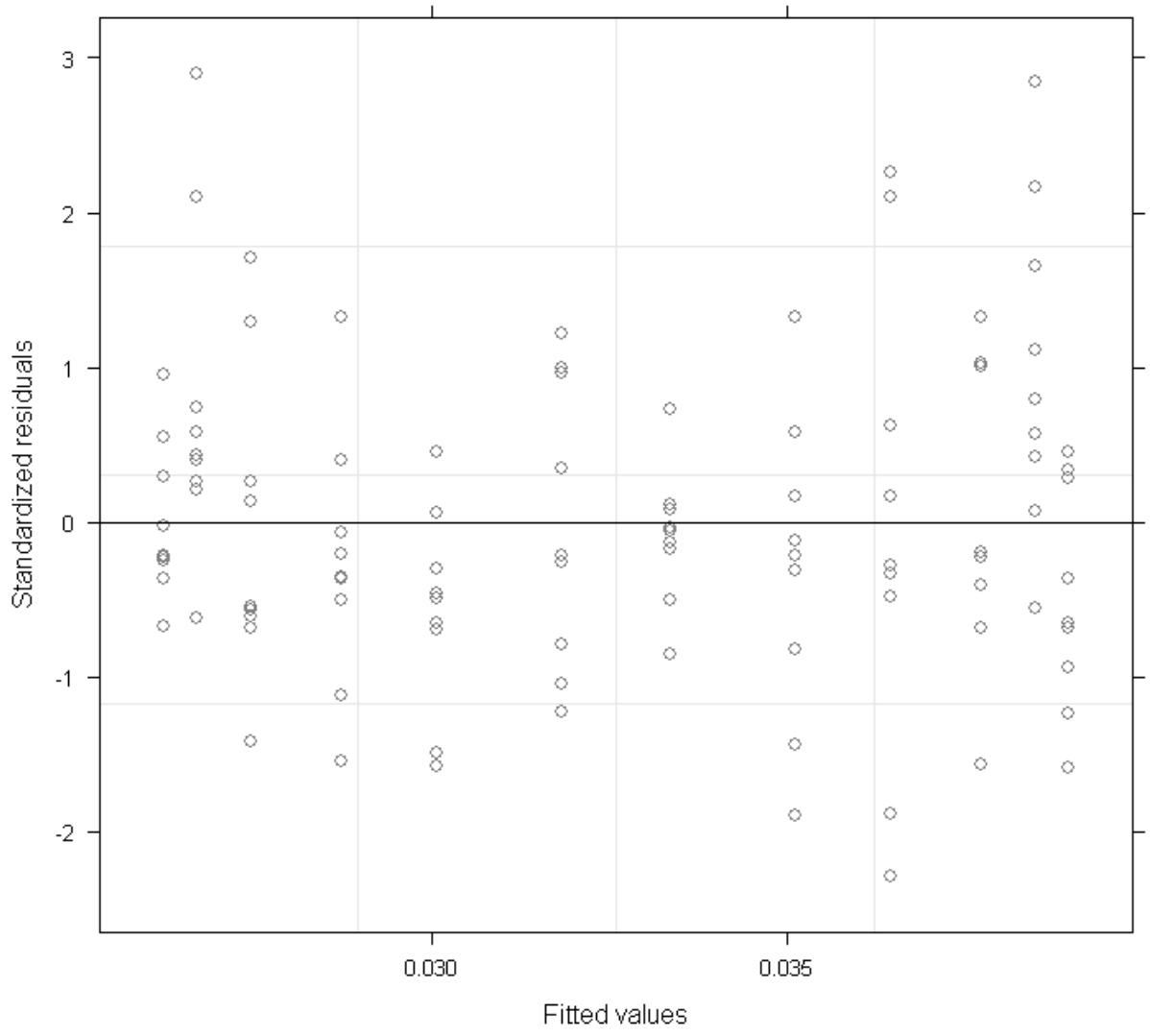


Figure 9.38 Scatter plot of residual and fitted values to assess homoscedascity, independence and variance of the residual distribution for model R0. There is clear variation in the variance of the distribution suggesting abnormal variance in the residuals suspicious of a lack of independence between observations.

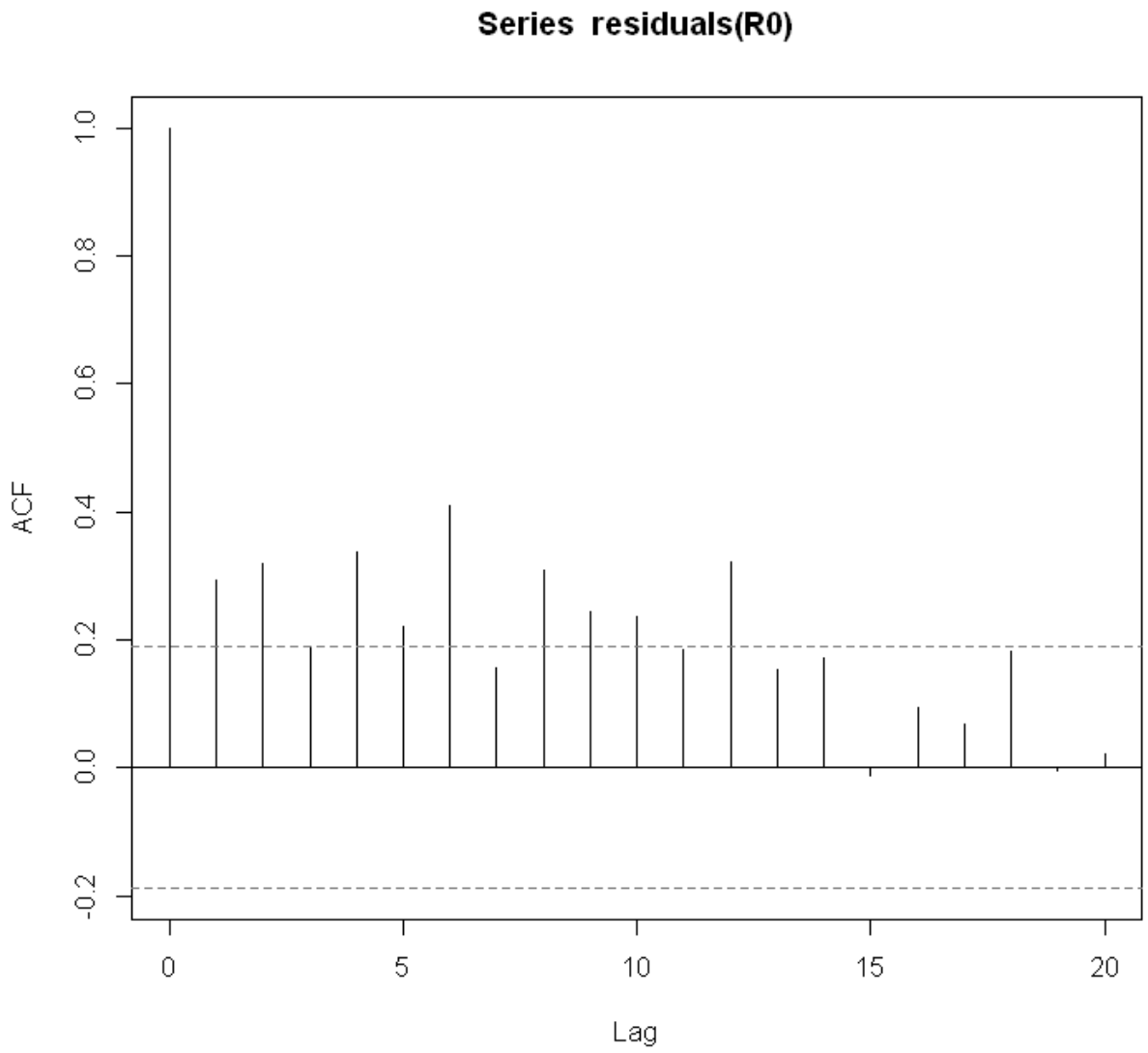


Figure 9.39 Autocorrelation plot illustrating significant autocorrelation in residuals of model R0. ***Dotted lines indicate cut-off for presences of significant autocorrelation. ACF – autocorrelation function (size of autocorrelation, lag indicates number of months between observations when correlation occurs).***

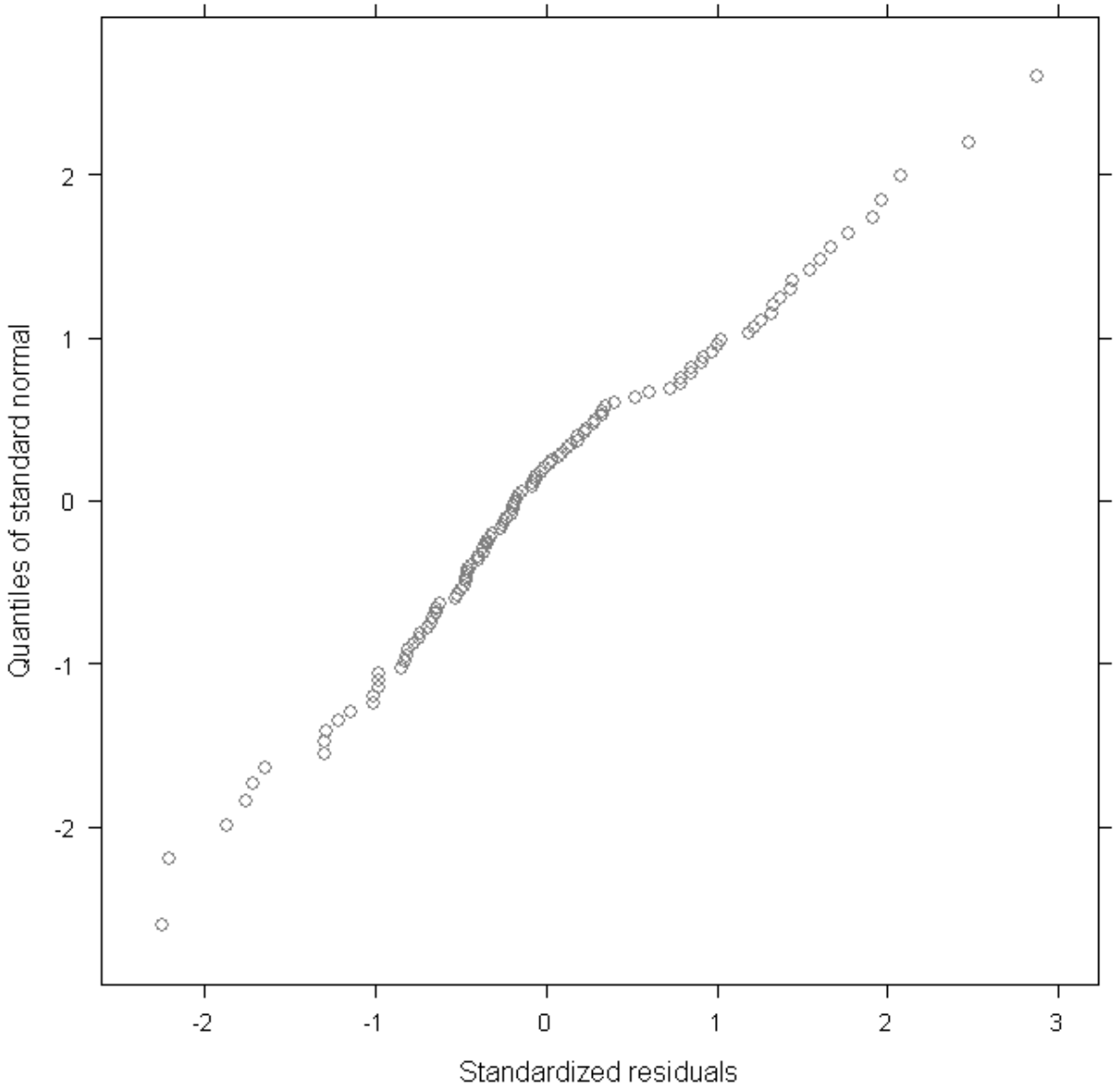


Figure 9.40 Normal probability plot of standardised residuals to assess distribution of errors and detect any evidence of non-normal residual distribution and abnormal variance from model R1 showing some mild overdispersion of extreme values.

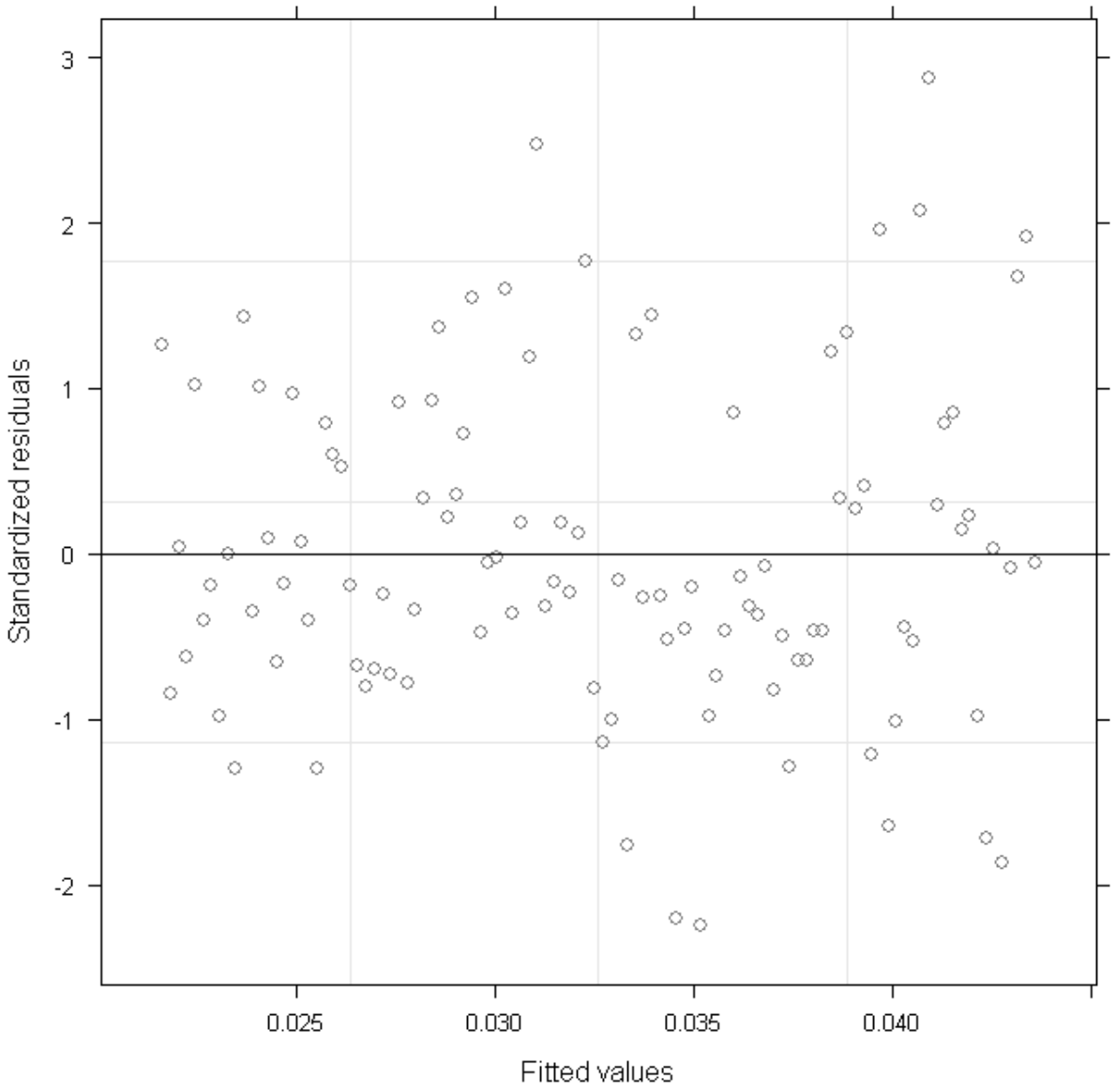


Figure 9.41 Scatter plot of residual and fitted values to assess homoscedascity, independence and variance of the residual distribution for model R1. The residuals have a normal and symmetrical distribution suggesting the model assumptions of independence are valid.

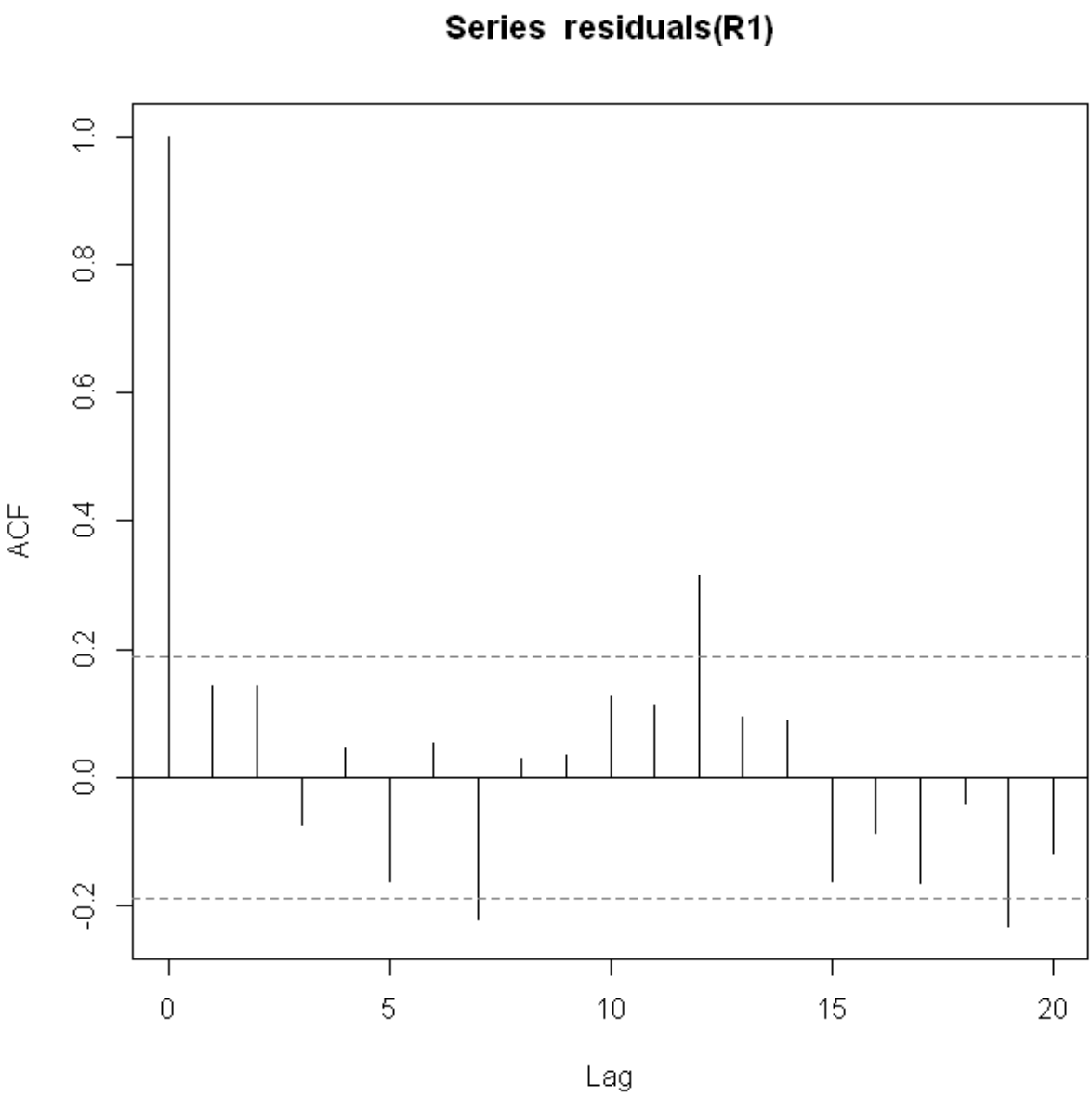


Figure 9.42 Autocorrelation plot illustrating presence of significant autocorrelation in residuals of model R1 at lags7, 12 and 19. ***Dotted lines indicate cut-off for presences of significant autocorrelation. ACF – autocorrelation function (size of autocorrelation, lag indicates number of months between observations when correlation occurs).***

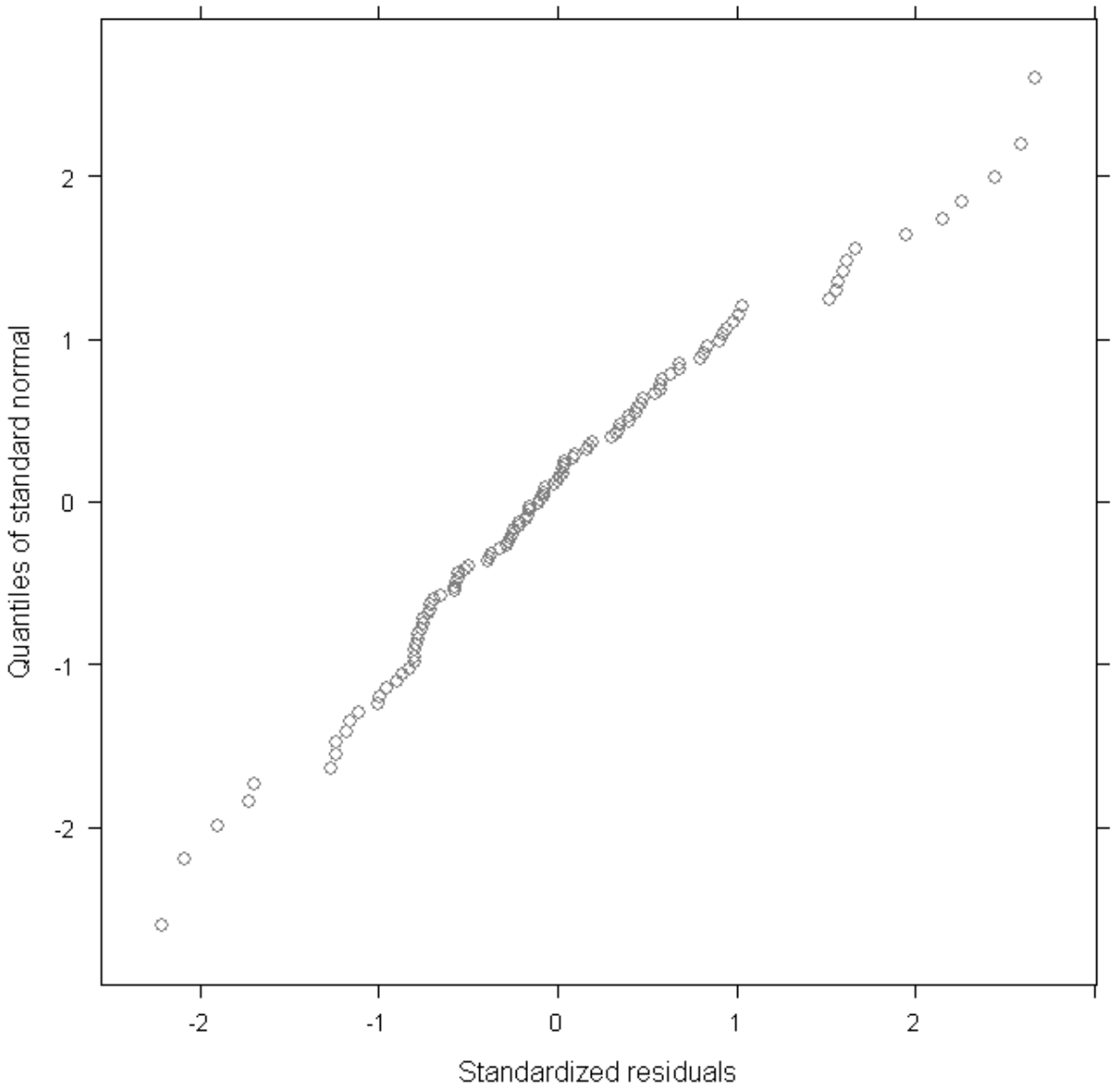


Figure 9.43 Normal probability plot of standardised residuals to assess distribution of errors and detect any evidence of non-normal residual distribution and abnormal variance from model R2. There is evidence of overdispersion of the residuals at the extreme values.

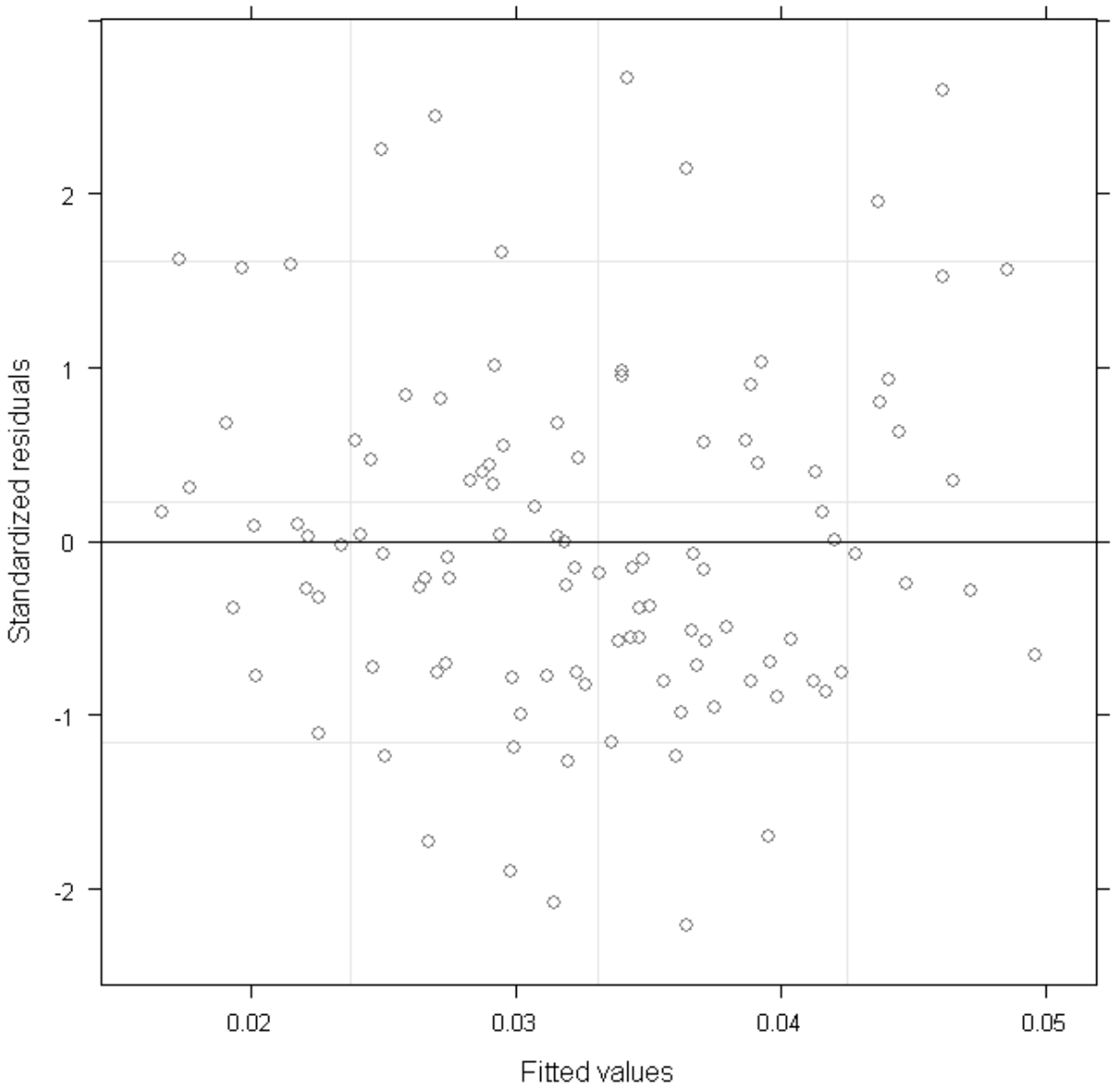


Figure 9.44 Scatter plot of residual and fitted values to assess homoscedascity, independence and variance of the residual distribution for model R2. The residual distribution is normal.

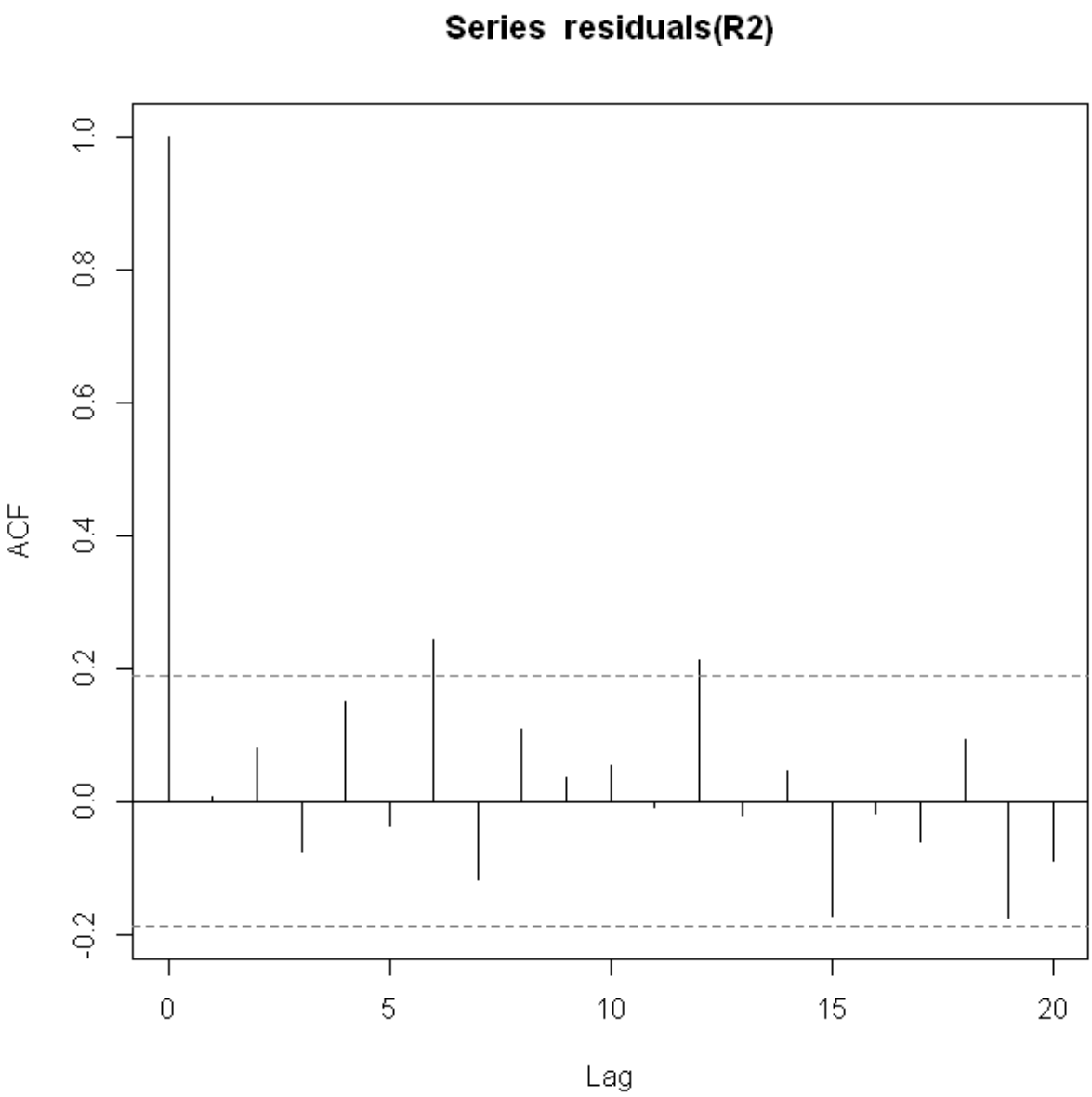


Figure 9.45 Autocorrelation plot illustrating presence of significant autocorrelation in residuals of model R2 at lags 7 and 12. ***Dotted lines indicate cut-off for presences of significant autocorrelation. ACF – autocorrelation function (size of autocorrelation, lag indicates number of months between observations when correlation occurs).***

9.7 Appendix G - Model diagnostic plots for definition of periodicity and cyclicity within ratio of empyema to pneumonia admissions (Model R3)

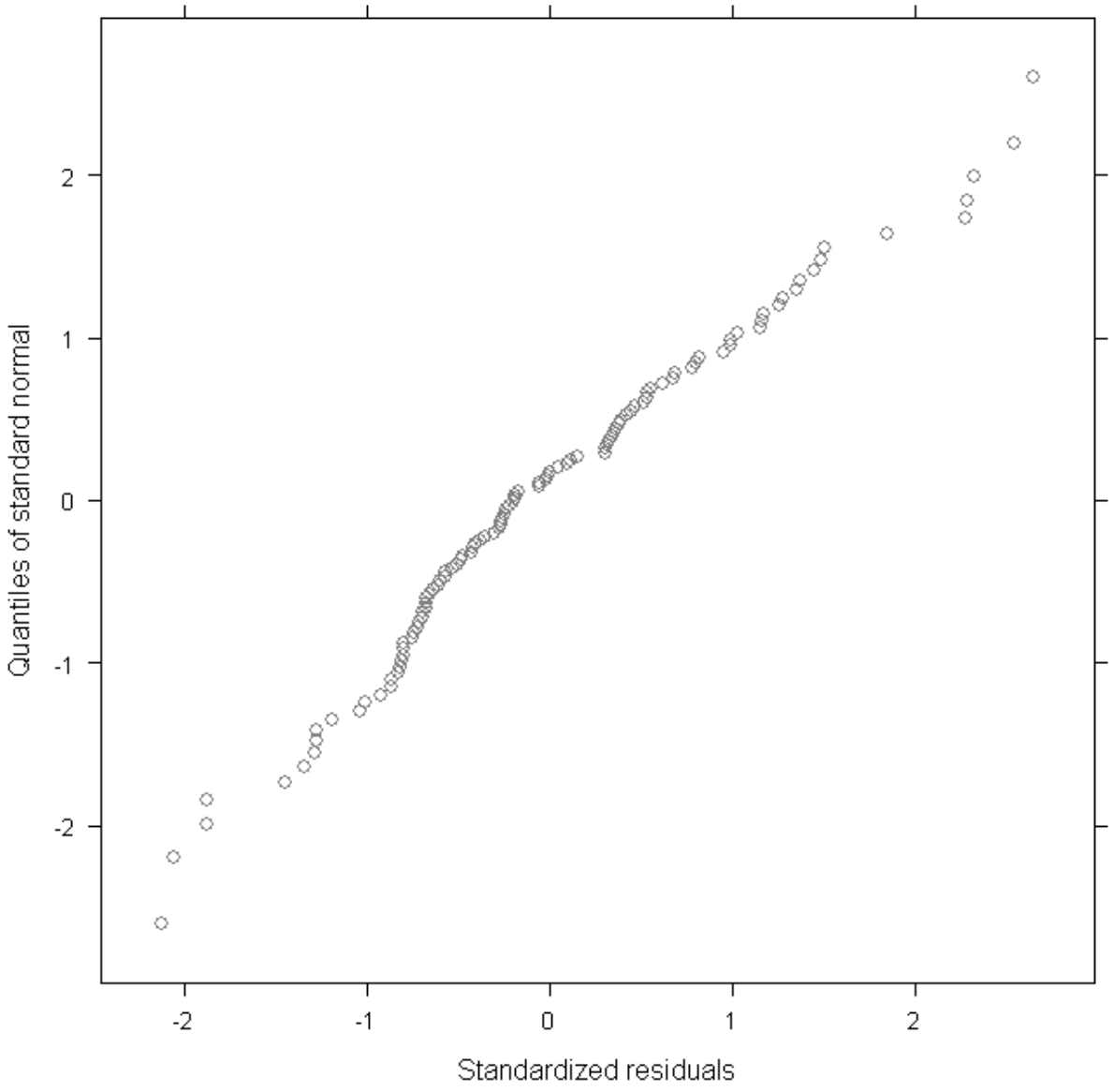


Figure 9.46 Normal probability plot of standardised residuals from model R3.

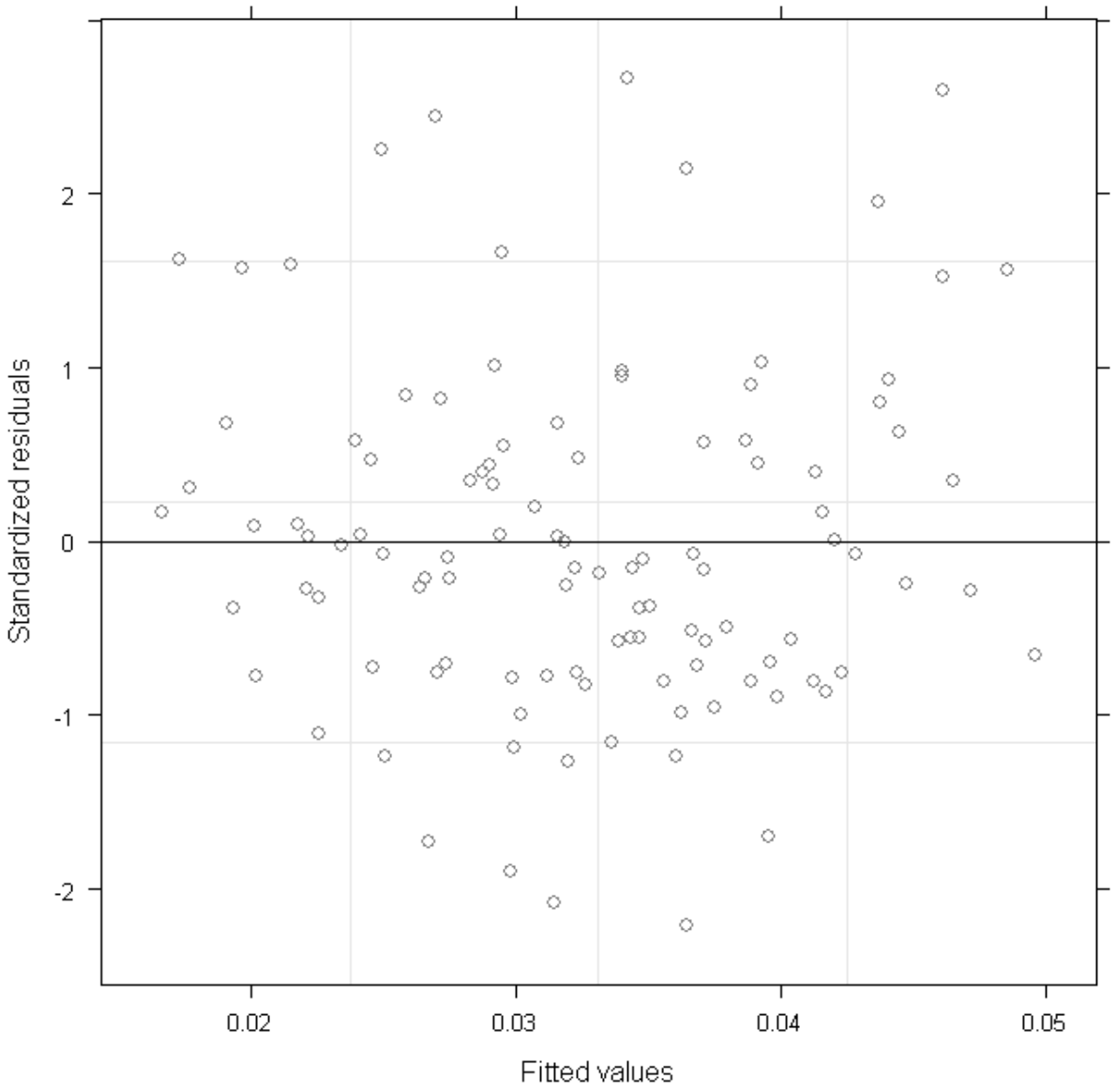


Figure 9.47 Scatter plot of residual and fitted values to assess homoscedascity, independence and variance of the residual distribution for model R3. The residual distribution is normal.

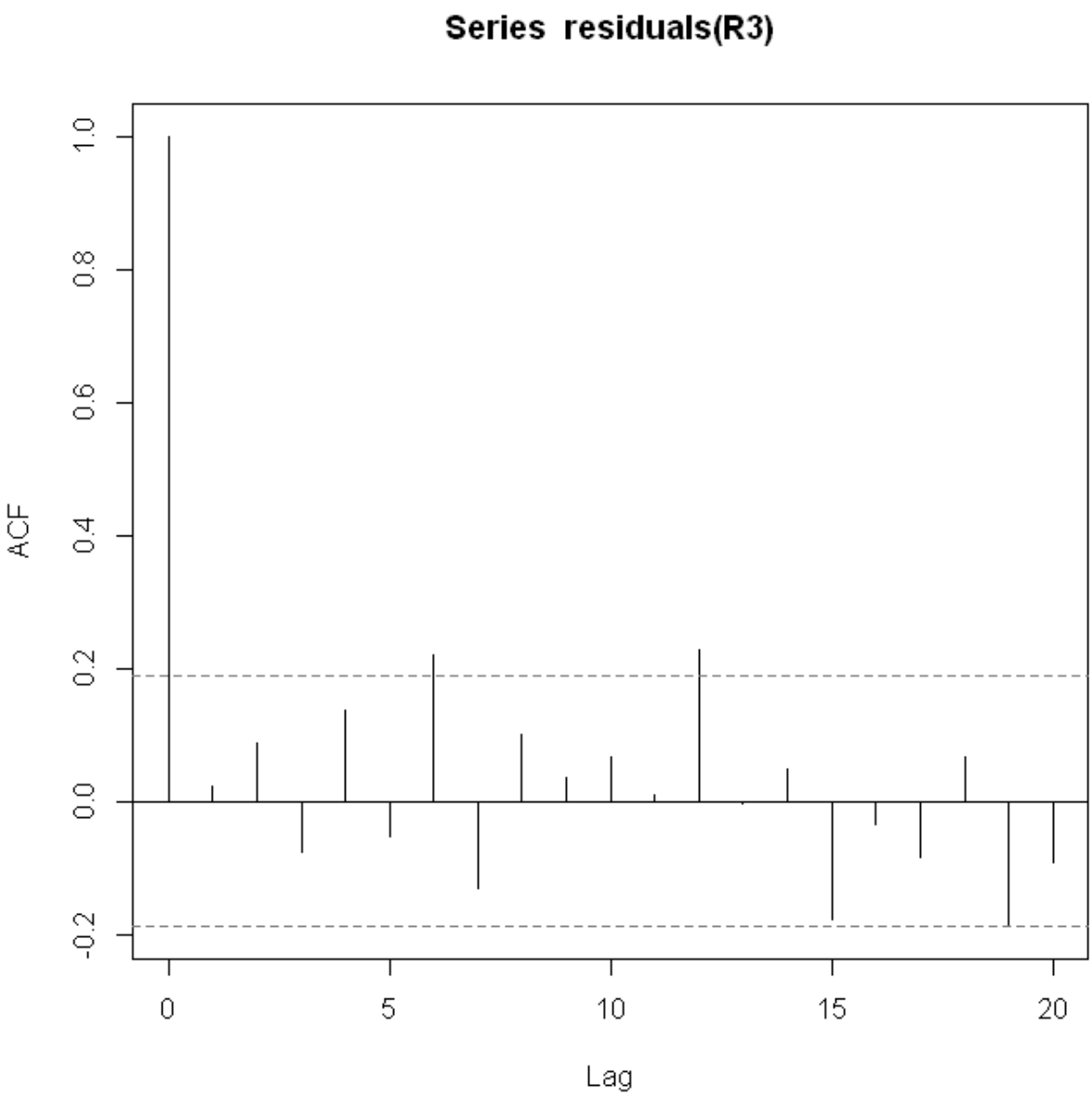


Figure 9.48 Plot illustrating presence of significant autocorrelation in residuals of model R3 at lags 6 and 12. ***Dotted lines indicate cut-off for presences of significant autocorrelation. ACF – autocorrelation function (size of autocorrelation, lag indicates number of months between observations when correlation occurs).***

9.8 Appendix H - Model diagnostic plots from investigation of relationship between empyema admissions and isolations of *S. pneumoniae, S. pyogenes, S. aureus* and *M. pneumoniae* (Model P0 and S0).

The residuals of the initial models which did not include any autocorrelative structure demonstrated clear evidence of serial dependence (Models P0 (**Figure 9.49**) and S0 (**Figure 9.50**)).

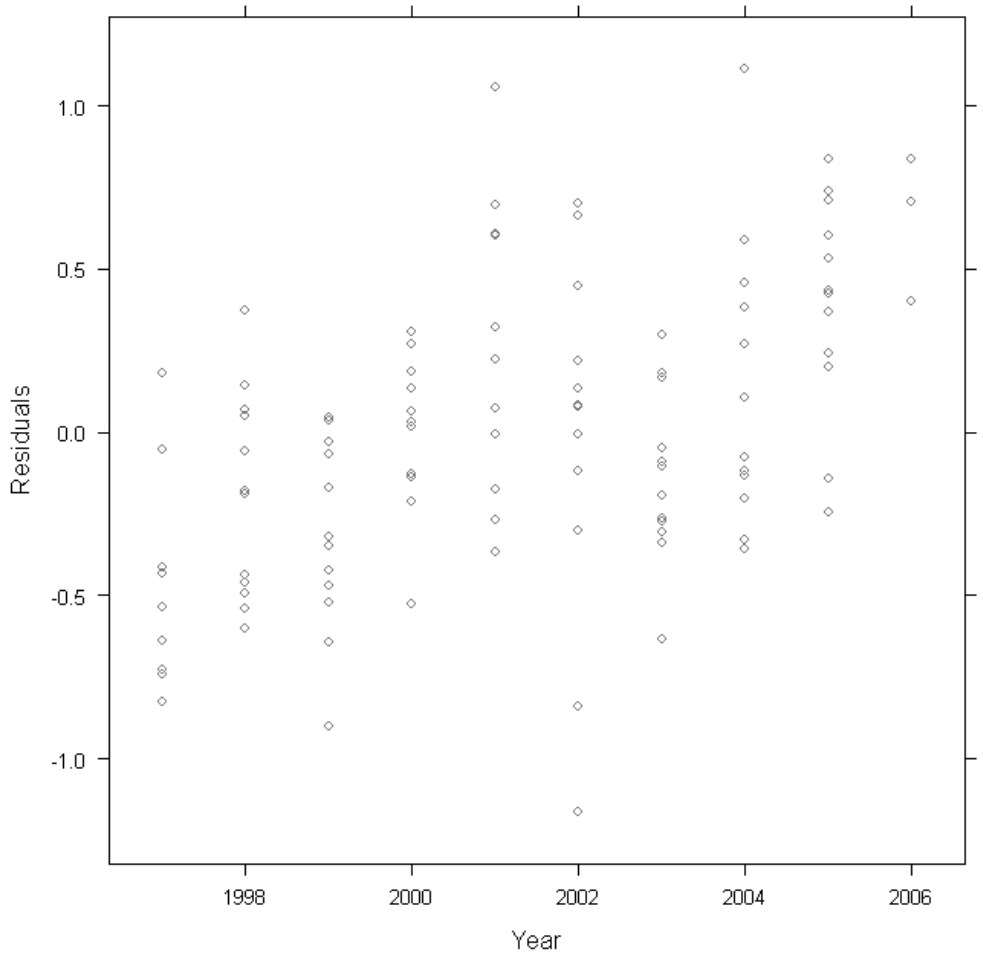


Figure 9.49 Residuals of initial time series model (P0) without autocorrelation structure relating empyema admissions to isolations of *S. pneumoniae, S. pyogenes, S. aureus* and *M. pneumoniae* from children (0-14 years) in England from April 1997 to April 2006 demonstrating significant serial dependence.

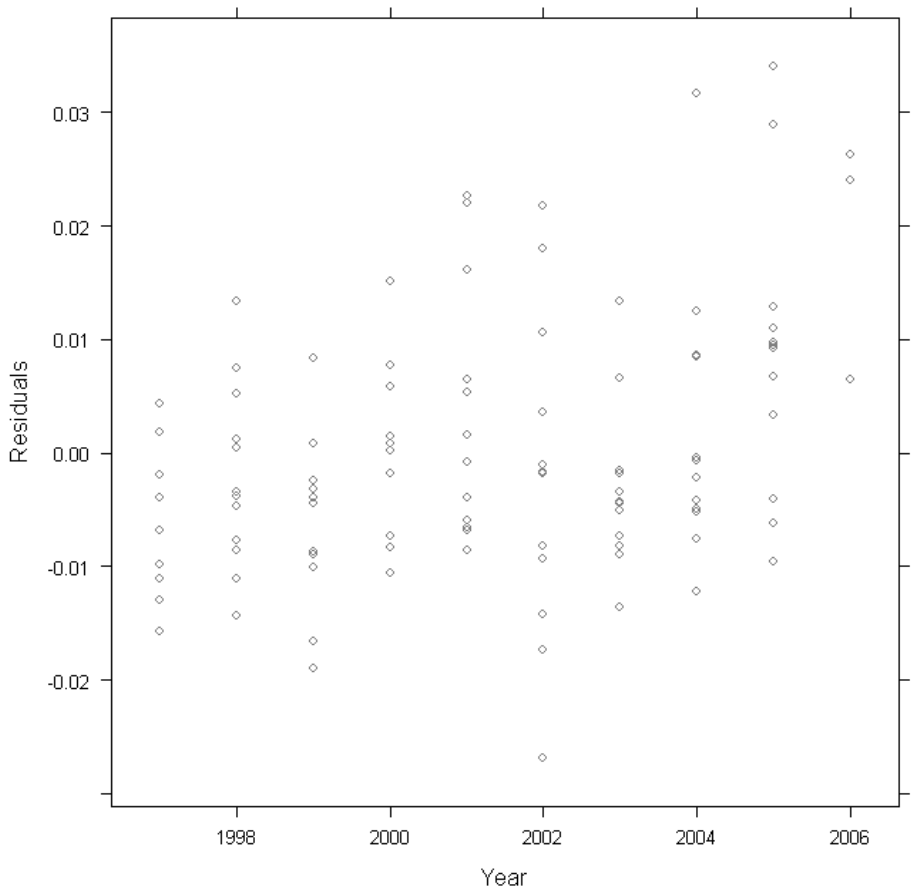


Figure 9.50 Residuals of initial time series model (S0) without autocorrelation structure relating ratio of empyema admissions to pneumonia admissions to isolations of *S. pneumoniae*, *S. pyogenes*, *S. aureus* and *M. pneumoniae* from children (0-14 years) in England from April 1997 to April 2006 demonstrating serial dependence.

Correlation structure	**AIC**	**Phi / Theta**
None	159.47	-
Compound symmetry	161.47	-3.44×10^{-18} (Rho)
Auto-regressive model of order 1	142.79	0.52
Auto-regressive moving average p = 1, q = 0	142.79	0.52
Auto-regressive moving average p = 2, q = 0	125.92	0.34 , 0.43
Auto-regressive moving average p = 3, q = 0	**125.58**	**0.26, 0.39, 0.18**
Auto-regressive moving average p = 1, q = 1	NA	NA
Auto-regressive moving average p = 2, q = 1	NA	NA
Auto-regressive moving average p = 3, q = 1	NA	NA
Auto-regressive moving average p = 1, q = 2	NA	NA
Auto-regressive moving average p = 1, q = 3	NA	NA
Auto-regressive moving average p = 2, q = 2	112.49	7.12×10^{-7}, 0.99, 0.22, -0.77
Auto-regressive moving average p = 2, q = 3	NA	NA
Auto-regressive moving average p = 3, q = 3	NA	NA
Auto-regressive moving average p = 0, q = 1	153.24	0.24
Auto-regressive moving average p = 0, q = 2	142.75	0.27, 0.34
Auto-regressive moving average p = 0, q = 3	140.29	0.32, 0.38, 0.17

Table 9.2 Results of modelling relating empyema admissions to isolations of *S. pneumoniae*, *S. pyogenes*, *S. aureus* and *M. pneumoniae* from children (0-14 years) in England from April 1997 to April 2006 and optimization of the residual autocorrelation structure – Model P0.

Correlation structure	**AIC**	**Phi / Theta**
None	-599.97	-
Compound symmetry	-597/97	3.78×10^{-9} (rho)
Auto-regressive model of order 1	-602.56	0.24
Auto-regressive moving average p = 1, q = 0	-602.56	0.24
Auto-regressive moving average p = 2, q = 0	**-604.65**	**0.24. 0.22**
Auto-regressive moving average p = 3, q = 0	-602.67	0.25, 0.23, -0.016
Auto-regressive moving average p = 1, q = 1	NA	NA
Auto-regressive moving average p = 2, q = 1	NA	NA
Auto-regressive moving average p = 3, q = 1	NA	NA
Auto-regressive moving average p = 1, q = 2	-609.02	$0.996.68 \times 10^{-9}$, -1
Auto-regressive moving average p = 1, q = 3	NA	NA
Auto-regressive moving average p = 2, q = 2	NA	NA
Auto-regressive moving average p = 2, q = 3	NA	NA
Auto-regressive moving average p = 3, q = 3	NA	NA
Auto-regressive moving average p = 0, q = 1	-601.09	0.16
Auto-regressive moving average p = 0, q = 2	-603.41	0.23, 0.21
Auto-regressive moving average p = 0, q = 3	-601.42	0.23, 0.21, 0.0086

Table 9.3 Results of modelling relating ratio of empyema admissions to pneumonia admissions to isolations of *S. pneumoniae, S. pyogenes, S. aureus* and *M. pneumoniae* from children (0-14 years) in England from April 1997 to April 2006 and optimization of the residual autocorrelation structure – Model S0.

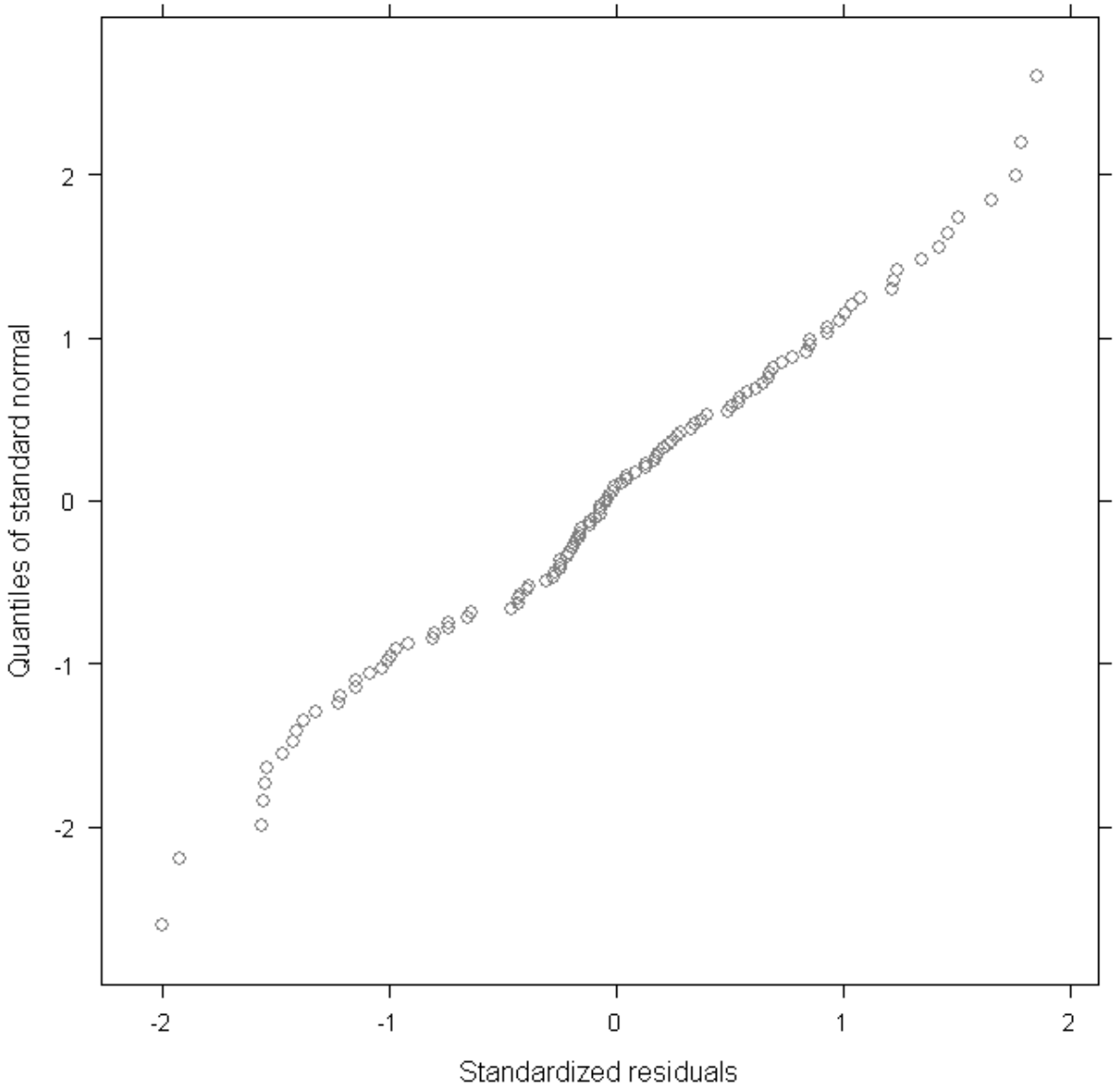


Figure 9.51 Normal probability plot of standardised residuals from model P0 which is within acceptable limits.

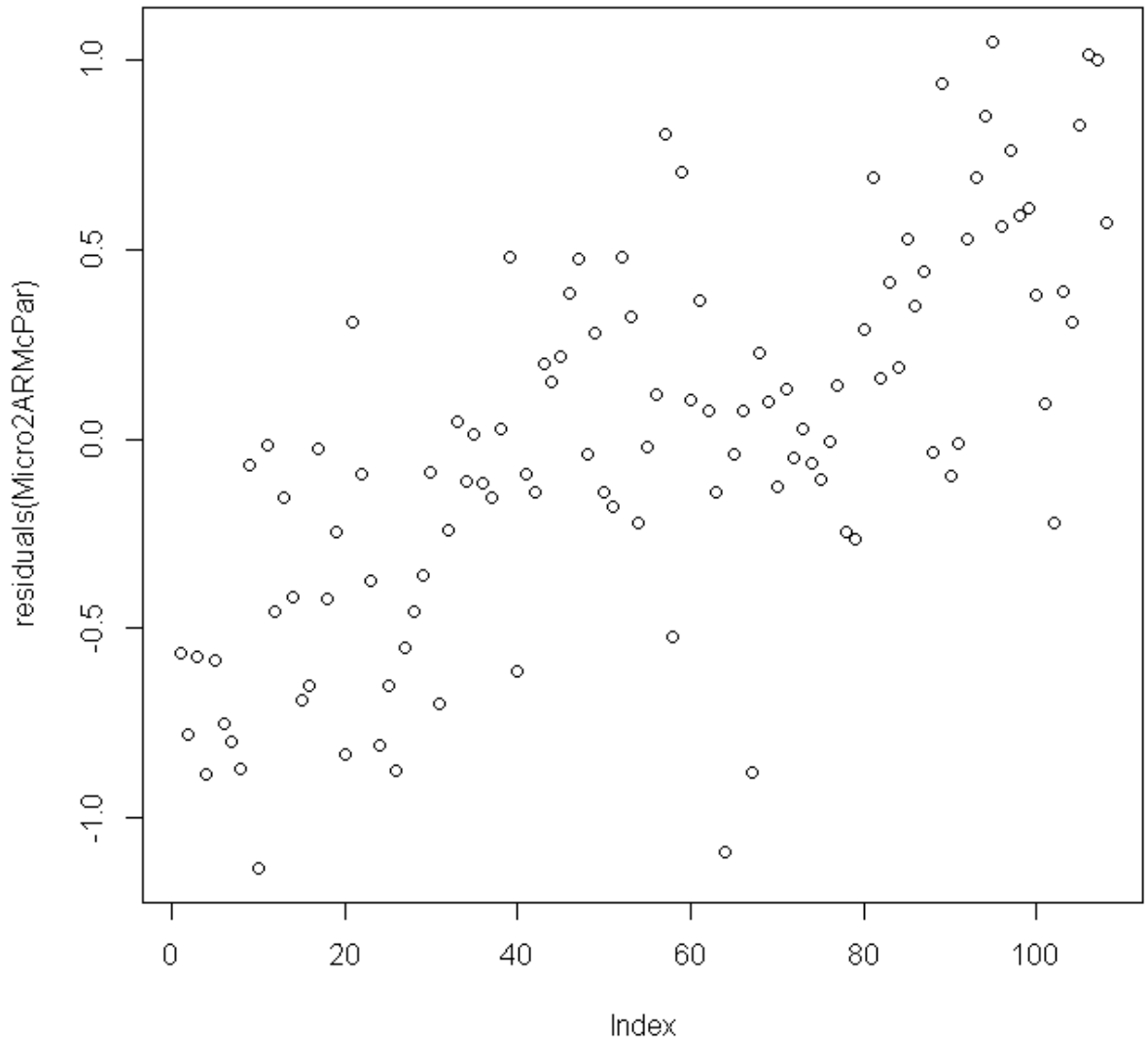


Figure 9.52 Scatter plot of residuals to assess homoscedascity, independence and variance of the residual distribution for model P0 showing clearly pattern to the residuals suggesting a lack of independence between the observed data.

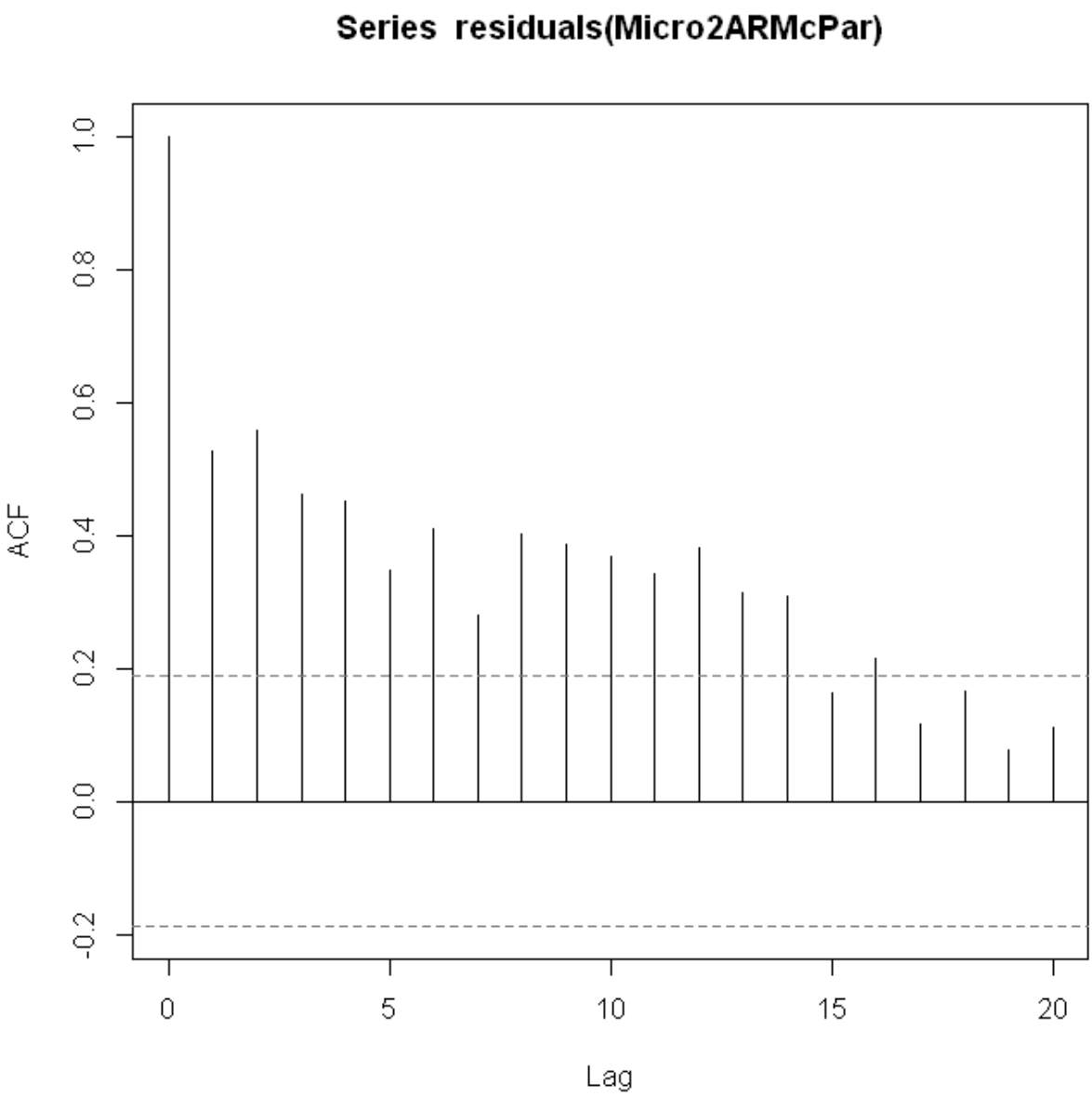


Figure 9.53 Autocorrelation plot of model P0 with autocorrelation present until lag 14. ***Dotted lines indicate cut-off for presences of significant autocorrelation. ACF – autocorrelation function (size of autocorrelation, lag indicates number of months between observations when correlation occurs).***

9.9 Appendix I - Model diagnostic plots for testing of relationship on monthly rate of progression from pneumonia to empyema and isolations of *S. pneumoniae, S. pyogenes, S. aureus* and *M. pneumoniae* (Model S0)

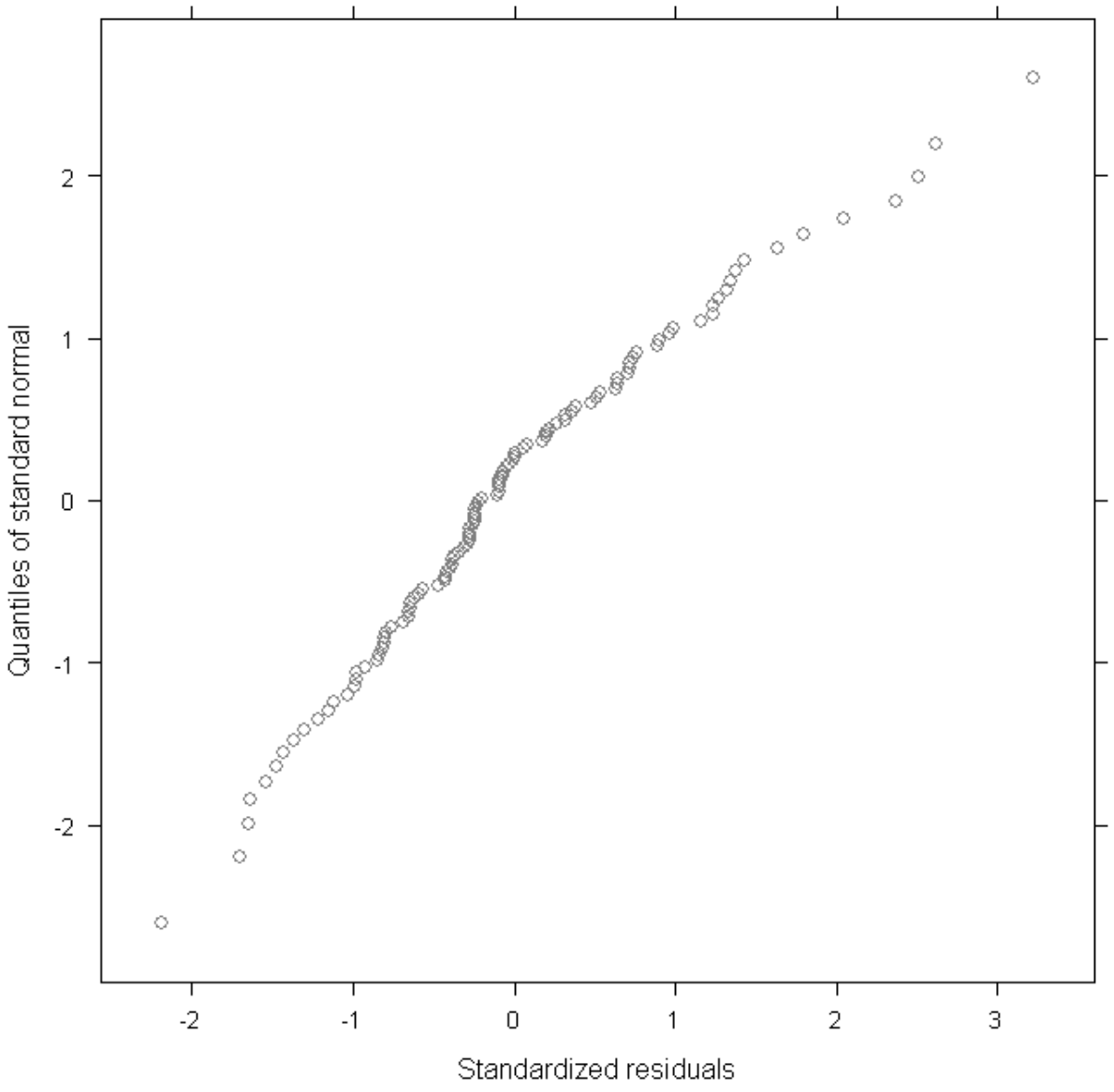


Figure 9.54 Normal probability plot of standardised residuals to assess distribution of errors and detect any evidence of non-normal residual distribution and abnormal variance from model S0 with overdispersion of the residuals at the extreme values.

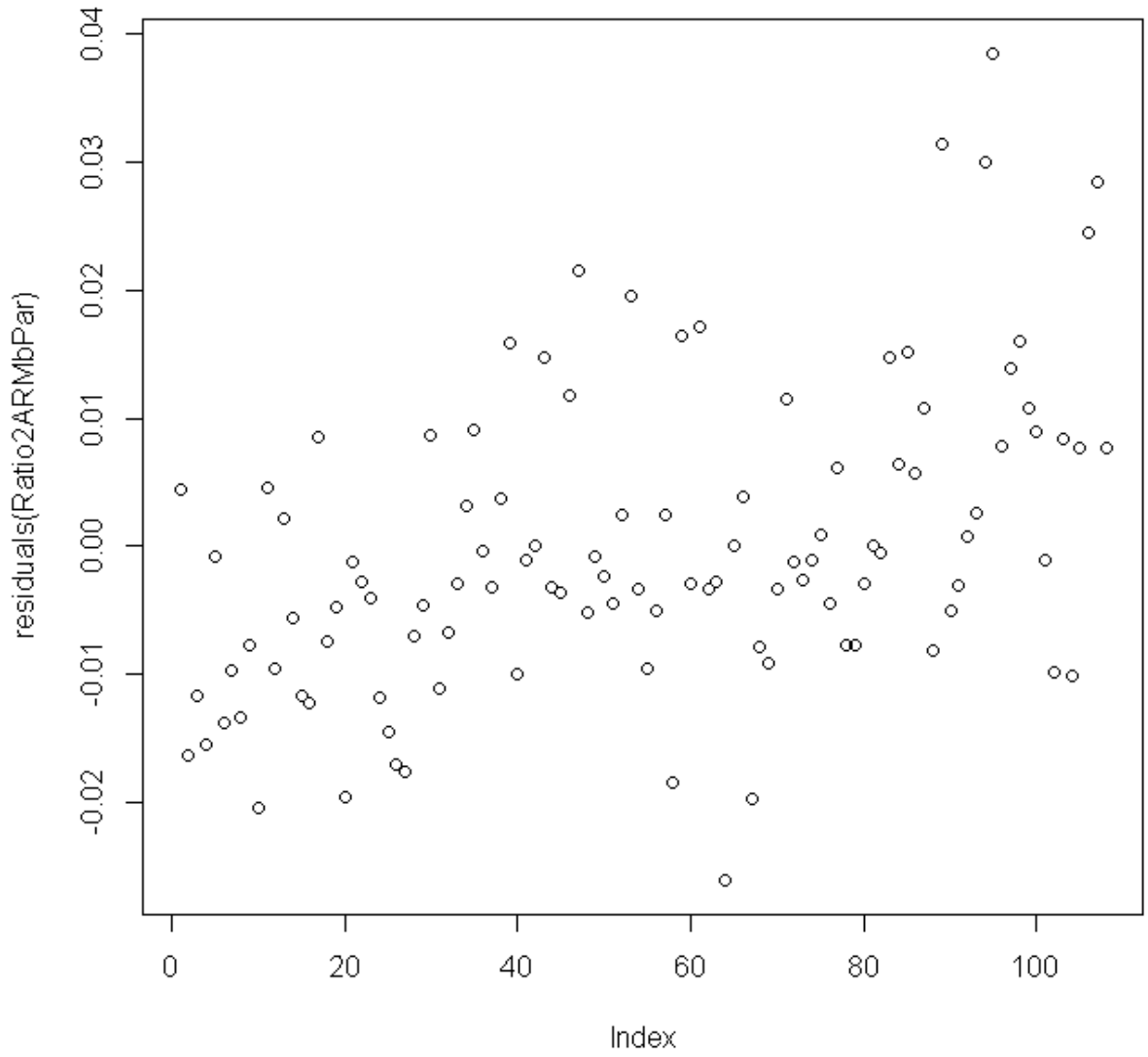


Figure 9.55 Scatter plot of residuals to assess homoscedascity, independence and variance of the residual distribution for model S0 showing clearly pattern to the residuals suggesting a lack of independence between the observed data.

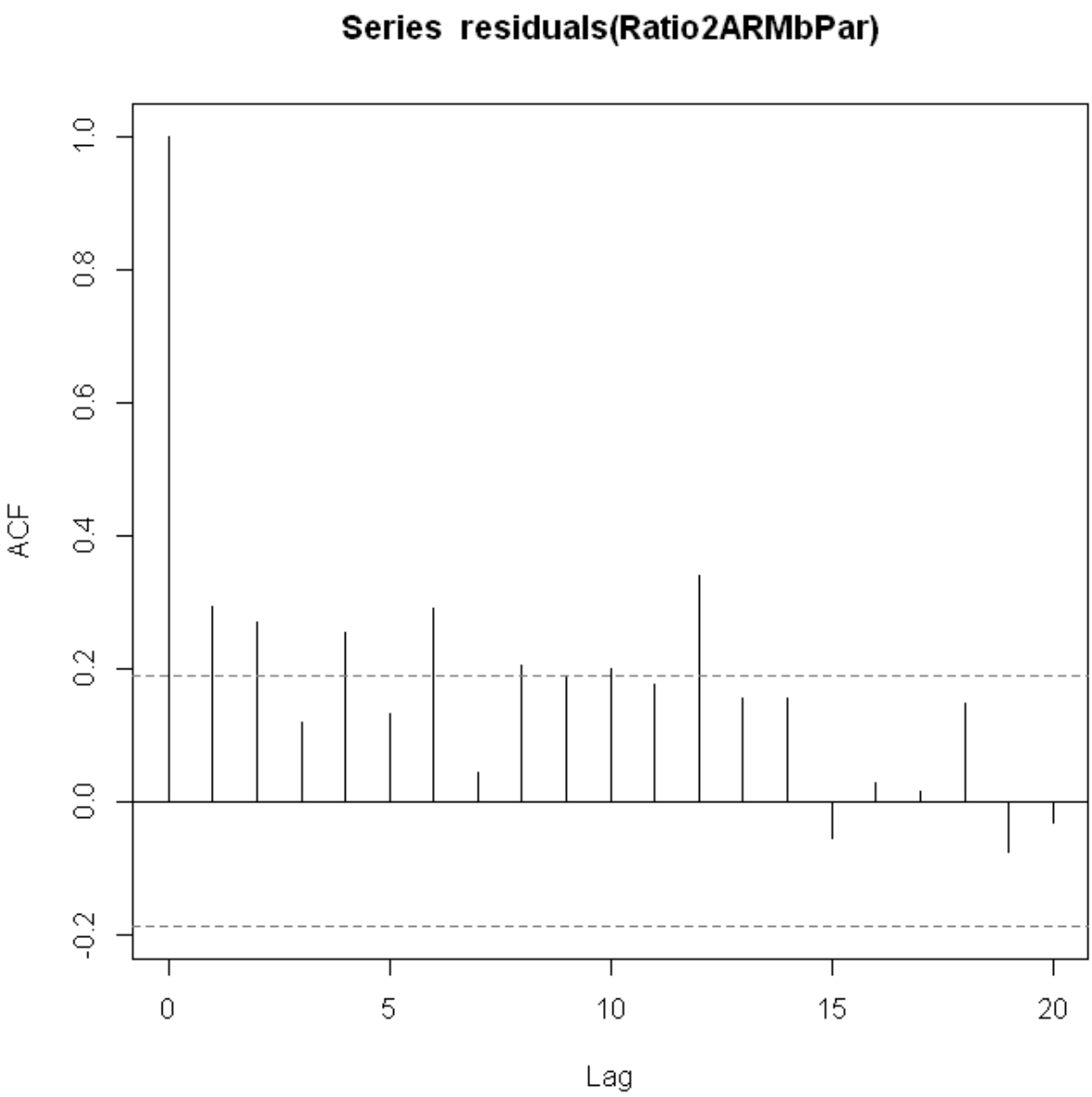


Figure 9.56 Autocorrelation plot of model P0 with autocorrelation present at lags 1,2,4,6,7,8,9,10 and 12 months. ***Dotted lines indicate cut-off for presences of significant autocorrelation. ACF – autocorrelation function (size of autocorrelation, lag indicates number of months between observations when correlation occurs).***

9.10 Appendix J - Distribution of length of stay including illness length

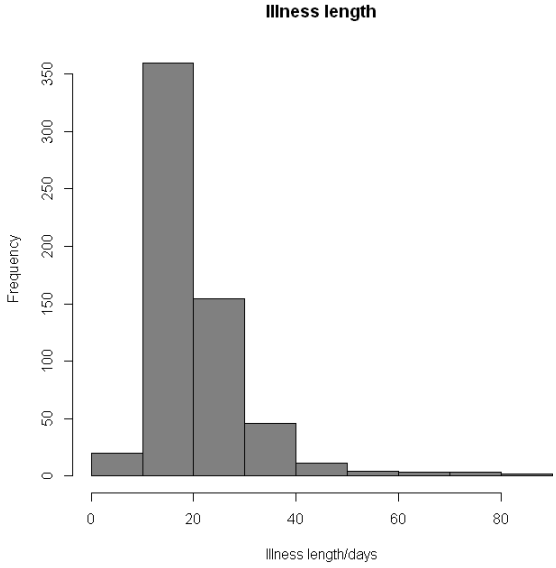


Figure 9.57 Frequency distribution of illness length in empyema cases illustrating non-normality (Shapiro-Wilk test for normality - W = 0.7653, p-value <0.001)

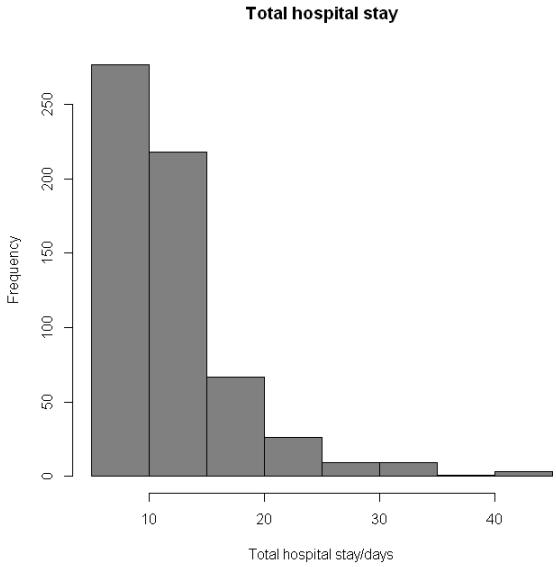


Figure 9.58 Frequency distribution of total hospital stay in empyema cases illustrating non-normality (Shapiro-Wilk test for normality - W = 0.8432, p-value <0.001)

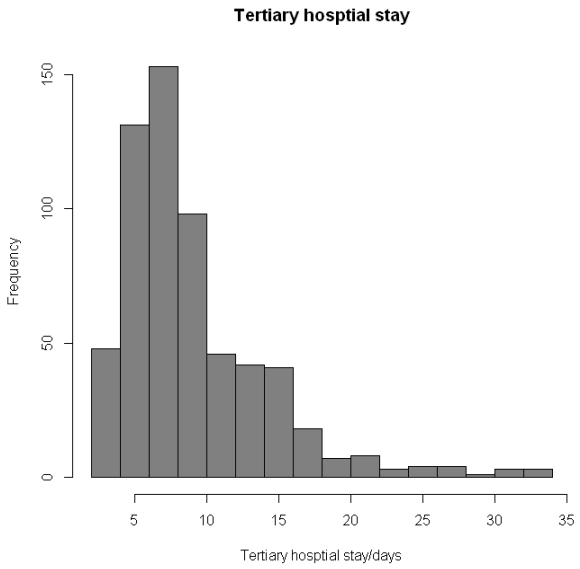


Figure 9.59 Frequency distribution of tertiary hospital stay in empyema cases illustrating non-normality (Shapiro-Wilk test for normality - W = 0.8375, p-value <0.001)

9.11 Appendix K - Schoenfeld residuals and assessment of proportionality of hazards organism and illness length model (Model Org)

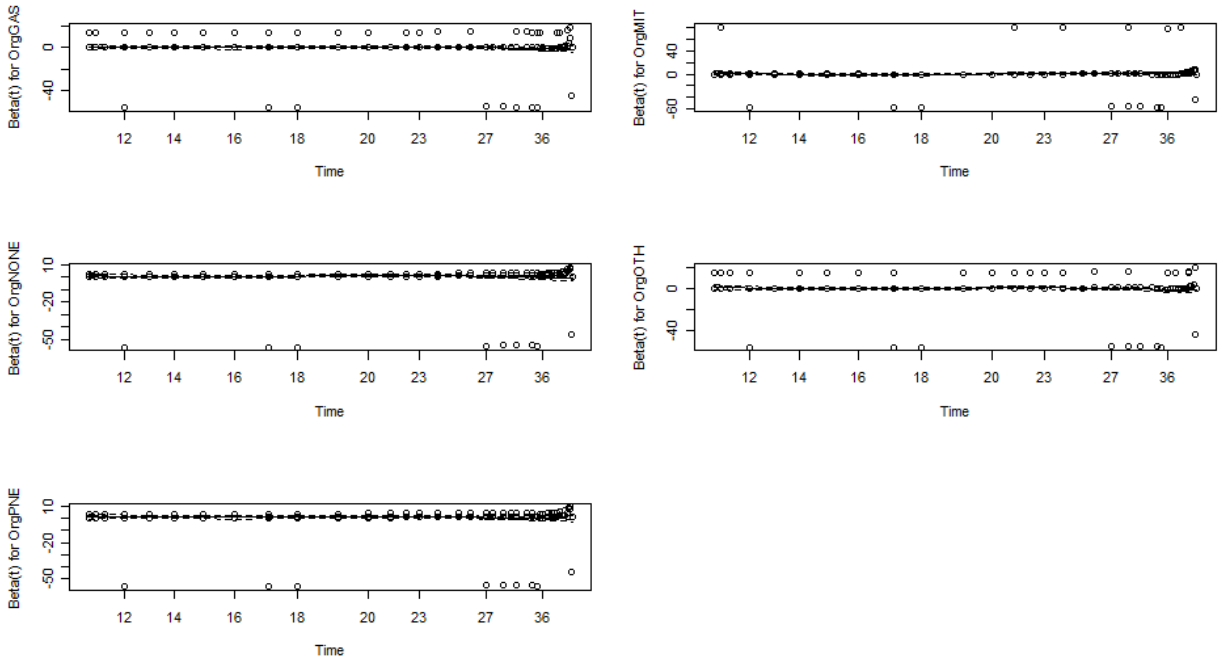


Figure 9.60 Plot of Schoenfeld residuals from analysis of relationship between organism and illness length.

Variable	Correlation	Chi-squared	p-value
S. aureus	-0.0187	0.193	0.660
S. pyogenes	0.0198	0.216	0.642
S. mitis	-0.0136	0.102	0.749
No organism detected	-0.0268	0.395	0.530
Other	-0.0416	0.946	0.331
Global	NA	6.621	0.250

Table 9.4 Assessment of proportionality of hazards over time for relationship between organism and illness length (correlation between smoothing function and scaled Schoenfeld residuals, chi-squared test for each covariate and overall chi-squared).

9.12 Appendix L - Schoenfeld residuals and assessment of proportionality of hazards pneumococcal serotype and illness length model (Model Sero)

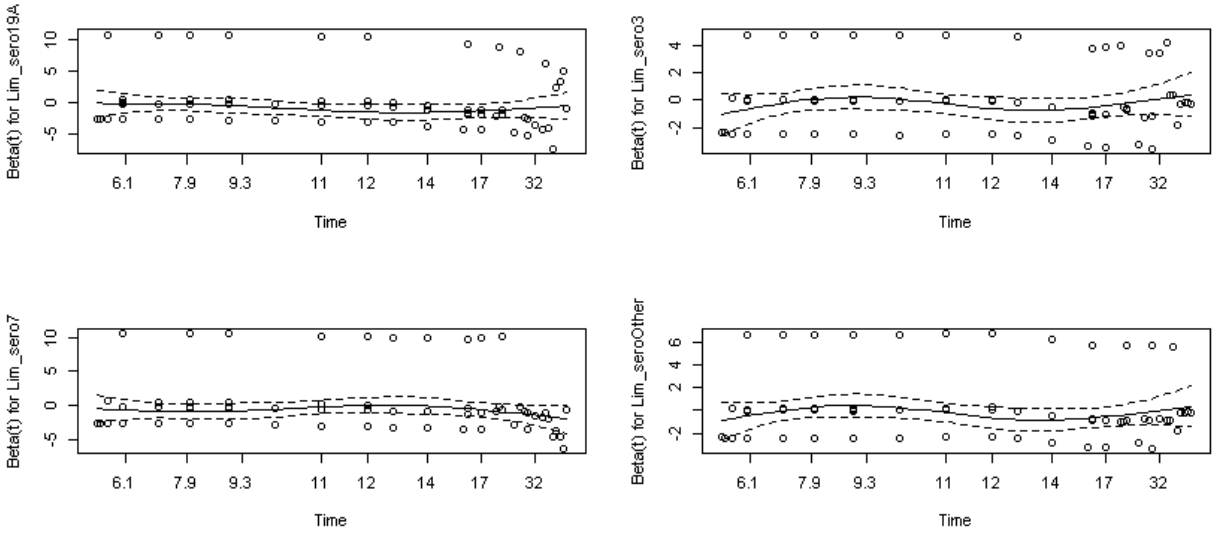


Figure 9.61 Plot of Schoenfeld residuals from serotype illness length model.

Variable	Correlation	Chi-squared	p-value
Serotype 19A	-0.1012	1.3237	0.2500
Serotype 3	0.0121	0.0194	0.8890
Serotype 7	-0.0182	0.0476	0.8270
Other serotypes	-0.0277	0.1020	0.7490
Global	NA	1.5558	0.8170

Table 9.5 Assessment of proportionality of hazards over time for illness length model (correlation between smoothing function and scaled Schoenfeld residuals, chi-squared test for each covariate and overall chi-squared).

9.13 Appendix M - Schoenfeld residuals and assessment of proportionality of hazards in treatment models (Models HSa/b and THSa/b)

Co-variate	Rho	Chi-squared	P-value
Model HSa			
PICU Admission	0.0487	1.6457	0.2000
Age (Years)	0.0033	0.0063	0.9370
Surgical treatment	-0.1282	29.5792	<0.0001
Global hazard	NA	31.3129	<0.0001
Model THSa			
PICU Admission	0.0359	0.8820	0.3480
Surgical treatment	-0.1568	39.0640	<0.0001
Global hazard	0.0359	0.8820	0.3480
Model HSb			
PICU Admission	0.0662	3.0103	0.0827
Age (Years)	0.0081	0.0387	0.8440
Drainage alone	-0.1045	13.5879	0.0002
Mini-thoracotomy	-0.1418	62.5865	<0.0001
VATS	-0.0368	0.9130	0.3390
Global hazard	NA	64.3161	<0.0001
Model THSb			
PICU Admission	0.0504	1.7100	0.1910
Drainage alone	-0.1064	12.5100	0.0004
Mini-thoracotomy	-0.1752	78.7900	<0.0001
VATS	-0.0398	1.1100	0.2930
Global hazard	NA	80.9000	<0.0001

Table 9.6 Results of regression of weighted residuals to test for proportionality of hazards over time for all four parsimonious models.

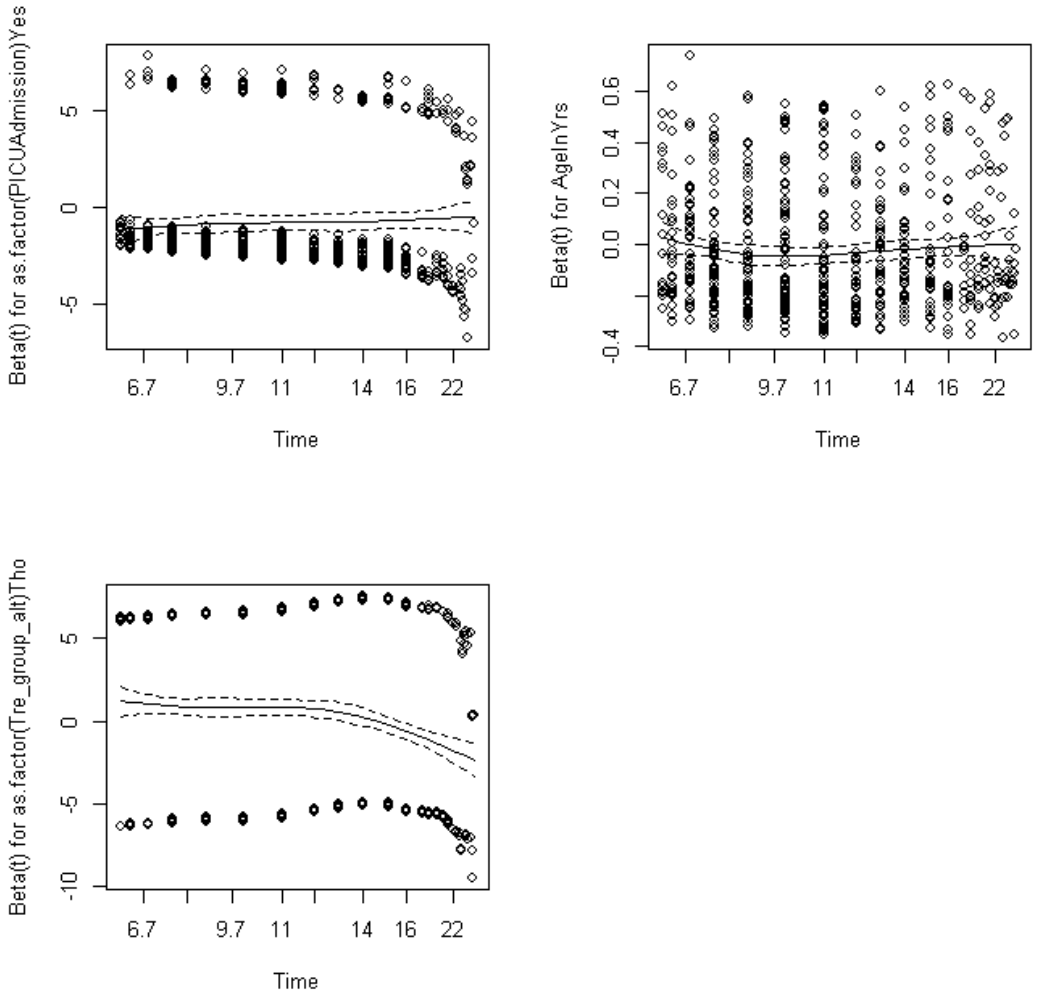


Figure 9.62 Plot of Schoenfeld residuals from serotype model HSa investigating the relationship between total hospital stay and surgical and non-surgical treatment methods.

Tre_group_alt)Tho refers to surgical treatment methods.

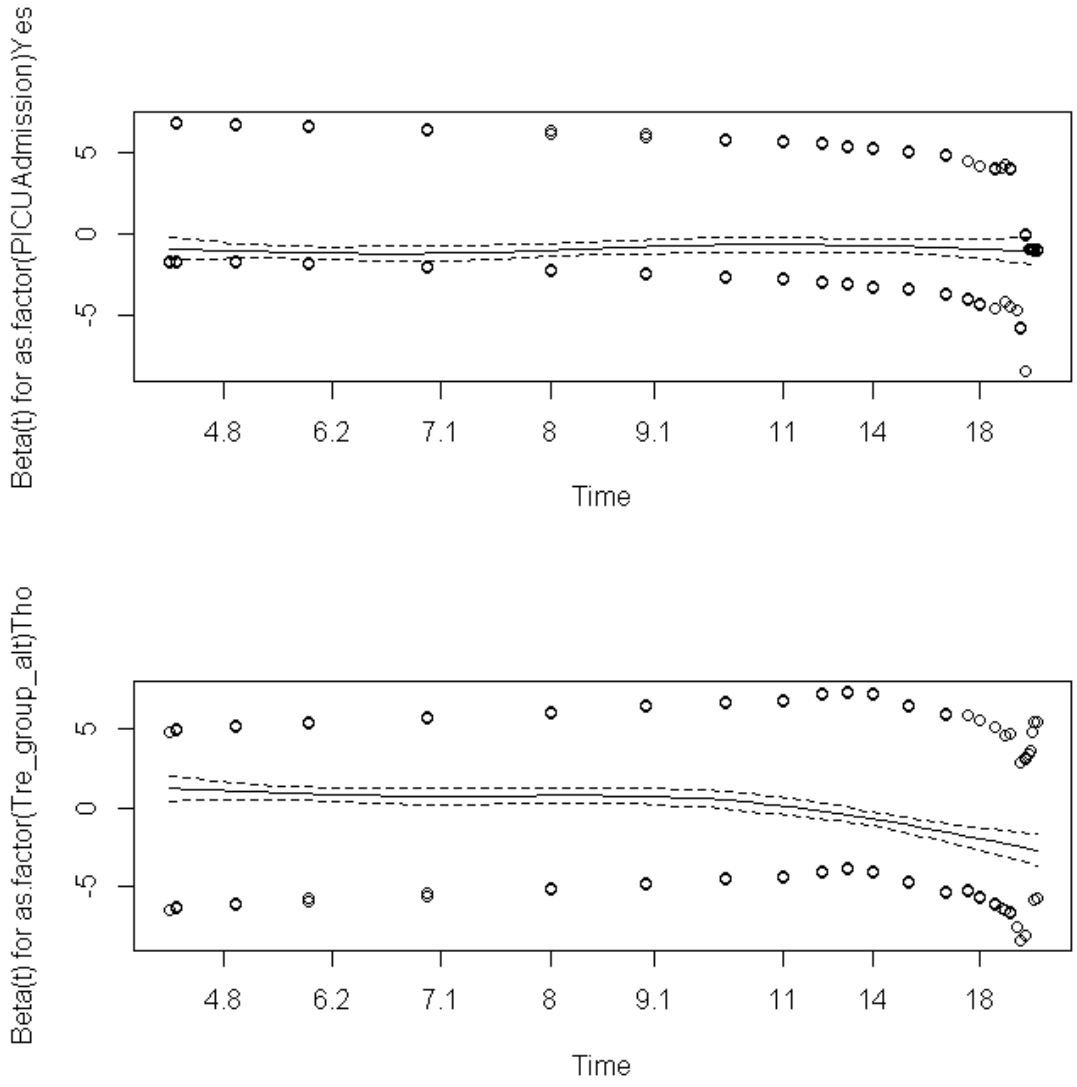


Figure 9.63 Plot of Schoenfeld residuals from serotype model THSa investigating the relationship between tertiary centre stay and surgical and non-surgical treatment methods. ***Tre_group_alt)Tho refers to surgical treatment methods.***

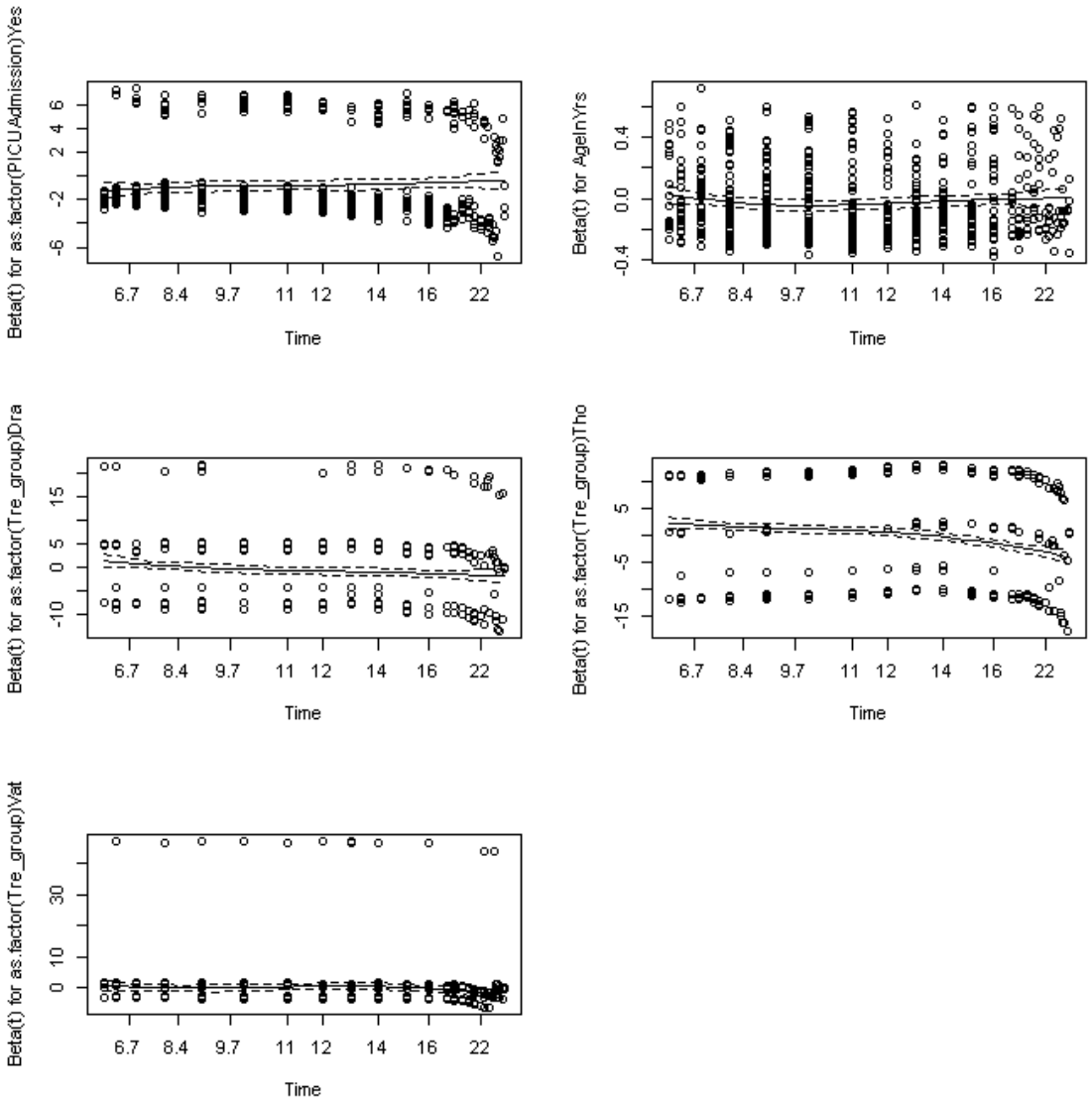


Figure 9.64 Plot of Schoenfeld residuals from serotype model HSb investigating the relationship between total hospital stay and the four treatment methods.

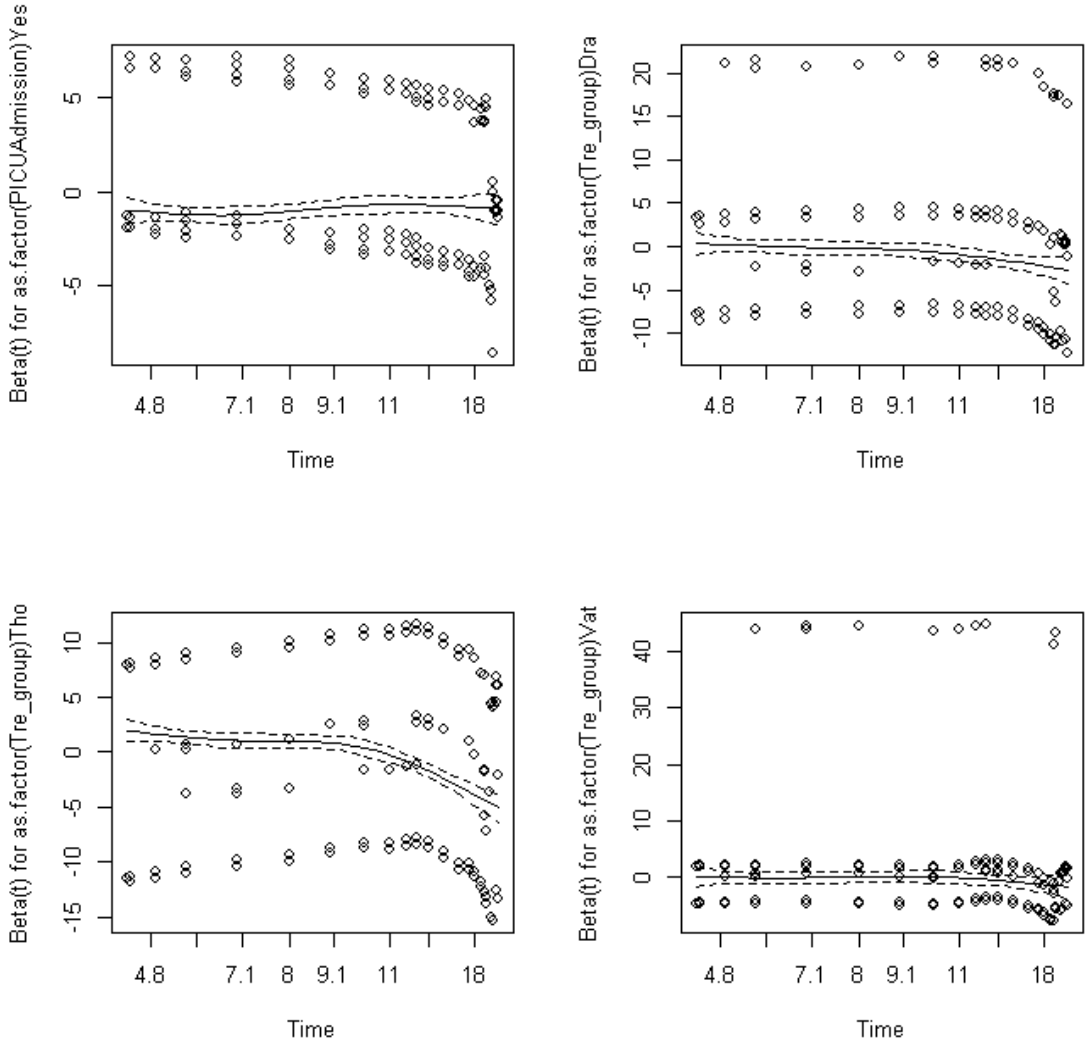


Figure 9.65 Plot of Schoenfeld residuals from serotype model HSb investigating the relationship between tertiary centre stay and the four treatment methods

9.14 Appendix N - List of free-text complications

Abscess + necrotic lung+bronchopleural fistula-VV ECMO BC19A
Air leak
Albumin 11, oedematous
Albumin 17
ARF secondary to pneumococcal HUS requiring PD then HD
broncho pleural fistula requring a second drain
Cellulitis at drain site
Chylothorax
Drain migrated
Drain migration
Drain recurrent infection 15/1, redo right decortication + wound debridement 20/1
Drain removed by patient
Glomerulonephritis
Group A strep sepsis-toxic shock, retropharyngeal abscess, celulitis R torso/neck, rsv positive
HUS-VA ECMO pre decortication. Died
Influenza A H1N1, strep A Toxic shock septicemia, Left pleural effusion, pneumonia and RSV positive
Left sided pleural effusion, Influenza A H1N1 low level positive pcr suggestive of recent infection
Loculated pneumothorax - had pleuroplasty as 2nd op & drain with Heimlich
Lung abscesses
Necrotising pneumonia-consolidation of left lower lobe, adenovirus low level positive
Pancytopenia, pneumothorax. Discharged to Rochdale for IVs
Pneumatocoele
Pneumothorax
Pneumothorax - drain reinserted 19/01/2010 Bronchopleural fistula
Pneumothorax (prior to drain)
Pneumothorax 1 mnth post decortication
Pneumothorax and surgical emphysema
Pneumothorax Chest drain reinserted, removed 2/1/11
Pneumothorax req 2nd chest drain
Pneumothorax x2, necrotic lung abscess, bronchopleural fistula, Heimlick valve needed
Pneumothorax, lung abscess
Presumed streptococcal vasculitis
Reaccumulation pleural effusion-chest drain with USS guidance. Died 26/12/2009
Readmitted - required further drainage
Readmitted with pneumothorax. Given 6x2 doses urokinase
Redo decortication
Required decortication
Residual pneumothorax
RSV pos,L pneumothorax and pleural effusion, septic shock, hypoglycemia and metabolic acidosis
Septic shock at presentation. Pneumothorax . extravasation injuries
Severe anaemia, requiring transfusion
Small pneumothorax
Small pneumothorax

Small pneumothorax -not drained
Small residual pneumothorax
Toxic shock syndrome. Chest drain re-inserted 31.03.11 removed 06.04.11 8 doses of Urokinase.
Toxic shock, facial tics
VATS was partial decortication, hypoalbuniaemia, lung abscess
Wound debridement & resuturing 1 wk post op 6 wks pneumothorax
x3 procedures (adult services) 3rd op. stapling necrotic lung. MRSA in wound

Table 9.7 All free text entries of complications included in analysis

9.15 Appendix O - Emergence of pneumococcal 19A empyema in UK children

Published ADC Online First, published on October 16, 2012 as 10.1136/archdischild-2012-301790

MF Thomas[1,2], CL Sheppard[3], M Guiver[4], MPE Slack[3], RC George[3], R Gorton[5], JY Paton[6], C Simmister[2], D Cliff[7], MA Elemraid[8,9], JE Clark[9], SP Rushton[1], DA Spencer[2]

[1] School of Biology, Newcastle University, Newcastle-Upon-Tyne, UK

[2] Department of Respiratory Paediatrics, Newcastle-Upon-Tyne Hospitals NHS Trust, Newcastle-Upon-Tyne, UK

[3] Respiratory and Systemic Infection Laboratory, Centre for Infections, Health Protection Agency, Colindale, London, UK

[4] Molecular Diagnostic Laboratory, Health Protection Agency North West, Manchester, UK

[5] Regional Epidemiology Unit, Health Protection Agency North East, Newcastle-Upon-Tyne, UK

[6] School of Medicine, University of Glasgow, Glasgow, UK

[7] Newcastle-Upon-Tyne Hospitals NHS Trust, Newcastle-Upon-Tyne, UK

[8] Institute of cellular medicine, Newcastle University, UK

[9] Department of Paediatric Infectious Disease and Immunology, Newcastle-Upon-Tyne Hospitals NHS Trust, Newcastle-Upon-Tyne, UK

Running Title: Pneumococcal 19A empyema in UK children

Key words: Epidemiology, pleural infection, pneumococcal vaccination, respiratory infection

Word Count: [Main text: 1180, Abstract: 150], Tables: 1, Figures: 1

Corresponding author:

Dr David Spencer

Department of Respiratory Paediatrics,

Great North Children's Hospital,

Newcastle-Upon-Tyne,

NE1 4LP, Tyne & Wear, UK

Tel: +44 (0) 191 2825089, Fax: +44 (0) 191 222 3082

E-mail: David.Spencer2@nuth.nhs.uk

ABSTRACT

Introduction – Invasive pneumococcal disease due to serotype 19A has become a major concern, particularly in the USA and Asia. We describe the characteristics of pneumococcal serotype 19A related empyema and changes in its incidence in the UK.

Methods –Data from paediatric empyema patients between September 2006 and March 2011 were collected from 17 respiratory centres in the UK. Pneumococcal serotypes were identified as part of the Health Protection Agency (HPA) enhanced paediatric empyema surveillance programme.

Results – Four serotypes accounted for over 80% of 136 cases (Serotype 1:43%, 3:21%, 7:11% and 19A:10%). The incidence of empyema due to serotype 19A quadrupled from 0.48 (0.16-1.13) cases per million children in 2006/7 to 2.02 (1.25-3.09) in 2010/11. Severity of disease was significantly increased in children with 19A infection when compared to other serotypes.

Conclusions – The incidence of empyema due to pneumococcal serotype 19A infection has increased significantly and is associated with substantial morbidity.

INTRODUCTION

Streptococcus pneumoniae is the leading cause of bacterial pneumonia in children worldwide, and is responsible for an estimated 700,000 to one million deaths annually in children under 5 years of age.(Reinert *et al.*, 2010) There are over 90 recognised serotypes of *S. pneumoniae,* but not all are commonly associated with human disease.(Reinert *et al.*, 2010) Conjugated pneumococcal vaccines currently provide protection against a limited number of serotypes with the seven valent conjugate vaccine (PCV-7) providing protection against seven serotypes (4, 6B, 9V, 14, 18C, 19F and 23F) and the 13-valent vaccine (PCV-13), which was recently introduced in the UK, active against an additional six serotypes (1, 3, 5, 6A, 7F and 19A).

The range of pneumococcal serotypes causing invasive disease is changing. In particular, the incidence of serotype 19A (Spn19A) has increased substantially in some countries, notably the USA, Spain and South Korea.(Reinert *et al.*, 2010) In the USA, Spn19A was present in only 2.5% of isolates from children under 5 years of age prior to the introduction of PCV-7 in 1998-1999 but was identified in 47.2% of isolates six years later and prior to the introduction of PCV-13 became the leading cause of invasive pneumococcal disease in children in that country.(Reinert *et al.*, 2010) In England and Wales, a recent report from the HPA has highlighted the emergence of Spn19A as a cause of invasive pneumococcal disease.(Miller *et al.*, 2011)

Historically, Spn19A has not been a recognised cause of paediatric empyema in the UK (Eastham *et al.*, 2004; McKee *et al.*, 2011) but it has been linked to the rise in incidence of empyema in the USA.(Reinert *et al.*, 2010) Complication rates in paediatric empyema have been shown to vary between pneumococcal serotypes but Spn19A has not been previously highlighted as a concern.(McKee *et al.*, 2011) In light of the increasing importance of Spn19A in pneumococcal disease we describe changes in the incidence of Spn19A infection in empyema in England and evaluate the resulting changes in disease severity.

METHODS

Pneumococcal serotype surveillance

From September 2006, culture negative pleural fluid samples from children (0-16 years) with empyema were forwarded to the HPA from microbiology laboratories across England as part of the programme of enhanced surveillance of pneumococcal empyema in UK children (UK-ESPE). Samples were tested with a pneumococcal PCR (pneumolysin and autolysin) and positive samples underwent serotyping. Serotyping involved a non-culture multiplex polysaccharide antigen detection assay that has been described previously.(Sheppard *et al.*, 2011) Serotyping of culture negative pleural fluid was used as the source of pneumococcal serotyping data as in the majority of cases (127 of the 136 (93%) UK-ESPE cases included) pleural fluid in children with pneumococcal empyema was sterile at the point of testing, presumably because of antibiotic use prior to referral.(Sheppard *et al.*, 2011)

Clinical characteristics

Detailed clinical data on children (0-16 years) with empyema requiring pleural drainage between September 2006 and March 2011 were collected from 17 collaborating UK-ESPE centres. A secure web-accessed pro-forma was completed on each patient by the clinical team at the relevant collaborating centre. These data were then matched to the national serotyping surveillance data. Additional patients from Scottish centres collaborating with the UK-ESPE study were added where serotyping results were available. Caldicott approval for the collection of clinical data was obtained.

Statistical analysis

Incidence calculations were based on total numbers of patients identified by the PCR based culture negative surveillance carried out by the HPA in England from 2006 onwards. Clinical characteristics were analysed using Kruskal-Wallis and Fisher's exact test for continuous and categorical variables respectively. Cox's proportional hazard models were used for length of stay analysis. All analyses were carried out using the R statistical package (v.11.2).

RESULTS

Cases of complicated pneumonia and empyema due to Spn19A in England increased from 5 in 2006/7 (the year following the introduction of PCV-7) to 21 cases in 2010/11 which was the year following the replacement of PCV-7 with PCV-13. The incidence increased from 0.48 cases per million children (95% confidence intervals (CI) 0.16-1.13) in 2006/7 to 2.02 (95% CI 1.25-3.09) in 2010/11, giving an incidence rate ratio of 4.17 (95% CI 1.53-14.2).

Of the 136 UK-ESPE cases (134 from England and 2 from Scotland) where a pneumococcal serotype was identified and clinical details were available, 14 (10%) were serotype 19A. Serotypes 1, 3 and 7 (43%, 20.6% and 11%) were the other common serotypes detected. Of Spn19A cases, 5 (36%) had a positive blood culture and all but one had culture negative pleural fluid (7%). One isolate was resistant to Penicillin, Tetracycline and Clindamycin but sensitive to Cephalosporins on standard testing. No evidence of antibiotic resistance was reported in any of the other isolates. Characteristics of children with Spn19A infection and those with other pneumococcal serotypes are shown in Table. Four of five children with Spn19A infection who developed complications had culture positive disease. Duration of hospital admission was increased by 50% in patients with 19A disease compared to all other serotypes (adjusting for age and sex - Hazard: 0.52, 95% CI 0.28-0.95, p=0.034). Survivorship curves are shown in the Figure.

DISCUSSION

The incidence of empyema due to Spn19A infection has increased dramatically in UK children. It is unclear exactly what has driven this change in incidence but antibiotic pressure, vaccine induced serotype replacement, existing secular trends and introduction of new clones are all potentially relevant factors.(Reinert *et al.*, 2010; Miller *et al.*, 2011) In the absence of molecular epidemiological and pre-vaccine data it is impossible to be certain which of these factors are relevant.

The most striking finding from our data was the severity of disease. Children with Spn19A related empyema stayed longer in hospital. They were significantly more likely to require intensive care and had higher complication rates when compared to children with empyema due to other pneumococcal serotypes. There were no differences in co-morbidity, pre-hospital antibiotic usage and surgical intervention rate suggesting that the increased severity was serotype related rather than due to patient or treatment factors. Furthermore, Picazo and colleagues found similar high rates of intensive care admission (63.6%) and complications (27.3%) in Spanish children with empyema due to Spn19A.(Picazo *et al.*, 2011) The pneumococcal capsule is an important virulence factor, and hence differences in disease severity between serotypes are well described.(Reinert *et al.*, 2010) However, this level of disease severity is in marked contrast to the traditional clinical picture in paediatric empyema and is of significant concern.(McKee *et al.*, 2011)

We found only limited evidence of antibiotic resistance within the Spn19A isolates, although molecular testing for resistance e.g. by use of pbp2b PCR assay was not available. These findings are in contrast to the experience in the USA and Asia where Spn19A isolates are frequently resistant to multiple antibiotics, possibly as a consequence of excess antibiotic usage in the community.(Reinert *et al.*, 2010)

Miller and colleagues have previously demonstrated that the epidemiology of invasive pneumococcal disease is changing both in children and in adults in the UK.(Miller *et al.*, 2011) Spn19A related empyema has increased in UK children and this change may have been driven by vaccine related serotype replacement disease following the introduction of PCV-7. PCV-13 introduced in 2010 in the UK contains antigen for 19A. This is likely to lead to a significant reduction in Spn19A disease but given the PCV-7 experience further changes in the clinical profile of pneumococcal disease are possible. Continued surveillance will be required to monitor for further evolutionary changes in this group of organisms.

Acknowledgements

The authors would also like acknowledge S Beaton, Glasgow University for his support of the web-based data collection, T Harrison, Health Protection Agency and all those at the UK-ESPE collaborating centres study: *C Wallis, V Rasiah, R Thursfield – Great Ormond Street Hospital, London; C Murray, J Mercy – Royal Manchester Children's Hospital, Manchester; C Pao – Royal London Hospital, London; R Ross-Russell – Addenbrooke's Hospital, Cambridge; J Bhatt, R Radcliffe, K Ward – Nottingham Children's Hospital; A Thomson – Children's Hospital, Oxford; I Balfour-Lynn – Royal Brompton Hospital, London; W Lenney – University Hospital of North Staffordshire, Stoke-on-Trent; J Clarke, I Brookes – Birmingham Children's Hospital, Birmingham; R Primhak – Sheffield Children's Hospital, Sheffield; E Carroll, J Clark – Alder Hey Children's Hospital, Liverpool; T Hilliard – Bristol Royal Hospital for Children, Bristol; A Cade – Derriford Hospital, Plymouth; A Prendiville, H Soloman, Royal Cornwall Hospital, Truro; S Cunningham, Royal Hospital for Sick Children, Edinburgh*

Disclosure

CLS has received support from Pfizer Inc. to attend overseas conferences. MPES is a member of advisory boards of GSK and Pfizer Inc, has received funding to her institution from GSK and Pfizer Inc. for investigation of multiple pneumococcal serotype carriage using DNA microarray, been an invited speaker at scientific meetings organized by GSK and Pfizer Inc., and has received support from both companies to attend conferences. RCG received funding to his institution from GSK and Pfizer Inc. for pneumococcal diagnostics and characterization studies. He also received support from GSK and Pfizer Inc. to attend overseas conferences. JEC has received pneumonia research funding from Pfizer Inc. The UK-ESPE study has been supported by unrestricted grants from GSK and Pfizer Inc. to DAS. GSK and Pfizer Inc. had no involvement in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

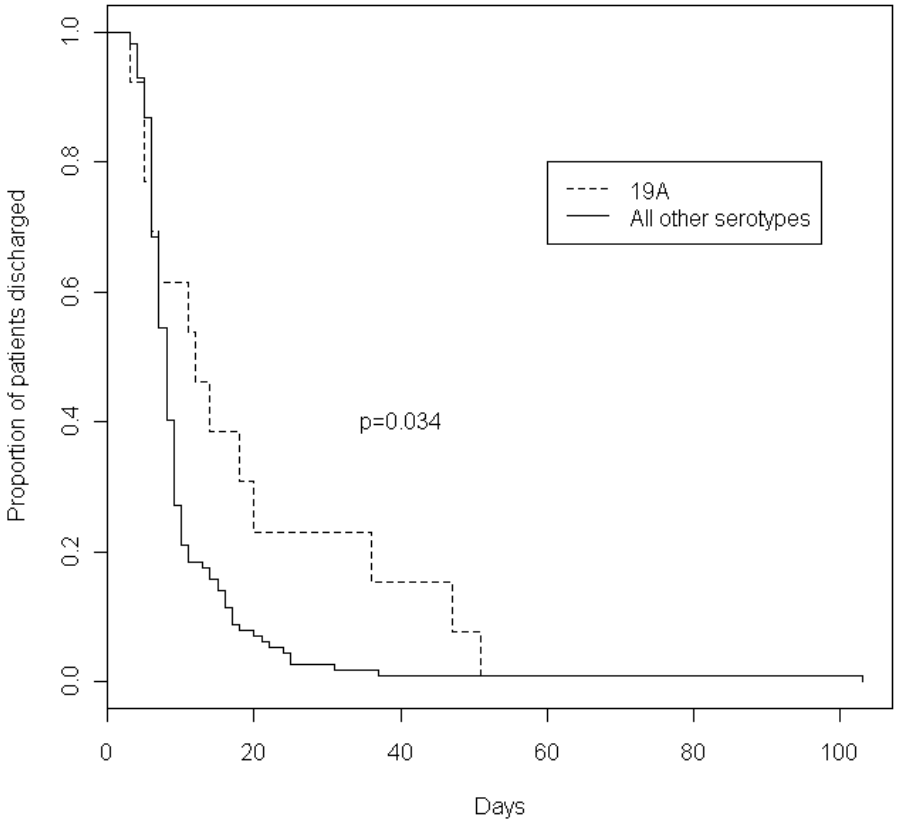


Figure: Survivorship curves of hospital stay of patients with empyema caused by pneumococcal serotype 19A vs. patients with empyema caused by other pneumococcal serotypes.

Characteristics	**Serotype 19A (n=14)**	**All other serotypes (n=122)**	**KW/Fisher's test, p value**
Median age of patients / years (Range)	2.0 (0.2-5.4)	4.3 (0.4-15.7)	0.002
Male % (n)	57% (8)	56% (68)	1
Median pre-hospital symptom duration / days (Range)*	7 (3-14)	6 (1-41)	0.703
Pre-hospital antibiotics % (n)*	25% (2)	39% (33)	0.706
Any co-morbidity % (n)	7% (1)	7% (8)	1
Surgery as primary drainage procedure % (n)	43% (6)	56% (68)	0.573
Any surgical procedure % (n)	57% (8)	64% (78)	0.402
Readmission % (n)*	0% (0)	5% (6)	1
Complications† % (n)	36% (5)	11% (13)	0.022
Intensive care admission % (n)*	36% (5)	12% (14)	0.019
Assisted ventilation % (n)*	36% (5)	9% (10)	0.007
Extra corporeal membrane oxygenation required	2	0	-
Deaths‡	1	0	-

†Complications (n): 19A – Pneumothorax (1); broncho-pleural fistula (1); lung abscess and lung necrosis (1); haemolytic-uraemic syndrome (2); Non-19A – Pneumothorax (7); broncho-pleural fistula (3); lung abscess and lung necrosis (1); cellulitis at drain site (1); streptococcal vasculitis (1)

*Indicates data unavailable for some patients; Pre-hospital symptom duration % non-respondents = 0%(19A), 4%(other serotypes); Pre-hospital antibiotics 50%,30%; Readmission 14%,10%; Intensive care admission 14%,7%; Assisted ventilation 14%,11%.

‡The child who died had no significant pre-existing co-morbidity

Table: Characteristics of patients with empyema caused by pneumococcal serotype 19A vs. patients with empyema caused by other pneumococcal serotypes.

References

1. Reinert R, Jacobs MR, Kaplan SL. Pneumococcal disease caused by serotype 19A: Review of the literature and implications for future vaccine development. Vaccine. 2010;28:4249-59 doi: 10.1016/j.vaccine.2010.04.020 [published Online First 21 April 2010.

2. Miller E, Andrews NJ, Waight PA, et al. Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: an observational cohort study. Lancet Infect Dis. 2011;11:760-8.

3. Eastham KM, Freeman R, Kearns AM, et al. Clinical features, aetiology and outcome of empyema in children in the north east of England. Thorax. 2004;59:522-5.

4. McKee AJ, Ives A, Balfour-Lynn IM. Increased incidence of bronchopulmonary fistulas complicating pediatric pneumonia. Pediatr Pulmonol. 2011;46:717-21.

5. Sheppard CL, Guiver M, Hartley J, et al. The use of a multiplexed immunoassay for detection of serotype-specific Streptococcus pneumoniae antigen in pleural fluid and CSF specimens. J Med Microbiol. 2011; doi: 10.1099/jmm.0.034975-0 [published Online First 11 August 2011.

6. Picazo J, Ruiz-Contreras J, Hernandez B, et al. Clonal and clinical profile of Streptococcus pneumoniae serotype 19A causing pediatric invasive infections: A 2-year (2007–2009) laboratory-based surveillance in Madrid. Vaccine. 2011;29:1770-6 doi: 10.1016/j.vaccine.2010.12.114 [published Online First 7 January 2011.

What is already known on this topic?

Pneumococcal 19A infection has become the commonest cause of invasive disease in children in the USA but has not been a significant problem in the UK

There are differences in the disease severity caused by infection with different pneumococcal serotypes

What this study adds?

The incidence of paediatric empyema due to pneumococcal serotype 19A has increased in the UK

Serotype 19A related empyema is associated with significantly higher levels of morbidity than empyema related to other pneumococcal serotypes

9.16 Appendix P - UK-ESPE study protocol

Summary

The incidence of paediatric empyema thoracis has risen dramatically over the last decade(Rees *et al.*, 1997)(Playfor *et al.*, 1997)(Thompson *et al.*, 1999; Byington *et al.*, 2002b; Hsieh *et al.*, 2004; Deiros Bronte *et al.*, 2006; Gupta and Crowley, 2006; Roxburgh and Youngson, 2007) and *Streptococcus pneumoniae* serotype 1 has been identified as the principal pathogen in the UK(Eltringham *et al.*, 2003; Eastham *et al.*, 2004; Fletcher *et al.*, 2006). Various suggestions have been proposed to explain the observed rise in incidence including changes in diagnostic criteria and referral practice, and alterations in pathogen virulence. At present there is insufficient evidence in favour of any single hypothesis.

The 7-valent pneumococcal conjugate vaccine was introduced into the national immunisation schedule September 2006. This vaccine does not offer protection against serotype 1 disease. It has been suggested that introduction of this vaccine *may* lead to a further increase in pneumococcal disease. This might occur as a result of progressive increase in serotype 1 disease as well as an increase in disease due to other non-vaccine pneumococcal serotypes. There is some evidence that this phenomenon is already beginning to occur in the United States.

This project is intended to monitor the changing incidence of pneumococcal empyema thoracis in children throughout the UK, to determine the prevalence of different pneumococcal serotypes over time and to explore epidemiological and genetic factors. It represents collaboration between clinicians, microbiologists, epidemiologists, genetic scientists and those with expertise in biological modelling in an effort to elucidate the mechanisms responsible for this phenomenon.

Research Questions

We propose to explore the following hypotheses:

1. That the incidence of pneumococcal empyema will continue to increase in British children. That the increase will occur as a result of a progressive increase in serotype 1 disease which commenced before the introduction of the conjugate vaccine, as well as increase in disease from other non-vaccine serotypes.

2. That antibiotic treatment in primary care influences disease progression.

3. That space-time patterning of empyema cases will be related to environmental exposure to air-born pollutants.

4. That spatial patterning of empyema cases will be related to socio-economic

factors.

5. That certain gene polymorphisms will be associated with an increased risk of development of pneumococcal empyema

In order to explore these hypotheses in detail it will be necessary to:

1. Investigate the existence of spatial and temporal variation in incidence and clustering of cases of empyema.

2. Identify and monitor potential risk factors predisposing to the development of empyema.

3. Closely monitor changes in the prevalence of different pneumococcal serotypes causing complicated pneumonia and empyema in children.

4. Monitor changes in the incidence and severity of empyema thoracis over time.

5. Relate these changes to the introduction of the 7 valent conjugate pneumococcal vaccine into the routine childhood immunisation programme in the UK, and to the possible introduction of future vaccines with extended coverage against other pneumococcal serotypes.

6. Relate these changes to the national guidance aimed at reducing the use of antibiotics for simple respiratory tract infections in primary care.

7. Investigate links between the incidence of empyema and socio-economic indicators and environmental predictors.

8. Develop data collection systems which will facilitate the performance of future multi-centre studies to compare current and future medical and surgical interventions in the management of complicated pneumonia and empyema in children.

9. Audit outcomes according to variations in clinical management across the UK.

10. Investigate gene polymorphisms that may influence susceptibility to the development of invasive pneumococcal disease (IPD) and empyema in children.

Epidemiology

Pneumonia kills more children than any other illness, accounting for over 2 million deaths in children under-five years of age worldwide every year(Rudan *et al.*, 2008a). Although death is rare in developed countries, the condition is associated with significant acute and chronic morbidity(Clark *et al.*, 2007b). Pleural disease is a relatively common complication of pneumonia, occuring in approximately 3-4% of children hospitalised with community acquired pneumonia in the UK(Clark *et al.*, 2007b). Fluid may accumulate within the potential pleural space and cause a spectrum of problems ranging from simple pleural effusion to complicated empyema loculated by the formation of fibrin bands.

The great majority of patients with empyema will not respond to antibiotics alone, and some form of surgical drainage is usually required. Various management options are employed including simple thoracocentesis, thoracocentesis with instillation of urokinase and either early or late thoracotomy with debridement ("decortication"). Thoracotomy may be performed traditionally using direct vision, or using video assisted thoracotomy (VATS). Opinion is divided regarding the management strategies, and there are wide regional and national variations in therapeutic approaches, which partially reflects differences in the availability of skills and resources(Balfour-Lynn *et al.*, 2005).

An increase in the incidence of paediatric empyema was first reported from the West Midlands, UK in 1997(Rees *et al.*, 1997). In this series there was a seven-fold increase in cases when compared to the previous three years. This observation has now been confirmed in many other centres in North America, Europe and Asia(Byington *et al.*, 2006; Spencer *et al.*, 2006b; Munoz-Almagro *et al.*, 2008; Obando *et al.*, 2008; Spencer and Cliff, 2008). We have recently reported a progressive increase in the number of cases since 1997 with those requiring surgical intervention increasing up to 60/ year(Spencer *et al.*, 2006b). We have estimated that there are currently 800-1,000 cases/year in the UK. The cause or causes for these changes are not known, however, various hypotheses have been proposed:

1. Changes in referral practice or management of cases
2. Changes in antibiotic prescribing practice in primary care
3. Effect of changes in UK vaccination policy
4. Antibiotic tolerance

5. Changes in the ecology of bacterial colonisation of the nasopharynx
6. Loss of herd immunity to *S. pneumoniae* serotype 1
7. Increase in inherent pneumococcal virulence

Most culture positive cases of childhood empyema thoracis in Europe and the USA are related to infection with *S. pneumoniae*(Eastham *et al.*, 2004; Byington *et al.*, 2006; Le Monnier *et al.*, 2006). However, these studies are hindered by the fact that routine bacterial culture of pleural fluid is usually negative, probably due to antibiotic use prior to hospital admission. Certain pneumococcal serotypes such as 1 and 5 are typically antibiotic sensitive whereas others such 6,9,14,19 and 23 are typically antibiotic resistant, so that undue reliance on bacterial culture results may give a false impression of the true pattern of disease in the community(Hausdorff *et al.*, 2005; Sanchez-Tatay *et al.*, 2008). Use of PCR and serotype specific ELISA in pleural fluid from children with culture negative disease has demonstrated that most UK cases are related to *S. pneumoniae,* with serotype 1 as the dominant serotype(Eltringham *et al.*, 2003). Serotype 1 has also been reported to be the dominant pathogen in both culture positive and culture negative disease in both Europe and the USA. (Table 1)

Table 1

Author	Proportion of cases of pneumococcal empyema attributable to serotype 1 (%)
Hardie et al	31
Byington et al	50
Buckingham et al	25
Eltringham et al	63
Tan et al	24.4
Eastham et al	68
Fletcher at al	67
Bekri et al	23
Obando et al	48

Serotype 1 is rarely isolated from the nasopharynyx of healthy individuals, raising questions regarding the mode of transmission and suggesting that isolation of this serotype is associated with a high attack rate(Sleeman *et al.*, 2006). We have recently reported a small school outbreak of pneumococcal serotype 1 disease in which transmission *may* have occurred via nasopharyngeal carriage from an adult(Gupta *et al.*, 2008).

It has been reported that the introduction of the conjugate pneumococcal vaccine into some communities in the USA has been associated with an *increase* in the incidence of empyema thoracis due to infection with serotypes that are not contained within the 7-valent vaccine, a phenomenon which has been referred to as ‘serotype replacement’ disease(Byington *et al.*, 2006).

The UK government commenced universal paediatric vaccination with the heptavalent pneumococcal conjugate vaccine (Prevenar© or Prevnar©) in September 2006. The serotypes contained within the vaccine are 4, 6B, 9V, 14, 18C, 19F and 23 F. The vaccine was first licensed for use in 2000 and was introduced into the universal immunisation programme in the USA in 2001. As the first country to introduce the vaccine, the USA provides the greatest breadth of experience of the impact of the introduction of the conjugate vaccine on serotype ecology and incidence.

The US Centres for Disease Control and Prevention (CDC) has reported surveillance data on the incidence of invasive pneumococcal disease (IPD) in 2005 compared with the pre-vaccine years. Overall IPD rates among children aged 5 and under decreased by 77% (incidence 98.7 cases per 100,000 during 1998-1999 compared to 23.4 per 100,000 in 2005) with an estimated 13,000 fewer cases('Invasive pneumococcal disease in children 5 years after conjugate vaccine introduction--eight states, 1998-2005,' 2008). The public health impact was described as "dramatic" with the additional benefits of herd immunity also being evident in the adult population('Invasive pneumococcal disease in children 5 years after conjugate vaccine introduction--eight states, 1998-2005,' 2008).However, several groups have now reported an increase in invasive pneumococcal disease (IPD) caused by non-vaccine serotypes, so called 'serotype replacement disease'. The most dramatic of these reports is from Singleton *et al* who reported an increase of 140% of non-vaccine serotypes causing invasive pneumococcal disease in native Alaskan children, which nearly ablates the net reduction in invasive pneumococcal disease related to vaccination(Singleton *et al.*, 2007). Other groups have shown more modest increases in IPD related to non-vaccine serotypes, with the balance following the introduction of Prevenar© still significantly towards an overall reduction in total IPD case numbers. Similar experiences of increase in IPD caused by non-vaccine serotypes and increase in incidence of empyema following the introduction of Prevenar© have also been reported from Spain[17].

There are significant differences in the reported serotype distribution between the USA and Europe(Hausdorff, 2002). This partially reflects the consequences of defining infection in terms of bacterial culture, and the fact that antibiotic sensitive serotypes are less likely to be cultured in patients who have received antibiotics prior to sampling(Hausdorff *et al.*, 2005). This further hampers our ability to extrapolate data on the impact of immunisation from America to Europe.

The Health Protection Agency (HPA) in the United Kingdom has been monitoring weekly trends in serotype incidence since the introduction of the conjugate pneumococcal vaccine in to the routine vaccination schedule. The latest data is shown below.

Figure 1 – Incidence of invasive pneumococcal disease caused by serotypes contained within Prevenar© conjugate vaccine 2003 – present in children under the age of 2 years(Agency, 2009b)

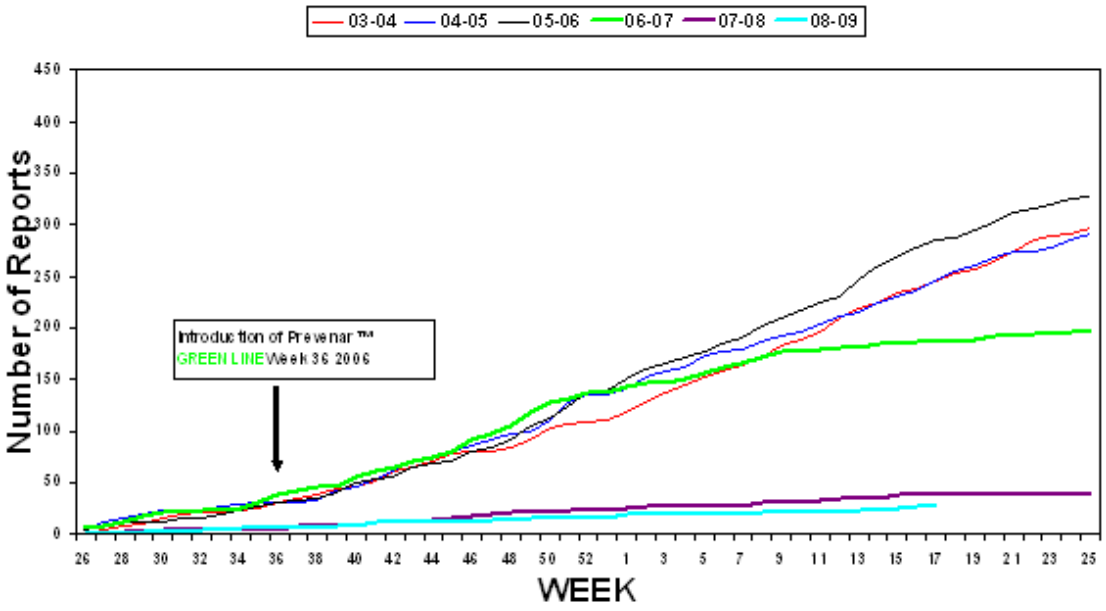
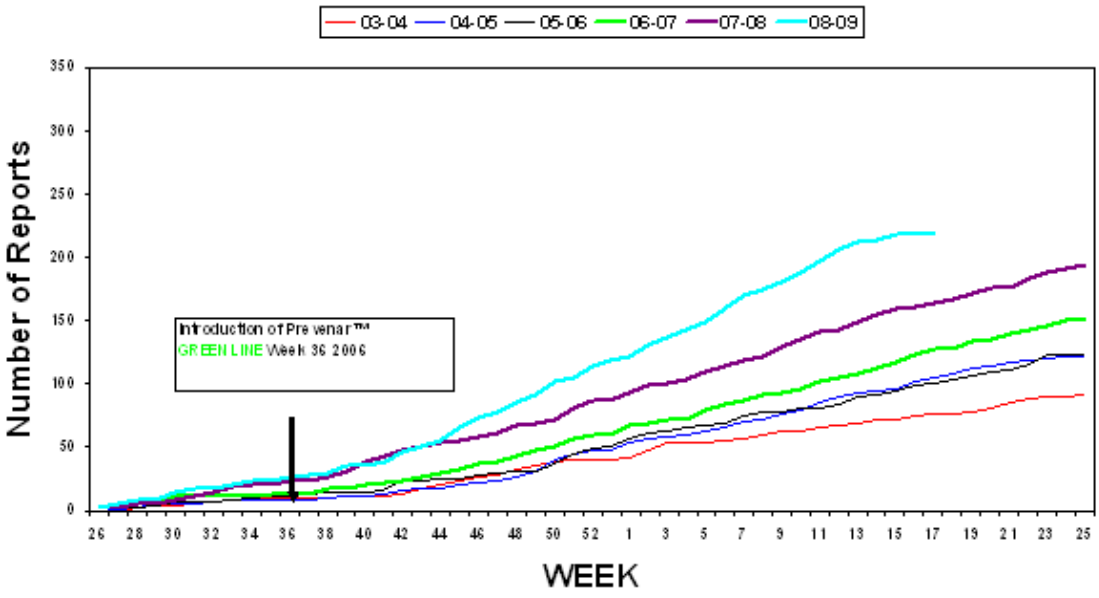


Figure 2 – Incidence of invasive pneumococcal disease caused by serotypes not contained within Prevenar© conjugate vaccine 2003 – present in children under the age of 2 years(Agency, 2009a).



These graphs demonstrate the benefits of vaccination in reducing invasive disease related to vaccine serotypes, but suggest that disease due to non-vaccine serotypes is emerging in the UK. This data has been produced from a passive screening programme of *culture positive* invasive pneumococcal disease and suffers from the intrinsic weaknesses of such a system. We estimate on the basis of extrapolating our comprehensive regional data that this represents in the region of only 5-10% of the total burden of invasive pneumococcal disease. This is because approximately 90% of disease is culture negative and that recruitment to the surveillance programme will vary considerably across regions and over time. It should also be appreciated that differences in the antibiotic sensitivity of individual pneumococcal serotypes will result in a bias towards reporting a disproportionately high percentage of disease attributable to antibiotic resistant serotypes when surveillance systems are based solely on bacterial culture. For these reasons the HPA agreed to introduce enhanced surveillance for paediatric pneumococcal empyema in England and Wales using culture negative techniques.

Further evidence of changing serotype ecology and prevalence related to immunisation is also demonstrated by studies of pneumococcal nasopharyngeal carriage. Groups in the Netherlands and the USA have demonstrated an increase in carriage of non-vaccine serotypes following mass immunisation, providing a possible mechanism by which serotype replacement disease may occur(Bogaert *et al.*, 2005; Hanage *et al.*, 2007).

Antibiotic prescribing

In recent years there has been considerable pressure exerted to reduce the use of antibiotics in primary care for presumed viral upper respiratory tract infection. Although the public health benefits of this policy are fully appreciated, there is some reasonable concern that a reduction in antibiotic prescribing in the community *could* lead to an increase in the incidence of some severe bacterial diseases presenting to hospital(Sharland *et al.*, 2005). The evidence base for these concerns is weak, largely because of the major epidemiological difficulties in proving such associations. The National Institute for Health and Clinical Excellence (NICE) has recently issued guidance aimed at further reducing inappropriate prescribing(NICE, 2008). It is likely that this will have a major impact on reducing the use of antibiotics prescribed for presumed viral upper respiratory tract infections in children.

The clinical features of pneumonia in children at first presentation are often rather non-specific. Not all children have respiratory symptoms and localising physical signs may be absent. The majority of children presenting with empyema have received antibiotics in primary care, which also casts some doubt as to the effectiveness of this intervention in preventing the progression to severe disease. It is conceivable that any further reductions in the prescription of antibiotics in primary care *could* lead to an increase in cases of complicated pneumonia in children. It would therefore be sensible and prudent to monitor long term changes in the incidence of paediatric empyema nationally, and to relate this data to antibiotic prescribing information available from the Prescription Prescribing Authority (PPA) and other evolving sources of national paediatric drug prescribing data.

Environmental variables

Invasive pneumococcal disease and childhood empyema both demonstrate marked seasonal variation in incidence, suggesting that environmental factors are of critical importance. Variability of climate, changes in seasonal prevalence of viral infections and exposure to airborne pollutants have all been implicated as possible explanations for these phenomena(Kim *et al.*, 1996; Dowell *et al.*, 2003; Murdoch and Jennings, 2009). The results of research in this field are difficult to interpret and contradictory results have frequently been reported.

Genetic Susceptibility to Invasive Pneumococcal Disease

Recent work has demonstrated that the genetic make-up of an individual plays an important role in determining their likelihood of developing IPD, including pneumococcal empyema. In addition to providing a comprehensive study of the clinical and environmental factors that underlie the development of empyema in childhood, the current research aims to identify gene variants that may determine why some individuals are particularly susceptible to IPD. Our previous research in this area has been successful in identifying a small number of susceptibility genes for IPD, including genes encoding mannose-binding lectin, C-reactive protein, PTPN22, Mal/TIRAP and NFKBIA(Roy *et al.*, 2002a; Roy *et al.*, 2002b; Chapman *et al.*, 2006; Khor *et al.*, 2007). Many more genes are likely to be involved, however, and their identification is dependent upon the use of larger sample collections of DNA from individuals with this disease. In particular, large DNA sample collections allow the use of 'genome-wide' approaches which study markers across all human genes; such an approach has recently proven extremely successfully in the identification of novel genetic polymorphisms responsible for common conditions such as asthma and diabetes, but has not previously been used in the study of pneumococcal disease.

Methods

In order to institute long term surveillance and research in to this problem it has been necessary to introduce a robust and efficient system of data collection which is able to record both laboratory and clinical data on individual patients in a manner which is compatible with the Data Protection Act and fulfils the Caldicott guidelines on recording patient identifiable data. The British Paediatric Respiratory Society (BPRS) initially established a web-based system for recording such information administered by Dr James Paton, Reader in Respiratory Paediatrics at the University of Glasgow. This system is secure and uses high-level encryption of data, which has been further enhanced by incorporation of the "Soundex" system for securing patient identifiable information. The system was initially used to record data for the national annual paediatric acute asthma audit, and has now been successfully modified to record data on paediatric empyema.

The BPRS research committee has formally endorsed this study and the HPA is collaborating actively. A clinical lead responsible for coordination of specimen and data collection has been established at each site. It is intended that data entry is performed during the course of the hospital stay or shortly thereafter. Clinical data obtained via the web based remote data collection system includes relevant past medical history, antibiotic history, treatment, length of stay, outcome and patient demographics. The HPA also contacts the general practitioner of positive cases that are within the age groups eligible to receive one or more doses of conjugate vaccine.

Most UK respiratory paediatricians do not routinely differentiate between post-pneumonic effusion and empyema, and these are now generally regarded as representing parts of spectrum of infection-related pleural disease. As such, all paediatric patients within the UK in whom pleural fluid is being removed as part of their routine management will be eligible for inclusion in the study. In practice this will be limited to patients receiving some form of surgical management (either simple thoracocentesis, thoracocentesis with installation of urokinase or decortication). It is not normal practice within the UK for patients being managed solely with antibiotics to have needle aspiration performed, and so this study is intentionally limited to patients with more severe disease. However, the great majority of patients with true empyema will not respond to antibiotics alone.

Saliva samples are also being collected from each patient for DNA analysis. Two different but complimentary genetic study approaches will be taken using the DNA samples:

1) Investigation of 'candidate' genes

2) A 'genome-wide' screen, which studies a very large number of genetic markers spread across the entire genome

DNA analysis is being performed in the laboratory of Professor Adrian Hill based at the Wellcome Trust Centre for Human Genetics, University of Oxford.

Study of candidate genes involves the specific examination of a relatively small number of genes which on the basis of existing information are strongly suspected to play a role in disease causation. A genome-wide screen, on the other hand, studies a very large number of common genetic markers spread across the entire genome in an attempt to capture information about the role of the majority of human genes in susceptibility to disease. For both approaches, polymorphism frequencies in patients with IPD will be compared with those of healthy individuals from the United Kingdom. These have previously been determined by the Sanger Institute in Cambridge and are freely available to investigators on-line. Such an approach has now been extensively used for genome-wide studies and is well-validated. Genetic variants that appear to affect susceptibility to IPD will also be examined within the patient group for a possible role in clinical outcome.

The collaborating team has complete control over the clinical, microbiological and genetics data with Dr David Spencer, the Chief Investigator being responsible for the overall conduct of the study.

The duration of the study is expected to be 10 years. Recruitment commenced in August 2006.

Samples

A specimen of pleural fluid from every patient recruited is sent to Drs Ray Borrow and Malcolm Guiver at the HPA laboratory in Manchester. An initial screening PCR for pneumolysin is performed and screen positives are then dispatched to Dr Robert George at the RSIL laboratory, Colindale, London, who performs a confirmatory PCR and tests positive samples for serotype-specific antigen using the Bioplex method as originally developed by Borrow and Leeming. This assay currently detects 13 serotypes, including serotype 1. If positive for one of these 13 this confirms pneumococcal infection and identifies the serotype. If negative then confirmation of pneumococcal infection is sought with an autolysin PCR.

A saliva sample is also collected from each patient in order to provide a DNA specimen.

A flow chart of progress through the study is attached in Appendix A.

Data validation and ascertainment

In order to effectively calculate rates of recruitment and ascertainment of the number of cases of empyema in children in individual hospitals, it is necessary to use hospital coding data. The coding departments of the hospitals who have recruited patients for the study will be approached to supply hospital coding data on the number of patients between the ages of 0-16 years with the diagnosis of empyema (International Classification of Disease - 10 codes – J86.0 and J86.9) and their demographics to allow cross-checking with study data.

Ethical approval

An application has been approved by the national research and ethics committee (REC). Informed consent is obtained for obtaining saliva samples from individual patients and patient information leaflets/consent forms are provided along with the sampling kit. Section 60 support from the Patient Information and Advisory Group has been granted to retrospectively use patient identifiable data without having to gain informed consent. Informed consent is obtained for prospective use of patient identifiable data. Caldicott approval has been granted at centres receiving patient identifiable information and is being sought at collaborating centres as required. All necessary measures have been taken to ensure compatibility with the Data Protection Act.

Statistical analysis

The data will allow us to establish changes in the incidence of empyema thoracis over time, and be used to test conceptual models of the epidemiology of empyema thoracis. The clinical data-base will be integrated with a Geographical Information System (GIS) which will allow linkage between residential address of cases with environmental and socio-economic covariates that are hypothesised to play a role in disease (temperature, rainfall, socio-economic status, exposure to air pollution and specific indices of deprivation)(Grant *et al.*, 2003).

Space-time clustering approaches will be used to investigate the extent to which disease arising from different serotypes is clustered in space and time. In addition, the geographical pattern in the spread of serotypes nationally and in relation to case age is being quantified with a view to testing hypotheses that there is spatial variation in serotype distribution amongst empyema thoracis cases. Point process models will be used to investigate how the pattern of cases of each serotype depends on other covariates including climatic variables, socio-economic status and antibiotic use.

Generalised Linear Modelling and Generalised Linear Mixed Effect Modelling is being used in combination with the GIS data to investigate the role of weather, climate, air pollution exposure and other covariates in determining cases of disease at the level of the regional centre.

Descriptive statistics will be used to analyse and interpret results from the clinical and microbiological data. We are also documenting differences in the length of admission between different UK centres and the influences of different treatment regimes. Survival analysis will be used to investigate the impacts of

socio-economic status, serotype, antibiotic treatment and other co-variates on the course of disease from initial diagnosis to discharge home.

The influence of antibiotic prescribing in primary care will be assessed using a proxy measurement of generic antibiotic prescribing amounts at SHA or PCT level, based on information available from the PPA. This will then be used in a time-series analysis against empyema cases to look at significant trends.

For the genetic component of the study, differences in the genotype distributions between empyema cases and healthy controls will be compared using Chi-squared testing. Logistic regression analysis will be performed to correct for the possible confounding effects of other genes and other recorded clinical variables on disease susceptibility. Correction will be made for the multiple independent comparisons tested as part of a genome-wide study. The majority of markers tested will not be associated with disease, but the large number of markers studied increases the risk of false positive associations. A *P* value of <0.000001 is generally recognised as an appropriate threshold for significance in the context of genome-wide analysis.

The total number of individuals with pneumococcal empyema who will be recruited to the genetics components of these studies is difficult to predict, as this will be determined largely by the success rate in obtaining informed consent. 2000 specimens will be required in order to obtain statistically meaningful results, which we estimate may take up to four years to obtain. Control genome-wide genotype results are already available from 2000 healthy individuals from the United Kingdom, determined by the Sanger Institute in Cambridge and freely available to investigators on-line. Sample size calculations for genome-wide genetic association studies are complex, and require consideration of variable allele frequencies, number of genetic markers tested, and extent of linkage disequilibrium between markers, reflecting their independence). Stringent *P* values are required for declaration of true association in order to correct for the very large number of independent comparisons performed on the same dataset. A sample size of approximately 1000 cases and the same or greater number of controls is generally recommended for genome-wide association studies; such a sample size should allow the detection of odds ratios of 1.5 for allele frequencies of 10% or more, with 80% power to detect a *P* value of <0.000001(Wang *et al.*, 2005). Some genome-wide association studies have been successfully performed on significantly fewer samples, however, although this limits the power of the study to detect relatively small genetic effects. Additional DNA sample collections from individuals with invasive pneumococcal disease are already available in

Professor Adrian Hill's laboratory in Oxford, and these will be studied alongside the described collection as additional replication groups for any positive genetic associations.

Conclusions

The incidence of paediatric thoracic empyema has increased dramatically over the last decade, but the reasons for these changes remain unclear.

It is very likely that the serotypes of important disease causing pneumococcal serotypes will continue to change as a consequence of both natural variation and possibly selection pressure exerted by the introduction of new conjugate pneumococcal vaccines. It is increasingly important to maintain long term and continuous surveillance of changes in the prevalence of disease related to different pneumococcal serotypes in order to plan future vaccine design and national vaccination strategies. The national UK programme of enhanced paediatric empyema surveillance will be an essential component of this surveillance (Whitney *et al.*, 2003).

9.17 Appendix Q – Ethical approvals

PIAG approval letter

Dr David Cliff

Sir James Spence Institute of Child Health

Royal Victoria Infirmary

Queen Victoria Road

Newcastle upon Tyne

NE1 4LP

12 December 2007

Dear Dr Cliff,

RE: Application for Section 60 support

PIAG 4-05(j)/2007: Enhanced surveillance of paediatric pneumococcal empyema in the UK

Thank you for applying for support under Section 60 of the Health and Social Care Act 2001 to process patient identifiable information without consent. This application for Section 60 support was considered by the Advisory Group at its meeting on 4th December 2007.

The Advisory Group was unable to approve all of your application at the meeting but some aspects of the application were approved conditionally. The main grounds for seeking support under Section 60 is where 1) anonymisation in not possible 2) obtaining consent is not practical.

Your application requests access to prospective and retrospective data and we will address these issues separately. PIAG Members are content that there is no reasonable alternative for the collection of retrospective data and in this instance, have approved your application for the collection of retrospective data.

However, in relation to prospective data, PIAG Members feel that a reasonable alternative to Section 60 exists, namely consenting patients, and therefore it would be inappropriate to allow Section 60 support for this purpose.

In summary:

- PIAG Members agree that Section 60 approval is warranted for the collection of retrospective data
- PIAG Members do not approve Section 60 support for the collection of prospective data.

Approval for the collection of retrospective data is subject to the following condition:

- Confirmation of appropriate security arrangements (Please note these are still to be checked by our security advisor, I will be back in contact with comments shortly)

In addition to this, please confirm that you meet the standard conditions of approval (attached overleaf)

As soon as I receive a satisfactory response, I will arrange for the Register of approved applications to be updated on our website www.advisorybodies.doh.gov.uk/piag to reflect this.

Please note, this approval is subject to our annual review process and you will be required to submit a report in 12 months time. I will send out a reminder and guidance document nearer the time.

Yours Sincerely

John Sheehan

Policy Support Manager

PIAG Secretariat

APPLICANT'S CHECKLIST

All studies except clinical trials of investigational medicinal products

REC Ref:	07/H0904/84
Short Title of Study:	Enhanced surveillance of paediatric pneumococcal empyema in the UK
CI Name:	Dr David/DA Spencer
Sponsor:	Newcastle upon Tyne Hospitals NHS foundation trust

Please complete this checklist and send it with your application

- Send ONE copy of each document (except where stated)
- ALL accompanying documents must bear version numbers and dates (except where stated)
- When collating please do NOT staple documents as they will need to be photocopied.

Document	Enclosed?	Date	Version	Office use
Covering letter on headed paper	○ Yes ◉ No			
NHS REC Application Form, Parts A&B	Mandatory	01/10/2007	5.4	
Site–Specific Information Form (for SSA)	○ Yes ◉ No			
Research protocol or project proposal (6 copies)	Mandatory	01/10/2007	1.0	
Summary C.V. for Chief Investigator (CI)	Mandatory	01/10/2007		
Summary C.V. for supervisor (student research)	◉ Yes ○ No	01/10/2007		
Research participant information sheet (PIS)	◉ Yes ○ No	01/10/2007	1.0	
Research participant consent form	◉ Yes ○ No	01/10/2007	1.0	
Letters of invitation to participants	○ Yes ◉ No			
GP/Consultant information sheets or letters	○ Yes ◉ No			
Statement of indemnity arrangements	○ Yes ◉ No			
Letter from sponsor	◉ Yes ○ No	01/10/2007		
Letter from statistician	○ Yes ◉ No			
Letter from funder	◉ Yes ○ No	01/10/2007		
Referees' or other scientific critique report	○ Yes ◉ No			
Summary, synopsis or diagram (flowchart) of protocol in non–technical language	○ Yes ◉ No			
Interview schedules or topic guides for participants	○ Yes ◉ No			
Validated questionnaire	○ Yes ◉ No			
Non–validated questionnaire	○ Yes ◉ No			
Copies of advertisement material for research participants, e.g. posters, newspaper adverts, website. For video or audio cassettes, please also provide the printed script.	○ Yes ◉ No			

WELCOME TO THE NHS RESEARCH ETHICS COMMITTEE APPLICATION FORM

An application form specific to your project will be created from the answers you give to the following questions.

1. Is your project an audit or service evaluation?

○ Yes ◉ No

2. Select one research category from the list below:

○ Clinical trials of investigational medicinal products
○ Clinical investigations or other studies of medical devices
◉ Other clinical trial or clinical investigation
○ Research administering questionnaires/interviews for quantitative analysis, or using mixed quantitative/qualitative methodology
○ Research involving qualitative methods only
○ Research limited to working with human tissue samples and/or data
○ Research tissue bank

If your work does not fit any of these categories, select the option below:

○ Other research

2a . Please answer the following questions:

a) Does the study involve the use of any ionising radiation?	○ Yes	◉ No
b) Will you be taking new human tissue samples?	◉ Yes	○ No
c) Will you be using existing human tissue samples?	○ Yes	◉ No

3. Is your research confined to one site?

○ Yes ◉ No

4. Does your research involve work with prisoners?

○ Yes ◉ No

5. Do you plan to include in this research adults unable to consent for themselves through physical or mental incapacity?

○ Yes ◉ No

6. Is the study, or any part of the study, being undertaken as an educational project?

◉ Yes ○ No

6a. Is the project being undertaken in part fulfilment of a PhD or other doctorate?

(●) Yes ( ) No

NHS Research Ethics Committee NHS
Application form

This form should be completed by the Chief Investigator, after reading the guidance notes. See glossary for clarification of different terms in the application form.

Short title and version number: (maximum 70 characters – this will be inserted as header on all forms)
Enhanced surveillance of paediatric pneumococcal empyema in the UK

Name of NHS Research Ethics Committee to which application for ethical review is being made:
Sunderland resesarch and ethics committee

Project reference number from above REC: 07/H0904/84
Submission date: 01/10/2007

PART A: Introduction

A1. Title of the research

Full title: The national programme for enhanced pneumococcal surveillance of complicated pneumococcal pneumonia and empyema in UK Children.

Key words: empyema, streptococcus pneumonia, incidence, epidemiology, immunisation

A2. Chief Investigator

Title:	Dr
Forename/Initials:	David/DA
Surname:	Spencer
Post:	Consultant in Respiratory Paediatrics
Qualifications:	MB BS(Hons) MRCP MD FRCPCH
Organisation:	Newcastle Upon Tyne Hospitals NHS Foundation Trust
Work Address:	Freeman Hospital
	Freeman Road, High Heaton
	Newcastle Upon Tyne
Post Code:	NE7 7DN
E–mail:	david.spencer2@nuth.nhs.uk
Telephone:	0191 244 8292
Fax:	0191 223 1099
Mobile:	07967379085

A copy of a current CV (maximum 2 pages of A4) for the Chief Investigator must be submitted with the application

A3. Proposed study dates and duration

Start date: 01/08/2007
End date: 01/08/2012
Duration: Years: 10 ; Months: 00

A4. Primary purpose of the research: *(Tick as appropriate)*

- ☐ Commercial product development and/or licensing
- ☐ Publicly funded trial or scientific investigation
- ☑ Educational qualification
- ☑ Establishing a database/data storage facility
- ☐ Other

Question(s) 5 disabled.

A6. Does this research require site−specific assessment (SSA)? (Advice can be found in the guidance notes on this topic.)

○ Yes ◉ No

If No, please justify:

The only samples we require that are considered over and above those of the routine management of the patient are saliva samples. Taking these is a low risk procedure within the professional competance of local collaborators.

If Yes, an application for SSA should be made for each research site on the Site−Specific Information Form and submitted to the relevant local Research Ethics Committee. Do not apply for SSA at sites other than the lead site until the main application has been booked for review and validated by the main Research Ethics Committee.

Management approval to proceed with the research will be required from the R&D office for each NHS care organisation in which research procedures are undertaken. This applies whether or not the research is exempt from SSA. R&D applications in England, Wales and Scotland should be made using the Site−Specific Information Form.

PART A: Section 1

A7. What is the principal research question/objective? *(Must be in language comprehensible to a lay person.)*

Empyema is a serious complication of pneumonia in which pus accumulates between the lung and the chest wall. The incidence of this problem has increased dramatically in children in the UK over the last decade and appears to be due to changes in the virulence of the pneumococcal bacteria of which there are over 90 'serotypes' (ie subgroups). In order to investigate the causes for this problem in more detail it is necessary to introduce a robust and efficient system of data collection on children with empyema which is able to record both laboratory and clinical data on individual patients. This will be compatible with the Data Protection Act and fulfil the Caldicott guidelines on recording patient identifiable data. The laboratory and clinical data will provide important epidemiological information and form the basis of a population based surveillance system that can be used when deciding on the introduction of future vaccines against this infection.

A8. What are the secondary research questions/objectives? *(If applicable, must be in language comprehensible to a lay person.)*

This data we collect will allow us to establish changes in the relative incidence of different pneumococcal serotypes over time, document the difference in the length of admission between different UK centres and the effect of different treatment regimes. It will also help us to identify potential risk factors for the condition.

The pneumococal bacteria itself has a number of different features that could explain the increase in the incidence of empyema including tolerance to certain antibiotics and 'virulence' factors. These make it more likely that there is either a poor response to treatment or an increase in severity of disease respectively. Both of these will be investigated as a potential cause of the problem.

The prevalence of empyema thoracis varies seasonally and changes in the climate could be a contributory factor to the rising number of cases. Investigating correlations between local changes in climate and incidence will form part of the research.

Recent studies have suggested that there may be a genetic predisposition that leads to to the development of complicated pneumococcal disease in adults and we will be collecting DNA samples from research participants to investigate this in children.

A9. What is the scientific justification for the research? What is the background? Why is this an area of importance?*(Must be in language comprehensible to a lay person.)*

Over the last decade there has been a significant increase in the incidence of pneumonia complicated by the formation of pus between the chest wall and the lung (an "Empyema") in UK children. It is now known that this phenomenon is largely due to an increase in the incidence of infection with one serotype (subgroup)of the common bacterium Streptococcus pneumoniae. The reason/s for this change are not currently understood. Children with this condition are often severely ill, and usually require surgical treatment. A vaccine designed to protect against certain strains of streptococcus ("prevenar") has recently been introduced into the routine UK vaccination schedule, but this does not cover this particular serotype. Indeed, it has been suggested that introduction of this vaccine in the USA might possibly be associated with an increase in the incidence of empyema thoracis due to strains not covered by the vaccine. It would therefore be prudent to monitor any changes in the incidence of empyema in UK children following introduction of the pneumococcal vaccine and to determine which serotypes are causing this problem each year.

Another major area of uncertainty concerns the factors that determine susceptibility to pneumococcal empyema: although exposure to the pneumococcus is widespread in children, only a minority go on to develop this severe infection. Recent work has demonstrated that an individual's genetic make–up plays an important role in determining their likelihood of developing respiratory infection. Previous research at the University of Oxford has been successful in identifying a small number of susceptibility genes for severe pneumococcal infection and thoracic empyema in adults, but many more genes are likely to be involved and their identification is dependent upon the use of larger sample collections of DNA from individuals with this disease. In particular, large DNA sample collections may allow the use of 'genome–wide' approaches which study markers across all human genes – such an approach has recently proven extremely successfully in the identification of genes responsible for common conditions such as asthma and diabetes, but has not previously been used in the study of pneumococcal disease or thoracic empyema.

A10–1. Give a full summary of the purpose, design and methodology of the planned research, including a brief explanation of the theoretical framework that informs it. It should be clear exactly what will happen to the research participant, how many times and in what order.

This section must be completed in language comprehensible to the lay person. It must also be self–standing as it will be replicated in any applications for site–specific assessment on the Site–Specific Information Form. Do not simply reproduce or refer to the protocol. Further guidance is available in the guidance notes.

The incidence of empyema thoracis complicating community acquired pneumonia has increased dramatically in children in the UK over the last decade. Investigation of the cause/s for this phenomenon have previously been complicated by the fact that the majority of cases have received antibiotics before referral to the regional respiratory centre making most cases culture negative and no causative organism identified. Using advanced microbiological techniques it has now been demonstrated that most cases are secondary to infection with Streptococcus pneumonia, and that most of these are of serotype (ie subgroup) 1, whereas until recently most cases of invasive pneumococcal infection in children were of serotype 14. The reasons for this apparent serotype switch are currently unknown.

A new pneumococcal vaccine covering seven of the most common serotypes was introduced into routine vaccination schedules in the USA in 2000. This vaccine does not protect against infection with serotype 1. Introduction of this vaccine has been associated with an increase in paediatric empyema cases in at least one region of the United States; this appears to be due to a progressive increase in serotype 1 disease along with the emergence of disease due to serotypes not covered by the vaccine. The possibility of disease developing from non–vaccine serotypes is a well–recognised potential problem and has been termed "replacement disease". This new vaccine ("prevenar") was introduced into the routine vaccination schedule for UK infants in September 2006. Given that the UK already has a problem with a rapidly increasing rate of empyema it is recognised that there is a possibility that the introduction of this new vaccine may be associated with further increases in the incidence of both serotype 1 and other non–vaccine pneumococcal serotype disease. In order to monitor this problem the British Paediatric Respiratory Society (BPRS) and the Health Protection Agency (HPA) have commenced on a collaborative programme of enhanced pneumococcal surveillance for children with empyema thoracis for the whole of the United Kingdom.

It is recognised that the environmental, climate and social environment including temperature, rainfall, pollution levels and socio–economic factors may influence the incidence of pneumococcal disease. In order to study the influence of these factors on the incidence of empyema in detail it will be necessary to document the patients' postcode, which will allow us to map each case against known data for that particular geographical location.

Each UK paediatric respiratory centre managing this type of patient has agreed to collaborate and contribute to this study, and a local collaborator has been established at each site. Clinical data, obtained routinely during the hospital stay will be entered into the web–based database using the case notes once the patient has been discharged. Data recorded on this system includes demographics, relevant past medical history, antibiotic history, treatment, length of stay and outcome as well as postcode.

We also intend to use historical case note data from patients admitted since commencement of the national immunisation programme September 2006. We will use these data to compare and contrast with the current population which have been offered vaccination.

The patient's ethnic background will also be recorded – this is an essential requirement for genetic association studies (described below), as genetic polymorphism frequencies vary naturally between different ethnic groups irrespective of disease status.

Pleural fluid is obtained for bacterial studies as part of routine patient management. Since September 2006 UK centres have been sending this pleural fluid to the HPA in Manchester. An initial screening test (called PCR) is performed for streptococcus pneumonia and screen positives are sent to the HPA laboratory at the Centre for Infection, Colindale London, who perform confirmatory tests and identify the specific serotype. The HPA also contact the general practitioner of every positive case to obtain data on vaccination history and antibiotic therapy in the month prior to admission. Positive cultures (Blood or pleural fluid) will also be dispatched to Newcastle and Glasgow for research work on factors responsible for causing the virulence of this infection and how its sensitivity to antibiotics may be changing. Data analysis from the above will be performed in collaboration with statisticians and epidemiologists working within the Sir James Spence Institute of Child Health, University of Newcastle upon Tyne.

We additionally propose to collect DNA samples from patients in order to try and identify gene variants (polymorphisms) that play a role in the development of empyema in children. We will ask patients to provide a single small sample of saliva, either by spitting into a small jar (adults and older children) or by briefly placing a soft small sponge inside their cheek (younger children). Saliva samples will be sent to the Wellcome Trust Centre For Human Genetics at the University of Oxford for storage, and DNA will be extracted at 6−monthly intervals throughout the study period. DNA samples and subsequent genetic data will be anonymised and assigned a trial number, and it will not be possible for researchers to trace genetic polymorphism data back to individual patients.

Two different but complimentary genetic study approaches may be taken using the anonymised DNA samples: investigation of 'candidate' genes and a 'genome−wide' screen. Study of candidate genes involves the specific examination of a relatively small number of genes which on the basis of existing information are strongly suspected to play a role in disease causation. Examples would include those genes that have been previously been found to play a role in susceptibility to empyema in adults. A genome−wide screen, on the other hand, studies a very large number of genetic markers spread across the entire genome in an attempt to capture information about the role of the majority of human genes in susceptibility to disease. For both approaches, polymorphism frequencies will be compared with those of healthy individuals from the United Kingdom, which have previously been determined by the Sanger Institute in Cambridge and which are freely available to investigators on−line. Such an approach has now been extensively used for genome−wide studies and is well−validated.

A10−2. In which parts of the research have patients, members of the public or service users been involved?

- ☐ As user−researchers
- ☐ As members of a research project group
- ☐ As advisor to a project
- ☐ As members of a departmental or other wider research strategy group
- ☑ None of the above

Please provide brief details if applicable:

A10−3. Could the research lead to the development of a new product/process or the generation of intellectual property?

○ Yes ◉ No ○ Not sure

A11. Will any intervention or procedure, which would normally be considered a part of routine care, be withheld from the research participants?

○ Yes ◉ No

A12. Give details of any clinical intervention(s) or procedure(s) to be received by research participants over and above those which would normally be considered a part of routine clinical care.*(These include uses of medicinal products or devices, other medical treatments or assessments, mental health interventions, imaging investigations and taking samples of human biological material.)*

Additional Intervention	**Average number per participant**		**Average time taken** (mins/hours/days)	**Details of additional intervention or procedure, who will undertake it, and what training they have received.**
	Routine Care	Research		
Other tissue/bodily sample	0	1	10 seconds	We will ask patients to provide a small sample of saliva, either by spitting into a small jar (older children) or by briefly placing a soft small sponge inside their cheek (younger children). This is a simple,

				low risk procedure which will be performed by the admitting clinician/local collaborator. This will take place when the patient is admitted to hospital and the diagnosis made in a private area where the consultation takes place.

A13. Give details of any non–clinical research–related intervention(s) or procedure(s). *(These include interviews, non–clinical observations and use of questionnaires.)*

Additional Intervention	Average number per participant	Average time taken (mins/hours/days)	Details of additional intervention or procedure, who will undertake it, and what training they have received.

A14. Will individual or group interviews/questionnaires discuss any topics or issues that might be sensitive, embarrassing or upsetting, or is it possible that criminal or other disclosures requiring action could take place during the study (e.g. during interviews/group discussions, or use of screening tests for drugs)?

○ Yes ◉ No

The Information Sheet should make it clear under what circumstances action may be taken

A15. What is the expected total duration of participation in the study for each participant?

The total duration will be the time taken for the clinical and demographic details to be taken and for the saliva sample to be provided.

A16. What are the potential adverse effects, risks or hazards for research participants either from giving or withholding medications, devices, ionising radiation, or from other interventions (including non–clinical)?

Taking the saliva (DNA) sample is a simple, low risk procedure not normally associated with any complications. Apart from this, the study will not require any additional samples above or beyond those obtained as part of the routine clinical care of a child with empyema thoracis.

A17. What is the potential for pain, discomfort, distress, inconvenience or changes to lifestyle for research participants?

Taking the saliva (DNA) sample is a simple, low risk procedure not associated with any discomfort or complications. Apart from this the study will not involve any additional discomfort/procedure above or beyond those that are part of the routine clinical care of the patient. Issues to do with confidentiality of genetic information are discussed in the answer to question 68; all samples will be anonymised and it will not be possible to trace genetic data back to individual study participants.

A18. What is the potential for benefit to research participants?

There will be no immediate benefit for those taking part in the study.

A19. What is the potential for adverse effects, risks or hazards, pain, discomfort, distress, or inconvenience to the researchers themselves? *(if any)*

A clinical lead/local collaborator established at each site has agreed to assist in collecting the sample of

saliva (DNA) and input the clinical data into the database. Inputting data into the database takes approximately 15 minutes per patient.

A20. How will potential participants in the study be (i) identified, (ii) approached and (iii) recruited?
Give details for cases and controls separately if appropriate:

All paediatric patients within the UK in whom a clinical diagnosis of empyema thoracis is made should have samples of fluid from around the lung taken as part of their routine management. All these children will be eligible for inclusion in the study. In practice this will be limited to patients receiving some form of surgical management as part of their care as described previously. These patients will be identified by clinicians at each participating site and once the diagnosis is made the clinical details will be recorded. The details will then be entered into the database.

Control data for the genetic analyses are already established in databases created from healthy volunteers in the United Kingdom (maintained by the Sanger Institute, Cambridge).

A21. Where research participants will be recruited via advertisement, give specific details.

☑ Not Applicable

If applicable, enclose a copy of the advertisement/radio script/website/video for television (with a version number and date).

A22. What are the principal inclusion criteria? *(Please justify)*

All paediatric patients with a clinical diagnosis of empyema within the UK in whom samples of fluid from around the lung are taken as part of their routine management will be eligible for inclusion in this study. In practice this will be limited to patients who are receiving some form of surgical management as part of their care.

A23. What are the principal exclusion criteria? *(Please justify)*

As this study involves establishing a database and analysing microbiololgical samples in conjunction with clinical data, no child presenting with empyema needs to be excluded.

A24. Will the participants be from any of the following groups? *(Tick as appropriate)*

- ☑ Children under 16
- ☐ Adults with learning disabilities
- ☐ Adults who are unconscious or very severely ill
- ☐ Adults who have a terminal illness
- ☐ Adults in emergency situations
- ☐ Adults with mental illness (particularly if detained under Mental Health Legislation)
- ☐ Adults with dementia
- ☐ Prisoners
- ☐ Young Offenders
- ☐ Adults in Scotland who are unable to consent for themselves
- ☐ Healthy Volunteers
- ☐ Those who could be considered to have a particularly dependent relationship with the investigator, e.g. those in care homes, medical students
- ☐ Other vulnerable groups

Justify their inclusion.

The study as previously described is specifically designed to investigate the increase in incidence of empyema thoracis in children, particularly in light of the introduction of the vaccine against Streptococcus pneumoniae.

☐ No participants from any of the above groups

Question(s) 24 1–5 disabled.

A25. Will any research participants be recruited who are involved in existing research or have recently been involved in any research prior to recruitment?

○ Yes ○ No ◉ Not Known

If Yes, give details and justify their inclusion. If Not Known, what steps will you take to find out?

The clinicians at each participating site will ask parents/carers if their child is involved in any existing research as part of their clinical evaluation. This should not mean that patients will be excluded from particating in this study.

A26. Will informed consent be obtained from the research participants?

◉ Yes ○ No

If Yes, give details of who will take consent and how it will be done. Give details of any particular steps to provide information (in addition to a written information sheet) e.g. videos, interactive material.

If participants are to be recruited from any of the potentially vulnerable groups listed in A24, give details of extra steps taken to assure their protection. Describe any arrangements to be made for obtaining consent from a legal representative.

If consent is not to be obtained, please explain why not.

Consent for the sample of saliva (DNA) will be obtained by the local collaborator or his/her designated assistant upon admission to hospital. An information sheet will be provided. If the child is deemed competent by the admitting clinician then they will be offered the opportunity to consent or refuse to be part of the study as per national guidelines. If they are not deemed competent then the parents/guardian alone will be asked to consent on their behalf. Regarding consent to use patient identifiable data, please refer to A31.

Copies of the written information and all other explanatory material should accompany this application.

A27. Will a signed record of consent be obtained?

◉ Yes ○ No

If Yes, attach a copy of the information sheet to be used, with a version number and date.

A28. How long will the participant have to decide whether to take part in the research?

The decision should be made after the patient has been assessed and the written information regarding the study provided. There will be the opportunity for any questions and more details if required. As the research is non interventional and purely data collection we expect that for the majority of the time decisions regarding consent will be immediate.

A29. What arrangements have been made for participants who might not adequately understand verbal explanations or written information given in English, or who have special communication needs? *(e.g. translation, use of interpreters etc.)*

There should not be any additional communication skills or tools required in terms of providing information to patients or parents above those that are needed as part of routine clinical management. Information leaflets will be translated in to commonly required foreign languages, and the hospital interpreting services will be used as deamed appropriate by the local clinician.

A30. What arrangements are in place to ensure participants receive any information that becomes available during the course of the research that may be relevant to their continued participation?

Not applicable

A30–1. What steps would you take if a participant, who has given informed consent, loses capacity to consent during the study? *Tick one option only.*

- ○ The participant would be withdrawn from the study. Data or tissue which is not identifiable to the research team may be retained. Any identifiable data or tissue would be anonymised or disposed of.
- ○ The participant would be withdrawn from the study. Identifiable data or tissue already collected with consent would be retained and used in the study.
- ○ The participant would continue to be included in the study.
- ○ Not applicable – informed consent will not be sought from any participants in this research.

Further details:

A31. Does this study have or require approval of the Patient Information Advisory Group (PIAG) or other bodies with a similar remit? *(see the guidance notes)*

◉ Yes ○ No

If Yes, give details

We are submitting an application to the Patient Information and Advisory Group for section 60 support to use patient identifiable data. The patient postcode will be used for purposes of analysis and DOB/surname for linking clinical and microbiological data. Reasons for using the post code in statistical analysis are described in A53. PIAG approval for the use of patient identifiable data is being sought for reasons described in A68

A32a. Will the research participants' General Practitioner (and/or any other health professional responsible for their care) be informed that they are taking part in the study?

○ Yes ◉ No

If Yes, enclose a copy of the information sheet/letter for the GP/health professional with a version number and date.

A32b. Will permission be sought from the research participants to inform their GP or other health professional before this is done?

○ Yes ◉ No

If No to either question, explain why not

The HPA has already established up an enhanced surveillance programme that involves contacting the general practitioner of every positive case to obtain data on vaccination history and antibiotic therapy in the month prior to admission. GPs will therefore be aware of the diagnosis of the subject and also that additional surveillance is being undertaken. There is therefore no need to inform the general practioners of the specifics of this research.

It should be made clear in the patient information sheet if the research participant's GP/health professional will be informed.

A33. Will individual research participants receive any payments for taking part in this research?

○ Yes ◉ No

A34. Will individual research participants receive *reimbursement of expenses* or any other *incentives or benefits* for taking part in this research?

○ Yes ◉ No

A35. Insurance/indemnity to meet potential legal liabilities

Note: References in this question to NHS indemnity schemes include equivalent schemes provided by Health and Personal Social Services (HPSS) in Northern Ireland.

A35–1. What arrangements will be made for insurance and/or indemnity to meet the potential legal liability of the *sponsor(s)* for harm to participants arising from the *management of the research*?

Note: Where a NHS organisation has agreed to act as the sponsor, indemnity is provided through NHS schemes. Indicate if this applies (there is no need to provide documentary evidence). For all other sponsors, describe the arrangements and provide evidence.

◉ NHS indemnity scheme will apply
○ Other insurance or indemnity arrangements will apply (give details below)

Please enclose a copy of relevant documents.

A35–2. What arrangements will be made for insurance and/or indemnity to meet the potential legal liability of the *sponsor(s) or employer(s)* for harm to participants arising from the *design of the research*?

Note: Where researchers with substantive NHS employment contracts have designed the research, indemnity is provided through NHS schemes. Indicate if this applies (there is no need to provide documentary evidence). For other protocol authors (e.g. company employees, university members), describe the arrangements and provide evidence.

◉ NHS indemnity scheme will apply to all protocol authors
○ Other insurance or indemnity arrangements will apply (give details below)

Please enclose a copy of relevant documents.

A35–3. What arrangements will be made for insurance and/or indemnity to meet the potential legal liability of *investigators/collaborators* and, where applicable, *Site Management Organisations*, arising from harm to participants in the *conduct of the research*?

Note: Where the participants are NHS patients, indemnity is provided through NHS schemes or through professional indemnity. Indicate if this applies to the whole of the study (there is no need to provide documentary evidence). Where non–NHS sites are to be included in the research, including private practices, describe the arrangements which will be made at these sites and provide evidence.

◉ All participants will be recruited at NHS sites and NHS indemnity scheme or professional indemnity will apply
○ Research includes non–NHS sites (give details of insurance/indemnity arrangements for these sites below)

Please enclose a copy of relevant documents.

A36. Has the sponsor(s) made arrangements for payment of compensation in the event of harm to the research participants where no legal liability arises?

○ Yes ◉ No

If Yes, give details of the compensation policy:

Please enclose a copy of relevant documents.

A37. How is it intended the results of the study will be reported and disseminated? *(Tick as appropriate)*

- ☑ Peer reviewed scientific journals
- ☐ Internal report
- ☑ Conference presentation
- ☐ Other publication
- ☑ Submission to regulatory authorities
- ☐ Access to raw data and right to publish freely by all investigators in study or by Independent Steering Committee on behalf of all investigators
- ☐ Written feedback to research participants
- ☐ Presentation to participants or relevant community groups
- ☑ Other/none e.g. Cochrane Review, University Library

If other/none of the above, give details and justify:

The research will form part of an MD and the resulting thesis will be stored at Newcastle University library

A38. How will the results of research be made available to research participants and communities from which they are drawn?

They will not be specifically informed.

A39. Will the research involve any of the following activities at any stage (including identification of potential research participants)? *(Tick as appropriate)*

- ☑ Examination of medical records by those outside the NHS, or within the NHS by those who would not normally have access
- ☑ Electronic transfer by magnetic or optical media, e–mail or computer networks
- ☑ Sharing of data with other organisations
- ☐ Export of data outside the European Union
- ☑ Use of personal addresses, postcodes, faxes, e–mails or telephone numbers
- ☐ Publication of direct quotations from respondents
- ☐ Publication of data that might allow identification of individuals
- ☐ Use of audio/visual recording devices
- ☐ Storage of personal data on any of the following:
  - ☐ Manual files including X–rays
  - ☐ NHS computers

- [x] Home or other personal computers
- [x] University computers
- [ ] Private company computers
- [ ] Laptop computers

Further details:

In order to institute long term surveillance and research in to this problem of empyema thoracis it will be necessary to introduce a robust and efficient system of data collection which is able to record both laboratory and clinical data on individual patients in a manner which is compatible with the Data Protection Act and fulfils the Caldicott guidelines on recording patient identifiable data. The British Paediatric Respiratory Society (BPRS) has established a web−based system for recording such information administered by Dr Jimmy Paton, Reader in Respiratory Paediatrics at the University of Glasgow. This system is secure and uses high−level encryption of data, which has been further enhanced by incorporation of the "Soundex" system for securing patient identifiable information. The system is already used to record data for the successful national annual paediatric acute asthma audit, and has now been modified to record data on paediatric empyema.
Data will have to be transferred electronically from the HPA laboratories, colindale London, to The University of Glasgow and Newcastle University for storage and interpretation (home computers will not be used to store this data). This process will require the use of patient identifiable data for purposes of linkage (DOB/Surname). All data transfer will be by password protected files and data entered into the database will be anonymised for subsequent analysis with the exception of the postcode for reasons described in A53.
We also intend to collect historical case note data. The clinical research fellow will be assisting the local collaborators in this and will obtain the appropriate authorisations (honorary contracts etc) to access patient case notes where necessary.
Anonymised genetic data resulting from the study will be stored on the Oxford University and personal computers of the chief investigator and research staff in secure locations. Participants will be identified only by trial number − no personal details such as name or date−of−birth will be stored, and it will not be possible to trace genetic polymorphism data back to individual participants.

A40. What measures have been put in place to ensure confidentiality of personal data? Give details of whether any encryption or other anonymisation procedures have been used and at what stage:

Clinical data will be stored on a secure web−based computer system maintained by The University of Glasgow that is compatible with the data protection act and fulfills the caldicott guidelines on use of patient identifiable data. The laboratory data will be added to this database after encryption using the soundex system to ensure patient anonymity. All data that is extracted from this database for subsequent analysis will be anonymised with the exception of the post code as this will be required for further statistical analysis (described in A53). Data will be stored at Newcastle University on password protected computers in secure swipe card protected buildings.
Anonymised genetic data resulting from the study will be stored on the Oxford University and personal computers of the chief investigator and research staff in secure locations. Participants will be identified only by trial number − no personal details such as name or date−of−birth will be stored, and it will not be possible to trace genetic polymorphism data back to individual participants.

A41. Where will the analysis of the data from the study take place and by whom will it be undertaken?

Data analysis will be performed in collaboration with statisticians and epidemiologists working within the Sir James Spence Institute of the Department of Child Health, University of Newcastle upon Tyne, and with genetic epidemiologists based at the Wellcome Trust Centre for Human Genetics, University of Oxford.

A42. Who will have control of and act as the custodian for the data generated by the study?

Drs David Spencer and Jimmy Paton for the information generated from the database and Professor Adrian Hill and Dr Stephen Chapman for the studies on DNA.

A43. Who will have access to research participants' or potential research participants' health records or other personal information? *Where access is by individuals outside the normal clinical team, justify and say whether consent will be sought.*

Only the clinicians treating the child will have immediate access to the personal health records. However, Dr David Cliff, Clinical Research Fellow will be assisting the local clinical teams/collaborators to collect historical case note data for reasons described previously. PIAG approval will be sought for this and he will also obtain the appropriate authorisations (honorary contracts etc) within the relevant institutions. All data will be anonymised for purposes of analysis with the exception of the post code (for reasons described in A53). Drs David Cliff and Steve Rushton will require access to this data for purposes of analysis.

A44. For how long will data from the study be stored?

10 Years **00** Months

Give details of where they will be stored, who will have access and the custodial arrangements for the data:

The database is maintained as described above at The University of Glasgow however, for purposes of analysis, data will be stored on secure computers at Newcastle University. The custodian will be Dr David Spencer and the research team analysing the data includes Drs Mark Pearce, Steve Rushton and David Cliff.

A45–1. How has the scientific quality of the research been assessed? *(Tick as appropriate)*

- ☐ Independent external review
- ☑ Review within a company
- ☐ Review within a multi–centre research group
- ☑ Review within the Chief Investigator's institution or host organisation
- ☑ Review within the research team
- ☑ Review by educational supervisor
- ☐ Other

Justify and describe the review process and outcome. If the review has been undertaken but not seen by the researcher, give details of the body which has undertaken the review:

Study has been approved by the Joint Research Executive Scientific Committee of the Newcastle Healthcare Charity and Newcastle Upon Tyne Hospitals Charity. It has also been reviewed by the primary funding organisation Wyeth Vaccines.

A45–2. How have the statistical aspects of the research been reviewed? *(Tick as appropriate)*

- ☐ Review by independent statistician commissioned by funder or sponsor
- ☐ Other review by independent statistician
- ☐ Review by company statistician
- ☑ Review by a statistician within the Chief Investigator's institution
- ☑ Review by a statistician within the research team or multi–centre group
- ☑ Review by educational supervisor
- ☐ Other review by individual with relevant statistical expertise

In all cases give details below of the individual responsible for reviewing the statistical aspects. If advice has been provided in confidence, give details of the department and institution concerned.

Title:	Forename/Initials:	Surname:
Dr	Mark	Pearce

Department:	Paediatric and Lifecourse Epidemiology Research Group
Institution:	Newcastle University
Work Address:	4th Floor, Sir James Spence Institute
	Royal Victoria Hospital
	Newcastle Upon Tyne Hospitals NHS Trust
Postcode:	NE1 4LP
Telephone:	0191 202 3082
Fax:	0191 202 3060
Mobile:	
E−mail:	mark.pearce@ncl.ac.uk

Please enclose a copy of any available comments or reports from a statistician.

Question(s) 46−47 disabled.

A48. What is the primary outcome measure for the study?

Changes in the incidence of empyema thoracis in UK children over time.

A49. What are the secondary outcome measures?*(if any)*

Changes in the relative incidence of different pneumococcal serotypes over time.
Difference in the length of admission between different UK centres and the effect of different treatment regimes. Identification of susceptibility genes for childhood empyema.

A50. How many participants will be recruited?
If there is more than one group, state how many participants will be recruited in each group. For international studies, say how many participants will be recruited in the UK and in total.

The incidence of empyema thoracis in UK children is approximately 800−1000 patients per year. We are intending to obtain data on the majority of these patients which can then be entered into the database.

A51. How was the number of participants decided upon?

This is the current incidence of emypema in children in the UK; it is expected that it may change with time for the reasons stated previously.

If a formal sample size calculation was used, indicate how this was done, giving sufficient information to justify and reproduce the calculation.

Not Applicable.

A52. Will participants be allocated to groups at random?

○ Yes ◉ No

A53. Describe the methods of analysis (statistical or other appropriate methods, e.g. for qualitative research) by which the data will be evaluated to meet the study objectives.

Descriptive statistics will be used to analyse and interpret the outcome measures. Any comparisons of categoric/continuous data will be performed using the appropriate tests based on the relevant statistical

assumptions and data distributions (for example t−tests or Mann−Whitney, ANOVA or Kruskal−Wallis). Various statistical packages eg Stata version 9.0 will be used to analyse the data. Space−time clustering and generalised linear modelling approaches will be used and allow us to investigate the extent to which incidence of the disease is clustered, to consider possible causes of clustering where it exists and investigate the role of social and environmental predictors in determining the incidence of infection. The full post code will be required for this analysis.
For the genetic aspect of the study, differences in the genotype distributions between empyema cases and healthy controls will be compared using Chi−squared testing. Logistic regression analysis will be performed to correct for the possible confounding effects of other genes and other recorded clinical variables on disease susceptibility. Correction will be made for the multiple independent comparisons tested as part of a genome−wide study (as the majority of markers tested will not be associated, but the large number of markers studied increases the risk of false positive associations).

A54. Where will the research take place? *(Tick as appropriate)*

- [x] UK
- [ ] Other states in European Union
- [ ] Other countries in European Economic Area
- [ ] Other

If Other, give details:

A55. Has this or a similar application been previously rejected by a Research Ethics Committee in the UK, the European Union or the European Economic Area?

○ Yes ◉ No

A56. In how many and what type of host organisations (NHS or other) in the UK is it intended the proposed study will take place?

Indicate the type of organisation by ticking the box and give approximate numbers if known:

	Number of organisations
[x] Acute teaching NHS Trusts	26
[ ] Acute NHS Trusts	
[ ] NHS Primary Care Trusts or Local Health Boards in Wales	
[ ] NHS Trusts providing mental healthcare	
[x] NHS Health Boards in Scotland	3
[x] HPSS Trusts in Northern Ireland	1
[ ] GP Practices	
[ ] NHS Care Trusts	
[ ] Social care organisations	
[ ] Prisons	
[ ] Independent hospitals	
[ ] Educational establishments	
[ ] Independent research units	
[ ] Other (give details)	

Other:

A57. What arrangements are in place for monitoring and auditing the conduct of the research?

Monitoring will take place on a monthly basis within the research team at Newcastle University. A steering committee will be convened under the guidance of the British Paediatric Respiratory Society.

A57a. Will a data monitoring committee be convened?

○ Yes ◉ No

If Yes, details of membership of the data monitoring committee (DMC), its standard operating procedures and summaries of reports of interim analyses to the DMC must be forwarded to the NHS Research Ethics Committee which gives a favourable opinion of the study.

What are the criteria for electively stopping the trial or other research prematurely?

Not applicable as this is a non interventional study.

A58. Has external funding for the research been secured?

◉ Yes ○ No

If Yes, give details of funding organisation(s) and amount secured and duration:

Organisation:	Wyeth Pharmaceuticals
Address:	Huntercombe Lane South, Taplow, Maidenhead,
Post Code:	SL6 0PH
UK contact:	Dr Ros Hollingsworth
Telephone:	01628 604377
Fax:	01628 414982
Mobile:	
E−mail:	
Amount (£):	150,000 Duration: 36 Months

A59. Has the funder of the research agreed to act as sponsor as set out in the Research Governance Framework?

○ Yes ◉ No

Has the employer of the Chief Investigator agreed to act as sponsor of the research?

◉ Yes ○ No

Lead sponsor *(must be completed in all cases)*

Name of organisation which will act as the lead sponsor for the research:

Newcastle upon Tyne Hospitals NHS foundation trust

Status:

(●) NHS or HPSS care organisation ( ) Academic ( ) Pharmaceutical industry ( ) Medical device industry ( ) Other

If Other, please specify:

Address:	R&D Dept 4th Floor, Leazes Wing Royal Victoria Infirmary,Queen Victoria Road,Newcastle upon Tyne
Post Code:	NE1 4LP
Telephone:	0191 2825213
Fax:	0191 282 0064
Mobile:	
E−mail:	Amanda.Tortice@nuth.nhs.uk

Sponsor's UK contact point for correspondence with the main REC *(must be completed in all cases)*

Title: Forename/Initials: Amanda Surname: Tortice

Work Address:	R&D Dept 4th Floor, Leazes Wing Royal Victoria Infirmary,Queen Victoria Road,Newcastle upon Tyne
Post Code:	NE1 4LP
Telephone:	0191 2825213
Fax:	0191 282 0064
Mobile:	
E−mail:	Amanda.Tortice@nuth.nhs.uk

Co−sponsors

Are there any co−sponsors for this research?

( ) Yes (●) No

A60. Has any responsibility for the research been delegated to a subcontractor?

( ) Yes (●) No

A61. Will individual *researchers* receive any personal payment over and above normal salary for undertaking this research?

( ) Yes (●) No

A62. Will individual *researchers* receive any other benefits or incentives for taking part in this research?

( ) Yes (●) No

A63. Will the host organisation or the researcher's department(s) or institution(s) receive any payment or benefits in excess of the costs of undertaking the research?

○ Yes ◉ No

A64. Does the Chief Investigator or any other investigator/collaborator have any direct personal involvement (e.g. financial, share−holding, personal relationship etc.) in the organisations sponsoring or funding the research that may give rise to a possible conflict of interest?

○ Yes ◉ No

A65. Research reference numbers: *(give any relevant references for your study):*

Applicant's/organisation's own reference number, e.g. R&D (if available):	3927
Sponsor's/protocol number:	3927
Funder's reference number:	102358
International Standard Randomised Controlled Trial Number (ISRCTN):	
Project website:	

A66. Other key investigators/collaborators *(all grant co−applicants or protocol co−authors should be listed)*

Title: Dr Forename/Initials: David Surname: Spencer

Post:	Consultant Respiratory Paediatrician
Qualifications:	MB BS(Hons) MRCP MD FRCPCH
Organisation:	Newcastle upon Tyne Hospitals NHS Foundation Trust
Work Address:	Newcastle upon Tyne Hospitals NHS Trust Freeman Hospital Newcastle upon Tyne, UK
Postcode:	NE7 7DN
Telephone:	
Fax:	
Mobile:	07967379085
E−mail:	David.Spencer2@nuth.nhs.uk

Title: Dr Forename/Initials: David Surname: Cliff

Post:	Clinical Research Fellow
Qualifications:	MB ChB MRCPCH BSc(Hons)
Organisation:	Newcastle upon Tyne Hospitals NHS Foundation Trust
Work Address:	Sir James Spence Institute, 4th floor Royal Victoria Hospital Newcastle upon Tyne Hospitals NHS Trust
Postcode:	NE1 4LP
Telephone:	0191 202 3099
Fax:	
Mobile:	07766052850

E−mail: david.cliff@ncl.ac.uk

Title: Dr Forename/Initials: Stephen Surname: Chapman

Post: Clinical Lecturer in Respiratory Medicine
Qualifications: BM BCh MRCP MA
Organisation: University of Oxford
Work Address: Wellcome Trust Centre for Human Genetic
Roosevelt Drive
Oxford
Postcode: OX3 7BN
Telephone: 01865 287592
Fax: 01865 287660
Mobile:
E−mail: schapman@well.ox.ac.uk

Title: Dr Forename/Initials: Mark Surname: Pearce

Post: Lecturer in life course epidemiology
Qualifications: BSc, MSc, CStat, PhD
Organisation: Newcastle University
Work Address: Sir James Spence Institute, 4th floor
Royal Victoria Hospital
Newcastle upon Tyne Hospitals NHS Trust
Postcode: NE1 4LP
Telephone: 0191 202 3082
Fax: 0191 202 3060
Mobile:
E−mail: mark.pearce@ncl.ac.uk

Title: Dr Forename/Initials: Steve SP Surname: Rushton

Post: Reader in Biological modelling
Qualifications: BA (Hons) Oxon, PhD
Organisation: Newcastle University
Work Address: IRES institute
Devonshire Building
Newcastle University
Postcode: NE1 7RU
Telephone: 0191 246 4836
Fax:
Mobile:
E−mail: steven.rushton@ncl.ac.uk

A67. What arrangements are being made for continued provision of the intervention for participants, if appropriate, once the research has finished? *May apply to any clinical intervention, including a drug, medical device, mental health intervention, complementary therapy, physiotherapy, dietary manipulation, lifestyle change, etc.*

Not applicable

PART A: Summary of Ethical Issues

A68. What are the main ethical issues with the research?

Summarise the main issues from the participant's point of view, and say how you propose to address them.

From the participants point of view the main issue is the use of patient identifiable information for purposes of linkage and analysis. We seek to do this with section 60 support from PIAG. The reasons for this are as follows:

1) As previously mentioned the pneumococcal vaccine was introduced into the national immunisation schedule in September of 2006. At the same time the HPA commenced the national enhanced surveillance programme (as previously described A10). We aim to obtain as much historical data on patients since this time as resources will allow. Comparative data over time is vital to our study as the national immunisation programme begins to have an effect and it is therefore essential we retrospectively obtain data. Considering there are 28 UK sites that are involved and potentially over 800 children per year (increasing) then collecting consent for retrospective analysis via pre prepared letter would be impracticable and not within our resources or those of the local collaborators.

2) As part of the process of obtaining ethical approval for the study it is necessary to inform the local research and development departments of all the individuals (as requested on the SSI form 14 and 23) who would be involved in taking consent and filling in the web−based database. This is alongside their qualifications and often with evidence of formal good clinical practice (GCP) training. We have established local collaborators (Consultant Respiratory Paediatricians and British Paediatric Respiratory Society members) at each of the 28 sites but it would be impossible and impracticable to provide local R and D offices with further information regarding which staff member would be the clinician available to take consent when a child presents or ensure that all of them have had formal GCP training. This is due to the considerable number of different clinicians admitting cases across the UK and the regular change over of staff. The work load of the individual local collaborators at each site is too great for them to commit to be available to take consent when cases present.

3)We feel that the aforementioned issues alongside the large number of centres involved would lead to significant under ascertainment and bias in our clinical dataset if consent were required. This is particularly relevant when retrospectively collecting data and combining information from two sources i.e the HPA and the British Paedaitric Respiratory database.

We appreciate that consent is required for the DNA specimen but are prepared to accept that one individual will be have to be responsible for this when patients are admitted and that not all children will have a specimen taken. Retrospective data is not necessary for the genetic aspect of the study.

With regard to the genetic studies, participants may be concerned regarding the confidentiality of any information relating to their genetic background. Others may be concerned about the potential for a genome−wide study to identify unsuspected genetic abnormalities which have implications for their future health − such information may be unwished for by the individual and may have implications for health insurance. In order to address both of these issues, all genetic samples will be anonymised and it will not be possible for researchers to trace genetic information back to individual patients. The identification of unsuspected genetic abnormalities which are know to cause serious disease (e.g. Huntingdon's disease) is in fact exceedingly unlikely, as these are rare mutations and will not be studied as part of a genome−wide approach. Genome−wide approaches instead study common genetic variation (single nucleotide polymorphisms), and the genetic variations associated with common disease typically have only small individual effects − a combination of multiple susceptible genetic variants and exposure to particular environmental factors then results in disease.

Indicate any issues on which you would welcome advice from the ethics committee.

We would welcome advice on whether the ethics committee agree with us that site specific assessment is not required. We feel that collecting saliva for DNA doesn't carry any appreciable risk to the patient and that those performing the procedure have considerable clinical experience in seeking informed consent from patients. All the local collaborators will have the support of the chief investigator and other key collaborators so named as required.

Question(s) 69 disabled.

PART A: Student Page

A70. Give details of the educational course or degree for which this research is being undertaken:

Name of student:

David Cliff

Name and level of course/degree:

MD

Name of educational establishment:

University of Newcastle upon Tyne
NE1 7RU
United Kingdom

Name and contact details of educational supervisor:

Dr David Spencer
Consultant in Respiratory Paediatrics
Newcastle upon Tyne Hospitals NHS Foundation Trust
Freeman Hospital
Newcastle upon Tyne
U.K
NE7 7DN
Direct Dect 0191 244 8292
Hospital Switch 0191 233 6161
Fax 0191 223 1099
email: david.spencer2@nuth.nhs.

A71. Declaration of educational supervisor

I have read and approved both the research proposal and this application for the ethical review. I am satisfied that the scientific content of the research is satisfactory for an educational qualification at this level. I undertake to fulfil the responsibilities of a supervisor as set out in the Research Governance Framework for Health and Social Care.

Signature: .....

Print Name: Dr David Spencer

Date: (dd/mm/yyyy)

A one–page summary of the supervisor's CV should be submitted with the application

PART B: Section 1 – Conduct of the research at local sites

From the answer given to question A6, it is assumed that:

- *Local Principal Investigators will not be appointed at each research site participating in this study.*
- *Applications for site–specific assessment by local Research Ethics Committees will not be required.*
- *There will be no requirement for individual research sites to be approved by the main REC as part of the ethical review.*

The following general information should be provided to the main REC about the local conduct of the study.

1. What research procedures will be carried out at individual research sites?

Concerning patient contact the only research procedures to be carried out at individual centres will be the input of clinical data into the database and the collection of saliva for DNA. All other relevant samples are sent as part of the routine management of the patient as described previously.
Investigation of 'virulence' factors and antibiotic tolerance will be performed at laboratories at The University of Glasgow and The Freeman Hospital, Newcastle Upon Tyne Hospitals NHS Foundation Trust respectively. These studies will use strains of the bacteria only from anonymised samples. The Health Protection Agency are supervising the work on serotyping and have their procedures in place.

2. Are any ethical issues likely to arise at individual sites that are not covered in the protocol for the study and if so how will these be addressed?
For example, a need for particular facilities, or to notify local clinicians or departments about the research, or to arrange additional local support for participants.

No

3. How will the Chief Investigator and his/her team supervise the conduct of the research at individual sites? What responsibilities will be delegated to local collaborators?

Clear guidelines will be given to each site concerning inputting patient information into the database and consent when collecting saliva for DNA. All the microbiological data to be analysed is already being collected as part of a national surveillance programme as previously outlined. The research fellow will maintain regular close contact by telephone and email with each collaborating centre and produce a monthly update email for all centres. The Chief Investigtor and clinical fellow are based at the Freeman Hospital and will directly supervise the studies on antibiotic tolerance. Regular monthly contact will be maintained with the team at Glasgow who have considerable experience working in the area of pneumococcal 'virulence' factors.

Management approval to proceed with the research will be required from the R&D office for each NHS care organisation in which research procedures are undertaken. The Site–Specific Information Form should be used to apply for R&D approval at NHS sites in England, Wales and Scotland.

PART B: Section 5 – Use of newly obtained human biological materials

1. What types of human tissue or other biological material will be included in the study?

A 1–2ml sample of saliva will be taken from each participant. Saliva samples will stored frozen, and DNA will be extracted at 6–month intervals throughout the study period.

2. Who will collect the samples?

Samples will be collected by the local collaborators (doctors) at each clinical site.

3. Will the samples be: *(Tick as appropriate)*

☑ Obtained primarily for research purposes?
☐ Surplus (i.e. left over from tissue taken in the course of normal clinical care for diagnostic or therapeutic purposes)?

4. Will informed consent be obtained from donors for use of the samples:

In this research?

◉ Yes ○ No

In future research?

◉ Yes ○ No

5. Will the samples be stored:

In fully anonymised form? *(link to donor broken)*

◉ Yes ○ No

In linked anonymised form? *(linked to donor but donor not identifiable to researchers)*

○ Yes ◉ No

If Yes, say who will have access to the code and personal information about the donor.

The only documentation which will contain both a trial number and the participant's name is the informed consent form. These will be held securely and separately from the online enrolment form (containing patient clinical information, identified only by trial number) and DNA sample/genetic data (identified only by trial number), and will not be available to researchers. There will be no code which links genetic data to an individual participant.

In a form in which the donor could be identifiable to researchers?

○ Yes ◉ No

If Yes, please justify:

6. What types of test or analysis will be carried out on the samples?

Study of variants spread across all human genes will be performed. Ethical issues relating to this are discussed in the answer to question 68.

7. Will the research involve the analysis of human DNA in the samples?

(●) Yes ( ) No

8. Is it possible that the research could produce findings of clinical significance for individuals? *(May include relatives as well as donors)*

(●) Yes ( ) No

9. If so, will arrangements be made to notify the individuals concerned?

( ) Yes (●) No ( ) Not applicable

If No, please justify. If Yes, say what arrangements will be made and give details of the support or counselling service.

It is theoretically possible but highly unlikely that the research will produce findings of clinical significance for individuals – as discussed in question 68, the genetic polymorphisms studied are common and usually have relatively small effects on disease causation. From an ethical perspective, however, it is important that there is no possibility of tracing genetic information back to individual patients and therefore all samples will be anonymised.

10. Give details of where the samples will be stored, who will have access and the custodial arrangements.

DNA samples will be stored frozen in Professor Adrian Hill's laboratory at the Wellcome Trust Centre For Human Genetics, University of Oxford. Only Professor Hill and Dr Stephen Chapman and their research team will have access to the samples.

11. What will happen to the samples at the end of the research?

( ) Destruction

( ) Transfer to research tissue bank
(If the bank is in England, Wales or Northern Ireland a licence from the Human Tissue Authority will be required to store the tissue for possible further research.)

( ) Storage by research team pending ethical approval for use in another project
(Unless the researcher holds a licence from the Human Tissue Authority, a further application for ethical review should be submitted before the end of this project.)

( ) Storage by research team as part of a new research tissue bank
(The bank will require a licence from the Human Tissue Authority. A separate application for ethical review of the tissue bank may also be submitted.)

(●) Not yet known

Please give further details of the proposed arrangements:

PART B: Section 7 – Declarations

Declaration by Chief Investigator

1. The information in this form is accurate to the best of my knowledge and belief and I take full responsibility for it.
2. I undertake to abide by the ethical principles underlying the Declaration of Helsinki and good practice guidelines on the proper conduct of research.
3. If the research is approved I undertake to adhere to the study protocol, the terms of the full application of which the main REC has given a favourable opinion and any conditions set out by the main REC in giving its favourable opinion.
4. I undertake to seek an ethical opinion from the main REC before implementing substantial amendments to the protocol or to the terms of the full application of which the main REC has given a favourable opinion.
5. I undertake to submit annual progress reports setting out the progress of the research.
6. I am aware of my responsibility to be up to date and comply with the requirements of the law and relevant guidelines relating to security and confidentiality of patient or other personal data, including the need to register when necessary with the appropriate Data Protection Officer.
7. I understand that research records/data may be subject to inspection for audit purposes if required in future.
8. I understand that personal data about me as a researcher in this application will be held by the relevant RECs and their operational managers and that this will be managed according to the principles established in the Data Protection Act.
9. I understand that the information contained in this application, any supporting documentation and all correspondence with NHS Research Ethics Committees or their operational managers relating to the application:

   – Will be held by the main REC until at least 3 years after the end of the study.

   – May be disclosed to the operational managers or the appointing body for the REC in order to check that the application has been processed correctly or to investigate any complaint.

   – May be seen by auditors appointed by the National Research Ethics Service to undertake accreditation of the REC.

   – Will be subject to the provisions of the Freedom of Information Acts and may be disclosed in response to requests made under the Acts except where statutory exemptions apply.

Optional – please tick as appropriate:

☐ I would be content for members of other RECs to have access to the information in the application in confidence for training purposes. All personal identifiers and references to sponsors, funders and research units would be removed.

Signature: .....

Print Name: David Spencer

Date: (dd/mm/yyyy)

Declaration by the sponsor's representative

If there is more than one sponsor, this declaration should be signed on behalf of the co−sponsors by a representative of the sponsor nominated to take the lead for the REC application.

I confirm that: *(tick as appropriate)*

- ☑ This research proposal has been discussed with the Chief Investigator and agreement in principle to sponsor the research is in place.
- ☑ An appropriate process of scientific critique has demonstrated that this research proposal is worthwhile and of high scientific quality.*
- ☑ Any necessary indemnity or insurance arrangements, as described in question A35, will be in place before this research starts.
- ☑ Arrangements will be in place before the study starts for the research team to access resources and support to deliver the research as proposed.
- ☑ Arrangements to allocate responsibilities for the management, monitoring and reporting of the research will be in place before the research starts.
- ☑ The duties of sponsors set out in the NHS Research Governance Framework for Health and Social Care will be undertaken in relation to this research.**

* Not applicable to student research (except doctoral research).
** Not applicable to research outside the scope of the Research Governance Framework.

Signature: .....

Print Name:

Post:

Organisation: Newcastle Upon Tyne Hospitals NHS Foundation Trust

Date: (dd/mm/yyyy)