University of Strathclyde
Department of Mathematics and Statistics

Modelling Genetic Effects in the Transmission of Pneumococcal Carriage

Karen E. Lamb

A thesis presented in fulfilment of the requirements of the degree of Doctor of Philosophy.

2010

For my Dad. You're a total legend!

Acknowledgements

I would like to thank both of my supervisors, Prof. Chris Robertson and Dr. David Greenhalgh, for their advice and encouragement during the course of my studies.

I would like to thank EPSRC and Wyeth for funding this project. In addition, I would like to thank Boyka Stoykova and the rest of the staff who assisted me at Wyeth pharmaceuticals for their guidance with the health economics work.

Thanks also to Prof. George Gettinby for the many conversations, lunches and coffees; the 7.47 girls (Kim, Amanda, Gail, Lesley and Matina) for putting up with my frequent moaning; Adam for much LaTeX assistance and just for generally letting me rant at him; Stephen for the cakes and chat. I would also like to thank all of my "stattos" (David, Claire, Hannah, Ruth and Robin) for their good chat and plentiful supply of wine. David- particular thanks go to you for responding to my frequent demands for computing assistance! Thanks also to Johnny for his computer help and to Carolyn- what the hell are we going to talk about now without the incessant Ph.D. chat?!

Furthermore, I would like to thank Marty for all of the wine in the early stages of the Ph.D. and Hannah for the hugs in the latter stages. You have both been a tremendous support. I am also incredibly indebted to my parents, my brother and Susan for always believing in me and to my best pal Colin- you now know more statistics than you've surely ever wished to know!

My final thanks go to Dave. Your support is always appreciated. I couldn't have done this without you.

Abstract

In this thesis mathematical, statistical and economic techniques involved in the decision making regarding the use of a vaccine intervention and the assessment of vaccine effectiveness are described and used. The central themes of the thesis are pneumococcal carriage and disease and prevention of both carriage and disease through the use of the 7-valent pneumococcal conjugate vaccine (PCV-7).

The thesis can be considered in four sections. The first section contains two mathematical modelling chapters in which differential equations are used to describe pneumococcal carriage. Through the use of these models potential problems for the long term efficacy of the vaccine are explored. These involve problems associated with increased non-vaccine type serotype carriage, or genetic concerns regarding sequence types which may manifest as a vaccine type and a non-vaccine type serotype.

The next section involves a review of techniques used in the economic analysis of health care technologies. Statistical techniques are used to model trends in pneumococcal diseases and infections in England and Wales to predict hospital episode figures assuming no intervention. These can be compared to the true figures following the introduction of PCV-7. This aids in the estimation of the effectiveness of the vaccine and revised parameter estimates can be fed into cost-effectiveness models for PCV-7.

The third section involves the analysis of routinely collected data on cases of invasive pneumococcal disease (IPD) in Scotland, with information available regarding the serogroup or serotype and multi-locus sequence type (MLST) of the bacterium responsible for the disease. Changes in trend of the serogroups or serotypes and MLSTs responsible for disease in Scotland prior to the introduction of PCV-7 are identified as well as associations between 30 day mortality and serogroup/type or MLST.

In the final section, single-level and multi-level modelling techniques are adopted to assess individual level and postcode district level factors affecting the uptake of PCV-7 in Scotland.

Contents

List of Tables

List of Figures

Chapter 1

The biology of *Streptococcus pneumoniae*

1.1 Introduction

In this thesis mathematical, statistical and economic techniques involved in the decision making regarding the use of a vaccine intervention and the vaccine effectiveness are described and used. In this chapter the biological background to the thesis will be discussed.

1.2 The bacterium

Streptococcus pneumoniae (*S. pneumoniae*), or pneumococcus, is a bacterium (a single celled microorganism), that was discovered in 1880 simultaneously, but independently, by the American physician George Miller Sternberg (Sternberg 1881) and the French chemist Louis Pasteur (Pasteur 1881). Since the discovery of this bacterium, over 90 different pneumococcal serotypes within 46 serogroups have been identified (Henrichsen 1995), with the most recent discovery of the 91st serotype, 6C, in 2007 (Park et al. 2007).

Serotypes are defined according to the structure of the polysaccharide capsule which encases the bacterium. This polysaccharide capsule protects the bacterium from the immune system, enabling it to cause infection and disease. Thus,

the serotype is a known virulence factor, with the level of virulence varying by serotype (Lindberg 1999). However, it is clear that other genetic factors must be important in the ability of the bacterium to cause disease (Hollingshead and Briles 2001).

Certain serotypes have similar antigens to one another and so are classified together in a serogroup. For example, serotypes 9A, 9L, 9N and 9V are all antigenically related and are all from serogroup 9. Certain serotypes are not related to any others and these serotypes are classified by number alone, with no associated letter (Kalin 1998).

S. pneumoniae may also be categorised by multi-locus sequence type (MLST). Pneumococcal sequence types are defined according to 7 house-keeping genes identified within the pneumococcal genetic material (Enright and Spratt 1998). These particular genes are used to determine the sequence type as they are not under a great deal of selective pressure and so remain relatively constant over time. There are hundreds of different pneumococcal MLSTs and it has been observed that some MLSTs are able to manifest in more than one serotype.

1.3 Serotypes and MLSTs

The relationship between pneumococcal serotypes and MLSTs is difficult to define as no direct correlation has been established. However, certain MLSTs have been shown to be more associated with particular serotypes than others. For example, it has been observed that MLST 9 commonly appears associated with serotype 14 (Brueggemann et al. 2003; Jefferies et al. 2004). Studies of *S. pneumoniae* primarily concentrate on the serotypes and the MLSTs are not often recorded so it is difficult to establish the importance of MLSTs in the disease potential of the bacterium. However, in recent years studies have been carried out to identify MLSTs common in invasive disease in an attempt to understand the relationship between serotypes and MLSTs.

A study of 501 pneumococcal isolates (150 from invasive disease isolates collected between 1995 and 2003, and 351 from various carriage studies between 1999 and

2001) from children under the age of five years in Oxford was carried out to determine whether it is the serotype or the MLST that is the important factor for the development of invasive pneumococcal disease (IPD) (Brueggemann et al. 2003). In this paper, it is stated that if serotypes are the only important factor in determining whether or not IPD will occur then a pneumococcal isolate of one serotype associated with a certain MLST should have the same ability to cause invasive disease as an isolate of the same serotype but different MLST. For example, if MLSTs play no part in determining whether or not invasive disease shall occur then a serotype 14 isolate should have the same potential to cause invasive disease regardless of whether or not the isolate is associated with MLST 9 or MLST 124. In this paper, it is concluded that there is in fact a correlation between pneumococcal MLSTs and serotypes, with the most invasive sequence types corresponding to the most invasive serotypes.

A study of 368 pneumococcal isolates recovered from invasive disease in Scotland also showed a strong correlation between MLSTs and serotypes involved in invasive disease (Clarke et al. 2004). In addition, an American study has also demonstrated a correlation between MLSTs and serotypes involved in IPD (Beall et al. 2006).

1.4 Pneumococcal carriage

S. pneumoniae may be carried asymptomatically in the nasopharynxes of both children and adults, with up to 30% of adults and 60% of children becoming colonised with no adverse effects (Ghaffar et al. 1999). Asymptomatic carriage is most common among very young children, with studies showing high carriage rates in children under the age of two years (Bogaert et al. 2006). Other studies demonstrate the relationship between carriage and age during the early years of life, with asymptomatic colonisation shown to be anywhere between approximately 10% during the early weeks following birth to around 100% by the end of the first year of life (Spratt et al. 2004).

Pneumococcal carriage studies have shown that the duration of carriage is also

associated with age, with the length of time the pneumococcus colonises an individual decreasing with increasing age. In one study, it was estimated that the average duration of carriage was 51 days for children and only 19 days for adults (Melegaro et al. 2004).

Multiple carriage of serotypes has been documented but appears rare, with a longitudinal carriage study of pneumococcus in 82 children showing carriage of 2 serotypes in 4% of the isolates examined and 3 serotypes in 0.3% (Gray et al. 1980). It has been suggested from other studies that the incidence of multiple colonisation could be up to about 10% (Crook et al. 2004).

Pneumococcal strains may be passed from person to person through direct contact or via airborne respiratory droplets, such as those from coughs and sneezes. As children are the primary carriers of *S. pneumoniae*, this group is deemed to be the predominant cause of the transmission of pneumococci within the population (Leiberman et al. 1999). Pneumococcal carriage rates and transmission within a community are dependent upon several factors such as frequent close contact with other individuals, particularly young children in environments such as child care centres, and high incidence of viral respiratory tract infections (Dagan et al. 2002).

Differences in pneumococcal carriage have been observed between the developed and developing countries. Within developing countries, studies have shown high carriage rates of two or three times the rates observed in developed countries, with pneumococcal colonisation occurring at an earlier age (Obaro and Adegbola 2002). However, as with developed countries, it has been observed that developing countries display reductions in carriage rates with increasing age (Lloyd-Evans et al. 1996). These reductions both in developed and developing countries may be due to decreased exposure to pneumococci as close contact with many individuals is a risk factor for carriage and this occurs particularly at a young age. Alternatively, this could be attributable to increased resistance of the immune response to the predominant pneumococcal strains with increasing age (Obaro and Adegbola 2002).

1.5 Pneumococcal disease and infection

Pneumococcus can cause a variety of infections such as otitis media (OM), an ear infection common in infants, sinusitis, and pneumonia. It is the most common cause of serious pneumonia (NHS Scotland 2003). In addition, various invasive diseases may be caused by *S. pneumoniae* such as meningitis and septicaemia. Pneumococcal disease occurs when the bacterium is carried to a normally sterile air space in the body such as the lungs, sinuses or middle ear. Disease is always preceded by nasopharyngeal colonisation (Gray et al. 1980).

Worldwide, there are approximately one and a half million deaths per year which are attributable to pneumococcal disease, with an estimated 700,000 to one million in children under the age of five years (World Health Organization 2007). For developed countries such as the United States of America (USA) and Europe, the observed incidence of IPD amongst young children is between 8 and 75 cases per 100,000 population each year whilst in developing countries the range in incidence is from 100 to over 500 cases per 100,000 which clearly shows the difference in disease burden between these groups of countries (Brueggemann et al. 2004).

It has been observed that, although over 90 types exist, the majority of pneumococcal disease is attributable to only around 20 or 30 types (George and Melegaro 2001). Approximately two thirds of adult pneumococcal disease and 80% of disease in children is attributable to between 8 and 10 serotypes (Salisbury et al. 2006).

It is known that pneumococcal disease and infection must follow colonisation but it is important to assess the relationship between the bacterial types commonly found in carriage studies and those primarily found in disease. This relationship was assessed in a meta-analysis carried out by Brueggemann et al. (2004). In this analysis, it was discovered that those serogroups or serotypes most commonly carried were the least likely types to cause IPD whilst those carried less often were identified more commonly in disease. It has been stated that the serotypes 1, 3, 4 and 5 are identified in IPD but are not commonly found in asymptomatic colonisation of children (Dagan et al. 2002). Furthermore, Dagan et al. state that in developed countries serotypes 1, 4 and 5 are more common in disease in

adults whilst studies in developing countries have shown serotype 1 in particular to feature prominently in disease (Saha et al. 2009; Williams et al. 2009). In contrast to the statement made by Dagan et al., a recent study of IPD in Germany showed serotype 1 to be amongst the most common disease causing serotypes in children (Imöhl et al. 2009).

1.6 Prevention and treatment

For many years antibiotics were used to treat pneumococcal infections and disease. However, this led to an increase in antibiotic resistant pneumococcal strains, with great resistance shown to develop in the USA primarily in the 1990s (Brueggemann and Doern 2000). In an effort to prevent the occurrence of pneumococcal disease and infection, pneumococcal conjugate vaccines (PCVs) were developed, with serotypes included in these vaccines, or vaccine related serotypes, identified as those most commonly associated with resistance (Finkelstein et al. 2003; Dagan et al. 2003).

Pneumococcal vaccines have been in use for the last thirty years. In 1977, a 14-valent pneumococcal polysaccharide vaccine (PPV) was licensed for use in the USA and in 1979 in the United Kingdom (UK). This vaccine was replaced in the USA in 1983, and in the UK in 1989, by the 23-valent polysaccharide vaccine (Pneumovax^{TM}II, PPV-23), produced by Merck Research Laboratories, USA, and this vaccine is still currently administered in both the USA and the UK, and in other countries throughout the world.

1.6.1 23-valent polysaccharide vaccine

PPV-23 was initially administered to anyone over 50 years of age believed to be at an increased risk of developing pneumococcal disease, such as those with an immune deficiency. The license was altered in 1984 in the USA (Immunization Practices Advisory Committee 1984), and introduced routinely in 2003 in the UK, to include any adult over the age of 65 years (Salisbury et al. 2006), regardless of the risk posed to them by pneumococcal disease. However, the introduction of PPV-23 was staggered in England with all individuals aged over

80 years administered the vaccine in August 2003; those aged 75 years and over administered the vaccine in April 2004 and those aged over 65 years who had not been vaccinated given the vaccine in April 2005. Routine immunisation for over 65 year old individuals in England commenced in 2006/07 (NHS Scotland 2003).

PPV-23 consists of purified capsular polysaccharide antigens from 23 different pneumococcal serotypes (1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19F, 19A, 20, 22F, 23F and 33F) which cause over 90% of the pneumococcal disease cases in the UK (George and Melegaro 2001). Early studies suggested PPV-23 has between 50 and 70% efficacy in preventing IPD due to the 23 serotypes contained in the vaccine (Fedson 1999). Recent studies have shown the vaccine to provide no protection against pneumonia (Huss et al. 2009), or pneumococcal pneumonia without septicaemia (Jackson et al. 2003).

In the UK, PPV-23 is recommended for one dose per lifetime for healthy elderly individuals. However, repeated doses may be offered in five year intervals to those whose antibody levels decline more quickly (NHS Scotland 2003). In addition to routine immunisation of the elderly, PPV-23 may be administered to anyone over the age of 2 years at an increased risk of developing pneumococcal disease and infection. However, PPV-23 cannot be administered to children under the age of 2 years as this age group do not have a good antibody response to polysaccharide vaccinations (Riley et al. 1981).

1.6.2 7-valent pneumococcal conjugate vaccine

In 2000, the 7-valent pneumococcal conjugate vaccine (Prevnar/Prevenar^{TM}, PCV-7), produced by Wyeth, was licensed for use in the USA to prevent pneumococcal disease in infants (Immunization Practices Advisory Committee 2000). The vaccine consists of the purified polysaccharide capsular antigens of 7 pneumococcal serotypes (4, 6B, 9V, 14, 18C, 19F and 23F) conjugated to a protein. It is due to the fact that the purified capsular antigens are conjugated to a protein that the vaccine is effective in preventing disease for those under 2 years of age. PCV-7 was licensed in Europe in 2001. Amongst the 27 European Union countries and other European countries, Croatia, Iceland, Norway, Switzerland and

Turkey, 24 had adopted, or decided to adopt, PCV-7 in their routine childhood immunisation schedule by the beginning of 2009. Of these 24, 12 introduced the vaccine in a schedule consisting of three primary doses and one booster dose, as used in the USA; 11 introduced a schedule of two primary doses followed by a booster dose; Switzerland adopted both of these schedules, with three primary doses administered to children at an increased risk of infection or disease and two primary doses administered to all other children (De Carvalho Gomes et al. 2009). Recent reports have stated that PCV-7 is available in over 90 different countries worldwide (Wyeth Pharmaceuticals 2009a). In addition to Europe and the USA, these include Australia, South America and Canada in three or four dose schedules (Beutels et al. 2007).

PCV-7 was introduced in the UK in 2002 for those children under 2 years of age most at risk of developing pneumococcal disease and in 2004 this was extended to all at-risk children under the age of 5 years. In 2006, PCV-7 was introduced to the routine childhood immunisation schedule in the UK (Salisbury et al. 2006). Early studies following the introduction of PCV-7 to the routine immunisation schedule showed over 90% effectiveness in preventing IPD attributed to the 7 serotypes included in the vaccine in healthy children (Whitney et al. 2006).

1.6.3 Developments in conjugate vaccination formulations

Recently new formulations of PCV have been licensed. A 10-valent pneumococcal vaccine, Synflorix^{TM}, produced by GlaxoSmithKline was licensed in Europe in 2009 (GlaxoSmithKline 2009). This vaccine includes purified polysaccharide capsules of the seven serotypes found in PCV-7 in addition to that of the serotypes 1, 5 and 7F (Knuf et al. 2008). These three serotypes are very invasive and becoming increasingly common (Brueggemann and Spratt 2003; Ihekweazu et al. 2008; Munõz-Almagro et al. 2008). Thus, it is believed that this vaccine could potentially have a greater impact on the burden of IPD than PCV-7.

In addition, a 13-valent vaccine, Prevenar 13^{TM}, developed by Wyeth has recently been approved by the Chilean Ministry of Health, the first government agency to approve the vaccine, and will be administered in four doses in the infant

immunisation schedule in Chile, with primary doses at 2, 4 and 6 months and a fourth dose administered between the ages of 12 and 15 months. This vaccine should provide protection against the 10 serotypes found in Synflorix^{TM}as well as serotypes 3, 6A and 19A and builds on the formulation of PCV-7. The vaccine is also currently undergoing trials for use in adults (Wyeth Pharmaceuticals 2009b)

1.6.4 Benefits of conjugate vaccination

The main advantage of the PCVs over the PPV, other than the fact that they are effective in preventing disease in one of the key risk groups, young children, is that conjugate vaccines have been shown to not only prevent invasive disease due to the serotypes included in the vaccine but also prevent carriage of these serotypes (Dagan et al. 1996; Dagan et al. 2002) and in the USA, approximately 80% of all carried serotypes are vaccine type (VT) serotypes (Butler et al. 1995). In preventing carriage, PCV-7 allows for the possibility of herd immunity to occur in the population.

Herd immunity occurs when the immunisation of selected individuals in a community elicits protection to the whole community. As mentioned previously, children under the age of 2 years commonly carry *S. pneumoniae* asymptomatically. Therefore, it is believed that, since the conjugate vaccine reduces the carriage of the 7 VT serotypes in children less than 2 years of age, the overall carriage of these serotypes will reduce. This is due to the fact that vaccination of this group will prevent the transmission of these serotypes to adults in the vaccinated population thus preventing VT pneumococcal disease in adults. Herd immunity effects were observed in the USA following the introduction of PCV-7, with reductions of between 8 and 32% observed in the incidence of IPD in unvaccinated age groups (Whitney et al. 2003). It is still too early to determine whether or not the UK will experience the same herd immunity impact.

1.7 Concerns about long term vaccine efficacy

There are potential problems associated with the use of pneumococcal vaccines that may prevent them from having long term efficacy. These problems are attributed to the fact that only a limited number of the total number of pneumococcal serotypes in existence may be contained in a vaccine. In Brueggemann et al. (2004) it is stated that the selection of serotypes for the vaccine was determined by assessing the ranking of the IPD serotypes. However, Brueggemann et al. relate their concerns about this method of selection by stating that the prevalence of a serotype in disease may not truly reflect the invasiveness of that serotype, with serotypes ranking highly in IPD perhaps only reflecting the fact that these serotypes are very common in colonisation in the population and occasionally cause disease, rather than the fact that they are the most invasive serotypes. This is of concern if truly invasive serotypes begin to thrive and become more prevalent following the introduction of conjugate vaccines.

1.7.1 Serotype replacement

A key problem which may impact upon the effectiveness of conjugate vaccines is that of serotype replacement, where non-vaccine type (NVT) serotypes replace the niche previously filled by the VT serotypes (Lipsitch 1999). Serotype replacement can occur in two possible ways; there could be an increase in the number of serotypes that had already been present in the population prior to the introduction of the vaccine, or serotypes that were previously absent from the population prior to vaccination as they were unable to compete with VT serotypes may begin to appear once the VT serotypes are eliminated from carriage (Spratt and Greenwood 2000). This is a problem for long term vaccine efficacy as it may lead to an increase in NVT invasive disease and infection.

Serotype replacement has been observed following routine implementation of PCV-7. Following the introduction of PCV-7 in Alaska, high vaccine uptake rates resulted in elimination of almost all IPD caused by PCV-7 serotypes. However, serotype replacement was observed to occur, with increases in serotype 19A

reported (Singleton et al. 2007). Increases in serotype 19A IPD have been observed following PCV-7 use in other states of the USA (Beall et al. 2006). In addition, reports have shown the appearance of serotypes 1, 3, 5, 6A and 7 in IPD following vaccine use, most of which are included in the recently developed conjugate vaccines (Albrich et al. 2007). Other reports have documented increases in serogroups 15 and 33 following PCV-7 use (Gonzalez et al. 2006; Kaplan et al. 2004; Schutze et al. 2004).

1.7.2 Capsular switching

Another potential problem that may arise due to the use of PCVs is that capsular switching may occur. Capsular switching is where an organism that is expressing one particular serotype is able to express another serotype through large recombinational exchanges with a different donor serotype (Coffey et al. 1998). It is likely that in order for capsular switch to take place, an individual must become dually colonised with two different pneumococcal serotypes (Brugger et al. 2009). Brugger et al. state that not enough is known about co-colonisation and its effect on disease due to the inability of current techniques for typing isolates to identify more than one type, and often only the most abundant type within a sample will be identified.

It is believed that the expansion of 19A disease following the introduction of PCV-7, mentioned previously, may in fact be partly attributable to a capsular switch event (Brueggemann et al. 2007). In this article by Brueggemann et al., it is stated that the 19A strains identified in disease following vaccine use were found with a MLST that was previously identified as only associated with serotype 4 which is found in PCV-7.

1.8 Thesis outline

The central theme of this thesis is PCV-7, with emphasis on its potential long term effectiveness. Mathematical, statistical and economic techniques involved in the development and assessment of pneumococcal vaccines will be explored.

The first component of the thesis explores the long term effect of a vaccine which prevents carriage of a pneumococcal serotype. This part addresses the issue, described previously, of MLSTs which are able to manifest as more than one serotype, where one is included in a vaccine and the other not, and uses deterministic modelling to determine the impact of vaccine use. It is important to consider MLSTs in assessing the impact of vaccination as capsular switch events have been documented where pneumococcal MLSTs are able, through genetic transformation, to manifest as a different serotype. Thus, if the MLST is important in determining the disease potential of the bacterium and MLSTs found associated with VT serotypes become more commonly associated with NVT serotypes this may reduce the efficacy of the vaccine over time.

As well as assessment of the effectiveness of a pneumococcal vaccine in preventing disease and infection, it is important to assess the cost of introducing such an intervention to the childhood immunisation schedule. The next section of work, Chapters 4 and 5, was carried out in collaboration with the health economic team at the UK branch of Wyeth pharmaceuticals and begins with a general introduction to the economic analysis of a health care intervention, with emphasis on the use of economic analyses in the assessment of PCV-7. This is followed by an analysis of trends in hospital episodes of IPD in England and Wales prior to the introduction of PCV-7. Through the comparison of the predicted numbers of hospital episodes obtained using these models, which assume no intervention, to the true number of episodes following the introduction of PCV-7, updated figures on the efficacy of the vaccine may be used in the existing model for the assessment of the cost-effectiveness of PCV-7.

The development of PCV-7 was based upon the prevalence of serotypes involved in disease. Therefore, it is important to continually monitor the serotypes found in IPD. PCV-7 was not introduced in the UK until 2006 but the formulation of the vaccine was decided upon prior to 2000. In Chapters 6 and 7 of the thesis, the serotypes found in IPD in Scotland prior to the introduction of PCV-7 are examined. Chapter 6 focusses on assessing trends in the serotypes found in disease to determine whether or not any serotypes are becoming more prevalent in IPD. In Chapter 7, associations between serotypes causing IPD and 30 day

mortality are examined to determine whether or not disease from certain serotypes is linked to an increased risk of fatality. Analyses such as these are important in determining the impact of vaccines in preventing disease and in the creation of future interventions.

A vaccine will only provide protection to a community through herd immunity should high levels of uptake occur. Thus, it is crucial to identify population characteristics which may prevent the receipt of vaccination, or which will cause the vaccine to be administered later than scheduled. The final chapter of analysis in the thesis looks at the routine uptake of PCV-7 in Scotland. Individual and area level variables are used to describe the uptake and timing of the vaccine in univariate and multivariate response models to identify key factors which determine whether or not a child will receive the vaccine on time or at all.

Chapter 2

Modelling pneumococcal carriage in children

2.1 Introduction

In this chapter, the theory behind the mathematical modelling of infectious diseases will be discussed and existing mathematical models of the carriage and transmission of *S. pneumoniae* will be communicated. Following this, ordinary differential equation models of pneumococcal carriage in children of age to receive PCV-7 routinely will be considered. These models explore the relationship between pneumococcal serotypes and MLSTs. The results for three different mathematical models of MLST carriage will be presented. The model discussed in this chapter addresses the carriage of only one MLST whilst the two subsequent models related in Chapter 3 will consider carriage of one or other of two MLSTs.

2.2 Mathematical modelling of infectious diseases

Mathematical modelling of infectious diseases began in 1760 when Daniel Bernoulli developed epidemic models to determine whether inoculation of healthy individuals with smallpox was an effective means of preventing the spread of the disease (Bernoulli 1760). Bernoulli was the first to represent the proportion of healthy

individuals that are susceptible to an infectious disease in terms of the force of infection and life expectancy (Dietz and Heesterbeek 2000).

Deterministic epidemic modelling began to be commonly used in the 20th century, with mathematicians such as Ross, Kermack and McKendrick contributing significantly to this field. Prior to the 20th century, a fundamental result was determined by Hamer who established that the progression of an epidemic is dependent upon the number of susceptible individuals in a population and the rate at which infectious individuals and susceptible individuals come into contact with one another (Hamer 1906). Early in the 20th century, Ross developed a basic deterministic epidemic model in which differential equations are used to describe changes in the number of susceptible and infectious hosts, as well as the total number of hosts in the population, over time (Ross 1916). Deterministic models provide reasonable approximations to the changes in the number of susceptible and infectious hosts over time when the numbers of both types of host are large. This basic model may be simply extended to consider other features of the disease under study.

2.2.1 Kermack and McKendrick model

In 1927, Kermack and McKendrick extended the basic model of Ross to attempt to represent the changes in the number of infected individuals observed in epidemics such as the plague and cholera (Kermack and McKendrick 1927). The Kermack and McKendrick model retains the basic structure of the model by Ross, with non-linear ordinary differential equations used to describe the rate of change of the number of susceptible (S) and infectious (I) hosts. However, a third class of host is considered in this model for recovered hosts (R). Recovered hosts are those individuals who recovered from the infection and developed an immunity and thus do not return to the susceptible class. The non-linear equations that correspond to this model may be described as follows:

$$\frac{dS}{dt} = -\beta SI,$$

$$\frac{dI}{dt} = \beta SI - \gamma I,$$

and

$$\frac{dR}{dt} = \gamma I.$$

In the model notation, β is the rate of infection and γ is the recovery rate. This model describes the changes in a closed population over time, as no births or deaths are considered (Mollison 1995). The Kermack and McKendrick model assumes that there is an instantaneous incubation period for the infection and that the population is homogeneously mixed.

2.2.2 Threshold theorem

The key result in determining what happens to the number of susceptible and infectious hosts in the population over time is the Threshold Theorem. This theorem states that an epidemic cannot occur on introduction of a small number of infectious hosts to a population if the number of susceptible hosts is beneath some critical value. However, if the number of susceptible hosts is above this particular value then an epidemic will take place. This epidemic would reduce the number of susceptible hosts to a level as far beneath the critical value as the number of susceptible hosts originally was above this value (Bailey 1957). The rate at which susceptible hosts become infectious hosts (βSI) must be greater than the rate at which infectious hosts recover from the infection (γI).

2.2.3 Basic reproductive number

An important quantity to consider in Susceptible-Infectious (SI) models or Susceptible-Infectious-Recovered (SIR) models is the basic reproductive number, R_0. The basic reproductive number is "the average number of secondary cases produced by a 'typical' infected (assumed infectious) individual during his/her entire life as infectious (infectious period) when introduced in a population of susceptibles" (Diekmann and Heesterbeek 2000). The basic reproductive number

is often considered as a threshold value that may be used to determine when an infection is able to persist in a population. In most deterministic epidemic models, an infection will only persist in a completely susceptible population if and only if the basic reproductive number is greater than 1 (Hethcote 2000).

2.3 Modelling of *S. pneumoniae*

In this section existing models of pneumococcal carriage and disease will be described. In total, twelve models will be discussed. The key features of these models will be described and the main conclusions and limitations will be conveyed. The review is structured in three sections. The first section looks at models which feature specific serotypes involved in colonisation, such as VT or NVT serotypes. The second section focusses on models which have been created to explore the problem of antibiotic resistance. The final section looks at the problem of capsular switch.

An overview of the main themes of the models under consideration is shown in Table 2.1. A distinction is made between those mathematical models which involve an analytical approach, those which involve a numerical approach and those which involve some type of probabilistic simulation or parameter estimation.

An approach is defined as analytical if the authors have adopted analytical mathematical techniques, in which no parameter estimation is required, to obtain the general properties of a system of equations, such as the basic reproductive number, the equilibrium solutions and stability analyses. This is the approach adopted in the models created and developed in this chapter and the subsequent chapter.

A numerical approach may provide a more realistic impression of what is occurring in a population as parameter estimates for the system under consideration are input into a mathematical model in an attempt to determine what happens to the population sizes over time. The drawback to this type of approach is that it is difficult to establish whether or not all possible solutions have been obtained as the solutions are only applicable to the particular parameter estimates used.

As with the numerical approach, simulations may be used as an alternative, or in

addition, to analytical solutions to establish what happens to the different classes of host over time. Often when it is not possible to obtain analytical solutions to complex mathematical systems simulations are used to obtain modelling results, again once suitable parameter estimates have been identified.

Both the numerical and simulation approach rely on parameter estimates to obtain solutions. Thus, the solutions are only as reliable as the specific choices of parameter value used in these models, highlighting the importance of obtaining reliable, informative data. Other publications considered in this review involve the use of statistical methods to obtain parameter estimates to use in the mathematical modelling.

2.3.1 Serotype colonisation

A key deterministic model of the carriage of *S. pneumoniae*, and the earliest published pneumococcal model described here, often referred to by others studying pneumococcal carriage and transmission, is the model created by Marc Lipsitch, shown in Figure 2.1. The Lipsitch model considers a conjugate vaccine which can prevent carriage, either full or partial, of the VT pneumococcal serotypes and focusses on the possibility of resulting serotype replacement in carriage, as discussed in the introductory chapter of this thesis.

In the two serotype model shown in Figure 2.1, X is used to represent hosts susceptible to pneumococcal carriage, Y_1 and Y_2 are those hosts carrying either serotype 1 or serotype 2 and Y_{12} represents coexistence of these two serotypes within a host. In this model, and subsequent models discussed in this chapter, serotypes 1 and 2 do not refer specifically to the actual serotypes '1' and '2' in existence in the population. Instead, the values 1 and 2 are used as identifiers for any serotype in the population, where 1 is assumed to be present in a vaccine which will completely prevent carriage of this type and the vaccine may or may not elicit protection against 2. f is the fraction of hosts who enter the susceptible vaccinated class rather than the susceptible unvaccinated class.

Lipsitch considers three scenarios in the modelling discussed in this paper. The first is that the vaccine completely prevents carriage of serotype 2, the second is

Author	Ana[a]	Num	Sim	PE	Serotype	AR[b]	CS[c]	Vaccine	Carriage	Disease
Lipsitch 1997	✓				✓			✓	✓	
Lipsitch 2001		✓				✓			✓	
Temime et al. 2003		✓	✓			✓			✓	
McCormick et al. 2003		✓		✓			✓			✓
Temime et al. 2004		✓				✓		✓	✓	
Huang et al. 2005a	✓								✓	
Temime et al. 2005		✓				✓		✓	✓	
Andersson et al. 2005	✓		✓						✓	✓
Wang et Lipsitch 2006	✓	✓				✓			✓	
Temime et al. 2008	✓	✓			✓		✓	✓	✓	✓
Sutton et al. 2008		✓	✓	✓				✓	✓	✓
Lipsitch et al. 2009	✓	✓			✓				✓	

Table 2.1: Key features of mathematical models of pneumococcal carriage and disease.

[a]Ana=Analytical; Num=Numerical; Sim=Simulation; PE=Parameter estimation.
[b]ABR=Antibiotic resistance.
[c]CS=Capsular switch.

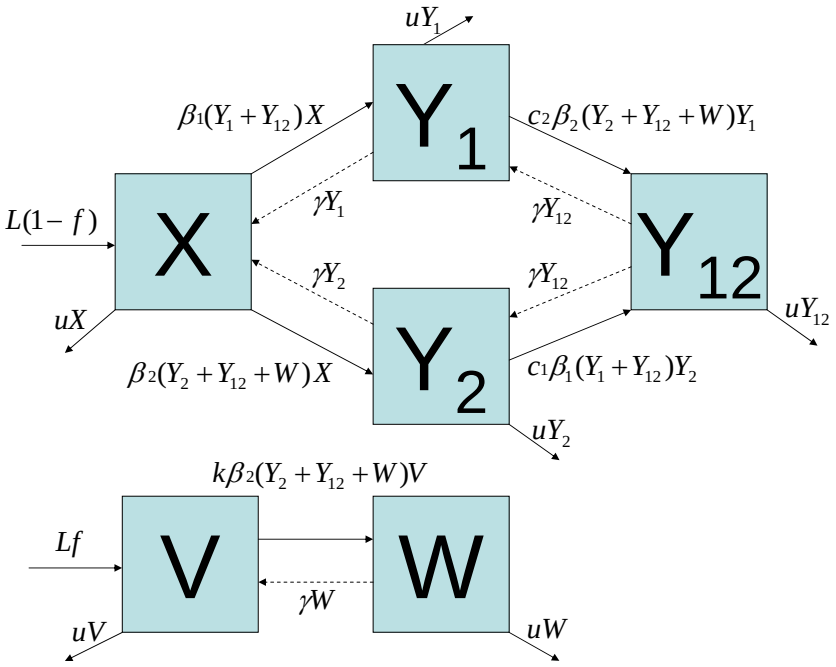

Figure 2.1: Model of carriage of two pneumococcal serotypes, vaccine completely effective against only one serotype (adapted from Lipsitch 1997).

that the vaccine partially prevents carriage of serotype 2 and, finally, the third is that the vaccine does not prevent carriage of serotype 2 at all. Vaccinated individuals are class V in the model and those vaccinated hosts carrying serotype 2 are class W. L is used to represent the constant rate at which susceptible hosts enter the population and u represents the per capita rate at which hosts leave the population. The total rate at which individuals leave a class is u multiplied by the number of individuals in that class. γ is the per capita rate at which hosts independently leave a carrying class to return to a susceptible class once again. The transitions between one class to another occur according to mass action processes, $\beta_1(Y_1+Y_{12})X$ and $\beta_2(Y_2+Y_{12}+W)X$ for unvaccinated hosts starting to carry serotypes 1 and 2 respectively; $k\beta_2(Y_2+Y_{12}+W)V$ for vaccinated hosts acquiring serotype 2. It is assumed that susceptible vaccinated hosts may go on to carry serotype 2 at a different rate to those susceptible unvaccinated hosts. In this model, $1-k$ is the fractional decrease in the rate at which a host will go on to carry serotype 2 if the host originated in the susceptible vaccinated class rather than the unvaccinated class. If $k=1$ in the model, the assumption is that

the vaccine offers complete prevention of carriage of serotype 1 but no prevention of serotype 2 carriage.

Lipsitch assumes that once a serotype is acquired it is more difficult to acquire a second serotype to progress to the dual colonised class, Y_{12}, than it is for a susceptible individual to acquire this other serotype. Thus, the parameters c_1 and c_2, both assumed to take values between 0 and 1, are entered in the model. These parameters account for the relative risk of the acquisition of a second serotype and may differ according to which serotype is first carried. Hosts are assumed to be unable to progress straight from the susceptible unvaccinated class to the dual carriage class. If $c_1 = c_2 = 0$, it is not possible for the two serotypes to coexist in the same host.

Lipsitch's main conclusion from this analysis is that if the vaccine specifically targets only one of the two serotypes in the population then the reduction in carriage of the VT serotype will result in an increase in carriage of the NVT serotype. However, it is surmised that the increase in carriage of the NVT serotype will be smaller than the size of the reduction of VT carriage. Thus, in this scenario, the vaccine will still ultimately be beneficial as overall carriage will be reduced. Unfortunately, this result is limited to the two serotype model. In situations where two or more serotypes are considered in the population and only one is a VT serotype, one of the NVTs may be able to replace carriage of the VT serotype.

In later work, not exclusively specific to *S. pneumoniae*, Lipsitch et al. revisit the issue of coexistence of serotypes in a population (Lipsitch et al. 2009). This paper stresses the importance of correctly modelling the possibility of a host being able to become simultaneously invaded with more than one strain, taking into account difficulties in obtaining a second strain if already colonised and considering acquired immunity of particular strains. Lipsitch et al. stress that models often involve the possibility of coexistence of strains without assumptions regarding the likelihood of such a situation arising being made clear. Coexistence often arises in such models due to the model structure rather than because any clear insight has been made about how coexistence should occur. In this paper, Lipsitch et al. present the notion of a neutral null model which requires two

criteria to be met in order to allow strain coexistence of two identical strains. For the first of these, referred to as "ecological neutrality", Lipsitch et al. state that the number of individuals in each class (the susceptible class and the classes carrying zero, one or more strains) should not depend on the particular strain type but should be dependent upon the state variables in the system, assuming that the strains are indistinguishable. The second condition, "population genetic neutrality", refers to the population equilibria in such a system of equations. This criteria alludes to the fact that no stable equilibria for the infective strain types should be possible in a model of this sort. Instead, it should be possible through the choice of initial conditions in the model to ensure that the prevalence of the strains is fixed constant for time $t \geqslant 0$.

In exploring the two requirements for a neutral null model, Lipsitch et al. conclude that the two serotype model considered in Lipsitch's 1997 paper does not meet these criteria. Not only can the model not be expressed independently of the type of serotype considered but the model also has an equilibrium solution when the two serotypes are assumed to be indistinguishable. In the 1997 model, Lipsitch considers only the possibility that hosts who have acquired one serotype can either cease to carry that serotype or go on to carry a second, different serotype. The possibility that an individual could be dually colonised with the same bacterium type is not considered. By approaching the modelling in this way, Lipsitch admits that one serotype will aid the prevalence of the other serotype in the population. This is due to the fact that for a set frequency of carriage of one of the serotypes, both serotypes will be able to become more commonly transmitted as the proportion of individuals carrying the other serotype increases. Thus, this model allows for the possibility of coexistence through the structure alone, rather than by consideration of the mechanism required for coexistence to occur in the population.

Lipsitch et al. adapt the two serotype model created in 1997 to obtain a model which will meet the criteria. This new model is shown in Figure 2.2. It can be noted that, unlike the 1997 Lipsitch model, hosts colonised with both serotypes 1 and 2 can return directly to the susceptible class without the necessity of losing one of the serotypes first. In this model, $\lambda_1 = \beta_1(Y_1+qY_{12})$ and $\lambda_2 = \beta_2(Y_2+qY_{12})$.

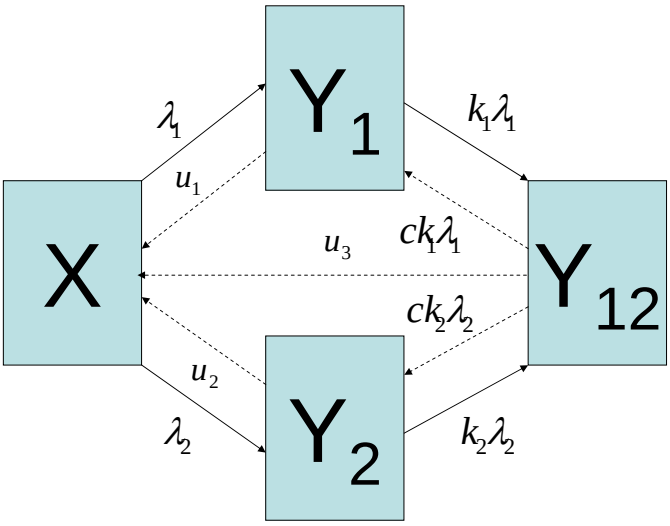

Figure 2.2: Altered model of carriage of two pneumococcal serotypes to account for coexistence criteria (adapted from Lipsitch et al. 2009).

q in the 1997 model was equal to 1 meaning those individuals carrying both serotypes are equally able to transmit these serotypes to others in the population as those carrying each of them individually. This model assumes that hosts that have acquired serotype 1 can go on to carry serotype 2 simultaneously at a rate $\lambda_1 k_1$. Similarly, for those who first acquire serotype 2, the rate is $\lambda_2 k_2$. This new model also addresses competing serotypes with the possibility that those in the Y_{12} class may return to a class where only one serotype is carried due to contact with this serotype. For example, if a host carrying serotypes 1 and 2 comes into contact with serotype 1 then this serotype could potentially eradicate the carriage of serotype 2, returning the host to a state of only serotype 1 carriage. This is assumed to take place at a rate c times as large as the rate of secondary carriage for a host who is only colonised by one of the bacterial types. In this attempt to obtain a neutral null model for the 1997 two serotype model, Lipsitch et al. state that the new model still does not eliminate the problem of expressing the model in terms of different serotypes. However, if $c = q = \frac{1}{2}$, the model will meet the necessary requirements.

In concluding discussions of the 1997 model in the 2009 paper, Lipsitch et al. state that this model is still valid for consideration as it is known that coexistence of serotypes occurs within individuals in the population. Thus, coexistence would inevitably have had to be built into a model. In addition, the model was created to observe what may happen in the population should only one of the serotypes be targeted by an intervention so the serotypes are not assumed to be indistinguishable.

Sutton et al. expand upon the ideas of the 1997 Lipsitch model to consider not only those colonised by *S. pneumoniae* but those who continue to develop pneumococcal infection following colonisation (Sutton et al. 2008). A diagram of this model is shown in Figure 2.3. In this figure, X and V represent susceptible hosts, unvaccinated and vaccinated, as in the 1997 Lipsitch model. C_X and C_V represent unvaccinated and vaccinated carriers and I_X and I_V are those hosts who go on to develop infections following carriage. Infectious hosts are assumed to return to the susceptible class at a per capita rate γ. Carrying hosts are assumed to cease carriage at a per capita rate α. Two per capita death rates are assumed in this model: one is η representing death due to infection, one is μ for natural death. Hosts are assumed to enter the susceptible, unvaccinated population at birth at a constant rate, λ. β is the per capita contact rate with a contact between a susceptible and a colonised host resulting in the susceptible host becoming colonised so that susceptible hosts become colonised at total rate $\frac{\beta X(C_X+C_V+I_X+I_V)}{N}$. Hosts can become colonised through contact with carriers or infected individuals. α is the per capita rate at which hosts cease to be colonised. Other parameters in the model are ρ, ϕ, ϵ, δ and l representing the rate of the loss of vaccine induced protection, rate of vaccine effectiveness, reductions in carriage due to vaccine, reduction in disease due to vaccine, and the proportion of susceptible hosts who are at risk. The parameter κ represents the rate of infection for those colonised hosts who are at risk of developing infection. In addition, Sutton et al. include parameters to take into account average infection rates and seasonality of infection.

Using Australian surveillance data, Sutton et al. use this model in conjunction with statistical methods, assuming model errors, to investigate vaccination effects

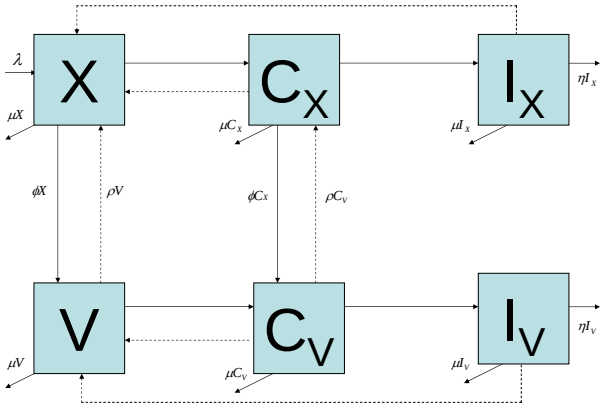

Figure 2.3: Model of pneumococcal colonisation and infection incorporating a vaccine effect (adapted from Sutton et al. 2008), with $\omega = \frac{C_X+C_V+I_X+I_V}{N}$.

in the population. Model results imply that reductions observed in pneumococcal infections are attributable to the conjugate vaccine preventing infection but also due to the reduction in the ability of pneumococci to colonise. In addition, a herd immunity effect is suggested in the model. However, Sutton et al. discuss the fact that although the model adequately reflects the disease picture for the earliest year data are available, 2005, it underestimates the burden of disease for 2006 and 2007. Thus, it appears that perhaps alterations need to be made to models to reflect changes in vaccine effect over time, changing host susceptibility to infection and, also, variations in circulating pneumococci. The authors conclude by stating that wide spread PCV-7 immunisation is likely to have a dramatic effect in the epidemiology of *S. pneumoniae* which will require close scrutiny in years following routine implementation.

2.3.2 Antibiotic treatment and resistance

The next series of models described address the issue of antibiotic resistance. The earliest published resistance model considered in this section was created by Lipsitch (Lipsitch 2001). This mathematical model was created to investigate how antibiotic resistance is related to previous antibiotic use.

The model shown in Figure 2.4 is used to describe a trial in which patients are

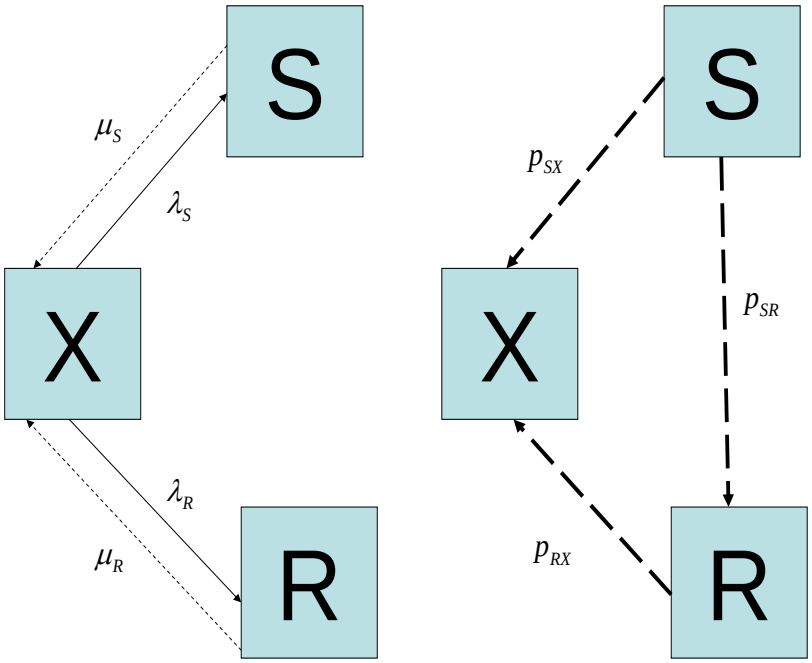

Figure 2.4: Model of penicillin resistance in sample population carrying both susceptible and resistant *S. pneumoniae* (adapted from Lipsitch 2001).

administered antibiotics and relates to a closed population in which no hosts may enter or leave the trial. In Figure 2.4, the left hand group of X, S and R hosts refer to those who have not received antibiotics. X represents hosts susceptible to carriage, S represents those carrying antibiotic susceptible (AS) serotypes and R represents those carrying resistant strains. λ_R and λ_S represent the rates at which carriage susceptible hosts may become carriers of antibiotic resistant and susceptible strains respectively. μ_R and μ_S are the rates at which carrying hosts cease to carry each of these type of pneumococcal strains. The groups on the right-hand side of Figure 2.4 represent the sample population immediately following treatment, where hosts with resistant strains are assumed to become decolonised during treatment with per capita probability p_{RX}; those hosts carrying AS strains can become decolonised during treatment with probability p_{SX}. p_{SR} is the per capita probability that hosts carrying AS serotypes immediately before treatment will become primarily colonised with resistant serotypes immediately after treatment.

The model shown in Figure 2.4 differs from other models of pneumococcal carriage considered in that the purpose of the model is to look at carriage in a sample of individuals rather than in the population. This sample is assumed to be a trial sample of patients being followed to assess the effect of antibiotics on carriage. The aim of the modelling is to determine how the chosen measure of the link between antibiotic use and resistance alters between the time of antibiotic treatment to the time at which a subject has a sample taken. Therefore, time is considered in the model with $t = 0$ assumed to be the point at which a subject is administered antibiotics.

Lipsitch discusses the fact that antibiotic treatment reduces the R_0 of AS serotypes and hosts carrying this type of serotype who receive treatment will have either a reduced infectiousness or reduced duration of carriage, or both. However, this will result in an increase in AR strains in the population. If antibiotic treatment is administered fairly often then this will result in an increase in antibiotic resistant serotypes.

The first of four pneumococcal models, three of which consider antibiotic resis-

tance, involving Temime is shown in Figure 2.5 (Temime et al. 2003). In this work, Temime et al. consider both *S. pneumoniae* and *N. meningitidis*, bacteria with similar antibiotic resistance mechanisms. The model shown in Figure 2.5 is used to assess the process of antibiotic resistance in a population. Hosts can be in one of two classes: those susceptible to carriage, i.e. uncolonised, and those carrying, colonised. Vaccine is not considered in this model. Instead, as interest is in antibiotic resistance, treatment with penicillin is assumed. σ is the probability of loss of the bacterium following antibiotic treatment. If the bacterium is not lost, replacement may occur with a bacterium with a higher minimum inhibitory concentration (MIC) replacing the original serotype. Temime et al. explain the assumption they have made that serotype replacement is only possible in those individuals administered treatment by stating that, although it is known that genetic events are plausible which could cause replacement, the competition that arises due to the fact that large numbers of bacteria colonise a host means that it is very unlikely that the natural bacterial population within a host will be replaced by genetically altered bacterial types without some intervention.

$\frac{1}{\lambda}$ is the mean duration of carriage in the absence of antibiotic treatment, irrespective of the MIC of the serotype, and β is the per capita infectious contact rate between susceptible and carrying hosts. Once a carrying host ceases to carry the serotype, a period in which the host cannot become a carrier again is assumed. This lasts for a mean duration of $\frac{1}{\theta}$. The model parameter α represents the frequency at which hosts move from the untreated class to the treated class at the onset of treatment with penicillin and the average duration of treatment is assumed to be $\frac{1}{\gamma}$.

From literature, Temime et al. obtained an estimate of the average pneumococcal carriage duration of 2.2 months. However, it is noted that the period in which a host cannot be colonised after ceasing to carry a serotype is more difficult to determine and thus an estimate of 2 weeks was chosen and varied between 4 days to 2 months. From French data, Temime et al. were able to estimate the average duration of penicillin treatment to be 8 days. Differences in carriage and the number of penicillin treatments were observed for different age groups and an attempt to reflect this information in the model was made, even though

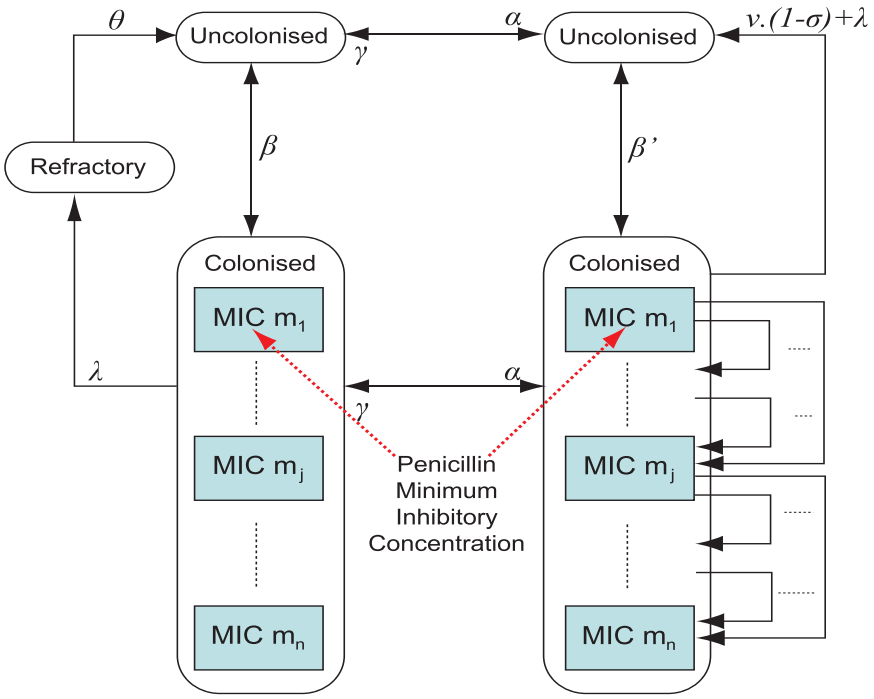

Figure 2.5: Compartmental model of penicillin resistance in carriage of both *S. pneumoniae* and *N. meningitidis* (adapted from Temime et al. 2003).

the model was not specifically structured by age group. This was carried out by weighting the number of antibiotic treatments using the probabilities of pneumococcal colonisation by age. Antibiotics of the same concentration were assumed to be administered, regardless of the MIC of the pneumococcal strain. Thus, the probability of the antibiotic being unable to prevent carriage is assumed to be described as a function of MIC, m, only. The formula for the probability is $\sigma_m = \frac{m^3}{0.05+m^3}$. β was assumed constant for those untreated with antibiotics and was taken to be 0.23 per week per person so that the proportion of hosts carrying serotypes reflects the 45% reported in the literature. The model assumes that in the presence of treatment, a susceptible host has a higher chance of becoming a carrier once in contact with a carrying host if the bacterium involved had a high MIC. Colonisation was assumed to be less likely when a AS bacterium is involved. Thus, β' is defined where $\beta' = \frac{2\beta m^3}{0.5+m^3}$ so that $\beta' < \beta$ when $m = 0$ but $\beta' > \beta$ when m is large.

The conclusions of the model regarding penicillin resistance were that the MIC of the strain is important. If a pneumococcal strain only has low resistance then it is likely that it can become eradicated from the population before it is able to increase MIC through genetic events. However, if a strain appears with high MIC then it should be able to remain in the population and will lead to the introduction of strains with even higher MIC levels. One of the main outcomes from the modelling was the conclusion that there is a great deal of variation in the time to selection of bacteria of given resistance levels. The model appears to adequately predict the bimodal distribution in resistance levels that appears in the French data. The critical parameters in modelling the continuation of resistance in the population were identified as: how often penicillin is administered, how long a strain may be carried for, how long a host is administered penicillin and the rate at which uncolonised hosts come into contact with colonised hosts. Varying the time at which an individual cannot become colonised after ceasing to carry a strain did not appear significant in determining the model outcome.

The limitations of the Temime et al. model appear to be that although penicillin intervention was considered, the effect of a vaccine intervention was not. In addition, Temime et al. did not consider the possibility of coexistence of strains

within a population or individual. The model only focusses on one serotype and considers varying degrees of resistance. It does not allow for variation in transmission or carriage duration for different pneumococcal serotypes.

In the same year as this paper was published, McCormick et al. published work on a model which also considered antibiotic resistance of pneumococcal serotypes (McCormick et al. 2003), shown in Figure 2.6. Unlike the 2003 Termime et al. model, McCormick et al. consider not only penicillin resistance but also erythromycin resistance. In Figure 2.6, X represents hosts susceptible to pneumococcal colonisation; S represents hosts with strains susceptible to both penicillin and erythromycin; P is the class of hosts with pneumococci resistant to penicillin; E is the erythromycin resistant pneumococci carrying class; D represents hosts with strains resistant to both penicillin and erythromycin. McCormick et al. are interested in causes of geographical variation in pneumococci resistant to these antimicrobial agents. The authors believe that there are various possible causes of geographical variation in resistance; one of the causes could be that in areas where higher selection pressure is exerted, i.e. areas which have higher antimicrobial use, more resistant serotypes will be present, or in areas with greater antibiotic use a higher proportion of resistance may be present within serotypes. A further possibility is simply that resistant pneumococci are more common in certain areas than in others. The first two cases are similar in that they are related to pressure exerted due to antimicrobial use, whilst the third case differs as it is assumed attributable to natural variation. The first two scenarios may reflect changes in serotype distribution for different areas.

In the modelling, data collected between 1996 and 1999, pre-PCV-7 introduction, by the Active Bacterial Core surveillance (ABCs) sites of the Centres for Disease Control and Prevention in the United States were used. The results suggest that differences in antimicrobial use are responsible for the differences in resistance for different geographical locations. In addition, this variation is mainly due to differences in the resistant proportions in each serotype. A third conclusion is that those serotypes unaffected by both penicillin and erythromycin have a survival advantage compared to those only able to resist one of the antimicrobial agents.

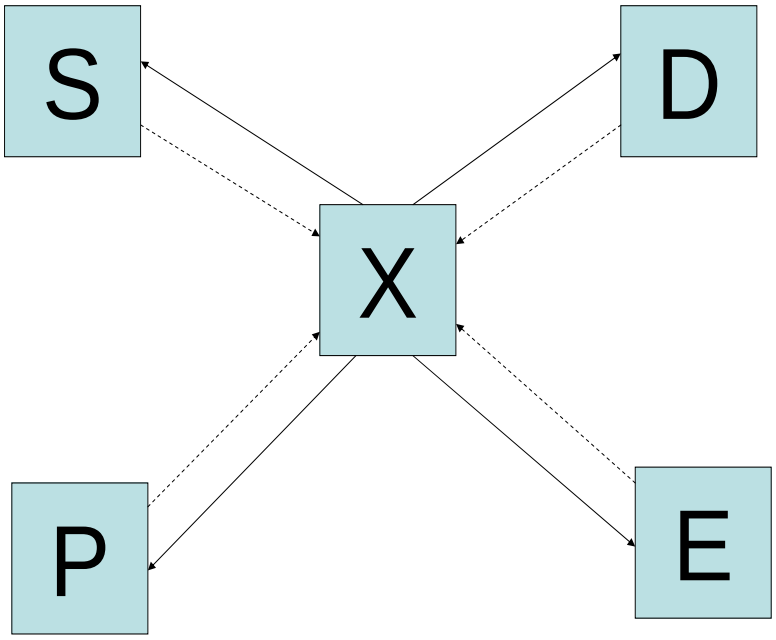

Figure 2.6: Compartmental model of penicillin and erythromycin resistance (adapted from McCormick et al. 2003).

Although the effect of the PCV is not explicit in the model, McCormick et al. do discuss the possible implications of vaccine introduction as this will negate the assumption of constant selection pressures on the pneumococci. The serotypes found in PCV-7 are amongst those with the highest antimicrobial resistance so would be expected to reduce the proportion of resistant strains more than those susceptible to antibiotics. However, the authors state that it is unlikely that vaccine use will alter the resistance within a serotype.

Continuing with models of antibiotic resistance, in subsequent work involving Temime, the 2003 Temime et al. model was extended to consider the vaccine effect (Temime et al. 2004). The compartmental model is shown in Figure 2.7. From assessment of this figure, it can be noted that this model is split into three age classes: those aged under 2 years; the vaccine targeted group; those aged 2 to 15 years, and those aged over 15 years. The class X_{np} represents the number of hosts who are susceptible to carriage who have not received penicillin whilst Y_{np} represents those hosts who are susceptible but have been treated. One of the model assumptions is that there are ten different levels of penicillin resistance so that $X_{V,1}$ and $Y_{V,1}$ are the number of individuals who are respectively untreated and treated with penicillin, are carrying a VT pneumococcal serotype and who have resistance level 1; similarly, $X_{V,2}$ and $Y_{V,2}$ represent those respectively untreated and treated with resistance level 2, and so on. The subscript NV represents hosts carrying NVT serotypes. As a subset of the under 2 years age group are vaccinated, separate additional compartments must be considered for this group. These are V representing those vaccinated who have not received penicillin and W for those vaccinated who have received penicillin. As the model assumes the vaccine is 100% effective in preventing colonisation with VT serotypes, the vaccinated children can only carry NVT strains. This model was created to address two key issues. The first of these involves the question of what happens to the *S. pneumoniae* population epidemiologically after the introduction of the conjugate vaccine. The second concerns penicillin resistance and, in particular, how the distribution of resistance levels changes in children and adults carrying pneumococcus following vaccine introduction.

In the Temime et al. model, hosts are assumed to enter the population at birth

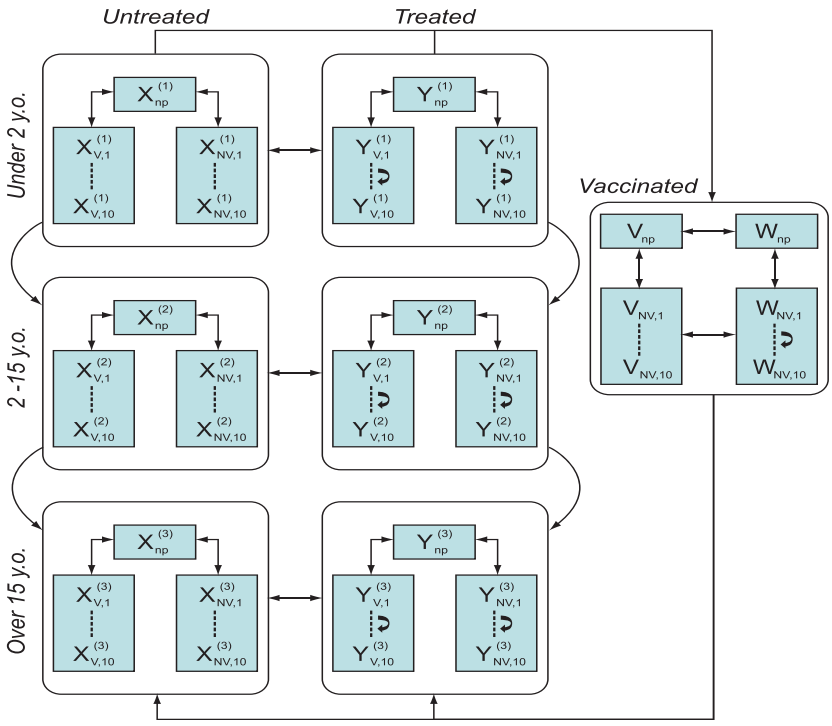

Figure 2.7: Compartmental model of the effect of vaccine on penicillin resistance (adapted from Temime et al. 2004).

at a constant rate μ_N. A proportion ν of hosts aged under 2 years receive the conjugate vaccine and are assumed to have a duration of protection represented by d_ν. This model does not consider coexistence of serotypes in carriage.

This study involved French data from the French Reference Centre for Pneumococci (Centre National de Réference des Pneumocoques) which included information about penicillin resistance for specific serotypes and age groups. The data were used to obtain a general picture about VT and NVT serotypes so that the information could be incorporated in the model.

The results from the model suggest a similar picture to that of the 1997 Lipsitch two serotype model. Temime et al. conclude that an increase in NVT pneumococcal carriage occurs as a result of the decrease in carriage of VT serotypes. However, as in the two serotype Lipsitch model, the increase in carriage of NVTs is not as great as the decrease in VT carriage. Thus, overall carriage in the population is reduced upon introduction of PCV. As for the penicillin resistance considered, results suggest that the replacement of VT carriage with NVT carriage occurs alongside increases in resistance. This means that, as the non-vaccine pneumococci increase in the population, the average resistance of these pneumococci also increases. This implies that penicillin resistance is unaffected by vaccine.

Temime et al. state that the important parameters in the model to determine overall rates of pneumococcal carriage in a population involving vaccinations are the duration of carriage and the rates at which contacts are made between susceptible and carrying hosts. Concerning the proportion of types resistant to penicillin, in addition to the parameters important for determining rates of carriage, the important variable is the duration of exposure to penicillin. Temime et al. admit that there may be some limitations to the model described in Figure 2.7. One point not addressed in the modelling is the possibility that non-vaccine serotypes are less able to colonise in a population than VT serotypes. However, differences between the two classes of serotypes were considered concerning the length of time VT and NVT serotypes are able to colonise. The results suggest that this does not impact on penicillin resistance but it may lessen how quickly serotype replacement occurs if the carriage of NVT strains is less than that of

VT strains.

The third antibiotic resistance model involving Temime focusses on one of the severe outcomes of pneumococcal colonisation, community-acquired bacterial meningitis (Temime et al. 2005). As with the 2004 Temime et al. paper, both antibiotic resistance and vaccine effect are considered in this model. This model has a similar structure to that shown in Figure 2.7 and builds on the initial model of Temime et al. on antibiotic resistance, Figure 2.5. The model is used to investigate how the level of antibiotic use in a population impacts the ability of the vaccine to prevent meningitis. Temime et al. use data from the Tracking Resistance in the United States Today (TRUST) surveillance study in the USA and the French Reference Centre for Pneumococci to investigate the effects of antibiotic use and the conjugate vaccine. The USA was chosen to represent low antibiotic coverage, whilst France represents countries with high antibiotic exposure. Antibiotic exposure with vaccine use was considered in three different ways: one way is that the use of vaccine reduces antibiotic exposure to all children, another is that the exposure is only less for vaccinated children, and the third is that the antibiotic exposure remains unaltered for all children. The authors hypothesise the length of time a vaccine provides immunity to disease to be 13 years. A constant proportion of all hosts carrying pneumococci are assumed to have bacterial meningitis independent of the serotype being carried. Temime et al. varied vaccine use from no coverage to complete coverage in the model.

The model results suggest that the length of time a host carries a pneumococcal serotype is the crucial parameter in determining meningitis incidence from antibiotic resistant strains. Antibiotic exposure also proved to be important, with high levels of antibiotic exposure leading to antibiotic resistant serotypes accounting for almost all of the cases of bacterial meningitis after 20 years of vaccine use, irrespective of the level of vaccine coverage. The length of time the vaccine provides immunity against disease does not appear to be of great importance and no significant difference was found in the number of meningitis occurrences for the three classes of antibiotic exposure that children could encounter.

Limitations to this model are the fact it does not consider coexistence of serotypes

within an individual so the possibility of serotype replacement was not addressed. In addition, the authors state that due to the age classes chosen in the modelling, the level of incidence of bacterial meningitis for those aged under 2 years is underestimated.

A different approach in the consideration of pneumococcal carriage is taken by Huang et al. (Huang et al. 2005a). It is known that young children are the primary carriers of *S. pneumoniae* and that high transmission rates are observed in places where many young children regularly congregate, such as nurseries and child care centres. Thus, the modelling carried out by Huang et al. considers a transmission model which takes into account the risk of higher rates for children who attend child care centres or who often are found to spend time with children who attend these centres. This carriage model does not consider coexistence as children are assumed to be able to carry only one serotype at a time. In addition, the model only considers one hypothetical community of young children and does not allow for new children to enter or existing children to leave this community. A diagram of the compartmental model is shown in Figure 2.8.

In Figure 2.8, X_C and X_N represent the proportion of children susceptible to pneumococcal carriage who do and do not attend child care centres respectively. Similarly, Y_C and Y_N represent the proportion of carriers for the same two groups. In this model, c and r are parameters included in the model to account for antibiotic treatment the children receive. c is the percentage of children for whom carriage is eliminated due to treatment with antibiotics and r is the percentage of children who receive antibiotic treatment each week. The parameters f and g refer to child care centre attendance, with f representing the fraction of all children who attend child care centres and g representing the average hours spent in child care centres. g is divided by 84 as 84 represents the number of hours that a child spends awake per week, assuming that a child is awake 12 hours per day. β_1 and β_2 are constant transmission rates, with β_1 being the rate for carriage transmission outwith child care centres and β_2 the rate within these centres. m is the parameter for weekly automatic clearance of pneumococcal carriage in the children. This model assumes that $\beta_2 > \beta_1$ as the transmission amongst children within a centre is expected to be higher than that outside. It is assumed that

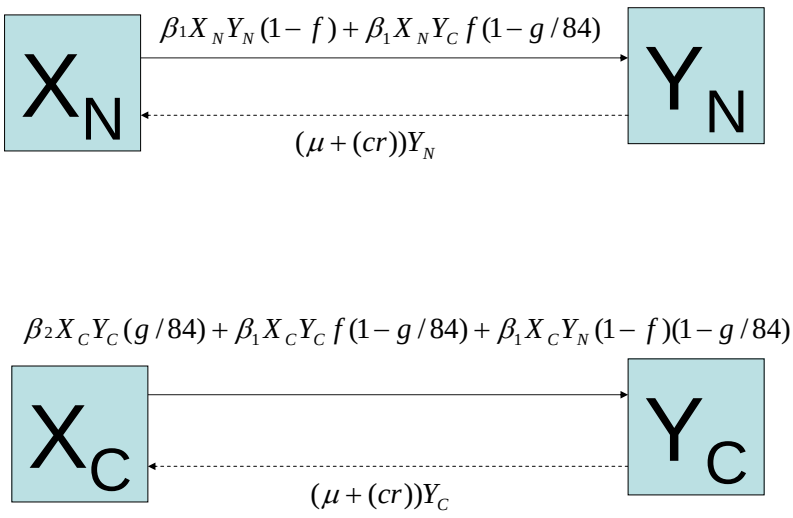

Figure 2.8: Serotype carriage model of children who are attendees and non-attendees of child care centres (adapted from Huang et al. 2005a).

children who do and do not attend centres are able to interact during hours the children are awake and attendees are not at the centres. The model also suggests that while children are attending centres, those who do not attend will not come into contact with as many other children as those within a centre.

The analysis carried out by Huang et al. involved data to obtain parameter estimates for the compartmental model. The data were taken from studies in Massachusetts communities. Obtaining equilibrium solutions, the model was used to estimate the prevalence of carriage, allowing f and g to vary. The results from the modelling stress the importance of child care centres. Conclusions are made that carriage prevalence increases with increasing child care attendance for both groups of children in the population as the model allows interactions between groups as well as within. Huang et al. state that the model considered leads to the impression that centre attendance possibly accounts for between 4% and 56% of the variability in the volume of carriage across communities. Thus, it is important to consider community-level risk factors, such as child care centre

attendance, when considering pneumococcal transmission. However, Huang et al. do admit that other risk factors were not considered in the model and that it was assumed that the carriage duration was fixed. In addition, there are other potential considerations to take into account in the model such as antibiotic resistant (AR) serotypes. Therefore, further work may have to be undertaken to fully understand the effect of child centre attendance, incorporating other risk factors and allowing other parameters in the model to vary.

In Sweden, a mathematical model was created to examine antibiotic resistance (Andersson et al. 2005). As with the Huang et al. model, Andersson et al. consider a model of children attending day-care centres (DCCs) as it is assumed that this type of close contact amongst the group known to be the primary pneumococcal carriers will have the greatest spread of carriage and infection in the population. Unlike the Huang et al. model, Andersson et al. specifically focus on resistance to antibiotics and also use a stochastic approach in the modelling. In addition, Andersson et al. do not consider the transmission of *S. pneumoniae* amongst the population outwith those attending day-care.

As antibiotic resistance was observed to be becoming increasingly more common following studies carried out in 1992, in 1995 an intervention project, the South Swedish Pneumococcal Intervention Project (SSPIP), was introduced to prevent resistant pneumococci from becoming more prevalent in the Malmöhus county in Sweden. This study observed no significant change in antibiotic strains found in this population during the first four years of intervention. However, it is difficult to determine if the intervention has no effect or if the unchanging situation is due to the intervention of the SSPIP on an ever increasing trend of resistant strains. Since no comparable population who had not received the intervention of the SSPIP was available, mathematical models were adopted to try and answer this question. Andersson et al. focus only on children attending day-care as there have been many studies focussing on this group and thus the authors were easily able to obtain appropriate parameter estimates of some of the parameters included in their model.

This model differs from the other models of antibiotic resistance discussed in

that hosts can be found in three classes: those susceptible to carriage or infection, those carrying pneumococcal serotypes and those who have a pneumococcal infection. Other resistance models consider only the possibility that hosts are uncolonised or colonised without the distinction between carriage and infection. All hosts must become carriers of pneumococci before becoming infected following an incubation time T. The three stochastic processes considered in the model are thus $S(t)$, the number of susceptible hosts at time t, $C(t)$, the number of carrying hosts at time t, and $I(t)$, the number of infected hosts at time t, where t is defined as the number of weeks since the introduction of pneumococci in the population. In addition, a fourth process $C(t, s)$ is considered which represents the number of pneumococcal carriers at time t who became carriers at time s. Other processes are used to describe the transitions between the various classes and also the number of children present in the DCC, both uncolonised and colonised, as children may be absent for various reasons.

Andersson et al. model seasonal variation by incorporating a parameter that accounts for the differences in transmission for each week of the year. The probability that a serotype is transmitted from a carrying host to a susceptible host is assumed to be proportional to this transmission parameter for that week. The size of the group in a DCC is also considered as a potentially important factor for transmission.

Figure 2.9 shows two diagrams depicting how the basic reproductive number varies with the size of the DCC and the week of the year, assuming two different incubation periods. Clearly, the size of the centre has an impact as the plots show higher R_0 values for larger groups of children. There is an apparent seasonal effect as R_0 is above the threshold value '1' for weeks in autumn and winter. R_0 appears higher with the shorter incubation period. Seasonality seems to be the most important feature. Week 36, the first week of September, is the point at which maximum transmission is estimated to take place.

The authors hypothesise that the increased transmission in the autumn periods could be attributable to increased contact between day-care attendees as they will be more likely to be kept indoors due to bad weather. In addition, other

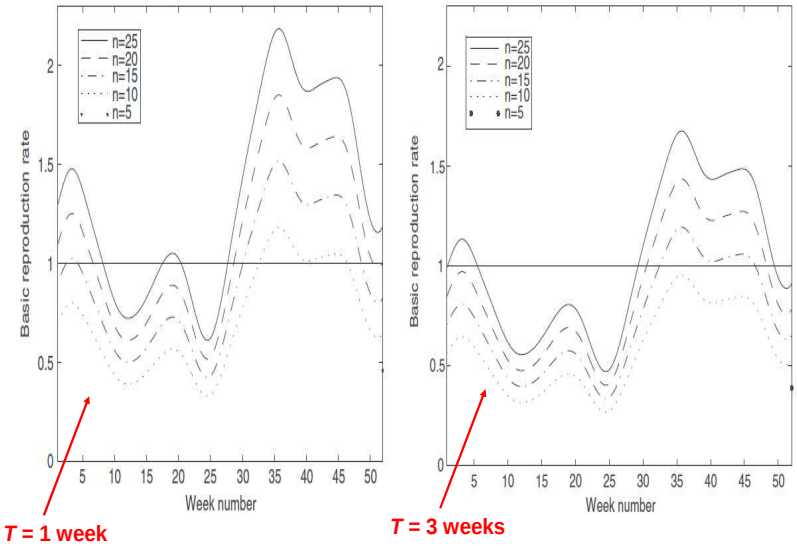

Figure 2.9: Plots of varying R_0 by week and group size (adapted from Andersson et al. 2005).

infections are common in these seasons. Thus, this could increase susceptibility to pneumococcal carriage and infection. As far as the intervention of the SSPIP is concerned, the intervention appears to be more effective in weeks towards the end of a year, in larger DCCs and for shorter incubation periods. Andersson et al. conclude by stating a belief that the intervention should prove effective under appropriate settings but it is difficult to determine whether or not the intervention will provide benefits to the wider population. This is due to the fact it is unknown how much of the transmission can be attributable to children attending DCCs.

Further modelling considering antibiotic resistance was published in 2006 (Wang and Lipsitch 2006). This model differs from other resistance models considered as Wang and Lipsitch look at the replacement of current, possibly ineffective antibiotics with new antibiotics to which serotypes may not have an established resistance. In the modelling, three types of pneumococci are considered: those susceptible to antibiotic treatment, those resistant to the old treatment but susceptible to the new treatment and those with resistance to both treatment types. Wang and Lipsitch are interested in assessing the effect of the timing of switching between treatments for the population.

The results from this model suggest that a compromise has to be met between attempting to have a successful antibiotic treatment and creating an environment in which resistance to the new antibiotic treatment can occur. This trade-off can be affected if there is a difference in the transmission ability of the resistant and susceptible pneumococci in the population. The model shows that if resistant serotypes have low transmissibility and there is a substantial risk that pneumococci will become resistant to the new treatment then an immediate switch to the new treatment will result in low transmission of resistant pneumococci. Early switch also should be adopted in the case where the old, existing antibiotic can cause resistant serotypes of both this treatment and the new antibiotic. Switching treatment should be put off until the existing treatment becomes ineffective if the new antibiotic has high transmissibility in the population but there is a reduced risk of pneumococcal serotypes becoming resistant to this drug. Wang and Lipsitch admit that there are some limitations to the model due to the fact that certain important factors in the spread of the bacterium were not incorporated

such as the age of the population and the effect of vaccine.

2.3.3 Capsular switching

Returning to the issue of the impact of pneumococcal vaccines in modelling without the issue of antibiotic resistance, capsular switching is considered in later modelling by Temime et al. (Temime et al. 2008). Figure 2.10 shows a representation of the model. In this figure, as with the Lipsitch model, Figure 2.1, X represents susceptible unvaccinated hosts, V represents susceptible vaccinated hosts, and Y_X and Y_V respectively represent unvaccinated and vaccinated hosts colonised with a NVT pneumococcal serotype. W_X and W_V are respectively classes of unvaccinated and vaccinated host who are colonised with NVT serotypes following capsular switch and, finally, Z represents hosts colonised with the VT serotype. As with the other colonised host types, there are separate Z classes for the vaccinated and unvaccinated hosts. Dual colonies are shown in the model. This model considers two age groups: those aged less than 2 years and thus targeted by vaccine and those aged over 2 years. The model assumes that a proportion of children are vaccinated, with the potential for immunity to continue to adulthood. The capsular switch which takes place amongst those vaccinated hosts who are colonised with both a vaccine and non-vaccine serotype is shown by the red arrow. Capsular switch due to the introduction of the vaccine is the only type of switch modelled. The natural switch process that occurs is not specifically defined in the model as Temime et al. state that amongst unvaccinated hosts, the switches to and from vaccine strains should be balanced and thus it is unnecessary to have an additional mechanism to account for this switch in the model.

The diagram of the Temime et al. model, Figure 2.10 appears strange in that classes of hosts are defined in which vaccinated hosts carry VT serotypes. This should not be possible since the authors state that the vaccine is supposed to be 100% effective in preventing colonisation by VT serotypes. Thus, in this case, the classes Z_V, YZ_V and ZW_V do not contain any individuals.

A competition effect was considered in the model by assuming that the probability

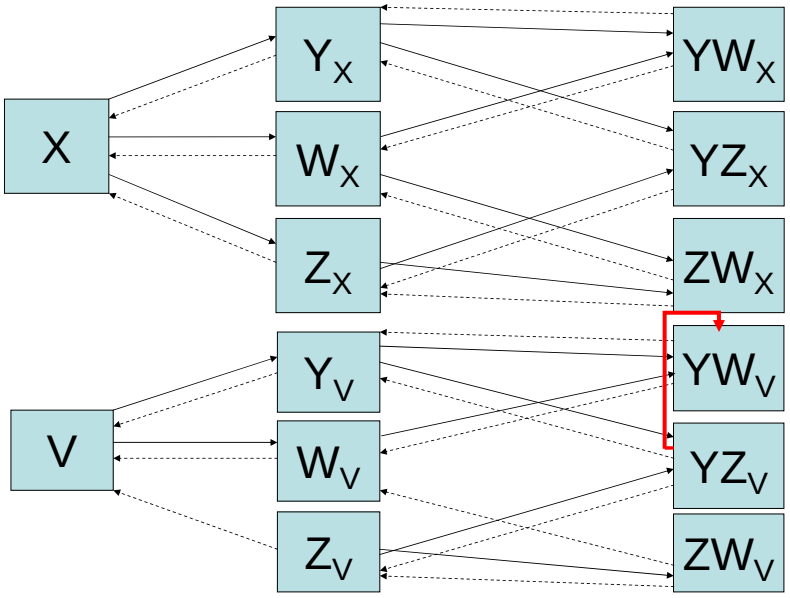

Figure 2.10: Capsular switch serotype carriage model of vaccinated and unvaccinated hosts (adapted from Temime et al. 2008).

of carrying a particular serotype was reduced by half if the host was already carrying one serotype. The model assumes that the average length of time a host carries a serotype is the same for all of the different serotypes.

In this numerical simulation, Temime et al. assume a vaccine coverage rate of 90% and used parameter estimates from a literature search to explore the effect that capsular switch has on the impact of vaccine. Results suggest that capsular switch should not significantly affect the benefits obtained through the use of vaccine, indicating that the reduction in disease incidence should not decrease due to switching. This conclusion is partially attributable to the competition effect considered, as switched serotypes compete with other NVT serotypes. However, the estimates of the occurrence of capsular switch are unreliable due to the lack of relevant data on this type of event. Another potential drawback to this modelling is that the assumption was made that capsular switching has no effect on the virulence of the bacterium, implying that NVT serotypes occurring from a switch will have the same virulence as the VT serotype. This may not be the case as

virulence factors are still being explored by biologists. Temime et al. conclude by stressing the importance of obtaining relevant data from further studies of *S. pneumoniae* in the population.

To summarise, of the twelve published pneumococcal models, six consider the problem of AR pneumococci. These models differ regarding the antibiotics considered to be resistant with some mathematical models created which consider more than one antibiotic treatment (McCormick et al. 2003; Wang and Lipsitch 2006). Some of the models focussed on assessing the effect of different antibiotic exposures on carriage and disease (Lipsitch 2001; McCormick et al. 2003), with McCormick et al. particularly addressing geographical variation in exposure. Only two of the models consider both vaccine and antibiotic resistance (Temime et al. 2004; Temime et al. 2005). In the AR study by McCormick et al. the possible implications of vaccine usage are discussed even though vaccine is not incorporated in the model. McCormick et al. state that the vaccine will have an effect on circulating pneumococcal strains by reducing the proportion of resistant strains since the vaccine includes the common resistant strains. Thus, it does appear from assessment of the various pneumococcal models that both antibiotic resistance and vaccination are important to consider in mathematical models as the vaccine is likely to reduce circulating resistant strains which may impact on the use of antibiotics in a population. However, it is possible that more AR strains may appear due to serotype replacement which can occur due to the use of conjugate vaccines which can only include a small number of the circulating serotypes. The results from the 2004 Temime et al. model emphasise this problem as the NVT serotypes will inevitably become more commonly carried, replacing VT serotypes in carriage in the population and the increase in carriage of NVTs was observed to coincide with increases in resistance.

Some of the models primarily consider future implications of vaccine use on pneumococcal carriage (Lipsitch 1997; Temime et al. 2008), addressing issues such as competition between VT and NVT serotypes (Lipsitch 1997) and capsular switching (Temime et al. 2008). The conclusions seem to suggest that an increase in prevalence of NVT serotypes should be expected after the introduction of vaccine but that the increase should not be as large as the decrease in prevalence of VT

serotypes. In addition, capsular switching should not have a substantial effect on the benefit obtained through the use of vaccines.

Clearly, there are a variety of ways in which mathematical models may be used to aid understanding of pneumococcal carriage and infection and the mechanisms involved in allowing certain types to become prevalent in the population. In addition, mathematical models assist in determining the benefits of various interventions against carriage and infection and in the development of these interventions.

The models described in this section adopt different mathematical and statistical approaches in order to obtain solutions to the biological processes under consideration, see Table 2.3. As mentioned previously, an analytical approach, such as that adopted by Lipsitch in his 1997 paper, is useful in that it enables generalisable results for all manner of parameter estimates. However, when the biological process being modelled, such as the capsular switch process defined by Temime et al. (2008), is complex, and analytical solutions are perhaps not possible to determine, solutions to the system may still be obtained through the use of simulations. Furthermore, statistical regression approaches may be adopted in conjunction with the differential equation mathematical models to aid in parameter estimation for these numerical or statistical simulations, or to obtain model results. For example, multinomial logistic regression is used by McCormick et al. within the framework of a four state transmission model to obtain their results. Thus, there are clearly a variety of approaches to take to obtain solutions to models of this nature.

In the next section, a discussion of the importance of considering pneumococcal MLSTs in addition to pneumococcal serotypes will be presented.

2.4 Multi-locus sequence types in mathematical models of *S. pneumoniae*

Mathematical models of *S. pneumoniae* have been created to consider possible adverse effects attributable to the introduction of PCV. These models, discussed

in the previous section of this chapter, involve concerns such as competition from NVT serotypes which can lead to serotype replacement resulting in an increase in NVT carriage and disease; the possibility of increasing NVT AR strains becoming prevalent and the rare phenomenon of capsular switching. Capsular switching is a process in which, through genetic transformation, pneumococci can exchange capsules (Musher 2006). This means that if genetic elements of the pneumococci, such as the MLST, associated with invasive disease and infection which manifest as VT serotypes become more commonly associated with NVT serotypes through capsular switch then an increase in NVT pneumococcal disease and infection is possible. Although this process of capsular switch involves consideration of the underlying genetic material of the pneumococcus, Temime et al. do not specifically refer to MLSTs in their mathematical model. Instead it is assumed that the capsular switch takes place when a vaccinated host becomes colonised with a VT and NVT serotype, without discussion of the underlying biological process involved in the switch.

Pneumococcal MLSTs have become a matter of interest for biologists in recent years and concerns have been raised about associations between MLSTs and disease. The serotype of the pneumococcus is considered to be the factor which causes the virulence of the bacterium. However, it has been observed that certain MLSTs are able to manifest in more than one serotype. This is a problem with consequences such as capsular switch. In addition, this is a problem if MLSTs that are associated with invasive disease more commonly manifest in NVT serotypes following the introduction of a vaccine as this could lead to an increase in non-vaccine serotype invasive disease and infection (McChlery et al. 2005).

In this chapter and the subsequent chapter, ideas from the two serotype differential equation model of Lipsitch (Lipsitch 1997) will be adapted to develop mathematical models involving the carriage of one or more pneumococcal sequence types. Lipsitch was interested in investigating the competition between NVT and VT serotypes in the models considered. In the models discussed in this thesis, interest is focussed on MLSTs that are associated with more than one serotype to incorporate the possibility of capsular switching into the model. The population of interest that is being considered is children aged under two years

as this is the group to which the new conjugate vaccine is administered.

All of the modelling presented is analytical, not requiring any parameter estimation. The analytical approach is preferred for the models considered in Chapters 2 and 3 due to the fact the solutions are more generalisable than those which could be obtained through the use of numerical techniques. Numerical solutions would provide results only for specific parameter values entered in the model. Thus, even if a variety of values are entered in the model which show similar results this is not a proof of what occurs in the model. However, this is not the case with analytical techniques which are generalisable to any parameter values.

For each mathematical model described, the differential equations representing the model are presented. The time independent solutions for the sizes of each class of host are found and the basic reproductive number, described earlier in this chapter, is determined. In addition, both local and global stability analyses are conducted to identify what happens to the number of carrying hosts in the long term.

The first model discussed in detail considers only one MLST associated with two different serotypes. This is the simplest model that will be described and was created to understand underlying properties of models involving the carriage of a MLST which can manifest as more than one serotype. By expansion of the structure of this model in the next chapter, more complex models may be explored. In Chapter 3, models involving two MLSTs which have the potential to manifest in more than one serotype will be considered.

The models will explore the effect of a PCV which prevents carriage of VT serotypes by examining the relationship between the MLSTs and serotypes, where a vaccine may not prevent carriage of a MLST if it is associated with a NVT in addition to a VT. In addition, the models will consider different transmission mechanisms.

2.5 Model of one MLST associated with two serotypes

The simplest analytical differential equation model considers only four possible classes of individuals; two classes of hosts susceptible to carriage of the pneumococci, one for unvaccinated children (X) and the other for vaccinated children (V), and two classes of carrying hosts, again one for those unvaccinated (T_1) and one for those vaccinated (V_{T_1}). In this model, and subsequent models, carriage refers to the MLST acquired and not specifically to the serotype. It is assumed that this MLST may manifest itself in either serotype 1 (Y_1) or serotype 2 (Y_2) with proportions PT_1 and $(1-P)T_1$ respectively in an unvaccinated population. It is assumed, for simplicity, that the proportion of children who receive vaccination is a constant.

Using the notation of Lipsitch 1997, the model assumes that susceptible hosts enter the population of interest at a constant rate L and that hosts leave the population at a per capita rate u. The transition from the susceptible class to the carrying class occurs according to a mass action process at a rate of $\beta_1 X(T_1+V_{T_1})$ for unvaccinated children and at a rate $\beta_1 V(T_1+V_{T_1})$ for vaccinated children. In this initial model, for simplicity of expression, transmission is assumed to be attributable to MLST and not serotype. However, it is biologically plausible that MLSTs play a part in the ability of the bacterium to transmit from one host to another. The possibility of transmission being attributable to serotype is explored in the modelling in Chapter 3. Carrying hosts may cease to carry the MLST, becoming susceptible again, independently of one another, at rate γ.

In this model, the vaccine is assumed to be 100% effective in preventing carriage of serotype 1 but ineffective in preventing carriage of serotype 2. Therefore, the vaccine will not eradicate carriage of the MLST as it is able to manifest in both serotypes 1 and 2. All vaccinated hosts carrying the MLST must be carrying it in the form of serotype 2. f is the proportion of children who receive the vaccine.

The ordinary differential equations (ODEs) that correspond to this model (Figure 2.11) are:

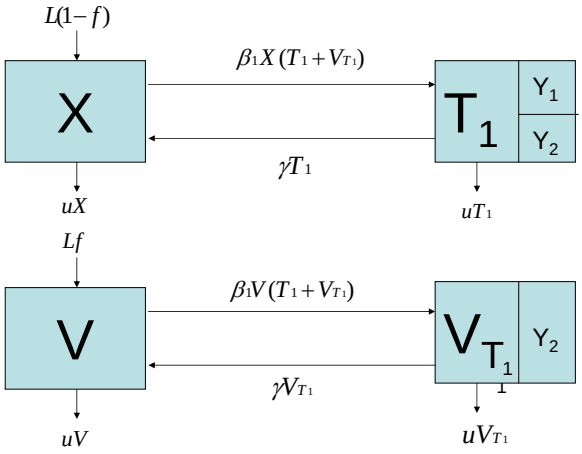

Figure 2.11: Model of one MLST which can manifest in two serotypes with vaccine effective against one serotype.

$$\frac{dX}{dt} = L(1-f) - uX - \beta_1 X(T_1 + V_{T_1}) + \gamma T_1,$$

$$\frac{dT_1}{dt} = \beta_1 X(T_1 + V_{T_1}) - (\gamma + u)T_1,$$

$$\frac{dV}{dt} = Lf - uV - \beta_1 V(T_1 + V_{T_1}) + \gamma V_{T_1},$$

and

$$\frac{dV_{T_1}}{dt} = \beta_1 V(T_1 + V_{T_1}) - (\gamma + u)V_{T_1}. \tag{2.1}$$

In the next section, the mathematical results for this model are described. To begin, an equilibrium analysis is carried out, followed by a discussion of the effective reproductive number. Then a local stability analysis is discussed, followed by a global stability analysis.

2.5.1 Results

Equilibrium solutions

By setting the time derivatives, $\frac{dX}{dt}$, $\frac{dV}{dt}$, $\frac{dT_1}{dt}$ and $\frac{dV_{T_1}}{dt}$, equal to zero, the particular steady state population sizes may be determined for each of the unique classes.

Adding the equilibrium equations corresponding to (2.1)(i) and (2.1)(ii), it can be deduced that at equilibrium $X + T_1 = (1 - f)\frac{L}{u}$. Similarly, from (2.1)(iii) and (2.1)(iv), it can be deduced that at equilibrium $V + V_{T_1} = f\frac{L}{u}$. This is intuitively sensible.

The total number of individuals in the population, N, is defined as $N = X + T_1 + V + V_{T_1}$. Therefore, $\frac{dN}{dt} = L - uN$ and it can easily be shown that $N \to \frac{L}{u}$ as $t \to \infty$. Thus, at equilibrium, the total number of hosts in the population is $\frac{L}{u}$. As f is the proportion of vaccinated hosts, at the steady state population it makes sense that the total number of vaccinated hosts must equal $f\frac{L}{u}$, i.e. $V + V_{T_1} = f\frac{L}{u}$. Similarly, since $1 - f$ is the proportion of hosts that have not been vaccinated, it makes sense that $X + T_1 = (1 - f)\frac{L}{u}$ at equilibrium.

Next, substitution of $X = (1 - f)\frac{L}{u} - T_1$ into $\frac{dT_1}{dt} = 0$ and $V = f\frac{L}{u} - V_{T_1}$ into $\frac{dV_{T_1}}{dt} = 0$ gives the equations

$$\beta_1 \left(\frac{L}{u}(1 - f) - T_1 \right) (T_1 + V_{T_1}) - (\gamma + u)T_1 = 0, \tag{2.2}$$

and

$$\beta_1 \left(\frac{L}{u}f - V_{T_1} \right) (T_1 + V_{T_1}) - (\gamma + u)V_{T_1} = 0. \tag{2.3}$$

The addition of (2.2) and (2.3) gives the following expression:

$$\beta_1 (T_1 + V_{T_1}) \left(\frac{L}{u}(1 - f) - T_1 + \frac{L}{u}f - V_{T_1} \right) - (\gamma + u)T_1 - (\gamma + u)V_{T_1} = 0,$$

so

$$\beta_1(T_1 + V_{T_1})\left(\frac{L}{u} - (T_1 + V_{T_1})\right) = (\gamma + u)(T_1 + V_{T_1}).$$

Therefore, either $T_1 + V_{T_1} = 0$, or

$$(\gamma + u) = \beta_1\left(\frac{L}{u} - (T_1 + V_{T_1})\right)$$

so

$$T_1 + V_{T_1} = \frac{L}{u} - \frac{\gamma + u}{\beta_1}.$$

When $T_1 + V_{T_1} = 0$, $T_1 = 0$ and $V_{T_1} = 0$ since $T_1 \geqslant 0$ and $V_{T_1} \geqslant 0$. Thus, $X = (1 - f)\frac{L}{u}$ and $V = f\frac{L}{u}$. Therefore, the carriage-free equilibrium (CFE) solution (X_e,T_{1_e},V_e,$V_{T_{1_e}}$) is:

$$\left((1 - f)\frac{L}{u}, 0, f\frac{L}{u}, 0\right).$$

This solution is intuitive since when there are no hosts carrying the MLST ($T_1 = 0$ and $V_{T_1} = 0$), all hosts, N, are susceptible to carriage. However, a proportion f of hosts receive the vaccine. Thus, the hosts are split between the two susceptible classes accordingly.

Next consider the case where $T_1 + V_{T_1} \neq 0$ and

$$T_1 + V_{T_1} = \frac{L}{u} - \frac{\gamma + u}{\beta_1}.$$

Now, $V = 0$ implies that $V_{T_1} = 0$ from the equilibrium version of (2.1)(iv). This contradicts the equilibrium version of (2.1)(iii). Hence, $V > 0$ at equilibrium and (2.1)(iv) now implies that $V_{T_1} > 0$. Dividing the equilibrium versions of (2.1)(ii) and (2.1)(iv) it can be deduced that

$$\frac{T_1}{V_{T_1}} = \frac{X}{V} = \frac{\frac{L}{u}(1 - f) - T_1}{\frac{L}{u}f - V_{T_1}},$$

since $X = (1-f)\frac{L}{u} - T_1$ and $V = f\frac{L}{u} - V_{T_1}$. Therefore,

$$\frac{L}{u}(1-f)V_{T_1} = \frac{L}{u}fT_1.$$

Thus, $(1-f)V_{T_1} = fT_1$ which gives $T_1 = (1-f)k$ and $V_{T_1} = fk$ for some k. Hence, $T_1 + V_{T_1} = k = \frac{L}{u} - \frac{\gamma+u}{\beta_1}$.

To summarise,

$$T_1 = (1-f)\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right),$$

$$V_{T_1} = f\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right),$$

$$X = \frac{L}{u}(1-f) - T_1 = (1-f)\frac{\gamma+u}{\beta_1},$$

and

$$V = \frac{L}{u}f - V_{T_1} = f\frac{\gamma+u}{\beta_1}.$$

The endemic, or carriage, equilibrium solution (X_e,T_{1_e},V_e,$V_{T_{1_e}}$) is:

$$\left((1-f)\frac{\gamma+u}{\beta_1}, (1-f)\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right), f\frac{\gamma+u}{\beta_1}, f\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right)\right).$$

This equilibrium will be biologically feasible and distinct from the CFE if and only if $\frac{L}{u} > \frac{\gamma+u}{\beta_1}$.

The carriage equilibriumn (CE) states that, at steady state, the total number of susceptible hosts $(X+V)$ is equal to the rate at which carrying hosts return to

the susceptible class, γ, added to the rate at which hosts leave the population, u, then divided by the carriage transmission rate β_1. The total number of carrying hosts (T_1+V_{T_1}) at equilibrium is simply the total population minus this number. To obtain the specific numbers for the vaccinated and unvaccinated classes, all that is required is to multiply by the parameter for the proportion vaccinated, f, or unvaccinated, $1-f$.

A special case of this model is the situation where no vaccine is present in the population. For this case the equilibrium number of susceptible and infected individuals at the two possible equilibria are obtained by setting $f = 0$ in the above expressions. Clearly, if no hosts receive vaccination then at the CFE all hosts, $\frac{L}{u}$, will be found in the unvaccinated susceptible class, X. At the CE equilibrium, $\frac{\gamma+u}{\beta_1}$ hosts will be found in the unvaccinated susceptible class and $\frac{L}{u} - \frac{\gamma+u}{\beta_1}$ in the unvaccinated carrying class, T_1. Carrying hosts may be colonised with the MLST manifested as serotype 1 or serotype 2 as there is no vaccine effect present to prevent carriage of serotype 1.

In addition, it is possible to consider a scenario where all children receive the vaccine, i.e. $f = 1$. In this situation, at the CFE all hosts are found in the vaccinated susceptible host class and at the CE $\frac{\gamma+u}{\beta_1}$ hosts are found in the vaccinated susceptible class, V, and $\frac{L}{u} - \frac{\gamma+u}{\beta_1}$ are found in the vaccinated carrying class, V_{T_1}. In this situation, as the vaccine is assumed to be 100% effective in preventing carriage of serotype 1, serotype 1 will not be present in the population. All carrying hosts will be colonised with serotype 2.

Returning to the model with vaccination, considering the model in terms of serotype, a proportion P of unvaccinated MLST carrying hosts, T_1, are assumed to be carrying serotype 1, Y_1, and a proportion $1-P$ are carrying serotype 2, Y_2. All vaccinated MLST carrying hosts, V_{T_1} are carrying serotype 2. Thus, at the CFE, no hosts are carrying either serotype as no MLSTs are carried. At the CE,

$$T_1 = (1-f)\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right).$$

Thus, from this,

$$Y_1 = P(1-f)\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right),$$

and the contribution to Y_2 from T_1 is

$$(1-P)(1-f)\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right).$$

However, as all V_{T_1} are manifest as Y_2,

$$\begin{aligned} Y_2 &= (1-P)T_1 + V_{T_1}, \\ &= (1-P)(1-f)\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right) + f\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right), \\ &= (1-P(1-f))\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right). \end{aligned}$$

In this chapter, P is assumed to take some fixed value. To explore the relationship between the MLST and the two serotypes associated with it, $P = 0$ and $P = 1$ are considered. When $P = 0$, no hosts are colonised with serotype 1 since this refers to the proportion of individuals carrying the MLST who are colonised with serotype 1. In this case, all hosts must be colonised with serotype 2. Thus, the vaccine would have no impact at all on the population as it is assumed 100% effective in preventing carriage of serotype 1 only. However, if it is assumed that $P = 1$, hosts may become colonised with either serotype 1 or serotype 2 and the CE is the same whether expressed in terms of MLSTs or serotypes. The model implicitly assumes that capsular switch is possible. If $V_{T_1}(0) > 0$ this is correct as there are initially serotype 2 pneumococci in existence in the population. If $V_{T_1}(0) = 0$ then it must be assumed that serotype 2 invades from outside the population for the model to be valid.

Effective reproductive number, R_e

The effective reproductive number has essentially the same definition as the basic reproductive number. It is called effective rather than basic as a vaccine effect has been incorporated in this model.

Consider the model discussed (Figure 2.11). The average duration of carriage of the MLST is $\frac{1}{\gamma+u}$ for both vaccinated and unvaccinated hosts. Let m_{ij} be the expected number of susceptible individuals of type i infected by a single type j carrying individual entering the CFE during his or her entire infectious period. For this model, type 1 refers to unvaccinated carriers of the MLST and 2 refers to vaccinated carriers of the MLST. Then,

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} \frac{\beta_1 L(1-f)}{u(\gamma+u)} & \frac{\beta_1 L(1-f)}{u(\gamma+u)} \\ \frac{\beta_1 Lf}{u(\gamma+u)} & \frac{\beta_1 Lf}{u(\gamma+u)} \end{bmatrix}.$$

For example, m_{11} is the expected number of unvaccinated carriers caused by a single carrying individual entering the CFE. At the CFE, there are $(1-f)\frac{L}{u}$ susceptible individuals. Each of these is infected by the original carrying individual at rate β_1 for time $\frac{1}{\gamma+u}$ so the expected number of unvaccinated carriers is

$$m_{11} = \beta_1 \times (1-f)\frac{L}{u} \times \frac{1}{\gamma+u} = \frac{\beta_1 L(1-f)}{u(\gamma+u)}.$$

The other entries of $\mathbf{M}$ are deduced similarly. The eigenvalues are the roots of the characteristic equation $\det(\mathbf{M} - \lambda\mathbf{I}) = 0$. Therefore,

$$\left(\frac{\beta_1 L(1-f)}{u(\gamma+u)} - \lambda\right)\left(\frac{\beta_1 Lf}{u(\gamma+u)} - \lambda\right) - \left(\frac{\beta_1 Lf}{u(\gamma+u)}\right)\left(\frac{\beta_1 L(1-f)}{u(\gamma+u)}\right) = 0.$$

It follows that

$$\lambda^2 - \frac{\beta_1 Lf}{u(\gamma+u)}\lambda - \frac{\beta_1 L(1-f)}{u(\gamma+u)}\lambda = 0,$$

so

$$\lambda\left(\lambda - \frac{\beta_1 L}{u(\gamma+u)}\right) = 0.$$

Hence, either $\lambda = 0$ or $\lambda = \frac{\beta_1 L}{u(\gamma+u)}$. R_e is defined to be the largest eigenvalue so $R_e = \frac{\beta_1 L}{u(\gamma+u)}$ since $\frac{\beta_1 L}{u(\gamma+u)} > 0$.

At equilibrium, the total rate at which hosts enter the population, L, may be assumed to be approximately equal to the total rate at which hosts leave the population, $u\hat{N}$, where $\hat{N}$ is the equilibrium population size. The population under consideration in these models is children under the age of two years. Thus, R_e may be expressed as

$$R_e = \frac{\beta_1 \hat{N}}{\gamma + u}.$$

This means that the value of the effective reproductive number is essentially determined by the transmission rate, and the rate at which carrying hosts cease to carry the MLST. Note that the CE is biologically feasible if and only if $R_e > 1$. Hence, for $R_e \leqslant 1$ there is only one equilibrium, the CFE, whereas for $R_e > 1$ there are two equilibria, namely the CE and the CFE.

It can be noted that R_e for this model is independent of f, the proportion of hosts that receive the vaccine. This is due to the fact that transmission is assumed to be attributable to MLST, not serotype, so the model relates specifically to the carriage of the MLST. Thus, since the vaccine does not prevent carriage of the MLST since it is associated with a VT and a NVT serotype, vaccine cannot have an impact on R_e.

Local stability analysis

To continue the investigation of the model involving only one MLST that is able to manifest in two serotypes, a stability analysis of the equilibrium solutions was carried out to determine what happens to the number of unvaccinated and vaccinated susceptible individuals and carrying hosts in the long term. Initially a local stability analysis (LSA) was carried out. Local stability (more formally local asymptotic stability) of an equilibrium means that if the initial population

sizes are sufficiently close to the equilibrium then, in the long term, the solutions will converge to the equilibrium. The drawback to the local stability analysis is that it is difficult to quantify "sufficiently close". However, a global stability analysis (GSA) was also performed on this model and this analysis does not require the assumption that the initial population sizes are sufficiently close to an equilibrium.

Consider the right-hand sides of equations (2.1),

$$\frac{dX}{dt} = f(X, T_1, V, V_{T_1}) = L(1-f) - uX - \beta_1 X(T_1 + V_{T_1}) + \gamma T_1,$$

$$\frac{dT_1}{dt} = g(X, T_1, V, V_{T_1}) = \beta_1 X(T_1 + V_{T_1}) - (\gamma + u)T_1,$$

$$\frac{dV}{dt} = h(X, T_1, V, V_{T_1}) = Lf - uV - \beta_1 V(T_1 + V_{T_1}) + \gamma V_{T_1},$$

and

$$\frac{dV_{T_1}}{dt} = j(X, T_1, V, V_{T_1}) = \beta_1 V(T_1 + V_{T_1}) - (\gamma + u)V_{T_1}.$$

The partial derivatives of these functions are shown in the following table:

	$\boldsymbol{\partial/\partial X}$	$\boldsymbol{\partial/\partial T_1}$	$\boldsymbol{\partial/\partial V}$	$\boldsymbol{\partial/\partial V_{T_1}}$
f	$-u - \beta_1(T_1 + V_{T_1})$	$-\beta_1 X + \gamma$	0	$-\beta_1 X$
g	$\beta_1(T_1 + V_{T_1})$	$\beta_1 X - \gamma - u$	0	$\beta_1 X$
h	0	$-\beta_1 V$	$-u - \beta_1(T_1 + V_{T_1})$	$-\beta_1 V + \gamma$
j	0	$\beta_1 V$	$\beta_1(T_1 + V_{T_1})$	$\beta_1 V - \gamma - u$

An equilibrium solution, $(X_e, T_{1_e}, V_e, V_{T1e})$ is locally asymptotically stable (LAS) if the real parts of the eigenvalues of the following matrix, $\mathbf{A}$, are negative.

$$\mathbf{A} = \begin{bmatrix} \frac{\partial f}{\partial X}(X_e, T_{1_e}, V_e, V_{T_{1_e}}) & \frac{\partial f}{\partial T_1}(X_e, T_{1_e}, V_e, V_{T_{1_e}}) & \frac{\partial f}{\partial V}(X_e, T_{1_e}, V_e, V_{T_{1_e}}) & \frac{\partial f}{\partial V_{T_1}}(X_e, T_{1_e}, V_e, V_{T_{1_e}}) \\ \frac{\partial g}{\partial X}(X_e, T_{1_e}, V_e, V_{T_{1_e}}) & \frac{\partial g}{\partial T_1}(X_e, T_{1_e}, V_e, V_{T_{1_e}}) & \frac{\partial g}{\partial V}(X_e, T_{1_e}, V_e, V_{T_{1_e}}) & \frac{\partial g}{\partial V_{T_1}}(X_e, T_{1_e}, V_e, V_{T_{1_e}}) \\ \frac{\partial h}{\partial X}(X_e, T_{1_e}, V_e, V_{T_{1_e}}) & \frac{\partial h}{\partial T_1}(X_e, T_{1_e}, V_e, V_{T_{1_e}}) & \frac{\partial h}{\partial V}(X_e, T_{1_e}, V_e, V_{T_{1_e}}) & \frac{\partial h}{\partial V_{T_1}}(X_e, T_{1_e}, V_e, V_{T_{1_e}}) \\ \frac{\partial j}{\partial X}(X_e, T_{1_e}, V_e, V_{T_{1_e}}) & \frac{\partial j}{\partial T_1}(X_e, T_{1_e}, V_e, T_{1_e}) & \frac{\partial j}{\partial V}(X_e, T_{1_e}, V_e, V_{T_{1_e}}) & \frac{\partial j}{\partial V_{T_1}}(X_e, T_{1_e}, V_e, V_{T_{1_e}}) \end{bmatrix}.$$

A LSA was carried out for each of the equilibrium solutions of this model.

1. Carriage-free equilibrium.

   At $(X_e, T_{1_e}, V_e, V_{T_{1_e}}) = \big((1-f)\frac{L}{u}, 0, f\frac{L}{u}, 0\big)$,

$$\mathbf{A} = \begin{bmatrix} -u & -\frac{\beta_1 L}{u}(1-f)+\gamma & 0 & -\frac{\beta_1 L}{u}(1-f) \\ 0 & \frac{\beta_1 L}{u}(1-f)-\gamma-u & 0 & \frac{\beta_1 L}{u}(1-f) \\ 0 & -\frac{\beta_1 L}{u}f & -u & -\frac{\beta_1 L}{u}f+\gamma \\ 0 & \frac{\beta_1 L}{u}f & 0 & \frac{\beta_1 L}{u}f-\gamma-u \end{bmatrix}.$$

   The eigenvalues of $\mathbf{A}$ are found by identifying when $\det(\mathbf{A}-\lambda\mathbf{I}) = 0$. Let $K = \frac{\beta_1 L}{u}$, then

$$\det(\mathbf{A}-\lambda\mathbf{I}) = \begin{vmatrix} -u-\lambda & -K(1-f)+\gamma & 0 & -K(1-f) \\ 0 & K(1-f)-\gamma-u-\lambda & 0 & K(1-f) \\ 0 & -Kf & -u-\lambda & -Kf+\gamma \\ 0 & Kf & 0 & Kf-\gamma-u-\lambda \end{vmatrix}.$$

   Using Maple, the eigenvalues of $\mathbf{A}$ were found to be $\lambda_1 = -(\gamma+u)$, $\lambda_2 = K-\gamma-u$ and $\lambda_3, \lambda_4 = -u$. λ_1, λ_3 and λ_4 are all negative. The CFE will be LAS if all of the eigenvalues have negative real parts. This will be true if $\lambda_2 < 0$, i.e. $\frac{\beta_1 L}{u} - \gamma - u < 0$. This inequality can be rearranged to get $R_e < 1$. Therefore, the CFE is LAS if $R_e < 1$ and is unstable if $R_e > 1$.

2. Carriage equilibrium.

   At

$$(X_e, T_{1_e}, V_e, V_{T_{1_e}}) = \left((1-f)\tfrac{(\gamma+u)}{\beta_1}, (1-f)\left(\tfrac{L}{u}-\tfrac{(\gamma+u)}{\beta_1}\right), f\tfrac{(\gamma+u)}{\beta_1}, f\left(\tfrac{L}{u}-\tfrac{(\gamma+u)}{\beta_1}\right)\right),$$

$$\mathbf{A} = \begin{bmatrix} \gamma - \frac{\beta_1 L}{u} & \gamma - (1-f)(\gamma+u) & 0 & -(1-f)(\gamma+u) \\ \frac{\beta_1 L}{u} - \gamma - u & (1-f)(\gamma+u) - \gamma - u & 0 & (1-f)(\gamma+u) \\ 0 & -f(\gamma+u) & \gamma - \frac{\beta_1 L}{u} & \gamma - f(\gamma+u) \\ 0 & f(\gamma+u) & \frac{\beta_1 L}{u} - \gamma - u & f(\gamma+u) - \gamma - u \end{bmatrix}.$$

Using Maple, the eigenvalues of $\mathbf{A}$ were found to be $\lambda_1 = -\frac{\beta_1 L}{u}$, $\lambda_2 = -\frac{\beta_1 L}{u} + \gamma + u$ and $\lambda_3, \lambda_4 = -u$. λ_1, λ_3 and λ_4 are all negative. Therefore, for the carriage equilibrium solution to have local asymptotic stability, it is necessary that $\lambda_2 < 0$, i.e. $-\frac{\beta_1 L}{u} + \gamma + u < 0$. This inequality can be rearranged to get $R_e > 1$. The endemic carriage equilibrium only exists if $R_e > 1$. Therefore, the endemic carriage equilibrium is always LAS when it exists.

To summarise the local asymptotic stability findings, when the effective reproductive number is less than or equal to one the CFE is the only equilibrium possible, and is LAS if $R_e < 1$. When the effective reproductive number is greater than one both equilibria are possible. However, in this situation, the CFE is unstable and the CE is LAS.

Global stability analysis

As mentioned previously, $N = X + T_1 + V + V_{T_1}$, where N is the total number of hosts in the population. So, $\frac{dN}{dt} = L - uN$ and $N \to \frac{L}{u}$ as $t \to \infty$.

Consider

$$\frac{d}{dt}(X+V) = L - u(X+V) - \beta_1(X+V)(T_1 + V_{T_1}) + \gamma(T_1 + V_{T_1}), \qquad (2.4)$$

and

$$\frac{d}{dt}(T_1 + V_{T_1}) = \beta_1(X+V)(T_1 + V_{T_1}) - (\gamma+u)(T_1 + V_{T_1}). \qquad (2.5)$$

Initially in the global stability analysis (GSA), $R_e \leqslant 1$ is considered. As men-

tioned previously, when $R_e < 1$ the CFE is the only equilibrium possible and is LAS. There are two cases to consider. First of all, what occurs in the long term to the number of susceptible and carrying hosts if there are no carrying hosts initially, i.e. $(T_1+V_{T_1})(0) = 0$. Secondly, what happens to the population sizes when there are some carrying hosts in the initial population, i.e. $(T_1 + V_{T_1})(0) > 0$.

- Case 1: $(T_1 + V_{T_1})(0) = 0$.

  $(T_1 + V_{T_1})(0) = 0 \Rightarrow \frac{d(T_1+V_{T_1})}{dt} = 0.$
  Therefore, $(T_1 + V_{T_1})(t) = 0 \ \forall \ t.$

  In this case the equation for $X + V$ is the same as the equation for N, i.e. $\frac{d(X+V)}{dt} = L - u(X + V)$. Therefore, regardless of the initial state of $X + V$, $X + V \to \frac{L}{u}$.

- Case 2: $(T_1 + V_{T_1})(0) > 0$.

  In this case it is required to prove that $(T_1+V_{T_1})(0) > 0 \Rightarrow (T_1+V_{T_1})(t) > 0$ $\forall \ t > 0$. This can be shown using a proof by contradiction. ? Suppose that

  $(T_1 + V_{T_1})(t) \not> 0 \ \forall \ t > 0.$

  Let $S = \{t > 0 : (T_1 + V_{T_1})(s) > 0 \text{ on } [0, t]\}$ and let $\xi_0 = \sup(S)$. If $\sup(S) = \xi_0 = \infty$, there is nothing to prove since in this case $(T_1+V_{T_1})(s) > 0 \ \forall \ s > 0$.

  Suppose that $\xi_0 < \infty$. Therefore, $(T_1 + V_{T_1})(s) > 0$ on $[0, \xi]$ for any $\xi < \xi_0$. From (2.5),

  $$\frac{1}{(T_1 + V_{T_1})} \frac{d(T_1 + V_{T_1})}{dt} = \beta_1(X + V) - (\gamma + u).$$

  Integrating both sides in the interval $[0, \xi]$ gives:

$$\int_0^{\xi} \frac{d}{dt}\big(\ln(T_1+V_{T_1})\big)ds = \int_0^{\xi} (\beta_1(X+V)-(\gamma+u))\,ds$$

so

$$\ln\left(\frac{(T_1+V_{T_1})(\xi)}{(T_1+V_{T_1})(0)}\right) = \int_0^{\xi} (\beta_1(X+V)-(\gamma+u))\,ds.$$

Therefore, $(T_1+V_{T_1})(\xi) = (T_1+V_{T_1})(0)\exp\left(\int_0^{\xi}(\beta_1(X+V)-(\gamma+u))\,ds\right)$ $\forall\ \xi \in [0,\xi_0)$. Note that $X+V$ is continuous on $[0,\xi_0]$ and so is bounded on $[0,\xi_0]$.

Let $\xi \to \xi_0$, then

$$\int_0^{\xi} (\beta_1(X+V)(s)-(\gamma+u))\,ds \to \int_0^{\xi_0} (\beta_1(X+V)(s)-(\gamma+u))\,ds.$$

Hence, by continuity of exponentials,

$$(T_1+V_{T_1})(\xi) \to (T_1+V_{T_1})(0)\exp\left(\int_0^{\xi_0}(\beta_1(X+V)(s)-(\gamma+u))\right)ds > 0$$

as $\xi \to \xi_0$.

But $T_1+V_{T_1}$ is continuous on $[0,\infty)$ so $(T_1+V_{T_1})(\xi) \to (T_1+V_{T_1})(\xi_0)$ as $\xi \to \xi_0$. Hence,

$$(T_1+V_{T_1})(\xi_0) = (T_1+V_{T_1})(0)\exp\left(\int_0^{\xi_0}(\beta_1(X+V)(s)-(\gamma+u))\right)ds > 0.$$

Hence, by continuity, $(T_1+V_{T_1})(t) > 0$ in $[0,\xi_0+\triangle\xi]$ for some $\triangle\xi > 0$. This is a contradiction so $\xi_0 = \infty$. Therefore, $(T_1+V_{T_1})(t) > 0\ \forall\ t$.

Now, to prove that the CFE is globally asymptotically stable (GAS) when $R_e \leqslant 1$, it is required to show that regardless of the initial number of carrying hosts in the population, the number will tend to zero in the long term.

First, the case where $R_e < 1$ is considered. To show that the number of carrying hosts tends to zero, it shall be shown that $T_1 + V_{T_1}$ is decreasing. Note that $T_1 + V_{T_1} \geqslant 0$.

Choose some arbitrary $\epsilon > 0$ such that $-k_0 = (\gamma + u)(R_e - 1) + \beta_1\epsilon < 0$. This is the case when, for example,

$$\epsilon = \frac{-(\gamma + u)}{2\beta_1}(R_e - 1).$$

$\exists\ t_0$ such that for $t \geqslant t_0$, $X + V \leqslant \frac{L}{u} + \epsilon$. Therefore for $t \geqslant t_0$,

$$\begin{aligned}
\frac{1}{T_1 + V_{T_1}} \frac{d(T_1 + V_{T_1})}{dt} &= \beta_1(X + V) - (\gamma + u), \\
&\leqslant \beta_1 \left(\frac{L}{u} + \epsilon \right) - (\gamma + u), \\
&= \beta_1 \frac{L}{u} - (\gamma + u) + \beta_1\epsilon, \\
&= -k_0 < 0.
\end{aligned}$$

Hence,

$$\int_{t_0}^{t} \frac{1}{T_1 + V_{T_1}} \frac{d(T_1 + V_{T_1})}{dt} ds \leqslant \int_{t_0}^{t} -k_0 ds$$

so

$$\ln \left(\frac{(T_1 + V_{T_1})(t)}{(T_1 + V_{T_1})(t_0)} \right) \leqslant -k_0(t - t_0).$$

It can be deduced that, $(T_1 + V_{T_1})(t) \leqslant (T_1 + V_{T_1})(t_0) \exp(-k_0(t - t_0))$. Therefore, $0 \leqslant (T_1 + V_{T_1})(t) \leqslant (T_1 + V_{T_1})(t_0) \exp(-k_0(t - t_0))$ so $T_1 + V_{T_1} \to 0$ as $t \to \infty$ (since $\exp(-k_0(t - t_0)) \to 0$ as $t \to \infty$). As a consequence, $X + V = N - (T_1 + V_{T_1}) \to \frac{L}{u}$ as $t \to \infty$.

Next, the case where $R_e = 1$ is considered. Suppose again that $\epsilon > 0$. It is known

that $\exists\ t_0$ such that for $t \geqslant t_0$, $N - \frac{L}{u} \leqslant \epsilon$.

For $t \geqslant t_0$,

$$\begin{aligned}\frac{d}{dt}(T_1 + V_{T_1}) &= \beta_1(N - (T_1 + V_{T_1}))(T_1 + V_{T_1}) - (\gamma + u)(T_1 + V_{T_1}),\\ &\leqslant \beta_1(\epsilon - (T_1 + V_{T_1}))(T_1 + V_{T_1}).\end{aligned}$$

Hence, for $T_1 + V_{T_1} \geqslant 2\epsilon$, $\frac{d(T_1+V_{T_1})}{dt} \leqslant -\beta_1\epsilon(T_1 + V_{T_1})$.

Lemma 2.5.1.1

$\exists\ t_2 > t_0$ such that for $t \geqslant t_2$, $0 \leqslant (T_1 + V_{T_1})(t) \leqslant 3\epsilon$.

Proof

If $(T_1 + V_{T_1})(t_0) > 3\epsilon$ then whilst $(T_1 + V_{T_1})(t) \geqslant 2\epsilon$,

$$(T_1 + V_{T_1})(t) \leqslant (T_1 + V_{T_1})(t_0)\exp[-\beta_1\epsilon(t - t_0)]$$

so $\exists\ t_1 \geqslant t_0$ such that $(T_1 + V_{T_1})(t_1) \leqslant 3\epsilon$, whether or not $(T_1 + V_{T_1})(t_0) > 3\epsilon$. Then, if $\triangle t$ is small and positive, $(T_1 + V_{T_1})(t_1 + \triangle t) \leqslant 3\epsilon$.

Let $\xi_1 = \sup\{t > t_1 : (T_1 + V_{T_1})(t) \leqslant 3\epsilon\}$ so $\xi_1 > t_1$. If $\xi_1 < \infty$, then by continuity $(T_1 + V_{T_1})(\xi_1) = 3\epsilon$, so

$$\begin{aligned}(T_1 + V_{T_1})(\xi_1 + \triangle t) &= 3\epsilon + \frac{d}{dt}(T_1 + V_{T_1})|_{\xi_1}\triangle t + o(\triangle t),\\ &\leqslant 3\epsilon - 3\beta_1\epsilon^2\triangle t + o(\triangle t),\\ &< 3\epsilon\end{aligned}$$

if $\triangle t > 0$ and $\triangle t$ is sufficiently small. This contradicts the definition of ξ_1. Thus, $\xi_1 = \infty$ and the lemma holds. Since $\epsilon > 0$ is arbitrary, as an immediate corollary it is deduced that $(T_1 + V_{T_1})(t) \to 0$ and $(X + V)(t) \to \frac{L}{u}$ as $t \to \infty$ as required, even when $R_e = 1$.

To summarise, when $R_e \leqslant 1$, $T_1 + V_{T_1} \to 0$ and $X + V \to \frac{L}{u}$ regardless of the

initial values of $T_1 + V_{T_1}$ and $X + V$. Since $T_1 \geqslant 0$ and $V_{T_1} \geqslant 0$, as $T_1 + V_{T_1} \to 0$, $T_1 \to 0$ and $V_{T_1} \to 0$ as $t \to \infty$.

To determine what X and V each tend to as $t \to \infty$, consider the equation for $\frac{dX}{dt}$ in the model described in Figure 2.11. Substituting $T_1 + V_{T_1} \to 0$ and $T_1 \to 0$ into the first equation shown in (2.1) gives

$$\frac{dX}{dt} \to L(1-f) - uX \text{ as } t \to \infty.$$

Consider some arbitrary $\epsilon > 0$. $\exists\ t_3$ such that for $t \geqslant t_3$,

$$\frac{dX}{dt} < L(1-f) - uX + \epsilon. \tag{2.6}$$

Hence, X must eventually fall beneath the level $\frac{L(1-f)}{u} + \frac{2\epsilon}{u}$, i.e., $\exists\ t_4 > t_3$ such that $X(t_4) < \frac{L(1-f)}{u} + \frac{2\epsilon}{u}$.

It shall be shown that once X goes beneath the level $\frac{L(1-f)}{u} + \frac{2\epsilon}{u}$ it can never rise above it. Suppose otherwise, and that

$$X(t_5) > \frac{L(1-f)}{u} + \frac{2\epsilon}{u} \text{ for some } t_5 > t_4.$$

Then $\exists\ t_6, t_7$ with $t_4 < t_6 < t_7 < t_5$ such that $X(t)$ is strictly monotone increasing in $[t_6, t_7]$ and

$$\frac{L(1-f)}{u} + \frac{2\epsilon}{u} < X(t_6) < X(t_7).$$

This contradicts (2.6) which implies that $X(t)$ is monotone decreasing in $[t_6, t_7]$. Hence

$$X \leqslant \frac{L(1-f)}{u} + \frac{2\epsilon}{u}, \quad \forall t \geqslant t_4.$$

Therefore, once X goes beneath the level $\frac{L(1-f)}{u} + \frac{2\epsilon}{u}$ it can never rise above it. Similarly, $\exists\ t_8$ such that for $t \geqslant t_8$

$$\frac{dX}{dt} > L(1-f) - uX - \epsilon. \tag{2.7}$$

Hence X must eventually rise above the level $\frac{L(1-f)}{u} - \frac{2\epsilon}{u}$, i.e. $\exists\ t_9 > t_8$ such that

$$X(t_9) > \frac{L(1-f)}{u} - \frac{2\epsilon}{u}.$$

It shall be shown that once X rises above the level $\frac{L(1-f)}{u} - \frac{2\epsilon}{u}$ it can never drop beneath it. Suppose otherwise and that

$$X(t_{10}) < \frac{L(1-f)}{u} - \frac{2\epsilon}{u}$$

for some $t_{10} > t_9$. Then there exists t_{11}, t_{12} with $t_9 < t_{11} < t_{12} < t_{10}$ such that $X(t)$ is strictly monotone decreasing in $[t_{11}, t_{12}]$ and

$$\frac{L(1-f)}{u} - \frac{2\epsilon}{u} > X(t_{11}) > X(t_{12}).$$

This contradicts (2.7) which implies that $X(t)$ is monotone increasing in $[t_{11}, t_{12}]$. Therefore, once X is above $\frac{L(1-f)}{u} - \frac{2\epsilon}{u}$ it can never drop below this level. Thus,

$$X \geqslant \frac{L(1-f)}{u} - \frac{2\epsilon}{u} \quad \text{for} \quad t \geqslant t_9.$$

Thus, for $t \geqslant \max(t_4, t_9)$,

$$\left| X(t) - \frac{L(1-f)}{u} \right| \leqslant 2\frac{\epsilon}{u}.$$

But, $\epsilon > 0$ is arbitrary . Therefore,

$$X \to \frac{L(1-f)}{u} \text{ as } t \to \infty.$$

Since $X + V \to \frac{L}{u}$ as $t \to \infty$,

$$V \to \frac{L}{u} - \frac{L(1-f)}{u} = \frac{Lf}{u} \text{ as } t \to \infty.$$

Therefore, when $R_e \leqslant 1$, $X \to \frac{L(1-f)}{u}$, $V \to \frac{Lf}{u}$, $T_1 \to 0$ and $V_{T_1} \to 0$ as $t \to \infty$. Thus, it has been shown that the CFE is GAS when $R_e \leqslant 1$.

Next, it is necessary to examine the case where $R_e > 1$. In this situation both the CFE and CE exist. Here, the CE is LAS and the CFE is unstable. Once again there are two cases to consider. Firstly, what happens to the number of susceptible and carrying hosts when there are no hosts carrying MLSTs initially and secondly, what occurs when there are initial carrying hosts in the population.

- Case 1: $(T_1 + V_{T_1})(0) = 0$.

  The argument for this case is the same as that when $(T_1 + V_{T_1})(0) = 0$ and $R_e \leqslant 1$. Therefore, irrespective of the initial starting state of $X + V$, $X+V \to \frac{L}{u}$ when there are no initial hosts carrying the MLST. Furthermore, as before $X \to \frac{L(1-f)}{u}$ and $V \to \frac{Lf}{u}$ as $t \to \infty$.

- Case 2: $(T_1 + V_{T_1})(0) > 0$.

  Following the same argument as that of Case 2 when $R_e < 1$, it can be shown that $(T_1 + V_{T_1})(0) > 0 \Rightarrow (T_1 + V_{T_1})(t) > 0 \; \forall \; t > 0$.

  Choose some arbitrary $\epsilon > 0$. $\exists \; t_0$ such that for $t \geqslant t_0$

  $$\frac{L}{u} - \epsilon \leqslant N \leqslant \frac{L}{u} + \epsilon.$$

  Once again, considering (2.5),

  $$\frac{1}{T_1 + V_{T_1}} \frac{d(T_1 + V_{T_1})}{dt} = \beta_1(X + V) - (\gamma + u),$$

so

$$\frac{1}{T_1+V_{T_1}}\frac{d(T_1+V_{T_1})}{dt} = \beta_1(N-(T_1+V_{T_1}))-(\gamma+u),$$

$$\leqslant \beta_1\left(\frac{L}{u}+\epsilon\right)-\beta_1(T_1+V_{T_1})-(\gamma+u),$$

$$= (\gamma+u)(R_e-1)+\epsilon\beta_1-\beta_1(T_1+V_{T_1}). \qquad (2.8)$$

As the CE is the LAS equilibrium when $R_e > 1$, it is necessary to consider both $T_1+V_{T_1}$ increasing and decreasing to determine whether this equilibrium has global stability since $T_1+V_{T_1}$ can take values less than or greater than the value it takes at equilibrium.

To first assess $T_1+V_{T_1}$ decreasing, consider $\frac{d(T_1+V_{T_1})}{dt} < 0$. This is the case when $(\gamma+u)(R_e-1)+\epsilon\beta_1-\beta_1(T_1+V_{T_1}) < 0$. This is true when

$$T_1+V_{T_1} \geqslant \frac{(\gamma+u)(R_e-1)}{\beta_1}+\epsilon.$$

In this circumstance, $T_1+V_{T_1}$ is decreasing (and will always remain decreasing).

Lemma 2.5.1.2

$\exists\ t_1$, where $t_1 > t_0$, such that for $t \geqslant t_1$

$$T_1+V_{T_1} \leqslant \frac{(\gamma+u)(R_e-1)}{\beta_1}+2\epsilon.$$

Proof

To begin, assert that $\exists\ t_1$ such that

$$(T_1 + V_{T_1})(t_1) \leqslant \frac{(\gamma + u)(R_e - 1)}{\beta_1} + 2\epsilon.$$

The result is true if

$$(T_1 + V_{T_1})(t_0) \leqslant \frac{(\gamma + u)(R_e - 1)}{\beta_1} + 2\epsilon.$$

Otherwise, provided that

$$T_1 + V_{T_1} > \frac{(\gamma + u)(R_e - 1)}{\beta_1} + 2\epsilon$$

$$\text{which is true for } t \geqslant t_0,$$

$$\frac{1}{T_1 + V_{T_1}} \frac{d(T_1 + V_{T_1})}{dt} \leqslant -\epsilon\beta_1,$$

so

$$\int_{t_0}^{t} \frac{d}{dt} \left(\ln(T_1 + V_{T_1})\right) ds \leqslant \int_{t_0}^{t} (-\beta_1\epsilon) ds.$$

It can be deduced that

$$\ln\left(\frac{(T_1 + V_{T_1})(t)}{(T_1 + V_{T_1})(t_0)}\right) \leqslant (-\beta_1\epsilon t + \beta_1\epsilon t_0),$$

so

$$0 \leqslant (T_1 + V_{T_1})(t) \leqslant (T_1 + V_{T_1})(t_0) \exp(-\beta_1\epsilon(t - t_0)).$$

Thus, eventually

$$(T_1 + V_{T_1})(t_1) \leqslant \frac{(\gamma + u)(R_e - 1)}{\beta_1} + 2\epsilon \text{ for some } t_1 > t_0.$$

If $\exists\ t_2 > t_1$ such that

$$(T_1 + V_{T_1})(t_2) > \frac{(\gamma + u)(R_e - 1)}{\beta_1} + 2\epsilon$$

then $\exists\ t_3, t_4$ with $t_1 < t_3 < t_4 < t_2$ such that $(T_1 + V_{T_1})(t)$ is strictly monotone increasing in $[t_3, t_4]$ and

$$\frac{(\gamma + u)(R_e - 1)}{\beta_1} + \epsilon < (T_1 + V_{T_1})(t_3) < (T_1 + V_{T_1})(t_4).$$

This contradicts the fact, implied by (2.8), that $T_1 + V_{T_1}$ is monotone decreasing in $[t_3, t_4]$.

Therefore,

$$T_1 + V_{T_1} \leqslant \frac{(\gamma + u)(R_e - 1)}{\beta_1} + 2\epsilon,\ \ \forall\ t \geqslant t_1.$$

Once $T_1 + V_{T_1}$ goes beneath the level $\frac{(\gamma+u)(R_e-1)}{\beta_1} + 2\epsilon$ it can never rise above it. This completes the proof of Lemma 2.5.1.2.

Next it is of interest to assess $T_1 + V_{T_1}$ increasing. For $t \geqslant t_0$, consider again

$$\begin{aligned}
\frac{1}{T_1 + V_{T_1}} \frac{d(T_1 + V_{T_1})}{dt} &= \beta_1(N - (T_1 + V_{T_1})) - (\gamma + u), \\
&\geqslant \beta_1 \left(\frac{L}{u} - \epsilon \right) - \beta_1(T_1 + V_{T_1}) - (\gamma + u), \\
&= (\gamma + u)(R_e - 1) - \epsilon\beta_1 - \beta_1(T_1 + V_{T_1}). \quad (2.9)
\end{aligned}$$

To assess $T_1 + V_{T_1}$ increasing, it is necessary for $(\gamma + u)(R_e - 1) - \epsilon\beta_1 - \beta_1(T_1 + V_{T_1}) > 0$. This is the case when

$$T_1 + V_{T_1} \leqslant \frac{(\gamma + u)(R_e - 1)}{\beta_1} - \epsilon.$$

Lemma 2.5.1.3

$\exists\ t_5$ such that for $t \geqslant t_5$,

$$T_1 + V_{T_1} \geqslant \frac{(\gamma + u)(R_e - 1)}{\beta_1} - 2\epsilon.$$

Proof

First, assert that $\exists\ t_5$ such that

$$(T_1 + V_{T_1})(t_5) \geqslant \frac{(\gamma + u)(R_e - 1)}{\beta_1} - 2\epsilon.$$

The result is true if

$$(T_1 + V_{T_1})(t_0) \geqslant \frac{(\gamma + u)(R_e - 1)}{\beta_1} - 2\epsilon.$$

Otherwise, provided that

$$T_1 + V_{T_1} < \frac{(\gamma + u)(R_e - 1)}{\beta_1} - 2\epsilon,$$

$$\frac{1}{T_1 + V_{T_1}} \frac{d(T_1 + V_{T_1})}{dt} \geqslant \epsilon\beta_1,$$

i.e.

$$\frac{d}{dt} \left(\ln(T_1 + V_{T_1})\right) \geqslant \epsilon\beta_1,$$

so

$$(T_1 + V_{T_1})(t) \geqslant (T_1 + V_{T_1})(t_0) \exp(\epsilon\beta_1(t - t_0)).$$

As $t \to \infty$, $(T_1 + V_{T_1})(t_0) \exp(\epsilon\beta_1(t - t_0)) \to \infty$ and eventually

$$(T_1 + V_{T_1})(t_5) \geqslant \frac{(\gamma + u)(R_e - 1)}{\beta_1} - 2\epsilon \text{ for some } t_5 > 0.$$

Next, it is shown that $T_1 + V_{T_1}$ remains above this level for $t \geqslant t_5$.

If $\exists\ t_6 > t_5$ such that

$$(T_1 + V_{T_1})(t_6) < \frac{(\gamma + u)(R_e - 1)}{\beta_1} - 2\epsilon$$

then $\exists\ t_7, t_8$ with $t_5 < t_7 < t_8 < t_6$ such that $(T_1 + V_{T_1})(t)$ is strictly monotone decreasing in $[t_7, t_8]$ and

$$\frac{(\gamma + u)(R_e - 1)}{\beta_1} - \epsilon > (T_1 + V_{T_1})(t_7) > (T_1 + V_{T_1})(t_8).$$

This contradicts the fact, implied by (2.9), that $T_1 + V_{T_1}$ is monotone increasing in $[t_7, t_8]$.

Therefore,

$$T_1 + V_{T_1} \geqslant \frac{(\gamma + u)(R_e - 1)}{\beta_1} - 2\epsilon \text{ for all } t \geqslant t_5.$$

Once $T_1 + V_{T_1}$ is above $\frac{(\gamma+u)(R_e-1)}{\beta_1} - 2\epsilon$, it can never drop below this level. This completes the proof of Lemma 2.5.1.3.

Hence, combining the results of Lemmas 2.5.1.2 and 2.5.1.3, for $t \geqslant \max(t_1, t_5)$,

$$\left|(T_1 + V_{T_1})(t) - \frac{(\gamma + u)(R_e - 1)}{\beta_1}\right| < 2\epsilon.$$

But $\epsilon > 0$ is arbitrary. Therefore,

$$T_1 + V_{T_1} \to \frac{(\gamma + u)(R_e - 1)}{\beta_1} \text{ as } t \to \infty \text{ and}$$

$$X + V = N - (T_1 + V_{T_1}) \to \frac{L}{u} - \frac{(\gamma + u)(R_e - 1)}{\beta_1} = \frac{\gamma + u}{\beta_1} \text{ as } t \to \infty.$$

Thus, in summary, it is known that

$$X + V \to \frac{\gamma + u}{\beta_1}$$

and

$$T_1 + V_{T_1} \to \frac{L}{u} - \frac{\gamma + u}{\beta_1}. \tag{2.10}$$

However, it has not yet been established what X, V, T_1 and V_{T_1} each tend to separately. To establish this, consider

$$\frac{d}{dt}(X + T_1) = L(1 - f) - u(X + T_1).$$

It can easily be shown that

$$X + T_1 \to \frac{L(1 - f)}{u} \text{ as } t \to \infty.$$

In addition, consider

$$\frac{d}{dt}(V + V_{T_1}) = Lf - u(V + V_{T_1}).$$

Therefore,

$$V + V_{T_1} \to \frac{Lf}{u} \text{ as } t \to \infty.$$

Let $X = (1-f)\bar{X}$, $T_1 = (1-f)\bar{T_1}$, $V = f\bar{V}$ and $V_{T_1} = f\bar{V_{T_1}}$. So,

$$\frac{d\bar{X}}{dt} = L - u\bar{X} - \beta_1\bar{X}(T_1 + V_{T_1}) + \gamma\bar{T_1},$$

and

$$\frac{d\bar{T_1}}{dt} = \beta_1\bar{X}(T_1 + V_{T_1}) - (\gamma + u)\bar{T_1}.$$

Since $X + T_1 \to \frac{L(1-f)}{u}$ as $t \to \infty$, then it can be seen that $(1-f)\bar{X} + (1-f)\bar{T_1} \to \frac{L(1-f)}{u}$. Therefore, $\bar{X} + \bar{T_1} \to \frac{L}{u}$. As a consequence,

$$\begin{aligned} \frac{d\bar{T_1}}{dt} &\to \beta_1\left(\frac{L}{u} - \bar{T_1}\right)\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right) - (\gamma+u)\bar{T_1}, \\ &= \beta_1\frac{L}{u}\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right) - \frac{\beta_1 L}{u}\bar{T_1}. \end{aligned}$$

Consider some arbitrary $\epsilon > 0$.

$\exists\ t_{10}$ such that for $t \geqslant t_{10}$

$$\frac{d\bar{T_1}}{dt} < \beta_1\frac{L}{u}\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right) - \frac{\beta_1 L}{u}\bar{T_1} + \epsilon. \tag{2.11}$$

Hence $\bar{T_1}$ must eventually fall beneath the level $\frac{L}{u} - \frac{\gamma+u}{\beta_1} + \frac{2\epsilon u}{\beta_1 L}$. It shall be shown that once $\bar{T_1}$ goes beneath the level $\frac{L}{u} - \frac{\gamma+u}{\beta_1} + \frac{2\epsilon u}{\beta_1 L}$ it can never rise above it. Suppose otherwise, and that

$$\bar{T_1}(t_{11}) < \frac{L}{u} - \frac{\gamma+u}{\beta_1} + \frac{2\epsilon u}{\beta_1 L}$$

and

$$\bar{T_1}(t_{12}) > \frac{L}{u} - \frac{\gamma + u}{\beta_1} + \frac{2\epsilon u}{\beta_1 L} \tag{2.12}$$

for some $t_{12} > t_{11}$. Then $\exists\ t_{13}, t_{14}$ with $t_{11} < t_{13} < t_{14} < t_{12}$ such that $\bar{T_1}(t)$ is strictly monotone increasing in $[t_{13}, t_{14}]$ and

$$\frac{L}{u} - \frac{\gamma + u}{\beta_1} + \frac{\epsilon u}{\beta_1 L} < \bar{T_1}(t_{13}) < \bar{T_1}(t_{14}).$$

This contradicts (2.11) which implies that $\bar{T_1}(t)$ is monotone decreasing in $[t_{13}, t_{14}]$. Hence,

$$\bar{T_1} \leqslant \frac{L}{u} - \frac{\gamma + u}{\beta_1} + \frac{2\epsilon u}{\beta_1 L},\ \forall\ t \geqslant t_{11}.$$

Therefore, once $\bar{T_1}$ goes beneath the level $\frac{L}{u} - \frac{\gamma+u}{\beta_1} + \frac{2\epsilon u}{\beta_1 L}$ it can never rise above it.

Similarly, $\exists\ t_{15}$ such that for $t \geqslant t_{15}$,

$$\frac{d\bar{T_1}}{dt} > \beta_1 \frac{L}{u}\left(\frac{L}{u} - \frac{\gamma + u}{\beta_1}\right) - \frac{\beta_1 L}{u}\bar{T_1} - \epsilon. \tag{2.13}$$

Hence, $\bar{T_1}$ must eventually rise above the level $\frac{L}{u} - \frac{\gamma+u}{\beta_1} - \frac{2u\epsilon}{\beta_1 L}$, i.e. $\exists\ t_{16} > t_{15}$ such that

$$\bar{T_1}(t_{16}) > \frac{L}{u} - \frac{\gamma + u}{\beta_1} - \frac{2u\epsilon}{\beta_1 L}.$$

It will be shown that once $\bar{T_1}$ rises above the level $\frac{L}{u} - \frac{\gamma+u}{\beta_1} - \frac{2u\epsilon}{\beta_1 L}$ it can never drop below this level. Suppose otherwise and that

$$\bar{T_1}(t_{17}) < \frac{L}{u} - \frac{\gamma + u}{\beta_1} - \frac{2u\epsilon}{\beta_1 L}$$

for some $t_{17} > t_{16}$. Then $\exists\ t_{18}$, t_{19} with $t_{16} < t_{18} < t_{19} < t_{17}$ such that $\bar{T_1}(t)$ is strictly monotone decreasing in $[t_{18}, t_{19}]$ and

$$\frac{L}{u} - \frac{\gamma + u}{\beta_1} - \frac{\epsilon u}{\beta_1 L} > \bar{T_1}(t_{18}) > \bar{T_1}(t_{19}).$$

This contradicts (2.13) which implies that once $\bar{T_1}$ is beneath $\frac{L}{u} - \frac{\gamma+u}{\beta_1} - \frac{\epsilon u}{\beta_1 L}$ it is monotone increasing. So $\bar{T_1} \geqslant \frac{L}{u} - \frac{\gamma+u}{\beta_1} - \frac{2\epsilon u}{\beta_1 L}$ for $t \geqslant t_{16}$.

Therefore, for $t \geqslant \max(t_{11}, t_{16})$,

$$\left| \bar{T_1} - \left(\frac{L}{u} - \frac{(\gamma + u)}{\beta_1} \right) \right| \leqslant \frac{2u\epsilon}{\beta_1 L}.$$

But $\epsilon > 0$ is arbitrary. Therefore,

$$\bar{T_1} \to \frac{L}{u} - \frac{(\gamma + u)}{\beta_1} \text{ as } t \to \infty,$$

so

$$T_1 \to (1 - f) \left(\frac{L}{u} - \frac{(\gamma + u)}{\beta_1} \right) \text{ as } t \to \infty.$$

Since $X + T_1 \to \frac{L(1-f)}{u}$, it can be deduced that $X \to (1 - f)\frac{(\gamma+u)}{\beta_1}$ as $t \to \infty$. From (2.10), it can further be deduced that

$$V_{T_1} \to f \left(\frac{L}{u} - \frac{\gamma + u}{\beta_1} \right) \text{and } V \to f \frac{\gamma + u}{\beta_1} \text{as } t \to \infty.$$

In summary, the following theorem has been proven:

Theorem 2.5.1.1

(i) When $R_e \leqslant 1$, $T_1(t) \to 0$, $V_{T_1} \to 0$, $T_2(t) \to 0$, $V_{T_2}(t) \to 0$, $X(t) \to \frac{L(1-f)}{u}$ and $V(t) \to \frac{Lf}{u}$ as $t \to \infty$.

(ii) When $R_e > 1$ and both $T_1(0)$ and $V_{T_1}(0)$ equal 0, $T_1(t) = 0$, $V_{T_1}(t) = 0\ \forall\ t$, $X(t) \to \frac{L(1-f)}{u}$ and $V(t) \to \frac{Lf}{u}$ as $t \to \infty$.

(iii) When $R_e > 1$ and $(T_1 + V_{T_1})(0) > 0$, $T_1(t) \to (1-f)\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right)$, $V_{T_1}(t) \to f\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right)$, $X(t) \to (1-f)\frac{\gamma+u}{\beta_1}$ and $V(t) \to f\frac{\gamma+u}{\beta_1}$ as $\to \infty$.

2.5.2 Conclusions

The model considered in this chapter is a preliminary model created to explore the role of MLSTs in the process of serotype replacement through capsular switch. It was discussed in Chapter 1 that coexistence of pneumococcal strains within a host is a necessary prerequisite for capsular switch to occur. However, to obtain insight into MLST colonisation, it is important to consider simpler initial models from which mathematical properties for the biological process may be obtained. The results of the analysis of the mathematical model involving the carriage of one MLST which manifests as two serotypes, one VT and the other NVT, show that there are two possible steady state population sizes for children aged under 2 years in this model.

One possible equilibrium involves no MLST carriers in the population. Thus, no hosts carry either serotype at this equilibrium. As all hosts are divided amongst the four possible classes (vaccinated susceptible, unvaccinated susceptible, vaccinated carriers and unvaccinated carriers) at this equilibrium, all hosts must therefore be susceptible to carriage. The numbers in each of the vaccinated and unvaccinated classes are split according to the proportion of children who receive vaccination upon entering the population. This steady state is the limiting population value when $R_e \leqslant 1$, regardless of the initial number of vaccinated and unvaccinated carrying hosts in the population, as carriage will die out in the long term.

However, it is known that pneumococcal serotypes and MLSTs have been in existence in the population for many years. Thus, since pneumococcal colonisation prevails, it is likely that R_e must be greater than 1.

When $R_e > 1$, if no children initially carry the MLST in the population, there never will be any carriers and thus the total number of hosts will be split between the vaccinated and unvaccinated susceptible classes as in the case where $R_e < 1$.

The other possible (stable) population does involve carriage of the MLST. This equilibrium only exists when $R_e > 1$. When this is the case and there are initial carriers of the MLST in the population, the number of carrying hosts tends to $\frac{L}{u} - \frac{\gamma+u}{\beta_1}$, the rate at which hosts cease to carry the MLST, γ, plus the rate at which hosts leave the population, u, divided by the rate at which the MLST colonises hosts, β_1, subtracted from the total size of the population at equilibrium, $\frac{L}{u}$. To obtain the numbers of carrying hosts at equilibrium for the vaccinated and unvaccinated groups, this number needs to be multiplied by f and $1 - f$ respectively.

The analysis of the model shows that MLST colonisation is unaffected by vaccine intervention since the vaccine acts on the pneumococcal serotype and not the MLST. At the endemic equilibrium both serotypes are present in the population with serotype 2 able to colonise hosts who have received the vaccine. The number of hosts colonised with each serotype at equilibrium is dependent upon the proportion with which the MLST manifests as each of the serotypes and the proportion of children who receive the vaccine, as described on pages 54 and 55. As the vaccine intervention does not appear to play a role in eradicating MLST colonisation, it is important to consider the parameter P which represents the proportion of MLST carriers who are carrying serotype 1. When it is assumed that $P = 0$, no hosts are colonised with serotype 1. Thus, the vaccine has no impact since it is assumed to be effective only in preventing serotype 1 carriage. When it is assumed that $P = 1$, all unvaccinated hosts carrying the MLST will be carrying serotype 1 whilst all vaccinated hosts will be carrying serotype 2. Thus, the vaccine will not prevent MLST carriage as the MLST is associated with more than one serotype.

If MLSTs play a role in the ability of the pneumococcus to cause invasive disease and are important in the transmissibility of the pneumococcus between hosts, from the analysis in this chapter, it can be observed that invasive disease should continue to feature in the population as the vaccine will not prevent carriage of the MLST when it is associated with more than one serotype, where only one is a VT serotype.

In the next chapter, this simple one MLST model will be extended to consider more than one circulating MLST which can colonise hosts in the population. Once again, transmission according to MLST will be considered. However, in addition, a model with transmission attributable to serotype will be examined to identify whether or not the vaccine effect will be included in R_e in this case.

Chapter 3

Modelling carriage of two MLSTs in children

3.1 Introduction

In this chapter, further modelling of the carriage of pneumococcal MLSTs will be discussed, developing the concepts introduced in Chapter 2. The models considered will extend the model analysed in the previous chapter in an attempt to make the model more realistic by looking at the possibility that a carrier could be colonised with one of two different MLSTs. One or other of the MLSTs is able to manifest in more than one serotype. Once again, the effect of a conjugate vaccine will be considered. In the previous chapter the MLST results were independent of the vaccine effect. In this chapter, the relationship between MLSTs and VT and NVT serotypes will be explored further to identify whether or not the vaccine has an effect in more complex models. Two models will be discussed, assuming different mechanisms of transmission. In the first model, the transmission is assumed to be attributable to MLST whilst in the second the transmission is assumed to be attributable to serotype.

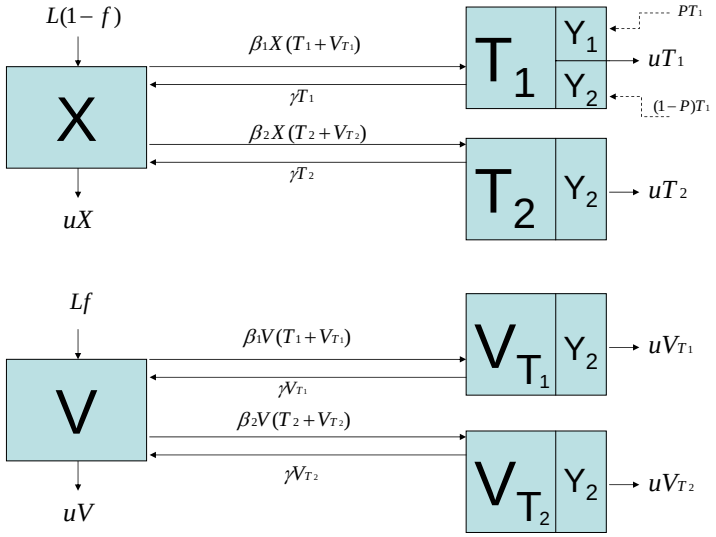

Figure 3.1: Model of two MLSTs; one associated with two serotypes, the other only one. Transmission is due to MLST and the vaccine is assumed to be effective against one serotype.

3.2 Model of two MLSTs with transmission due to MLST

This model assumes that vaccinated and unvaccinated susceptible hosts, X and V, may become colonised with either MLST 1 or MLST 2. T_1 and V_{T_1} respectively represent the number of unvaccinated and vaccinated hosts carrying MLST 1, whilst T_2 and V_{T_2} respectively represent those hosts carrying MLST 2. The vaccine is 100% effective in preventing carriage of serotype 1 but ineffective in preventing carriage of serotype 2. It is assumed that, without intervention, MLST 1 is able to manifest in either serotype 1, Y_1, or serotype 2, Y_2, with proportions PT_1 and $(1-P)T_1$ respectively; MLST 2 is associated only with serotype 2. Therefore, the vaccine will not eradicate carriage of either MLST 1 or 2. Those vaccinated hosts carrying MLST 1 must all be carrying this MLST as serotype 2. In addition, in this model it is assumed that the transmission parameter, β, is determined by MLST and not serotype.

The six ODEs that correspond to this model (Figure 3.1) are:

$$\frac{dX}{dt} = L(1-f) - uX - \beta_1 X(T_1 + V_{T_1}) + \gamma(T_1 + T_2) - \beta_2 X(T_2 + V_{T_2}),$$

$$\frac{dT_1}{dt} = \beta_1 X(T_1 + V_{T_1}) - (\gamma + u)T_1,$$

$$\frac{dT_2}{dt} = \beta_2 X(T_2 + V_{T_2}) - (\gamma + u)T_2,$$

$$\frac{dV}{dt} = Lf - uV - \beta_1 V(T_1 + V_{T_1}) + \gamma(V_{T_1} + V_{T_2}) - \beta_2 V(T_2 + V_{T_2}),$$

$$\frac{dV_{T_1}}{dt} = \beta_1 V(T_1 + V_{T_1}) - (\gamma + u)V_{T_1},$$

and

$$\frac{dV_{T_2}}{dt} = \beta_2 V(T_2 + V_{T_2}) - (\gamma + u)V_{T_2}.$$

In the next section, the analysis of this model will be presented. As with the model presented in Chapter 2, the equilibrium solutions and effective reproductive number are determined and both local and global stability analyses are carried out.

3.2.1 Results

Equilibrium solutions

Using a similar approach to that described in the previous chapter, the equilibria for this two MLST model were found. The details are omitted for brevity but are similar to the derivation of the equilibrium results for the case when transmission depends on serotype explained in detail in the second model in this chapter. For this model, three steady state populations are possible: one carriage-free and two carriage equilibria. Assuming that $\beta_1 \neq \beta_2$, these equilib-

ria, $(X_e, T_{1_e}, T_{2_e}, V_e, V_{T_{1_e}}, V_{T_{2_e}})$, are:

$$\left(\frac{L}{u}(1-f), 0, 0, \frac{L}{u}f, 0, 0\right), \tag{3.1}$$

$$\left((1-f)\frac{\gamma+u}{\beta_1}, (1-f)\left(\frac{L}{u}-\frac{\gamma+u}{\beta_1}\right), 0, f\frac{\gamma+u}{\beta_1}, f\left(\frac{L}{u}-\frac{\gamma+u}{\beta_1}\right), 0\right), \tag{3.2}$$

and

$$\left((1-f)\frac{\gamma+u}{\beta_2}, 0, (1-f)\left(\frac{L}{u}-\frac{\gamma+u}{\beta_2}\right), f\frac{\gamma+u}{\beta_2}, 0, f\left(\frac{L}{u}-\frac{\gamma+u}{\beta_2}\right)\right). \tag{3.3}$$

The first equilibrium (the CFE) is always feasible. The second equilibrium (a CE) is feasible if and only if $\frac{\beta_1 L}{u(\gamma+u)} \geqslant 1$ and the third equilibrium (also a CE) is feasible if and only if $\frac{\beta_2 L}{u(\gamma+u)} \geqslant 1$.

One CE refers only to carriage of MLST 1 and the other involves only carriage of MLST 2. Thus, at equilibrium the two MLSTs cannot coexist in the population. The number of susceptible hosts in each of these CE solutions only differ in terms of the transmission parameter, with the CE for carriage of MLST 1 dependent upon the transmission of MLST 1, β_1, and the CE for carriage of MLST 2 dependent upon transmission of MLST 2, β_2.

In the case where $\beta_1 = \beta_2 = \beta$, one equilibrium is again

$$\left(\frac{L}{u}(1-f), 0, 0, \frac{L}{u}f, 0, 0\right).$$

Additionally, if $\frac{\beta L}{u(\gamma+u)} > 1$ then any solution of the form

$$\left((1-f)\frac{\gamma+u}{\beta}, (1-f)\alpha\xi, (1-f)(1-\alpha)\xi, f\frac{\gamma+u}{\beta}, f\alpha\xi, f(1-\alpha)\xi\right)$$

with $\xi = \frac{L}{u} - \frac{\gamma+u}{\beta}$ and $0 < \alpha < 1$ is a feasible CE. Thus, in this case neither

MLST dominates and the two MLSTs can coexist in the population.

Effective reproductive number

R_e can take one of two possible values for this two MLST model. These values differ in terms of the transmission parameter. R_e is defined to be:

$$R_e = \max(\beta_1, \beta_2)\frac{L}{u(\gamma + u)}.$$

To see this, the matrix $\mathbf{M}$ is calculated as in Section 2.5.1 in the previous chapter with four types of carrying hosts: unvaccinated carriers of MLST 1 ($i = 1$); unvaccinated carriers of MLST 2 ($i = 2$); vaccinated carriers of MLST 1 ($i = 3$); vaccinated carriers of MLST 2 ($i = 4$). Then

$$\mathbf{M} = \begin{bmatrix} \frac{\beta_1 L(1-f)}{u(\gamma+u)} & 0 & \frac{\beta_1 L(1-f)}{u(\gamma+u)} & 0 \\ 0 & \frac{\beta_2 L(1-f)}{u(\gamma+u)} & 0 & \frac{\beta_2 L(1-f)}{u(\gamma+u)} \\ \frac{\beta_1 Lf}{u(\gamma+u)} & 0 & \frac{\beta_1 Lf}{u(\gamma+u)} & 0 \\ 0 & \frac{\beta_2 Lf}{u(\gamma+u)} & 0 & \frac{\beta_2 Lf}{u(\gamma+u)} \end{bmatrix}.$$

The largest eigenvalue of $\mathbf{M}$ is R_e above. For the purpose of later discussion, let $R_{e1} = \frac{\beta_1 L}{u(\gamma+u)}$ and $R_{e2} = \frac{\beta_2 L}{u(\gamma+u)}$.

Local stability analysis

If it is assumed that $\beta_1 > \beta_2 > 0$ then $R_{e1} > R_{e2}$. Thus, in this situation, R_{e1} is the effective reproductive number. The mathematics of carrying out a LSA was discussed in depth in the previous chapter for the simpler one MLST model. Therefore, here the mathematical detail will be omitted and a summary of the findings are provided.

There are three possible cases to consider for this analysis; $R_{e1} < 1$ and $R_{e2} < 1$, $R_{e1} > 1 > R_{e2}$ and $R_{e1} > R_{e2} > 1$. In the first scenario, the CFE, (3.1), is the only equilibrium possible and is LAS. In the second situation, the only equilibria possible are the CFE and the CE involving carriage of only MLST 1, (3.2). In

this case the CFE is unstable and the CE is LAS. In the third case, all equilibria are possible. Both the CFE and the CE involving carriage of MLST 2, (3.3), are unstable and the CE involving carriage of MLST 1 is LAS.

Therefore, in each case there is just one LAS equilibrium. The CE referring to carriage of MLST 1 is always LAS when it exists and the CE for carriage of MLST 2 is always unstable when it exists.

A similar argument follows when $\beta_2 > \beta_1 > 0$, where the CE referring to carriage of MLST 2 is always LAS when it exists and that of MLST 1 is unstable when it exists.

To summarise, if the effective reproductive number is less than 1, the CFE is the only possible equilibrium and is LAS. If it is greater than 1, the CFE is unstable. If $R_{e1} > R_{e2}$, then when $R_e > 1$, the equilibrium corresponding to carriage of MLST 1, is LAS and the equilibrium corresponding to carriage of MLST 2 either does not exist if $R_{e2} \leqslant 1$, or is unstable if $R_{e2} > 1$. Similarly, if $R_{e2} > R_{e1}$ and $R_{e2} > 1$, then the equilibrium corresponding to carriage of MLST 2 is LAS and the equilibrium corresponding to carriage of MLST 1 either does not exist or is unstable.

Global stability analysis

As discussed in Chapter 2, a GSA is preferred to a LSA to avoid the issue of determining how close the initial population sizes must be to the sizes shown in the equilibrium for the population sizes to tend to that steady state. Global stability can be difficult to prove. However, fortunately a GSA is possible for this model.

Assume once again that $\beta_1 > \beta_2$. Thus, $R_{e1} > R_{e2}$ so $R_e = R_{e1} = \frac{\beta_1 L}{u(\gamma+u)}$. Initially, as with the GSA described in previous chapter, vaccinated and unvaccinated susceptible and carrying hosts are grouped together. Thus, for this model, three ODEs are formed as this model also considers hosts carrying a second MLST.

These ODEs are described below.

$$\frac{d(X+V)}{dt} = L - (X+V)(u - \beta_1(T_1 + V_{T_1}) - \beta_2(T_2 + V_{T_2})) + \gamma(T_1 + T_2 + V_{T_1} + V_{T_2}),$$

$$\frac{d(T_1 + V_{T_1})}{dt} = \beta_1(X+V)(T_1 + V_{T_1}) - (\gamma + u)(T_1 + V_{T_1}),$$

and

$$\frac{d(T_2 + V_{T_2})}{dt} = \beta_2(X+V)(T_2 + V_{T_2}) - (\gamma + u)(T_2 + V_{T_2}). \tag{3.4}$$

The analysis involves the same approach as that described in Chapter 2. Thus, the first scenario to consider is when $R_e \leqslant 1$. As usual in this case, the CFE is the only equilibrium possible and is LAS if $R_e < 1$ and it is of interest to consider various cases involving the number of initial carrying hosts. The cases considered are what happens when there are no carrying hosts initially; what happens when there are no hosts carrying MLST 1 initially but there are carriers of MLST 2; what happens when there are hosts carrying MLST 1 initially but no carriers of MLST 2; what happens when there are hosts carrying MLST 1 and hosts carrying MLST 2 in the initial population.

The argument for the first case is very similar to that of the one MLST model as $(T_1 + V_{T_1})(0) = (T_2 + V_{T_2})(0) = 0 \Rightarrow (T_1 + V_{T_1})(t) = (T_2 + V_{T_2})(t) = 0 \ \forall$ t. As it is known that T_1, T_2, V_{T_1} and V_{T_2} are all greater than or equal to zero since they represent numbers of hosts, when $T_1 + V_{T_1} = 0$, T_1 and V_{T_1} must both equal zero. The same is true of T_2 and V_{T_2} when $T_2 + V_{T_2} = 0$. By substituting $T_1 = V_{T_1} = T_2 = V_{T_2} = 0$ into the ODE for $X+V$, it can be seen that $X+V \to \frac{L}{u}$ as $t \to \infty$. In this case, the same argument as that adopted in the one MLST model can be used to identify what X and V each tend to. This time $T_2+V_{T_2} = 0$ must be substituted into the equation for $\frac{dX}{dt}$ as well as $T_1 + V_{T_1} = 0$. Therefore, it can be concluded that $X \to \frac{L(1-f)}{u}$ and $V \to \frac{Lf}{u}$ as $t \to \infty$.

In the second case when there are no hosts carrying MLST 1 initially but there are carriers of MLST 2, $(T_1 + V_{T_1})(0) = 0$ implies that $(T_1 + V_{T_1})(t) = 0 \ \forall \ t$ so

that $T_1(t) = 0$ and $V_{T_1}(t) = 0 \; \forall \; t$ as above. In this situation, substituting $T_1 = 0$ and $V_{T_1} = 0$ into the first ODE and the third ODE, (3.4), leaves the following equations:

$$\frac{d(X+V)}{dt} = L - u(X+V) - \beta_2(X+V)(T_2 + V_{T_2}) + \gamma(T_2 + V_{T_2}), \qquad (3.5)$$

and

$$\frac{d(T_2 + V_{T_2})}{dt} = \beta_2(X+V)(T_2 + V_{T_2}) - (\gamma + u)(T_2 + V_{T_2}). \qquad (3.6)$$

Considering these equations, it can be seen that the proof for this part of the GSA is the same as that for Case 2 of the GSA in Chapter 2, with T_2 replacing T_1 in the arguments and V_{T_2} replacing V_{T_1}. It can be shown, using the same approach as in Case 2, that $(T_2 + V_{T_2})(0) > 0$ implies that $(T_2 + V_{T_2})(t) > 0 \; \forall \; t$, a requirement for this analysis. Following the rest of the proof in Case 2, it can be concluded that for this situation when $R_e \leqslant 1$, $T_2 + V_{T_2} \to 0$ and $X + V \to \frac{L}{u}$ as $t \to \infty$, regardless of the initial values of $T_2 + V_{T_2}$ and $X + V$. Hence, $T_2 \to 0$, $V_{T_2} \to 0$ and, from the results from the GSA in Chapter 2, $X \to \frac{L(1-f)}{u}$ and $V \to \frac{Lf}{u}$ as $t \to \infty$.

Similarly, in the third case when there are no hosts carrying MLST 2 initially but there are carriers of MLST 1, if $(T_2 + V_{T_2})(0) = 0$ and $(T_1 + V_{T_1})(0) > 0$, substituting $T_2 = 0$ and $V_{T_2} = 0$ leaves two ODEs in terms of $X + V$ and $T_1 + V_{T_1}$. The results in this situation are analogous to those detailed for $(T_1 + V_{T_1})(0) = 0$ and $(T_2 + V_{T_2})(0) > 0$.

Next, in the case where both MLSTs are present in the initial population, $(T_1 + V_{T_1})(0) > 0$ and $(T_2 + V_{T_2})(0) > 0$, is assessed. By using the same ideas as in Case 2 of the GSA in Chapter 2, it can be shown that $(T_1 + V_{T_1})(t) > 0$ and $(T_2 + V_{T_2})(t) > 0 \; \forall \; t$. For this part of the GSA, all carrying hosts must be considered together. Let $T = T_1 + V_{T_1} + T_2 + V_{T_2}$ and for $R_e < 1$ choose $\epsilon > 0$ such that $-k_0 = (\gamma + u)(R_e - 1) + \beta_1\epsilon < 0$. $\exists \; t_0$ such that $X + V \leqslant \frac{L}{u} + \epsilon$ for $t \geqslant t_0$. Since $\beta_1 > \beta_2$, for $t \geqslant t_0$

$$\frac{dT}{dt} \leqslant \beta_1 \left(\frac{L}{u} + \epsilon \right) T - (\gamma + u)T,$$

so

$$\frac{1}{T}\frac{dT}{dt} \leqslant \beta_1 \left(\frac{L}{u} + \epsilon \right) - (\gamma + u) = -k_0 < 0.$$

Then

$$\int_{t_0}^{t} \frac{1}{T}\frac{dT}{dt} ds \leqslant \int_{t_0}^{t} -k_0 ds.$$

$0 \leqslant T(t) \leqslant T(t_0)\exp(-k_0(t - t_0))$. Hence, as $t \to \infty$, $T(t) \to 0$. Therefore, $T_1 \to 0$, $V_{T_1} \to 0$, $T_2 \to 0$ and $V_{T_2} \to 0$ as $t \to \infty$.

When $R_e = 1$ and $(T_1 + V_{T_1})(0) > 0$ and $(T_2 + V_{T_2})(0) > 0$, it is straightforward to modify the argument given in Case 2 of the GSA in Chapter 2 to show that here also $T_1 \to 0$, $V_{T_1} \to 0$, $T_2 \to 0$ and $V_{T_2} \to 0$ as $t \to \infty$.

Since $X + V = N - T$, $X + V \to \frac{L}{u}$ as $t \to \infty$. Once again, following the previous argument, $X \to \frac{L(1-f)}{u}$ and $V \to \frac{Lf}{u}$.

To this point, the GSA for $R_e \leqslant 1$ has been shown. Next $R_e > 1$ is considered in the GSA. In this situation, the CFE is not the only possible equilibrium. The same three cases as considered for $R_e \leqslant 1$ must be considered once again for $R_e > 1$. That is, when there are no carrying hosts initially, when there is one type of MLST being carried initially and when both MLSTs are present in the initial population.

In the case of no initial carrying hosts the results are the same as those obtained when $R_e \leqslant 1$ and there are no hosts colonised with either MLST initially.

When $(T_1 + V_{T_1})(0) > 0$ but $(T_2 + V_{T_2})(0) = 0$ it can be shown, as in the previous analysis, that there never will be any hosts carrying MLST 2. In this situation, the global stability arguments of the one MLST model when $R_e > 1$ and $(T_1 +$

$V_{T_1})(0) > 0$ hold. Thus, $X \to (1-f)\frac{\gamma+u}{\beta_1}$, $T_1 \to (1-f)(\frac{L}{u} - \frac{\gamma+u}{\beta_1})$, $V \to f\frac{\gamma+u}{\beta_1}$ and $V_{T_1} \to f(\frac{L}{u} - \frac{\gamma+u}{\beta_1})$ as $t \to \infty$. When there are hosts initially carrying MLST 2 but not MLST 1 the situation is slightly different as it has been assumed that $R_e = R_{e1}$ which corresponds to transmission of MLST 1. Thus, when hosts initially carry MLST 2, it is necessary to consider both $R_{e2} \leqslant 1$ and $R_{e2} > 1$. When there are no hosts carrying MLST 1 initially, there never will be hosts carrying MLST 1. When $R_{e2} \leqslant 1$, $X + V \to \frac{L}{u}$ and $T_2 + V_{T_2} \to 0$. Therefore, following previous analysis, $X \to (1-f)\frac{L}{u}$, $V \to f\frac{L}{u}$, $T_2 \to 0$ and $V_{T_2} \to 0$ as $t \to \infty$. From the GSA of the one MLST model, when $R_{e2} > 1$, $X \to (1-f)\frac{\gamma+u}{\beta_2}$, $T_2 \to (1-f)(\frac{L}{u} - \frac{\gamma+u}{\beta_2})$, $V \to f\frac{\gamma+u}{\beta_2}$ and $V_{T_2} \to f(\frac{L}{u} - \frac{\gamma+u}{\beta_2})$ as $t \to \infty$.

When there are hosts present in the initial population carrying both of the two MLSTs and $\beta_1 > \beta_2$, then $X \to (1-f)\frac{\gamma+u}{\beta_1}$, $V \to f\frac{\gamma+u}{\beta_1}$, $T_1 \to (1-f)\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right)$, $V_{T_1} \to f\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right)$, $T_2 \to 0$ and $V_{T_2} \to 0$ as $t \to \infty$. The proof is complicated and is given in Appendix A.1 for simplicity of presentation.

Finally it is possible to show that if $R_{e1} = R_{e2} > 1$ and at least one MLST is initially present, X, T_1, V, V_{T_1} and V_{T_2} approach an equilibrium point where

$$X = (1-f)\left(\frac{L}{u} - \xi\right),\ T_1 = (1-f)\alpha\xi,\ T_2 = (1-f)(1-\alpha)\xi,$$
$$V = f\left(\frac{L}{u} - \xi\right), V_{T_1} = f\alpha\xi \text{ and } V_{T_2} = f(1-\alpha)\xi \text{ for}$$
$$\alpha = \frac{1}{1+k} \text{ and } 0 \leqslant \alpha \leqslant 1.$$

Here $\xi = \frac{L}{u} - \frac{\gamma+u}{\beta}$ where $\beta = \beta_1 = \beta_2$ and k is given in terms of the initial conditions by $\frac{T_2(0)+V_{T_2}(0)}{T_1(0)+V_{T_1}(0)}$. The proof is omitted for reasons of brevity but it is a simplified version of the corresponding proof for the second model in the chapter which is given in Appendix A.2.

To summarise, the following results have been proved:

Theorem 3.2.1.1

(i) When the effective reproductive number is less than or equal to one, the CFE

is the only possible equilibrium. In this situation, regardless of the number of hosts carrying either MLST 1 or MLST 2, the number of carrying hosts will tend to zero in the long term. The number of susceptible and vaccinated individuals will tend to their CFE values.

(ii) a) When the effective reproductive number is greater than one, if there are no hosts initially carrying either of the MLSTs, there will never be any hosts carrying either MLST.

(ii) b) If $R_e = R_{e1} > 1 \geqslant R_{e2}$ then if there are hosts carrying MLST 2 but not MLST 1, there will never be any hosts carrying MLST 1 and the number of hosts carrying MLST 2 will tend to zero. If $R_e = R_{e1} > R_{e2} > 1$ then under the same initial conditions the population sizes tend to those shown in the CE for carriage of MLST 2, (3.3).

(ii) c) However, if $R_{e1} > R_{e2}$ and there are hosts carrying MLST 1 initially, regardless of whether there are any hosts carrying MLST 2 initially, in the long term the number of hosts carrying MLST 2 will tend to zero and the population will tend to the population sizes described in the CE for carriage of MLST 1, (3.2).

(ii) d) If $R_{e2} > R_{e1}$ the situations above are reversed.

(ii) e) If $R_{e1} = R_{e2} > 1$ then the CE is a line of equilibria and in the long term coexistence of both MLSTs will occur along this line if both are initially present, see above for details.

3.2.2 Conclusions

The analysis of this two MLST carriage model shows that there are three possible equilibrium solutions when $\beta_1 > \beta_2$ or $\beta_2 > \beta_1$. One of the equilibria involves only susceptible hosts, the other two involve carriage of one or other MLST but not both. The two MLSTs are only able to coexist in the population should the transmission parameter for MLST 1, β_1, equal that of MLST 2, β_2.

The results show that the effective reproductive number is dependent upon the transmission parameters in the model. If MLST 1 has a higher transmission

than MLST 2 then MLST 1 will remain in the population in the long term if $R_{e1} > 1$ and there are hosts initially carrying MLST 1. This result is irrespective of whether or not there are hosts initially carrying MLST 2. If MLST 1 has higher transmission and there are no hosts carrying MLST 1 in the initial population but carriers of MLST 2 exist then MLST 2 will remain in the population if $R_{e2} > 1$. Thus, it is clear that the coexistence of MLSTs in the population is not possible for this model, unless $R_{e1} = R_{e2} > 1$. If MLST 2 has higher transmission then the situation is reversed. It is known that hosts may become dually colonised with pneumococcal serotypes and this phenomenon was considered in the Lipsitch two serotype model discussed in Chapter 2 (Lipsitch 1997). The Lipsitch model showed coexistence of the serotypes within the population. It is plausible that if a host can be dually colonised with pneumococcal serotypes that a host may be dually colonised with pneumococcal MLSTs. Thus, perhaps a model involving coexistence of the MLSTs should be considered. The main conclusion from the modelling appears to be that in general in order to have coexistence of MLSTs within the population it is necessary to have coexistence of MLSTs within an individual host.

Considering the serotypes with which the MLSTs are associated, MLST 2 is only associated with serotype 2. Thus, as the carriage equilibria show elimination of one or other MLST, when $R_e > 1$ and MLST 2 has higher transmission than MLST 1 then serotype 1 will be eliminated from the population. If MLST 1 remains present in the population then both serotypes shall remain as although the vaccine eradicates carriage of serotype 1, those unvaccinated hosts may still be colonised with MLST 1 in the form of serotype 1.

A special case of this two MLST model involves no intervention. Assuming that no individuals receive vaccine, setting $V = 0$, $V_{T_1} = 0$ and $V_{T_2} = 0$ will give the three ODEs for $\frac{dX}{dt}$, $\frac{dT_1}{dt}$ and $\frac{dT_2}{dt}$ that describe this system. Once again, this two MLST model involves three equilibria; the CFE and two CE. One of the CEs involves carriage of MLST 1 and the other MLST 2. Thus, in general this model also results in elimination of one or other serotype when $R_e > 1$ and there are initial carrying hosts. The equilibria for this case can be formally derived from the equilibria when there is vaccination by setting $f = 0$.

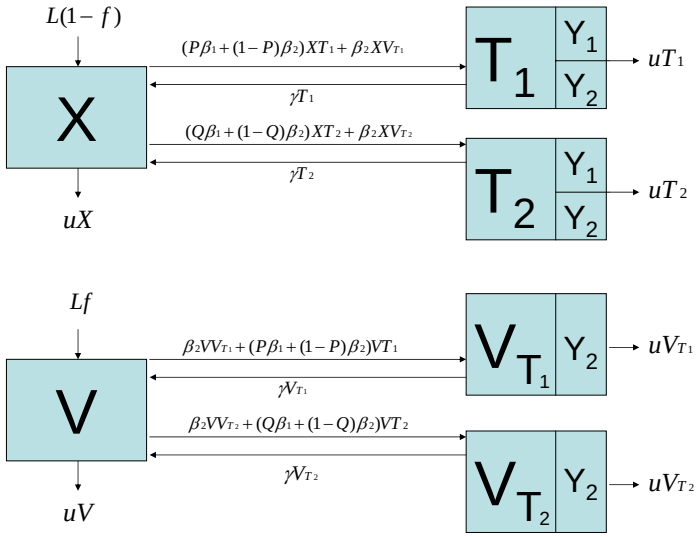

Figure 3.2: Model of two MLSTs; both associated with two serotypes. Transmission is due to serotype and the vaccine is assumed to be effective against one serotype.

3.3 Model of two MLSTs with transmission due to serotype

As with the previous model discussed in this chapter, it is assumed that hosts can become colonised with one or other of the two sequence types, MLST 1 and MLST 2. However, in this model it is assumed that MLST 1 and MLST 2 are both associated with serotypes 1 and 2. MLST 1 can manifest as each serotype with proportions PT_1 and $(1-P)T_1$ as in the previous model while MLST 2 is found as serotype 1 with proportion QT_2 and as serotype 2 with proportion $(1-Q)T_2$. However, as the vaccine is assumed to be 100% effective in preventing carriage of serotype 1, all vaccinated carriers of MLST 1 and 2 must be carrying serotype 2. Once again, a proportion f of hosts entering the population are assumed to receive the vaccine. The key difference between this model and the previous two MLST model is that it is assumed that transmission is attributable to serotype, not MLST. This can be seen on comparison of Figure 3.2 to Figure 3.1 for the previous model.

The six ODEs that correspond to this model (Figure 3.2) are:

$$\begin{aligned}\frac{dX}{dt} &= L(1-f) - uX - (P\beta_1 + (1-P)\beta_2)XT_1 - \beta_2 XV_{T_1} \\ &\quad - (Q\beta_1 + (1-Q)\beta_2)XT_2 - \beta_2 XV_{T_2} + \gamma(T_1 + T_2),\end{aligned}$$

$$\frac{dT_1}{dt} = (P\beta_1 + (1-P)\beta_2)XT_1 + \beta_2 XV_{T_1} - (\gamma + u)T_1,$$

$$\frac{dT_2}{dt} = (Q\beta_1 + (1-Q)\beta_2)XT_2 + \beta_2 XV_{T_2} - (\gamma + u)T_2,$$

$$\begin{aligned}\frac{dV}{dt} &= Lf - uV - \beta_2 VV_{T_1} - (P\beta_1 + (1-P)\beta_2)VT_1 - \beta_2 VV_{T_2} \\ &\quad - (Q\beta_1 + (1-Q)\beta_2)VT_2 + \gamma(V_{T_1} + V_{T_2}),\end{aligned}$$

$$\frac{dV_{T_1}}{dt} = \beta_2 VV_{T_1} + (P\beta_1 + (1-P)\beta_2)VT_1 - (\gamma + u)V_{T_1},$$

and

$$\frac{dV_{T_2}}{dt} = \beta_2 VV_{T_2} + (Q\beta_1 + (1-Q)\beta_2)VT_2 - (\gamma + u)V_{T_2}. \tag{3.7}$$

In the next section, the results for this model are presented. As before, these include the equilibrium solutions, effective reproductive number, local and global stability.

3.3.1 Results

Equilibrium solutions

In this model, as with the previous two MLST model, there are three equilibria. The CFE is identical to that of the earlier model. The two carriage equilibria, (X_e, T_{1_e}, T_{2_e}, V_e, $V_{T_{1_e}}$, $V_{T_{2_e}}$), are shown below:

$$\left(\frac{(1-f)(\gamma+u)}{(1-f)(P\beta_1+(1-P)\beta_2)+f\beta_2}, (1-f)\left(\frac{L}{u}-\frac{(\gamma+u)}{(1-f)(P\beta_1+(1-P)\beta_2)+f\beta_2}\right), 0,\right.$$

$$\left. f\frac{(\gamma+u)}{(1-f)(P\beta_1+(1-P)\beta_2)+f\beta_2}, f\left(\frac{L}{u}-\frac{(\gamma+u)}{(1-f)(P\beta_1+(1-P)\beta_2)+f\beta_2}\right), 0\right), \tag{3.8}$$

and

$$\left(\frac{(1-f)(\gamma+u)}{(1-f)(Q\beta_1+(1-Q)\beta_2)+f\beta_2}, 0, (1-f)\left(\frac{L}{u}-\frac{(\gamma+u)}{(1-f)(Q\beta_1+(1-Q)\beta_2)+f\beta_2}\right),\right.$$

$$\left. f\frac{(\gamma+u)}{(1-f)(Q\beta_1+(1-Q)\beta_2)+f\beta_2}, 0, f\left(\frac{L}{u}-\frac{(\gamma+u)}{(1-f)(Q\beta_1+(1-Q)\beta_2)+f\beta_2}\right)\right). \tag{3.9}$$

Once again, it can be observed that one CE corresponds to carriage of MLST 1, the other to carriage of MLST 2.

As the derivation of these equilibrium solutions differs from that of those shown in Chapter 2, the process is described. As usual, the time derivatives are set equal to zero and solved simultaneously to obtain the solutions.

From (3.7), $\frac{dT_1}{dt} = 0$ is rearranged to obtain an expression for T_1 in terms of X and V_{T_1}. For ease of expression, let $R = P\beta_1 + (1-P)\beta_2$. Then

$$T_1 = \frac{\beta_2 X V_{T_1}}{(\gamma+u) - RX}. \tag{3.10}$$

(3.10) is substituted into $\frac{dV_{T_1}}{dt} = 0$ to obtain

$$\beta_2 V V_{T_1} + \frac{VR\beta_2 X V_{T_1}}{(\gamma+u) - RX} - (\gamma+u)V_{T_1} = 0,$$

so

$$V_{T_1}\left(\beta_2 V + \frac{VR\beta_2 X}{(\gamma+u)-RX} - (\gamma+u)\right) = 0.$$

Therefore, either $V_{T_1} = 0$ or

$$V\left(\beta_2 + \frac{R\beta_2 X}{(\gamma+u)-RX}\right) = (\gamma+u),$$

i.e.

$$V\left(\frac{\beta_2(\gamma+u)}{(\gamma+u)-RX}\right) = (\gamma+u).$$

Thus,

$$V = \frac{(\gamma+u)-RX}{\beta_2}.$$

Let $K = Q\beta_1 + (1-Q)\beta_2$ and substitute $V = \frac{(\gamma+u)-RX}{\beta_2}$ into $\frac{dV_{T_2}}{dt} = 0$. This gives

$$\beta_2\left(\frac{(\gamma+u)-RX}{\beta_2}\right)V_{T_2} + \left(\frac{(\gamma+u)-RX}{\beta_2}\right)KT_2 - (\gamma+u)V_{T_2} = 0.$$

Hence, clearly,

$$V_{T_2} = \left(\frac{(\gamma+u)-RX}{\beta_2}\right)\frac{KT_2}{RX}.$$

Substituting this expression for V_{T_2} into $\frac{dT_2}{dt} = 0$ gives

$$XKT_2 + \beta_2 X\left(\frac{(\gamma+u)-RX}{\beta_2}\right)\frac{KT_2}{RX} - (\gamma+u)T_2 = 0.$$

It can be deduced that either $T_2 = 0$ or $XK + \frac{(\gamma+u)K}{R} - XK - (\gamma+u) = 0$. In the latter case this gives $(\gamma+u)\left(\frac{K}{R}-1\right) = 0$, i.e. $\frac{K}{R} = 1$ since $(\gamma+u) \neq 0$. This means $K = R$, i.e. $Q\beta_1 + (1-Q)\beta_2 = P\beta_1 + (1-P)\beta_2$. Thus, to find the equilibrium solutions, $V_{T_1} = 0$, $T_2 = 0$ and $K = R$ all must be considered.

When $V_{T_1} = 0$, $T_1 = 0$ or $V = 0$ from $\frac{dV_{T_1}}{dt} = 0$. When $V_{T_1} = V = 0$, this implies that $V_{T_2} = 0$ from $\frac{dV_{T_2}}{dt} = 0$. However, at equilibrium $V + V_{T_1} + V_{T_2} = \frac{Lf}{u}$. Thus,

this cannot be the case. Therefore, $V \neq 0$ so $T_1 = 0$ is the solution. When $V_{T_1} = T_1 = 0$, the equilibrium equations to consider are:

$$L(1-f) - uX - KXT_2 - \beta_2 XV_{T_2} + \gamma T_2 = 0,$$

$$KXT_2 + \beta_2 XV_{T_2} - (\gamma + u)T_2 = 0,$$

$$Lf - uV - \beta_2 VV_{T_2} - KVT_2 + \gamma V_{T_2} = 0,$$

and

$$\beta_2 VV_{T_2} + KVT_2 - (\gamma + u)V_{T_2} = 0. \tag{3.11}$$

Hence, $X + T_2 = \frac{L(1-f)}{u}$, $V + V_{T_2} = \frac{Lf}{u}$ and

$$\frac{T_2}{V_{T_2}} = \frac{X}{V} = \frac{\frac{L}{u}(1-f) - T_2}{\frac{L}{u}f - V_{T_2}}.$$

It can be seen that, after simplifications, $fT_2 = (1-f)V_{T_2}$. Therefore, $T_2 = (1-f)\upsilon$ and $V_{T_2} = f\upsilon$, where $T_2 + V_{T_2} = \upsilon$ and so $X = \frac{L(1-f)}{u} - (1-f)\upsilon$ and $V = \frac{Lf}{u} - f\upsilon$. Substituting these expressions involving υ for T_2, V_{T_2}, X and V into (3.11)(ii) added to (3.11)(iv) gives

$$\left(\frac{L}{u} - \upsilon\right) K(1-f)\upsilon + \beta_2 \left(\frac{L}{u} - \upsilon\right) f\upsilon - (\gamma + u)\upsilon = 0.$$

Thus, $\upsilon = 0$ or $\beta_2 f\left(\frac{L}{u} - \upsilon\right) + K(1-f)\left(\frac{L}{u} - \upsilon\right) - (\gamma + u) = 0$. Let $y = \frac{L}{u} - \upsilon$. Then

$$y = \frac{\gamma + u}{K(1-f) + \beta_2 f}.$$

This means that

$$\upsilon = \frac{L}{u} - \frac{(\gamma + u)}{K(1-f) + \beta_2 f}. \tag{3.12}$$

When $\upsilon = 0$, since $T_2 + V_{T_2} = 0$, $T_2 = 0$ and $V_{T_2} = 0$. This proves one equilibrium solution is the CFE. When υ takes the value shown, (3.12),

$$T_2 = (1-f)\left(\frac{L}{u} - \frac{(\gamma+u)}{K(1-f)+\beta_2 f}\right),$$

$$V_{T_2} = f\left(\frac{L}{u} - \frac{(\gamma+u)}{K(1-f)+\beta_2 f}\right),$$

$$X = \frac{(1-f)(\gamma+u)}{K(1-f)+\beta_2 f},$$

and

$$V = \frac{f(\gamma+u)}{K(1-f)+\beta_2 f}.$$

This is the CE shown previously, (3.9).

When $T_2 = 0$, either $X = 0$ or $V_{T_2} = 0$. However, $X \neq 0$ as $X + T_1 + T_2 = \frac{L}{u}(1-f)$ and $X = 0$ implies that $T_1 = 0$. Thus, $T_2 = V_{T_2} = 0$ and the procedure for finding the equilibrium solution (3.8) is the same as that for the other CE solution.

Finally, $K = R$ must be considered. Once again, consider the fact that at equilibrium $X + T_1 + T_2 = \frac{L}{u}(1-f)$ and $V + V_{T_1} + V_{T_2} = \frac{L}{u}f$. Then,

$$\frac{T_1}{V_{T_1}} = \frac{X}{V}$$

unless $T_1 = V_{T_1} = 0$. Similarly, $\frac{T_2}{V_{T_2}} = \frac{X}{V}$ unless $T_2 = V_{T_2} = 0$. When T_1 or $V_{T_1} \neq 0$ and T_2 or $V_{T_2} \neq 0$,

$$\frac{T_1}{V_{T_1}} = \frac{T_2}{V_{T_2}} = \frac{X}{V} = \pi.$$

So,

$$\pi = \frac{X + T_1 + T_2}{V + V_{T_1} + V_{T_2}} = \frac{1-f}{f},$$

or,

$$\pi = \frac{X}{V} = \frac{T_1 + T_2}{V_{T_1} + V_{T_2}} = \frac{1-f}{f}.$$

So, if $T_1 + T_2 = (1-f)\xi$, where $\xi \geqslant 0$, and $V_{T_1} + V_{T_2} = f\xi$,

$$\xi = T_1 + T_2 + V_{T_1} + V_{T_2} = \frac{L}{u} - (X + V) = \frac{L}{u} - X\left(1 + \frac{f}{1-f}\right).$$

Therefore, $X = (1-f)\left(\frac{L}{u} - \xi\right)$ and $V = f\left(\frac{L}{u} - \xi\right)$. Substituting these expressions into $\frac{dX}{dt} = 0$ in (3.7) gives

$$\begin{aligned} L(1-f) - u(1-f)\left(\frac{L}{u} - \xi\right) - (1-f)\left(\frac{L}{u} - \xi\right)K(1-f)\xi \\ - \beta_2(1-f)\left(\frac{L}{u} - \xi\right)f\xi + \gamma(1-f)\xi = 0, \end{aligned}$$

since $K = R$. This simplifies to

$$u\xi - (1-f)\left(\frac{L}{u} - \xi\right)K\xi - \beta_2\left(\frac{L}{u} - \xi\right)f\xi + \gamma\xi = 0.$$

Therefore, $\xi = 0$ or

$$\xi(K(1-f) + \beta_2 f) = \frac{L}{u}K(1-f) + \frac{\beta_2 L}{u}f - (u + \gamma).$$

When $\xi = 0$, $X = \frac{(1-f)L}{u}$ and $V = \frac{fL}{u}$. In addition, in this case $T_1 = T_2 = V_{T_1} = V_{T_2} = 0$. This is the CFE for this model. Otherwise, $\frac{L}{u}(K(1-f) + \beta_2 f)$ must be greater than or equal to $\gamma + u$.

It is known that $X = (1-f)\left(\frac{L}{u} - \xi\right)$, $V = f\left(\frac{L}{u} - \xi\right)$, $T_1 + T_2 = (1-f)\xi$ and $V_{T_1} + V_{T_2} = f\xi$.

Choose α so that $V_{T_1} = \alpha f\xi < f\xi$ for $\alpha < 1$. Then

$$V_{T_2} = f\xi - \alpha f\xi = f\xi(1-\alpha),$$

$$T_1 = \frac{1-f}{f} V_{T_1} = (1-f)\alpha\xi,$$

and

$$T_2 = (1-f)(1-\alpha)\xi.$$

These values of X, V, T_1, T_2, V_{T_1} and V_{T_2} satisfy all the equilibrium equations hence form an equilibrium for any value of α. In the situation where $T_2 = V_{T_2} = 0$, T_1 or $V_{T_1} \neq 0$.

$$\frac{T_1}{V_{T_1}} = \frac{X}{V} = \pi.$$

So,

$$\pi = \frac{X + T_1}{V + V_{T_1}} = \frac{1-f}{f}.$$

Hence, if $T_1 = (1-f)\xi$ and $V_{T_1} = f\xi$, $\xi = T_1 + V_{T_1} = \frac{L}{u} - (X+V)$. Thus, $\xi = \frac{L}{u} - X\left(1 + \frac{f}{1-f}\right)$, $X = (1-f)\left(\frac{L}{u} - \xi\right)$ and $V = f\left(\frac{L}{u} - \xi\right)$. Therefore, substituting once again into $\frac{dX}{dt} = 0$ from (3.7),

$$L(1-f) - u(1-f)\left(\frac{L}{u} - \xi\right) - (1-f)\left(\frac{L}{u} - \xi\right) K(1-f)\xi$$
$$- \beta_2(1-f)\left(\frac{L}{u} - \xi\right) f\xi + \gamma(1-f)\xi = 0.$$

So, as before $X = (1-f)\left(\frac{L}{u} - \xi\right)$, $V = f\left(\frac{L}{u} - \xi\right)$, $T_1 = (1-f)\xi$ and $V_{T_1} = f\xi$ is the equilibrium solution. Similarly, when $T_1 = V_{T_1} = 0$, T_2 or $V_{T_2} \neq 0$ and $X = (1-f)\left(\frac{L}{u} - \xi\right)$, $V = f\left(\frac{L}{u} - \xi\right)$, $T_2 = (1-f)\xi$ and $V_{T_2} = f\xi$.

Effective reproductive number

Again, as with the other two MLST model, R_e takes one of two possible values. R_e takes the larger of R_{e1} and R_{e2} where

$$R_{e1} = \frac{((1-f)(P\beta_1 + (1-P)\beta_2) + f\beta_2)L}{u(\gamma + u)}$$

and

$$R_{e2} = \frac{((1-f)(Q\beta_1 + (1-Q)\beta_2) + f\beta_2)L}{u(\gamma + u)}.$$

The derivation of R_e for this model is not presented here as it is similar to that of the previous model.

Clearly, which value R_e takes is dependent upon the transmission parameter. The difference between the two values for R_e is due to the differing proportions with which each of the MLSTs can manifest as each of the serotypes.

Local stability analysis

As with the previous models, a LSA was carried out for this model. Once again the stability matrix for the model must be identified. To do this, the equations (3.7) are expressed as

$$\frac{dX}{dt} = f(X, T_1, T_2, V, V_{T_1}, V_{T_2}),$$

$$\frac{dT_1}{dt} = g(X, T_1, T_2, V, V_{T_1}, V_{T_2}),$$

$$\frac{dT_2}{dt} = h(X, T_1, T_2, V, V_{T_1}, V_{T_2}),$$

$$\frac{dV}{dt} = i(X, T_1, T_2, V, V_{T_1}, V_{T_2}),$$

$$\frac{dV_{T_1}}{dt} = j(X, T_1, T_2, V, V_{T_1}, V_{T_2}),$$

and

$$\frac{dV_{T_2}}{dt} = k(X, T_1, T_2, V, V_{T_1}, V_{T_2}).$$

The partial derivatives of these functions are shown in Table 3.1, with $R = P\beta_1 + (1-P)\beta_2$ and $K = Q\beta_1 + (1-Q)\beta_2$.

An equilibrium solution, $(X_e, T_{1_e}, T_{2_e}, V_e, V_{T_1 e}, V_{T_2 e})$ is LAS if the real parts of the eigenvalues of the 6×6 matrix $\mathbf{A}$, an extension of the 4×4 matrix $\mathbf{A}$ in Chapter 2, Section 2.5.1, are negative.

A LSA was carried out for each of the equilibrium solutions of this model. The full details of the LSA are not presented here but the results show that when $R_e < 1$, the CFE is LAS but is unstable if $R_e > 1$. Recall that the CFE is the only equilibrium solution if $R_e \leqslant 1$. If $R_e = R_{e1} > \max(1, R_{e2})$ (i.e. $P\beta_1 + (1-P)\beta_2 > Q\beta_1 + (1-Q)\beta_2$) the first CE, (3.8), is LAS and if it exists (i.e. if $R_{e2} > 1$) then the second CE is unstable. In the reverse situation, if $R_e = R_{e2} > \max(1, R_{e1})$ then the first CE, (3.8), is unstable if it exists (i.e. if $R_{e1} > 1$). In this situation the second CE is always stable. If $R_{e1} = R_{e2}$ then any of the line of endemic equilibria are neutrally stable. A solution is defined as neutrally stable if a "differential shift in the initial state is preserved in time" (Öktem 2005). This means that the solution is stable but not attracting.

Global stability analysis

To carry out the GSA assume, without loss of generality, that $(P\beta_1 + (1-P)\beta_2)(1-f) + \beta_2 f \geqslant (Q\beta_1 + (1-Q)\beta_2)(1-f) + \beta_2 f$. Therefore, $R_e = R_{e1}$ and $R_e \geqslant R_{e2}$. The GSA can then be carried out using a similar approach to that described for the previous models by first of all considering the various initial states for $R_e \leqslant 1$ and then again for $R_e > 1$. For brevity, the proofs are not presented in this thesis. However, the analysis is available on request.

	$\boldsymbol{\partial/\partial X}$	$\boldsymbol{\partial/\partial T_1}$	$\boldsymbol{\partial/\partial T_2}$	$\boldsymbol{\partial/\partial V}$	$\boldsymbol{\partial/\partial V_{T_1}}$	$\boldsymbol{\partial/\partial V_{T_2}}$
f	$-u-RT_1-KT_2-\beta_2(V_{T_1}+V_{T_2})$	$-RX+\gamma$	$-KX+\gamma$	0	$-\beta_2 X$	$-\beta_2 X$
g	$RT_1+\beta_2 V_{T_1}$	$RX-(\gamma+u)$	0	0	$\beta_2 X$	0
h	$KT_2+\beta_2 V_{T_2}$	0	$KX-(\gamma+u)$	0	0	$\beta_2 X$
i	0	$-RV$	$-KV$	$-u-RT_1-KT_2-\beta_2(V_{T_1}+V_{T_2})$	$-\beta_2 V+\gamma$	$-\beta_2 V+\gamma$
j	0	RV	0	$\beta_2 V_{T_1}+RT_1$	$\beta_2 V-(\gamma+u)$	0
k	0	0	KV	$\beta_2 V_{T_2}+KT_2$	0	$\beta_2 V-(\gamma+u)$

Table 3.1: Partial derivatives for the stability matrix for Model 3.7.

As usual, the analysis is begun by considering $R_e \leqslant 1$. When $T_1(0) = V_{T_1}(0) = T_2(0) = V_{T_2}(0) = 0$ then $T_1(t) = V_{T_1}(t) = T_2(t) = V_{T_2}(t) = 0$, $\forall\ t$. In this case $X \to \frac{L(1-f)}{u}$ and $V \to \frac{Lf}{u}$ as $t \to \infty$. Considering $T_1(0) = V_{T_1}(0) = 0$ but $T_2(0) > 0$ or $V_{T_2}(0) > 0$, $T_1(0) = V_{T_1}(0) = 0$ which implies that $T_1(t) = V_{T_1}(t) = 0$, $\forall\ t$. Substituting $T_1 = 0$ and $V_{T_1} = 0$ in the four ODEs for X, T_2, V and V_{T_2} gives:

$$\frac{dX}{dt} = L(1-f) - uX - (Q\beta_1 + (1-Q)\beta_2)XT_2 - \beta_2 XV_{T_2} + \gamma T_2,$$

$$\frac{dT_2}{dt} = (Q\beta_1 + (1-Q)\beta_2)XT_2 + \beta_2 XV_{T_2} - (\gamma + u)T_2,$$

$$\frac{dV}{dt} = Lf - uV - \beta_2 VV_{T_2} - (Q\beta_1 + (1-Q)\beta_2)VT_2 + \gamma V_{T_2},$$

and

$$\frac{dV_{T_2}}{dt} = \beta_2 VV_{T_2} + (Q\beta_1 + (1-Q)\beta_2)VT_2 - (\gamma + u)V_{T_2}.$$

Unlike the GSA for the other two MLST model where the vaccinated and unvaccinated susceptible and carrying hosts were grouped to carry out the analysis, for this analysis ODEs involving grouped classes of unvaccinated hosts are considered as follows:

$$\frac{d(X + T_2)}{dt} = L(1-f) - u(X + T_2).$$

So,

$$X + T_2 \to \frac{L(1-f)}{u} \text{ as } t \to \infty.$$

Similarly,

$$V + V_{T_2} \to \frac{Lf}{u} \text{ as } t \to \infty.$$

This means that given $\epsilon > 0$, $\exists\ t_0$ such that for $t \geqslant t_0$,

$$X \leqslant X + T_2 \leqslant \frac{L(1-f)}{u} + \epsilon \text{ and } V \leqslant V + V_{T_2} \leqslant \frac{Lf}{u} + \epsilon.$$

Therefore,

$$\begin{aligned} \frac{d}{dt}\begin{pmatrix} T_2 \\ V_{T_2} \end{pmatrix} \quad \leqslant \quad & \begin{pmatrix} (Q\beta_1 + (1-Q)\beta_2)\left(\frac{L(1-f)}{u} + \epsilon\right) & \beta_2\left(\frac{L(1-f)}{u} + \epsilon\right) \\ (Q\beta_1 + (1-Q)\beta_2)\left(\frac{Lf}{u} + \epsilon\right) & \beta_2\left(\frac{Lf}{u} + \epsilon\right) \end{pmatrix}\begin{pmatrix} T_2 \\ V_{T_2} \end{pmatrix} \\ - \quad & (\gamma + u)\begin{pmatrix} T_2 \\ V_{T_2} \end{pmatrix}. \end{aligned}$$

Let the matrix $\mathbf{B}$ be defined by

$$\mathbf{B} = \begin{pmatrix} (Q\beta_1 + (1-Q)\beta_2)\left(\frac{L(1-f)}{u} + \epsilon\right) & \beta_2\left(\frac{L(1-f)}{u} + \epsilon\right) \\ (Q\beta_1 + (1-Q)\beta_2)\left(\frac{Lf}{u} + \epsilon\right) & \beta_2\left(\frac{Lf}{u} + \epsilon\right) \end{pmatrix}.$$

Recall that the spectral radius of a matrix is the largest absolute value of an eigenvalue of that matrix. The spectral radius of $\mathbf{B}^T$ is

$$\rho(\mathbf{B}^T) = (Q\beta_1 + (1-Q)\beta_2)\left(\frac{L}{u}(1-f) + \epsilon\right) + \beta_2\left(\frac{Lf}{u} + \epsilon\right).$$

If $R_{e2} < 1$, ϵ may be chosen to be small enough so that $\rho(\mathbf{B}^T) < \gamma + u$. By Lemma 2.1 of Nold (1980), there is an $\mathbf{e} = (e_1, e_2) > 0$ such that $\mathbf{B}^T\mathbf{e}^T = \rho(\mathbf{B}^T)\mathbf{e}^T$, i.e. $\mathbf{eB} = \rho(\mathbf{B}^T)\mathbf{e}$.

So, if

$$\mathbf{T} = \begin{pmatrix} T_2 \\ V_{T_2} \end{pmatrix},$$

then

$$\begin{aligned} \frac{d}{dt}[\mathbf{eT}] &\leqslant \mathbf{eBT} - \mathbf{e}(\gamma + u)\mathbf{T}, \\ &= (\rho(\mathbf{B}^T) - (\gamma + u))(\mathbf{eT}), \\ &= -k_0\mathbf{eT}, \end{aligned}$$

where $k_0 = (\gamma + u) - \rho(\mathbf{B}^T) > 0$. Hence $0 \leqslant \mathbf{eT}(t) \leqslant \mathbf{eT}(0)\exp(-k_0 t) \to 0$ as $t \to \infty$ so $\mathbf{eT}(t) = e_1 T_2(t) + e_2 V_{T_2}(t) \to 0$ as $t \to \infty$. Therefore, since $e_1 > 0$ and $e_2 > 0$, $T_2(t) \to 0$ and $V_{T_2}(t) \to 0$ as $t \to \infty$. So, $X \to \frac{L(1-f)}{u}$ and $V \to \frac{Lf}{u}$ as $t \to \infty$.

When $R_{e2} = 1$, define $\mathbf{B}(\epsilon) \equiv \mathbf{B}$. $\rho(\mathbf{B}(0)^T) = (\gamma + u)R_{e2} = (\gamma + u)$. By Lemma 2.1 of Nold (1980) there is an $\mathbf{e} = (e_1, e_2) > 0$ such that $\mathbf{B}(0)^T\mathbf{e}^T = \rho(\mathbf{B}(0)^T)\mathbf{e}$, i.e. $\mathbf{eB}(0) = \rho(\mathbf{B}(0)^T)\mathbf{e}$.

It is known that

$$\frac{d}{dt}\begin{pmatrix} T_2 \\ V_{T_2} \end{pmatrix} \leqslant \begin{pmatrix} (Q\beta_1 + (1-Q)\beta_2)\left(\frac{L(1-f)}{u} + \epsilon - T_2\right) & \beta_2\left(\frac{L(1-f)}{u} + \epsilon - T_2\right) \\ (Q\beta_1 + (1-Q)\beta_2)\left(\frac{Lf}{u} + \epsilon - V_{T_2}\right) & \beta_2\left(\frac{Lf}{u} + \epsilon - V_{T_2}\right) \end{pmatrix} \times \begin{pmatrix} T_2 \\ V_{T_2} \end{pmatrix} - (\gamma + u)\begin{pmatrix} T_2 \\ V_{T_2} \end{pmatrix}.$$

So,

$$\frac{d}{dt}[\mathbf{eT}] \leqslant \mathbf{e}\left(\mathbf{B}(0) + \begin{pmatrix} (Q\beta_1 + (1-Q)\beta_2)(\epsilon - T_2) & \beta_2(\epsilon - T_2) \\ (Q\beta_1 + (1-Q)\beta_2)(\epsilon - V_{T_2}) & \beta_2(\epsilon - V_{T_2}) \end{pmatrix}\right)\mathbf{T}$$
$$- \mathbf{e}(\gamma + u)\mathbf{T}, \text{ where } \mathbf{T} = \begin{pmatrix} T_2 \\ V_{T_2} \end{pmatrix},$$

$$= (e_1, e_2)\begin{pmatrix} (Q\beta_1 + (1-Q)\beta_2)(\epsilon - T_2) & \beta_2(\epsilon - T_2) \\ (Q\beta_1 + (1-Q)\beta_2)(\epsilon - V_{T_2}) & \beta_2(\epsilon - V_{T_2}) \end{pmatrix}\mathbf{T},$$

$$= \Big((Q\beta_1 + (1-Q)\beta_2)[e_1(\epsilon - T_2) + e_2(\epsilon - V_{T_2})],$$
$$\beta_2[e_1(\epsilon - T_2) + e_2(\epsilon - V_{T_2})]\Big)\begin{pmatrix} T_2 \\ V_{T_2} \end{pmatrix},$$

$$= ((Q\beta_1 + (1-Q)\beta_2)T_2 + \beta_2 V_{T_2})(e_1(\epsilon - T_2) + e_2(\epsilon - V_{T_2})),$$

$$= ((Q\beta_1 + (1-Q)\beta_2)T_2 + \beta_2 V_{T_2})((e_1 + e_2)\epsilon - e_1 T_2 - e_2 V_{T_2}).$$

Hence, for $\mathbf{eT} > (e_1 + e_2)\epsilon$, $\frac{d}{dt}(\mathbf{eT})$ is negative so $\exists\ t_1 \geqslant t_0$ such that for $t \geqslant t_1$, $0 \leqslant e_1 T_2 + e_2 V_{T_2} \leqslant 2(e_1 + e_2)\epsilon$. But $\epsilon > 0$ is arbitrary and $e_1 > 0$ and $e_2 > 0$ so T_2 and $V_{T_2} \to 0$ as $t \to \infty$. So, again, $X \to \frac{L(1-f)}{u}$ and $V \to \frac{Lf}{u}$ as $t \to \infty$.

The same arguments may be used when $T_2(0) = V_{T_2}(0) = 0$ and $T_1(0) > 0$ and $V_{T_2}(0) > 0$, replacing $Q\beta_1 + (1-Q)\beta_2$ with $P\beta_1 + (1-P)\beta_2$. Thus, $T_1(t) \to 0$, $V_{T_1} \to 0$, $X \to \frac{L(1-f)}{u}$ and $V \to \frac{Lf}{u}$ as $t \to \infty$.

Finally, the case where $(T_1 + V_{T_1})(0) > 0$ and $(T_2 + V_{T_2})(0) > 0$ when $R_e \leqslant 1$ is discussed. First, consider

$$\frac{d}{dt}(X + T_1 + T_2) = L(1-f) - u(X + T_1 + T_2).$$

So,

$$X + T_1 + T_2 \to \frac{L(1-f)}{u} \text{ as } t \to \infty.$$

Next, consider

$$\frac{d}{dt}(V + V_{T_1} + V_{T_2}) = Lf - u(V + V_{T_1} + V_{T_2}).$$

Hence,

$$V + V_{T_1} + V_{T_2} \to \frac{Lf}{u} \text{ as } t \to \infty.$$

This means that given $\epsilon > 0$, $\exists\ t_0$ such that for $t \geqslant t_0$,

$$X \leqslant X + T_1 + T_2 \leqslant \frac{L(1-f)}{u} + \epsilon \text{ and } V \leqslant V + V_{T_1} + V_{T_2} \leqslant \frac{Lf}{u} + \epsilon.$$

Thus,

$$\begin{aligned} \frac{d}{dt}\begin{pmatrix} T_1 \\ V_{T_1} \end{pmatrix} &\leqslant \begin{pmatrix} (P\beta_1 + (1-P)\beta_2)\left(\frac{L(1-f)}{u} + \epsilon\right) & \beta_2\left(\frac{L(1-f)}{u} + \epsilon\right) \\ (P\beta_1 + (1-P)\beta_2)\left(\frac{Lf}{u} + \epsilon\right) & \beta_2\left(\frac{Lf}{u} + \epsilon\right) \end{pmatrix}\begin{pmatrix} T_1 \\ V_{T_1} \end{pmatrix} \\ &- (\gamma + u)\begin{pmatrix} T_1 \\ V_{T_1} \end{pmatrix}. \end{aligned}$$

Therefore, by following the same argument as that for the case where $T_1(0) = V_{T_1}(0) = 0$ and $T_2(0) > 0$ and $V_{T_2}(0) > 0$, it can be shown that $T_1 \to 0$ and $V_{T_1} \to 0$. A similar argument can be used to show that both $T_2(0) \to 0$ and $V_{T_2}(0) \to 0$ as $t \to \infty$. Hence $X \to \frac{L(1-f)}{u}$ and $V \to \frac{Lf}{u}$ as $t \to \infty$.

As with the global stability analyses for the other models, next $R_e > 1$ is considered. When there are no carrying hosts initially, it can be shown that $X \to \frac{L(1-f)}{u}$ and $V \to \frac{Lf}{u}$ as $t \to \infty$ and there never will be carrying hosts.

If there are no hosts initially carrying MLST 2 but there are hosts carrying MLST 1, it is easily shown that no hosts will ever carry MLST 2 and that $X + T_1 \to \frac{L(1-f)}{u}$ and $V + V_{T_1} \to \frac{Lf}{u}$ as $t \to \infty$. The next part of the analysis involves combining the ODEs for susceptible and carrying hosts. Vaccinated and unvaccinated susceptible hosts are grouped together in a single ODE and vaccinated and unvaccinated carrying hosts are grouped together in another ODE. The ODE for the carrying hosts is complicated by the fact that transmission is attributable to serotype and not MLST. The analysis for this part of the GSA is shown in Appendix A.2. The results as $t \to \infty$ if $R_{e1} > 1$ are $T_2 \to 0$ and $V_{T_2} \to 0$ and

$$X \to \frac{(1-f)(\gamma+u)}{(P\beta_1 + (1-P)\beta_2)(1-f) + \beta_2 f},$$

$$T_1 \to (1-f)\left(\frac{L}{u} - \frac{\gamma+u}{(P\beta_1 + (1-P)\beta_2)(1-f) + \beta_2 f}\right),$$

$$V \to \frac{f(\gamma+u)}{(P\beta_1 + (1-P)\beta_2)(1-f) + \beta_2 f},$$

and

$$V_{T_1} \to f\left(\frac{L}{u} - \frac{\gamma+u}{(P\beta_1 + (1-P)\beta_2)(1-f) + \beta_2 f}\right)$$

When there are only hosts carrying MLST 2 initially and $R_e = R_{e1}$, if $R_{e2} \leqslant 1$ then the GSA follows the same argument as that when $R_e \leqslant 1$. Thus, $T_2 \to 0$, $V_{T_2} \to 0$, $X \to \frac{L(1-f)}{u}$ and $V \to \frac{Lf}{u}$ as $t \to \infty$. When $R_{e2} > 1$, the analysis follows the same argument as that of $T_2(0) = V_{T_2}(0) = 0$ when there are hosts initially carrying MLST 1, with $P\beta_1 + (1-P)\beta_2$ replaced by $Q\beta_1 + (1-Q)\beta_2$. In this scenario, if $P\beta_1 + (1-P)\beta_2 > Q\beta_1 + (1-Q)\beta_2$, then $R_e = R_{e1}$ and as $t \to \infty$,

$$X \to \frac{(1-f)(\gamma+u)}{(Q\beta_1+(1-Q)\beta_2)(1-f)+\beta_2 f},$$

$$T_2 \to (1-f)\left(\frac{L}{u}-\frac{\gamma+u}{(Q\beta_1+(1-Q)\beta_2)(1-f)+\beta_2 f}\right),$$

$$V \to \frac{f(\gamma+u)}{(Q\beta_1+(1-Q)\beta_2)(1-f)+\beta_2 f},$$

and

$$V_{T_2} \to f\left(\frac{L}{u}-\frac{\gamma+u}{(Q\beta_1+(1-Q)\beta_2)(1-f)+\beta_2 f}\right).$$

Finally, the last scenario to consider is when there are hosts initially carrying both MLSTs. The analysis is not displayed in this chapter but may be found in Appendix A.2. In this scenario, if $P\beta_1+(1-P)\beta_2 > Q\beta_1+(1-Q)\beta_2$, then $R_e = R_{e1}$ and as $t \to \infty$, $T_2 \to 0$, $V_{T_2} \to 0$ and

$$X \to \frac{(1-f)(\gamma+u)}{(P\beta_1+(1-P)\beta_2)(1-f)+\beta_2 f},$$

$$T_1 \to (1-f)\left(\frac{L}{u}-\frac{\gamma+u}{(P\beta_1+(1-P)\beta_2)(1-f)+\beta_2 f}\right),$$

$$V \to \frac{f(\gamma+u)}{(P\beta_1+(1-P)\beta_2)(1-f)+\beta_2 f},$$

and

$$V_{T_1} \to f\left(\frac{L}{u}-\frac{\gamma+u}{(P\beta_1+(1-P)\beta_2)(1-f)+\beta_2 f}\right).$$

However, if $Q\beta_1+(1-Q)\beta_2 > P\beta_1+(1-P)\beta_2$, then $R_e = R_{e2}$ and $T_1(t) \to 0$,

$V_{T_1}(t) \to 0$ and

$$X \to \frac{(1-f)(\gamma+u)}{(Q\beta_1+(1-Q)\beta_2)(1-f)+\beta_2 f},$$

$$T_2 \to (1-f)\left(\frac{L}{u}-\frac{\gamma+u}{(Q\beta_1+(1-Q)\beta_2)(1-f)+\beta_2 f}\right),$$

$$V \to \frac{f(\gamma+u)}{(Q\beta_1+(1-Q)\beta_2)(1-f)+\beta_2 f},$$

and

$$V_{T_2} \to f\left(\frac{L}{u}-\frac{\gamma+u}{(Q\beta_1+(1-Q)\beta_2)(1-f)+\beta_2 f}\right).$$

When $P\beta_1+(1-P)\beta_2 = Q\beta_1+(1-Q)\beta_2$, the situation is slightly more complex and X, T_1, T_2, V, V_{T_1} and V_{T_2} approach the equilibrium point where

$$X = (1-f)\left(\frac{L}{u}-\xi\right), T_1 = (1-f)\alpha\xi, T_2 = (1-f)(1-\alpha)\xi,$$

$$V = f\left(\frac{L}{u}-\xi\right), V_{T_1} = \alpha f\xi, \text{ and } V_{T_2} = f\xi(1-\alpha).$$

In these equations, $\xi = \frac{L}{u} - \frac{(\gamma+u)}{(P\beta_1+(1-P)\beta_2)(1-f)+\beta_2 f}$. Here, $\alpha = \frac{1}{1+k}$, where k is given in terms of the initial conditions by

$$\frac{(P\beta_1+(1-P)\beta_2)T_2(0)+\beta_2 V_{T_2}(0)}{(P\beta_1+(1-P)\beta_2)T_1(0)+\beta_2 V_{T_1}(0)}.$$

This analysis is described in full in Appendix A.

A summary of the results are shown in the following Theorem:

Theorem 3.3.1.1

(i) When the effective reproductive number is less than or equal to one, the CFE

is the only possible equilibrium. In this situation, regardless of the number of hosts initially carrying either MLST 1 or MLST 2, the number of carrying hosts will tend to zero in the long term. The number of susceptible and vaccinated individuals will tend to their CFE values.

(ii) a) When the effective reproductive number is greater than one, if there are no hosts initially carrying either of the MLSTs, there will never be any hosts carrying either MLST.

(ii) b) If $R_e = R_{e1} > 1 \geqslant R_{e2}$ then if initially there are hosts carrying MLST 2 but not MLST 1, then there will never be any hosts carrying MLST 1 and the number of hosts carrying MLST 2 will tend to zero. If $R_e = R_{e1} > R_{e2} > 1$ then under the same initial conditions the population sizes tend to those shown in the CE for carriage of MLST 2, (3.9).

(ii) c) However, if $R_{e1} > R_{e2}$ and there are hosts carrying MLST 1 initially, regardless of whether there are any hosts carrying MLST 2 initially, in the long term the number of hosts carrying MLST 2 will tend to zero and the population will tend to the population sizes described in the CE for carriage of MLST 1, (3.8).

(ii) d) If $R_{e2} > R_{e1}$ and $R_{e2} > 1$ the situations above are reversed.

(ii) e) If $R_{e1} = R_{e2} > 1$ then the CE is a line of equilibria and in the long term coexistence of both MLSTs is possible along this line if both are initially present. Provided that at least one MLST is initially present X, T_1, T_2, V, V_{T_1} and V_{T_2} approach the equilibrium point where

$$X = (1-f)\left(\frac{L}{u} - \xi\right), T_1 = (1-f)\alpha\xi, T_2 = (1-f)(1-\alpha)\xi,$$

$$V = f\left(\frac{L}{u} - \xi\right), V_{T_1} = f\alpha\xi \text{ and } V_{T_2} = f(1-\alpha)\xi,$$

for $\alpha = \frac{1}{1+k}$ and $0 \leqslant \alpha \leqslant 1$. Here, $k = \frac{(P\beta_1+(1-P)\beta_2)T_2(0)+\beta_2 V_{T_2}(0)}{(P\beta_1+(1-P)\beta_2)T_1(0)+\beta_2 V_{T_1}(0)}$ and $\xi = \frac{L}{u} - \frac{\gamma+u}{(P\beta_1+(1-P)\beta_2)(1-f)+\beta_2 f}$.

3.3.2 Conclusions

The results for this two MLST model show three possible equilibria, as with the previous two MLST model. Once again, there is the CFE and two possible CE; one corresponding to carriage of MLST 1, the other MLST 2. There is no possibility of hosts carrying MLST 1 and others carrying MLST 2 at equilibrium apart from the special case where $R_{e1} = R_{e2}$.

As transmission is assumed to be dependent on serotype for this model and not MLST, the effective reproductive number involves the proportions vaccinated and unvaccinated. However, the difference between the two possible values of R_e is attributable to the different proportions with which each MLST manifests as each of the serotypes as clearly the difference between the two values is that one involves $P\beta_1 + (1-P)\beta_2$ whilst the other involves $Q\beta_1 + (1-Q)\beta_2$. Thus, it is apparent that one of the R_e values corresponds to carriage of MLST 1, the other carriage of MLST 2.

The results of the GSA were summarised previously but the main conclusions are that if $P\beta_1 + (1-P)\beta_2 > Q\beta_1 + (1-Q)\beta_2$ then $R_e = R_{e1}$ and if this value is less than or equal to 1 then the number of hosts carrying either of the MLSTs will tend to zero in the long term, irrespective of the initial number of carrying hosts. This makes sense as when $R_e \leqslant 1$, the CFE is known to be the only possible equilibrium. When $R_e = R_{e1} > 1$, the population sizes will tend to those of the CE for carriage of MLST 1, (3.8), as MLST 1 dominates when $P\beta_1 + (1-P)\beta_2 > Q\beta_1 + (1-Q)\beta_2$. However, this result is dependent upon there being hosts initially carrying MLST 1 in the population. If no hosts carry either MLST and $R_e > 1$, the populations will tend to the sizes of the CFE which is the same as that of the first two MLST model, (3.1). Similarly, if there are children carrying MLST 2 initially but not MLST 1 but $R_{e2} < 1$, the limiting steady state population is described by the CFE. However, if there are initial individuals colonised by MLST 2 but not MLST 1 and $R_{e2} > 1$ then MLST 2 prevails in the population and the CE for carriage of MLST 2, (3.9) is both locally and globally asymptotically stable.

Concerning the two serotypes, as both MLSTs are able to manifest in both

serotypes then, even though MLSTs have been shown to be unable to coexist in the population, the two serotypes should be able to coexist in the population assuming that there are unvaccinated hosts. If $f = 1$, then all children will receive the conjugate vaccine, assumed to be 100% effective in preventing carriage of serotype 1. In this case, regardless of whether or not individuals are colonised with MLST 1 or 2, all hosts must be carrying serotype 2.

In the case of no vaccine intervention, i.e. $V = 0$, $V_{T_1} = 0$ and $V_{T_2} = 0$, all hosts fall into the unvaccinated susceptible class, X, unvaccinated MLST 1 carrier class, T_1 and the unvaccinated MLST 2 carrier class, T_2. The ODEs for these classes are as described for this model with vaccine, replacing the terms for vaccinated hosts in the expressions with zero. At the CFE, all hosts, $\frac{L}{u}$, are in the susceptible class X. The CEs for X, T_1 and T_2 are the same as those shown in (3.8) and (3.9), with $f = 0$. Thus, coexistence of MLSTs in the population at equilibrium is not possible unless $R_{e1} = R_{e2}$, as in the situation with vaccine intervention. Obviously, V, V_{T_1} and V_{T_2} should no longer be considered in these equilibria. The values for R_0 are the same as those described for the model with vaccine with $f = 0$ and the stability analysis shows similar results to that of the model with vaccine. When $R_0 \leqslant 1$, the CFE is the only equilibrium which exists and is globally stable, irrespective of the initial number of carrying hosts. If $R_0 > 1$ and no hosts are carriers initially, the CFE is the globally stable equilibrium. If $R_0 > 1$ and $P\beta_1 + (1-P)\beta_2 > Q\beta_1 + (1-Q)\beta_2$ and there are hosts initially carrying MLST 1 then the population will tend to that of the CE for carriage of MLST 1, regardless of the initial number of MLST 2 carriers. When $P\beta_1 + (1-P)\beta_2 > Q\beta_1 + (1-Q)\beta_2$, if there are no hosts carrying MLST 1 initially but hosts carry MLST 2, as long as the R_0 value corresponding to carriage of MLST 2 is greater than one then the population will tend to the MLST 2 carriage equilibria, otherwise the CFE shows the stable population sizes. If $Q\beta_1 + (1-Q)\beta_2 > P\beta_1 + (1-P)\beta_2$ the above situations are reversed.

3.4 Modelling conclusions

In the current chapter and the preceding chapter, three different models for the carriage of MLSTs in children under the age of two years are discussed. In

each of these models, at least one MLST is assumed to be associated with more than one serotype. The model in Chapter 2 considers only the possibility of colonisation with a single MLST which is associated with two serotypes with differing proportions. In this chapter, the models look at the possibility of a child becoming a carrier of one of two different MLSTs. The first model in this chapter assumes that only one of the MLSTs corresponds to two serotypes with differing proportions, whilst the other manifests only as one serotype. The second addresses the possibility that each MLST is associated with the same two serotypes but that the proportions with which each MLST manifests as each of the serotypes differ. The second model in this chapter can be used to obtain other models involving different combinations of serotypes and MLSTs. For example, by setting $Q = 0$ or $Q = 1$ in this model, the first model in this chapter is obtained but with transmission dependent on serotype rather than MLST.

Each of the three models consider a vaccine intervention which completely eliminates carriage of the VT serotype. In the models, only one of the serotypes is assumed to be a VT serotype. Vaccinated hosts are still able to become colonised with the MLST considered in the first model, or either of the MLSTs considered in the two MLST models, as the vaccine is serotype specific, not MLST specific. The vaccine would result in elimination of the MLST only if it was solely associated with the VT serotype. The reason that the models did not consider MLSTs solely associated with a VT serotype is that the models were created with a view to assessing what could occur in the population following a vaccine intervention should MLSTs be able to manifest as both a virulent VT and a NVT serotype.

The assumption that the vaccine is 100% effective in preventing carriage of VT serotypes is likely to be an overestimate of the vaccine effectiveness, with studies in France and the USA showing a reduction in carriage of VT serotypes but not total elimination (Cohen et al. 2006; **?**). Similarly, in a South African trial of PCV-7, the carriage of VT serotypes observed for vaccinated hosts was half of that observed for unvaccinated hosts but VT serotypes were still carried by vaccinated hosts (Mbelle et al. 1999). Thus, it appears unlikely that the vaccine will prevent all hosts from carrying VT serotypes and therefore this should possibly be taken into account in the mathematical models. However, other published mathematical

models of pneumococcus make assumptions that the vaccine is 100% effective in preventing colonisation with VT strains (Lipsitch 1997; Temime et al. 2004). Thus, it seems reasonable to have considered this in the MLST models analysed in this thesis. As the models described in this chapter do consider the possibility of a fraction of those aged under two years remaining unvaccinated, this allows for VT serotypes to remain in the population at equilibrium. Thus, by varying this parameter, it is possible to consider a reduced vaccine effect in the model.

A further issue explored in this chapter is the differences which occur when transmission is assumed to be attributable to serotype rather than to MLST. The first model considered in Chapter 2 assumes transmission is due to MLST as does the first model in this chapter, whilst the third model discussed addresses transmission by serotype. In other pneumococcal carriage models, such as the 1997 model by Lipsitch, transmission is assumed to be attributable to serotype. However, MLSTs were not considered in any of the other published pneumococcal models which all focus on serotype or strain, rather than addressing the possibility of classifying the pneumococcal isolates according to both serotype and MLST. As many of the modelling approaches discussed in Chapter 2 involve the examination of penicillin resistance by pneumococcal strains, it makes sense to model from the serotype perspective as the serotype is a known virulence factor of the bacterium and penicillin resistance has been observed for different serotypes involved in disease. In allowing the mechanism of transmission to differ for the two models considered in this chapter, differences are observed in the effective reproductive numbers for the models. In the first model of transmission by MLST, R_e is determined by the size of the transmission parameter between the susceptible class to the carrier classes for MLST 1 and 2, with the larger transmission parameter present in R_e. The expressions for R_e for the model where transmission is assumed attributable to serotype involves the proportions for which each MLST is able to manifest in each serotype. This model is the only one to have the parameter for the proportion of hosts vaccinated present in the expression for R_e.

Of the two models analysed in this chapter, neither show the possibility of coexistence of MLSTs in the population at equilibrium except for special parameter values. However, coexistence of MLSTs or serotypes within an individual host

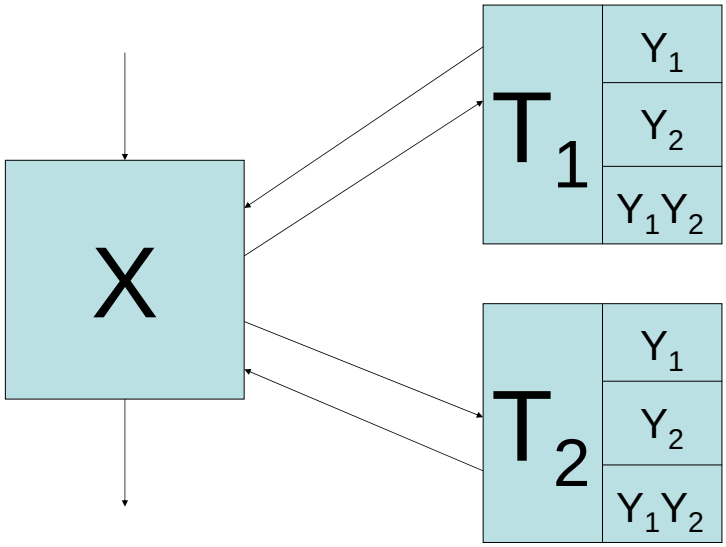

Figure 3.3: Model of two MLSTs; both associated with two serotypes with serotype coexistence.

was not considered in these models. Dual colonisation has been shown to occur within an individual in carriage; a longitudinal study of 82 children during the first 24 months in the USA showed 4.3% of all samples involved multiple carriage, 4% with two serotypes and 0.3% with three (Gray et al. 1980). Thus, coexistence of serotypes within a host should be considered as a possibility in carriage models. Therefore, the models discussed in this chapter could be extended by the inclusion of coexistence of either serotype or MLST. Possible models which could be considered in future analysis are shown in Figures 3.3 and 3.4. Figure 3.3 shows coexistence of serotypes but not MLSTs as the serotypes coexisting within an individual are assumed to have the same MLST. The other model, Figure 3.4, addresses both serotype and MLST coexistence. A host can be dually colonised with serotypes which have the same MLST, serotypes with different MLSTs and MLSTs with the same serotype. In these models, the transmission could be considered to be attributable to either MLST or serotype, as in the models considered in this chapter.

Coexistence of MLSTs or serotypes within an individual appears to be a necessary prerequisite for coexistence of MLSTs within the population. By excluding the possibility of coexistence within an individual in the three models assessed,

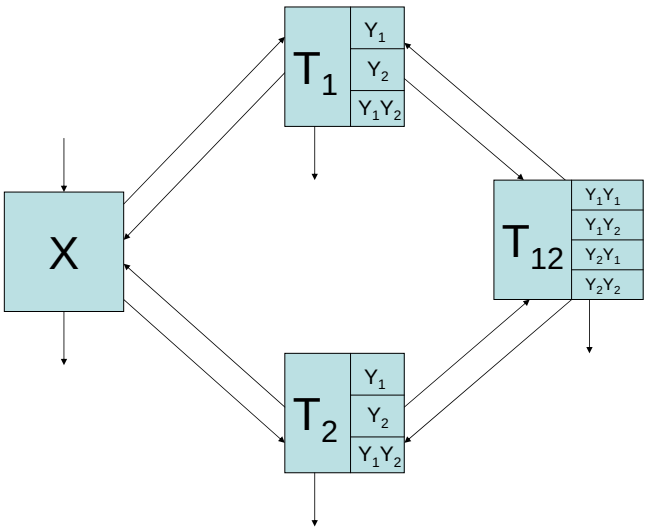

Figure 3.4: Model of two MLSTs; both associated with two serotypes with MLST coexistence resulting in serotype coexistence.

coexistence of the two MLSTs is not possible in the population at equilibrium. Of course, it is possible that either host heterogeneity or some type of stochasticity, or indeed something else which has not been incorporated in the model structure, could cause coexistence.

As coexistence is not incorporated into the model structures, a parameter to account for competition between the two MLSTs in the model, as included in the two serotype Lipsitch model, could not be included in the two MLST models. In the model with transmission attributable to MLST considered in this chapter, two MLSTs appear to compete for survival due to differences in their abilities to transmit. Whichever has the higher transmission thrives in the population when $R_e > 1$ and the other is eventually eliminated, assuming that there are initial carrying hosts of that MLST in the population. It is known that MLSTs can coexist in the population. Thus, this once again emphasises the necessity to include coexistence within an individual host in the MLST models.

The models analysed have a variety of limitations as described in this section. However, the conclusion drawn from the model that coexistence of MLSTs or serotypes within an individual host is necessary for coexistence within a popula-

tion is a significant conclusion to reach.

Further studies of MLSTs in pneumococcal carriage are required to obtain a better understanding of the importance of these stable genetic elements in describing the pneumococcal population. Considering coexistence in the models described would have greatly complicated the system of equations and thus computationally intensive measures would have been required to obtain equilibrium solutions and stability analyses. Longitudinal carriage studies have assessed the serotypes which commonly colonise but no studies with appropriate MLST and serotype data could be identified to obtain an idea of what happens from the genetic perspective in dual colonisation. This type of longitudinal study is required to establish which of the models should be considered in further analysis.

Chapter 4

Economic evaluation of health technologies

4.1 Introduction

In this next section of the thesis, interest lies in assessing how decisions are made about whether or not the benefits provided by a new treatment or health care intervention, such as PCV-7, outweigh the cost of introducing such an intervention. PCV-7 is the most expensive paediatric vaccine to become licensed (Beutels et al. 2007). However, to obtain a license for use, health economic analyses must be carried out to assess whether or not, through the use of an expensive product such as this, the burden to the health care provider will be reduced.

This chapter begins with a general introduction to the ideas behind economic assessment of a health care intervention and describes some of the methodology adopted in such an analysis. The focus in later sections of the chapter will return to the pneumococcus, with a description of the published analyses of the cost-effectiveness of PCV-7. Following this, the next chapter contains a statistical analysis of hospital episode statistics (HES) of the pneumococcal conditions septicaemia, meningitis and pneumonia for use in a cost-effectiveness model of PCV-7.

4.2 Background

Economic evaluation within the health sector is necessary, as with other sectors such as transport and defence, as there are only finite resources available with which to provide health care technologies. The health sector does not have the means with which to provide all possible health care technologies to suit the needs of all individuals. Therefore, economic evaluation is adopted to enable decisions to be made as to which technologies should be implemented. These decisions are based on the costs of providing the technology and the benefits, or effects, obtained through the use of the technology.

Health economics is a branch of economics created to enable better decision making when determining whether or not health care technologies are cost-effective. Health care from an economic perspective was considered as far back as the 17th century when the English economist and scientist Sir William Petty addressed important issues in health economy such as what value should be placed on a life. However, health economic theory and modelling was not explored in depth until the late 1960s when economists began to assess how government money should be spent in the health sector. The 1970s saw the introduction of the techniques of cost-benefit analysis (CBA) and cost-effectiveness analysis (CEA) to determine how resources should be allocated within and throughout the health sector (Mills 1997). Economic studies were not utilised to decide whether or not governments should adopt new treatments until the 1990s when, in 1993, Australia created a policy whereby any new drug required not only an efficacy trial but also some type of cost-effectiveness analysis before it would be considered for reimbursement (Kobelt 2002). Following this, a number of countries adopted various methods of incorporating cost-effectiveness evidence into the process of policy-making in health care. Canada introduced compulsory cost-effectiveness analyses on all new drugs in 1995 in Ontario and in 1996 in British Columbia, whilst in Europe, Finland, Norway and the Netherlands have had obligatory health economic analyses on all new drugs since 1998, 2002 and 2003 respectively (Kobelt 2002).

In the UK, the National Institute for Clinical Excellence (NICE) was created in 1999 to provide guidance on the clinical efficacy and cost-effectiveness of new technologies (such as medicines, medical devices and diagnostic techniques). NICE

was renamed the National Institute for Health and Clinical Excellence in 2005 when the role of the Health Development Agency (HDA) was adopted by the institute. One of the primary reasons for the establishment of NICE was to "reduce variation in the quality of care provided by the National Health Service" (Littlejohns 2001). One of the primary concerns of the 1997 Labour government was that throughout the UK there was a vast difference in the quality of health care provided and the government wanted to put an end to postcode prescribing where different health authorities were able to offer different treatments (Griffin and O'Grady 2006). In January 2002, it was made a requirement for England and Wales that National Health Service (NHS) organisations fund the various medical technologies that were recommended by NICE (NICE 2005). NICE recommendations are controversial due to the fact the new technologies are assessed according to evidence of their cost-effectiveness, not just their clinical efficacy, to determine whether or not the NHS should fund them. A key concern is that NICE recommendations could lead to an increase in the government spending required for the NHS to have sufficient funds for these new technologies. Another concern is that NICE will not actually have the effect intended, as it may not stop postcode prescribing since local health authorities are able to apply clinical discretion as to which current treatments will be replaced by the new treatments recommended by NICE (Griffin and O'Grady 2006).

In Scotland, the Scottish Medicines Consortium (SMC) provides recommendations as to which new health technologies should be adopted. The SMC was created in 2001 and represents all Health Boards (HBs) and Area Drug and Therapeutic Committees (ADTCs) in Scotland.

There are a number of differences in the way the SMC and NICE operate. SMC is responsible for assessing all drugs that have been licensed for market in Scotland whilst it is the responsibility of the Department of Health and the Welsh Assembly to decide which health technologies, not exclusively drugs, are to be assessed by NICE. The SMC provides a faster assessment of new technologies on their introduction whilst NICE provides a much longer and more comprehensive review of the evidence surrounding the new technology. Whilst it is policy for Health Authorities in England and Wales to adopt NICE recommended treatments, in

Scotland the ADTC receive guidance from the SMC but are not required to follow their recommendations unless it is an exceptional circumstance, such as in the case where there is no other treatment available for a particular condition. With both organisations, the emphasis is on assessing new treatments whilst treatments in use prior to the establishment of NICE and SMC had no requirement for a cost-effectiveness analysis and are not now considered by these agencies retrospectively (Cairns 2003).

It may appear rather odd that cost-effectiveness analyses only became commonly used in health care decision making in the 1990s as economic analyses have been used to enable decision making in other government sectors, such as transport or security, for a great number of years. However, economic analyses are much more complex when dealing with the health sector. Health must be considered in a different way to that of other sectors as economic analyses generally consider investments but health cannot be considered solely in terms of an investment. In addition, unlike other sectors involving consumers, in the case of health care the consumer relies on the advice of others, such as general practitioners (GPs), to establish what condition he or she is suffering from and which treatment should be adopted. In addition, it is difficult to determine the output in an economic evaluation of health as it is difficult to place a value on the improvement in health (Griffin and O'Grady 2006). There are other problems associated with economic analyses in the health sector attributed to weak or complicated methodology used in the cost-effectiveness analyses. Some complications attributed to the methodology will be discussed in this chapter.

4.3 Types of economic evaluation

Economic evaluations are defined according to the approach taken when costing or the benefit measure used in the analysis. There are four main types of analysis that may be used in the evaluation and comparison of health technologies: the cost-effectiveness analysis (CEA), cost-utility analysis (CUA), the cost-benefit analysis (CBA) and the cost-consequence analysis (CCA). In practice, the CEA, CUA or CBA are the preferred approaches. This is due to the fact that in a CCA, all costs and consequences, measured in the appropriate units, are presented to

the decision maker. This makes it difficult to judge whether or not the treatment is truly cost-effective as it is difficult to determine the trade-offs (Drummond et al. 2005). The other three types, the CEA, CUA and CBA, are defined according to the type of benefit measure used in the analysis.

4.3.1 Cost-effectiveness analysis

A cost-effectiveness analysis (CEA) is used when there are at least two technologies, a new technology and a current technology in use, to be compared in terms of their cost-effectiveness. For each technology, the costs for one unit of improved outcome, whether it be an additional year of life of a patient gained or an episode free day for an asthma patient (Drummond et al. 2005), are determined and compared. In a cost-effectiveness analysis the benefit measure is a natural unit which is relatively simple to determine such as number of lives, or life years saved, or number of cases of disease detected, for example. To compare the technologies the cost per life year gained, or cost per number of cases of a disease detected, may be calculated for each technology. This cost analysis must be carried out in conjunction with evidence of clinical efficacy as a new technology cannot be adopted purely to reduce costs if it will not prove as effective in practice as the current technology. A special case of the CEA is a cost-minimisation analysis (CMA). A CMA may be carried out if there is no difference in the health benefits obtained from the various health technologies being compared (Briggs et al. 2006). Therefore, the health technology adopted should be the technology which would cost least to implement. Cost-minimisation analysis is rarely used in practice as it is highly unlikely that two health technologies will have precisely the same health benefit, particularly when a new technology has been developed to improve the health benefit (Kobelt 2002).

A cost-effectiveness ratio is calculated to enable comparisons of technologies. The extra cost that one technology has over the alternative (usually the new technology proves more expensive than the existing technology), otherwise known as the incremental cost, is determined and the incremental cost effectiveness ratio (ICER) is computed. The ICER is the added expense of obtaining a further unit of outcome, such as life year gained, for one technology compared to another (Kobelt 2002).

The ICER is only calculated when the more expensive of the technologies has also been shown to have greater efficacy. In the case where technology B is more effective but more expensive than A, the calculation for the ICER is (Kobelt 2002):

$$\text{ICER} \;=\; \frac{\text{Cost}_B - \text{Cost}_A}{\text{Effect}_B - \text{Effect}_A}.$$

The better, more expensive technology is adopted if the willingness to pay of those determining which technology to utilise is greater than the incremental cost. If the better technology is also the less expensive technology there is no need to work out the ICER as clearly this technology should be the one adopted. The ICER may be displayed on the cost-effectiveness plane (Figure 4.1).

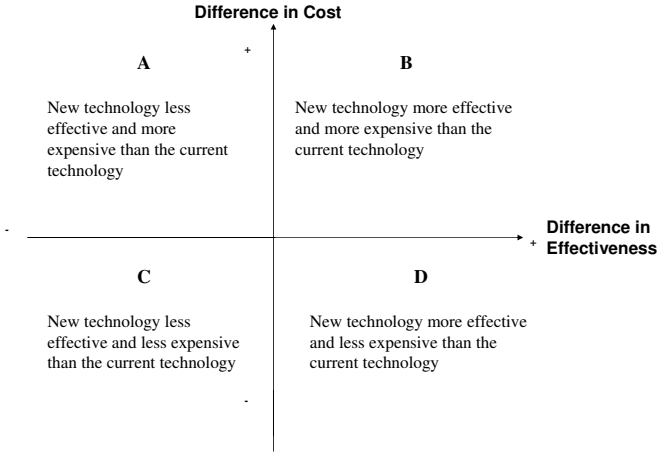

Figure 4.1: The cost-effectiveness plane (adapted from Drummond et al., 2005).

The horizontal-axis on the cost-effectiveness plane shows the difference in the efficacy of the new technology and the current technology that it is being compared to. The vertical-axis shows the difference in the cost of the two technologies (Drummond et al. 2005). The decision as to which technology to use is simple if the outcome lies in either category A or D, as if the new technology is less effective and more expensive than the current technology it should not be adopted to

replace the current technology and if the new technology is more effective and less expensive then it should replace the technology currently in use. The decision of which technology to use if the outcome lies in categories B and D is more difficult and is based upon the ICER and willingness to pay. Alternatively, the ICER may be displayed in a 3×3 table (Figure 4.2).

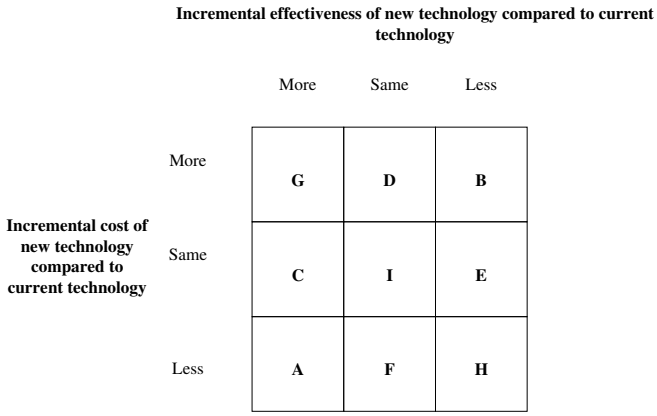

Figure 4.2: The cost-effectiveness table (adapted from Drummond et al., 2005).

A cost-effectiveness acceptability curve may be used to portray the information used in the CEA as it shows what proportion of estimates of the ICER are satisfactory for a series of willingness-to-pay values. Here, the willingness-to-pay values are determined by how much a policymaker is prepared to pay for the technology. An example of cost-effectiveness acceptability curves from an assessment of the cost-effectiveness of a rotavirus vaccine in Vietnam is shown in Figure 4.3 (Kim et al. 2009). These curves show the cost-effectiveness results when assessing the rotavirus vaccine from the societal perspective and from the health care payer perspective. Further details about perspectives in a health economic analysis will be provided later in this chapter.

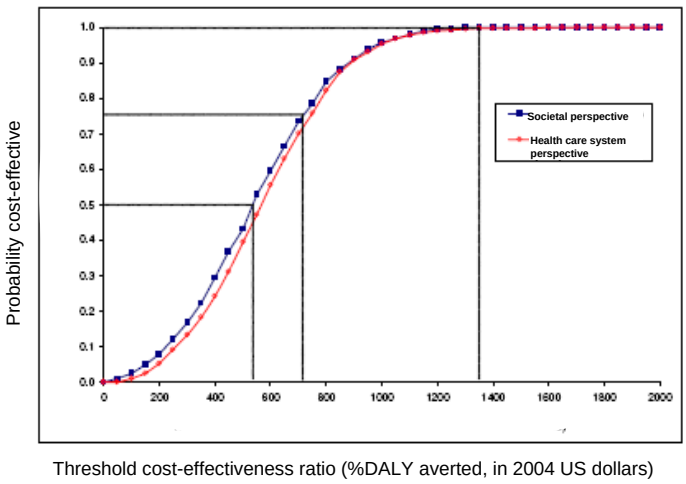

Figure 4.3: Cost-effectiveness acceptability curves (adapted from Kim et al., 2009).

The difficulty with a CEA analysis that uses natural units as a benefit measure is that it is often very disease specific, thus making it difficult to compare cost-effectiveness across all conditions. For example, in a published CEA of a vaccine for prevention of hepatitis A and B infections, the outcome benefit measure was the number of cases of hepatitis A and B infections prevented (Szucs 2000). By using a benefit measure of this type, it is difficult to determine whether the cost-effectiveness for this vaccine is better, worse or equal to the cost-effectiveness of a vaccine which prevents something else, such as influenza, for example.

Another example where this type of analysis may not be the most appropriate is a cost-effectiveness analysis of the use of the varicella vaccine in Canada (Brisson and Edmunds 2002). The drawback of this analysis is not that the outcome measure is not comparable across analyses but it does concern the outcome measure adopted. In this analysis the outcome measure is life-years gained. However, Brisson and Edmunds state that a more appropriate outcome measure to have used in the analysis is the cost per quality-adjusted life year (QALY) saved due to the fact that varicella is a mild disease resulting in little mortality and it is known that the vaccine results in reduced morbidity. A definition of a QALY is

provided in the next section of the chapter. The authors state that unfortunately this technique could not be used in the analysis as no methodology had been developed to measure QALYs lost due to mild diseases with short-term duration in very young children.

4.3.2 Cost-utility analysis

The preferred type of health economic evaluation is the cost-utility analysis (CUA), where the outcome to measure improvements in health care provided is the QALY and the cost for each QALY gained is determined for each technology. QALYs measure the desire of individuals to be in various states of health. To determine the quality-adjusted life expectancy, the total number of years that an individual spends in a particular health condition is multiplied by the score associated with that particular condition. The health scores generally range from 0 to 1, with scores approaching 0 reflecting worsening states of health as 0 represents death whilst 1 represents perfect health. It is possible to give a negative score if the health state is considered to be worse than death.

To help clarify the definition of a QALY, consider someone who lives for twenty years. In the first ten years the person has no health problems whatsoever and thus has a utility of 1 for these years. In the next ten years, the person suffers from a health condition which reduces the utility by half before the person dies at the end of these ten years. Thus, the person lived for twenty years but had a quality-adjusted life expectancy of $(10 \times 1) + (10 \times 0.5) = 15$ QALYs (Palmer 2005).

A threshold for the maximum cost per QALY for a health care intervention is defined for a health economic analysis. In the UK, for example, if a treatment costs more than $£20,000 - 30,000$ per QALY then it is not cost-effective (NHS NICE 2009).

In addition, a similar alternative measure to the QALY, often adopted in a CUA, is the disability-adjusted life year (DALY). The DALY differs from the QALY in that rather than determining individual preferences for living with a disability,

values are taken from decisions made by a panel of health care workers who looked at person trade-off scores. In addition, the weights used in DALYs consist of only nine distinct values, including one for healthy and one for dead whilst the QALY weights can hypothetically take any value between 0 and 1. Finally, the DALY involves weighting according to age whilst the QALY does not (Drummond et al. 2005).

The benefit of using a CUA over a CEA is that QALYs are not disease-specific so the cost-effectiveness of various technologies across disease areas can be compared. The problem in using a CUA is that difficulties can arise in obtaining truly representative utility weights necessary to calculate a QALY. Another issue is that different cost-effectiveness studies assess the technologies according to different perspectives. Therefore, the technologies cannot be compared in terms of the cost per QALY gained. For example, some cost-effectiveness analyses are carried out from the NHS perspective whilst others could be assessed from the point of view of those to receive the technology (Towse et al. 2002).

QALYs are a commonly used measure in analyses of the cost-effectiveness of vaccines. Recently, QALYs have been used as the outcome measure in analyses of the cost-effectiveness of the vaccine created to prevent cervical cancer, HPV (Colantonio et al. 2009; Dasbach et al. 2008; Jit et al. 2008; Ginsberg et al. 2007). Considering the analysis by Colantonio et al., in this analysis the cost-effectiveness of the vaccine in five Latin American countries was considered. The model consists of twelve different health states which are mutually exclusive and follows a female cohort from the age of 11 years until death. Each of these twelve states are allocated a cost and utility for spending one cycle, a year in this case, in this state. Thus, the model outcomes can vary as individuals can differ in the length of time spent in one state. Colantonio et al. then calculated the difference between scenarios for both health outcomes and total costs and were able to estimate the ICER per QALY saved. In this analysis an intervention is considered to be cost-effective when one QALY gained costs less than three times the Gross Domestic Product per capita in each country. The authors find the vaccine to be cost-effective under these conditions.

CUAs have also been used in assessment of other vaccines. For example, a CUA was adopted in an analysis of a rotavirus vaccine (Goossens et al. 2008). Rotavirus is the main cause of severe diarrhoea and vomiting in very young children and Goossens et al. investigate the cost-effectiveness of introducing a vaccine against rotavirus in the Netherlands. As with the Colantonio et al. HPV model, Goossens et al. consider cycles in which children can move between various health states. However, in this model the cycles are of much shorter duration, one month, due to the nature of the conditions caused by rotavirus. In the model it is assumed, for simplicity, that a child who did not die during a cycle following the first five years of life will have a utility of 1 for the rest of his or her life. In order to calculate the QALYs for this model, the utility values were based on a study of rotavirus infection outcomes which used the EQ5D questionnaire to make the valuation. The EQ5D questionnaire is a standardised instrument used to calculate utility values for different health states.

4.3.3 Cost-benefit analysis

A third type of health economic evaluation that may be undertaken is a cost-benefit analysis (CBA). In CBA, both the costs and the health benefit are measured in monetary terms. This necessitates the valuation of the usual measures of outcome, such as life-years gained, number of disease cases prevented or QALYs gained, in monetary terms. This differs from the CEA, where the health outcome measure is the therapeutic effect and the CUA where the health outcome measure is the quality of life of the patient. The advantage of this type of evaluation is that, since the benefit measure is money, it allows the government funding decisions within the health sector to be compared with the funding decisions of the government in other sectors, such as transport or security. The problem with this type of analysis is the difficulty of assigning a monetary value to life. There are two main ways in which health may be given a monetary value.

The first of these is in terms of willingness-to-pay (WTP). To assess the WTP for a CBA, contingent valuation studies are carried out in which questions are asked about how much the subjects taking the survey are willing to spend on the benefits obtained from a health technology. A contingent valuation can consist

of two types of question: open-ended and discrete. With open-ended questions, those taking the questionnaire are requested to state their maximum WTP whilst with the discrete questioning approach, respondents have to state whether or not they would be willing to pay certain amounts that are put before them (Kobelt 2002). In order to assess WTP, it is important that the respondents are able to make informed decisions based on all available scientific information about the health technology (Sloan 1996).

An example of this type of analysis involves the assessment of the WTP for PCV-7 (Prosser et al. 2004). In this analysis, data were collected on the WTP to prevent six conditions (simple OM, complex OM, moderate pneumonia, severe pneumonia, bacteremia, and meningitis) which are preventable through the use of PCV-7. The results from 30 minute telephone interviews of a sample of parents of children who had experienced at least one of the six outcomes and from a general community sample in the USA show both groups to assign relatively high values to preventing the more severe outcomes such as meningitis and pneumonia. Using the information in a health economic model resulted in a cost-effectiveness ratio being obtained which is comparable to those obtained for other health interventions commonly used.

The second method of valuing benefits is a market based approach looking at human capital. The human capital approach assesses the effect of lost productivity of workers were they to suffer from ill health (Gold et al. 1996). Therefore, in this approach, a good state of health is considered as an investment since a person in good health is assumed to be able to make a greater contribution in the workplace. In this way human capital may be measured in terms of earnings and the current value of possible future earnings attributed to the health care measure of interest is determined for comparison with that of current health measures. There are difficulties with assigning monetary values to health in this way. For example, there may be inconsistencies in the treatment of staff in the workplace due to discrimination that make a generic value for wage difficult to assign (Drummond et al. 2005). In addition, there are difficulties in assigning values for those who have retired (Sloan 1996).

This approach was adopted in an analysis of the varicella vaccine in Taiwan, where it is mentioned that this measures the value of the vaccine intervention by the effect it has on the lifetime earnings of the patients (Hsu et al. 2003). To obtain estimates for the monetary benefit, a cross-sectional study of varicella cases in Taiwan was carried out and information was collected about the income of the parents or families of the patient who were unable to work because they had to care for the patient. Thus, the monetary value placed on the benefit of preventing the case of varicella was calculated by multiplying the loss of working days of each adult by his or her daily income. The conclusions from the study suggest that the vaccination implementation is only worthwhile to administer routinely from the societal perspective with a benefit-cost ratio of 2.06, not from the health care payer perspective, where a benefit-cost ratio of 0.34 is obtained.

Additionally, Hsu et al. carried out the analysis using the WTP approach. To obtain an estimate of the WTP, in the cross-sectional study the participants were asked to quote the most that they would be willing to pay for a routine varicella vaccination programme that reduces the risk of death and long term disability. A list of the potential disabilities was provided. The results showed the vaccine not to be cost-effective using the WTP approach.

4.3.4 Cost-consequence analysis

Finally, a fourth type of economic evaluation which may be adopted is the cost-consequence analysis (CCA). In a CCA, the various health technologies for a particular health state are compared by listing all of the costs and health benefits, or consequences, of each and a decision is made about which technology to adopt by comparing these lists (Gold et al. 1996). A CCA is used when there are a variety of benefits from each technology, therefore rendering other types of economic evaluations difficult to use (Drummond et al. 2005). In a CCA, the costs and consequences are not combined, as with other types of economic evaluation, and the importance of the various consequences is not determined. The decision of which technology is most appropriate to adopt is made purely by the decision maker on examining the lists of costs and consequences.

A CCA was carried out in an analysis of a treatment for cancer of the urinary bladder (Marchetti et al. 2000). In this paper, Marchetti et al. provide estimates of the expected costs of administering the intervention of interest, Valrubicin, considering costs of potential adverse effects to the treatment. The authors simply conclude the analysis with the statement of the expected cost, making no conclusions about whether or not the treatment is cost-effective as, in this type of analysis, the decision ultimately has to be made by a higher authority who determines whether or not the intervention should be implemented.

4.4 Costs

Various factors have led to an increase in the cost of health care. These include an increasing demand for high standards of living, the ease with which information about health care technologies may be obtained in the modern world, technological advances and the ageing population in the developed world (Kobelt 2002).

In the economic evaluation of health technologies, money plays an important role but is not considered as a resource for providing health. However, it is important as it enables the purchase of necessary resources required to attempt to improve health such as doctors, hospital beds and drugs (Griffin and O'Grady 2006). The resources required in providing the health technology must be considered and the costs of each of these must be evaluated for the economic analysis. Costs involved in health care are not all necessarily defined in monetary terms but should still be considered in an economic evaluation.

There are various types of costs that may be incurred by a patient when a health technology is being implemented. There are direct costs which can be medical or non-medical. A direct medical cost is the cost associated with the health technology such as the price of a treatment whilst a direct non-medical cost could be, for example, the travel expenses the patient must pay in order to receive the treatment. Without paying these travel expenses the patient cannot receive treatment. Therefore, this expense is considered as a direct cost. Indirect costs are costs to the patient through, for example, the loss of wages whilst taking time

off work to receive treatment.

In all economic analyses the opportunity costs must be identified. The opportunity cost is defined as the "value of the forgone benefits because the resource is not available for its best alternative use" (Drummond et al. 2005, page 57). Opportunity costs can occur when a resource being used is not thought of in monetary terms, in that it does not have a market price, such as a family member voluntarily providing care for another. It is important to determine the opportunity costs for all technologies being compared in order to correctly determine which is the most cost-effective.

In determining the resources to include in the economic evaluation, it is important to include both resources that may have a relatively small individual cost but will have substantial usage, as well as resources associated with large costs which will have fairly infrequent use (Gold et al. 1996). The decision of which costs should be included in the analysis will entirely depend upon the perspective with which the economic evaluation is being carried out. Economic evaluations are carried out to enable decisions to be made about which health technologies should be adopted. In the decision making process, the perspective of the analysis is very important. One possible perspective that may be adopted is the societal perspective. Under this perspective, all individuals on whom the health technology is likely to have an impact, whether it be in terms of health or cost, are considered (Gold et al. 1996). However, there are various other perspectives which can be adopted in an economic evaluation, such as the third party payer, and the perspective selected is dependent upon the aim of the evaluation. A third party payer could be, for example, the government or an insurance company. With a third party payer perspective, only costs incurred by this payer are taken into account in the analysis. For example, a health insurance company is likely to be interested only in the direct medical costs incurred and will not take into account indirect costs such as those attributable to absence from work due to sick leave.

In the UK, economic evaluations of health technologies are carried out from the NHS and Personal Social Services perspective with the main aim to be cost containment. As mentioned, analyses carried out from perspectives other than the

societal viewpoint will only consider a subset of the factors considered in analyses from the societal perspective. For example, in a CEA of a health technology adopting a societal perspective costs incurred in purchasing and administering the technology, additional hospital costs which require to be taken into account when providing this technology, costs of services such as home health care or nursing home expenses related to the technology and patient costs are all considered, whereas if a CEA is carried out from a hospital perspective then only the costs of purchasing and administering the technology and any additional hospital costs require to be considered (Sloan 1996).

There are two main categories of costing: micro-costing and gross-costing. Gross-costing is relatively simple to carry out as an overall sum is determined for the services involved in implementing the health technology, such as costs incurred during a stay in hospital. However, there is a disadvantage to adopting this approach of costing in that there may be a lack of sensitivity in the analysis (Raftery 2000). In the UK gross-costing is adopted as hospital costs are estimated by averaging information on treatment costs from hospitals throughout the country. Micro-costing involves a greater degree of research work in order to detail all resources a patient may require and all costs incurred through the process of providing the intervention (Gold et al. 1996). Micro-costing may be appropriate when an adjustment is being made to an existing health care service (Raftery 2000). The disadvantage of micro-costing is that it is often expensive and time-consuming to obtain the detailed expenses required in this costing approach.

In considering costing, discounting must be taken into account. Discounting is carried out in economic evaluations when costs and benefits occur at different times. Discounting is important as, with some health care interventions, it may take time to observe benefits of the health technology but the cost of implementing the technology may have to be paid immediately (Kobelt 2002). The correction factor to adjust for discounting is $1/(1+r)^t$, where t is the number of years and r is the discount rate. In the UK, a discount factor of 3.5% per annum is used for both costs and benefits (Department of Health 2007). Thus, $£1,000$ now is $\left(\frac{1}{1+0.035}\right) \times £1,000$ in one year. The Treasury is responsible for deciding the discount rate to

be adopted in economic evaluations and research is carried out on the societal time preference for health benefits, i.e. the preference to receive benefits now rather than in the future, to enable this decision to be made. The rate of inflation may be used to guide this decision but discounting effectively has little to do with inflation. High discount rates reflect the preference to have higher benefits now whilst lower rates will reflect a higher emphasis on future benefits (Sloan 1996). Discounting is not always essential in an economic evaluation. For example, if effects of a health technology are experienced reasonably quickly, such as within a year or two of the technology being implemented, it may not be necessary to use discounting in the analysis. Considering vaccines once again, generally the beneficial effects are obtained in the long term. For example, with PCV-7 long term beneficial effects should be obtained through reductions in disease in those unvaccinated due to herd immunity. However, some instantaneous effects are observed such as a reduction in mortality to vaccine attributable disease in those vaccinated.

4.5 Uncertainty

Identifying and assessing the uncertainties involved in an economic evaluation plays a key role in decision making. Uncertainties may be assessed in two main ways. The first of these involves data on the resource use of individual patients to use in models to quantify uncertainty.

The other technique used in determining uncertainty is a sensitivity analysis. There are four types of sensitivity analysis that may be adopted in an economic evaluation: a one-way sensitivity analysis, a probabilistic analysis, a scenario analysis and a threshold analysis. These four methods are described below.

The one-way sensitivity analysis is the simplest of the four types. In this type of analysis, model parameters are varied one at a time to discover the impact of each on the outcome of the model. This enables identification of the variables which are important in determining the outcome. This technique also enables the discovery of errors in the model. The disadvantage of this type of sensitivity analysis is the difficulty in quantifying the combined effect of several potentially

sensitive variables on the model outcome (Drummond et al. 2005). Therefore, multiway analyses are often preferred such as two-way or three-way analyses, in which two or three parameters are varied simultaneously. Analyses with greater dimensions may be adopted. However, these analyses become very difficult to carry out and interpret.

A second method of sensitivity analysis is a probabilistic analysis. This method can be adopted when there are a variety of parameters which may prove important in the sensitivity analysis as, for each of the identified uncertain parameters, a probability distribution is assumed and the effect these parameters have on the model is determined through Monte Carlo simulation. Monte Carlo simulation will be discussed in the Methodology section in this chapter. Commonly, distributions such as the Gamma distribution or the Beta distribution are adopted for uncertain utilities (Sculpher 2004).

A third method is a scenario analysis. In this type of analysis, the parameters are chosen to reflect best case and worst case scenarios so that the decision maker can assess costs involved in each of these circumstances. A base case scenario which reflects the best estimate of the analyst of the parameter values is also considered for comparison.

Finally, threshold analysis may be adopted in determining uncertainty in an economic evaluation. Threshold analysis enables decisions of which price should be chosen for the technology if all other variables in the model are assumed to have a greater degree of certainty. In such an analysis, the decision maker may state a maximum cost for the health technology, above which the technology should not be adopted. In this situation, an analyst will vary the parameters in the model to determine which combinations of estimates will cause the health technology to exceed the threshold cost proposed by the decision maker (Drummond et al. 2005).

There are various uncertainties that must be considered in an economic evaluation such as structural uncertainty in which the model structure may not be suitable. This could arise where models incorrectly specified crucial steps in a disease progression. Variable uncertainty is also important to assess whether variables

with a high variability are important in deciding the outcome of the model.

4.6 Methodology

As yet, no mention has been made of the methodology adopted in the evaluation process. Commonly, modelling techniques are adopted. These techniques will be discussed in this section.

Modelling is important in the economic evaluation of health care technologies as it allows conclusions about the effectiveness of the technology, both in terms of costs and benefits, to be made outwith the scope of the data used in the analysis. The modelling methodologies used in economic evaluations include decision trees and Markov chains, Monte Carlo simulations and Bayesian techniques. Research identified all but the Bayesian approach in published economic evaluations of the 7-valent pneumococcal vaccine (PCV-7). For example, decision trees were adopted in an Australian CEA and CUA of the cost-effectiveness of PCV-7 (Butler et al. 2004); Markov models were used in a Norwegian study looking at a CEA and CUA (Wisløff et al. 2006); Monte Carlo simulations were adopted to carry out a sensitivity analysis in a UK CUA in conjunction with decision trees to model disease outcomes and costs (Melegaro and Edmunds 2004a). Descriptions of each of the usages will be provided in the relevant sections. In addition, an example for the Bayesian approach will be provided for a non-pneumococcal intervention.

The first step to be taken when carrying out a health economic evaluation is the decision, and clear specification, of the question that is to be addressed. There are key factors that must be taken into account such as the alternatives for comparison in the analysis and the individuals to whom the treatments are to be given. It is necessary to identify the perspective with which the analysis is to be carried out. For example, does interest lie in addressing the benefits to society or the benefits to the third party payer? The third party payer could be the government or a particular health care organisation (Kobelt 2002). If the societal benefits are addressed in the study then all costs, such as costs incurred by the health care service, patients and costs incurred by others in society through

production loss, must be included in the analysis. However, if the analysis is to be carried out from a health care organisation perspective, or any other third party payer perspective, then the only costs to be included are those for the resources that the health care organisation directly pays for.

To illustrate, consider the analysis by Butler et al. (2004). Butler et al. carried out an analysis to address the benefits of the introduction of PCV-7 to children not at high-risk of pneumococcal disease in Australia, since prior to this analysis only high-risk groups were administered the vaccine. The intention of the study is to assess the benefits and costs of adopting PCV-7 for use amongst young healthy children to prevent pneumococcal disease and infection.

Concerning the perspective, the authors state that it is apparent in Australia which perspective should be adopted in an analysis as there is a clearly defined system in place for new health care interventions in which only direct costs are included in the analysis. To estimate the unit costs of vaccination and treatment of various disease states, a societal perspective is adopted.

Considering the potential alternative treatments, Butler et al. estimated the costs for the alternative treatments required should various pneumococcal outcomes occur that PCV-7 could potentially prevent. For example, concerning meningitis, the authors identified two possible costs. The first of which is the pre-hospital admission treatment cost; the second, the average cost of treatment involving hospitalisation due to meningitis. These depend on the severity of the meningitis. The Butler et al. study will be discussed again later in this chapter.

A further issue that is important to consider is the structure that the model should have. Firstly, clinical events must be considered. For example, if interest lies in the cost-effectiveness of a particular treatment for a disease then it is necessary to determine clinical events such as the number of cases of disease and stages in the disease progression, as well as other diseases the virus or bacterium may cause. This information must be utilised in the structure of the model. If these are not considered then it is possible that incorrect conclusions may be made about the cost-effectiveness of the treatment. The potential effects of the treatment of interest must be determined, as well as the effects of all other interventions that can

be used in treating the disease. For example, in assessment of HPV introduced to prevent cervical cancer, it is important to calculate costs other than those avoided through the prevention of cancer due to the use of the vaccine. These costs include those incurred through cervical screening (as potentially the duration between routine examinations could increase following vaccine introduction, thus reducing the overall cost of screening), and costs of treating other vaccine preventable health conditions attributable to the virus, such as genital warts, such as in the analysis by Jit et al. (2008). If the other preventable costs are not considered the vaccine may prove too costly to be introduced. Jit et al. state that a reduction in treatment for warts accounts for half of the discounted cost savings to the health service.

It is essential to decide a time frame for the model in order to correctly identify the events to be incorporated into the model. For example, in models of HPV, the time frame selected can be the entire lifetime as cervical cancer can occur at any point in adult life. This is the case in the model by Colantonio et al. where a cycle of one year is considered from 11 years of age until death due to the nature of the virus and due to the fact that the vaccine should be administered to children of 12 years of age.

In order to carry out some of the modelling techniques adopted in a health economic analysis, data are required. The clinical effectiveness information utilised in a health economic analysis can come from a variety of sources and there is a defined hierarchy of information that may be used. The best, most valid and reliable evidence to use in an analysis is accrued through a systematic review of the literature on the effectiveness of interventions for a health condition of interest. Second in the hierarchy is a well-designed pragmatic randomised controlled trial. When such a trial is not possible then an observational study will have to be carried out to obtain the information required. Finally, if all other measures fail to provide the required information, expert opinion must be sought. However, at this low level in the hierarchy there is a high chance of error and thus it is preferred to rely on alternatives if at all possible. It has been argued that this hierarchy is not always relevant as in some situations it is extremely difficult, if not impossible, to carry out randomised controlled trials. For example, if the

severe health outcome being assessed occurs rarely then it may be better to use an observational study for a cost-effectiveness analysis of a treatment for this outcome rather than a randomised controlled trial (Evans 2003).

4.6.1 Decision trees

The decision tree is a commonly used method in health economic analyses. A diagram of a simple decision tree is shown in Figure 4.4. There are seven principal components: decision and chance nodes, branch probabilities, pathways, pathway probabilities, costs and expected values. The decision node is the first component of the model and illustrates the decision that is to be considered for the patients in the analysis. In Figure 4.4, the decision being dealt with in the model could be assessing which of the interventions A or B is more cost-effective in preventing disease without negative consequences. In the analysis of PCV-7 by Butler et al., the decision node represents whether or not children receive the vaccine, see Figure 4.5.

From the decision node, the tree progresses to the chance nodes which portray the outcomes of the various primary health care interventions being considered. An example of a chance node from the Butler et al. decision tree is the possibility of getting invasive pneumococcal disease, pneumonia or otitis media for those unvaccinated children. Chance nodes occur moving across the tree, from left to right, showing subsequent events following further interventions. These nodes represent uncertain outcomes as it is not evident which event will occur. There may be positive or negative outcomes resulting from the use of a particular intervention depending on the individual to which it is administered.

Joining the nodes to one another are the branches of the decision tree and these link the nodes to the potential effects of the interventions. Probabilities are associated with the branches that correspond to the likelihood of the individuals experiencing the various events. The sum of the probabilities of all the branches leaving each chance node must equal 1 since these probabilities represent the proportion of the cohort of individuals that follow each branch and all individuals must be accounted for in the decision tree (Global Forum for Health Research 2008). Conditional probabilities occur when moving from one chance node to

other chance nodes due to the fact that the probabilities of these further events occurring are dependent upon the fact that the patient experienced a previous event. For example, returning to the Butler et al. decision tree example, Figure 4.5, to obtain the outcome 'with tympanostomy', a patient must first have severe OM.

The pathways in the decision tree are mutually exclusive and consist of the various branch combinations. The probabilities associated with each pathway may be calculated from the branch probabilities by multiplying the branch probabilities associated with each chance node in the pathway. Pathway costs may also be calculated by determining the costs for each of the events experienced in the pathway and summing these costs. The costs involved could be, for example, the cost of the intervention itself or the cost of hospitalisation. Finally, the expected values in the decision tree must be calculated. For example, the expected cost for each primary intervention can be calculated by multiplying the pathway costs by their respective probabilities and summing across the pathways that correspond to each intervention (Drummond et al. 2005).

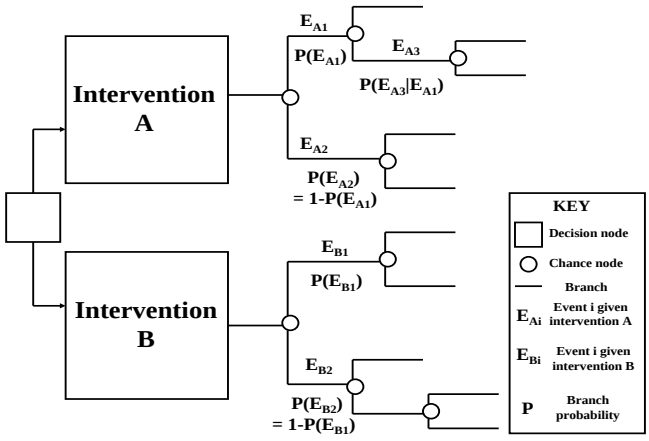

Figure 4.4: Decision tree.

The main advantage of using decision trees in evaluating the cost-effectiveness of a particular intervention is that the methodology is relatively simple to carry

out. The researcher is able to identify the information required to run the model and can determine when additional assumptions or expert opinion are required if information is not available for a particular parameter in the model. As mentioned previously, expert opinion is at the bottom of the hierarchy of requirements for information to be incorporated into a cost-effectiveness analysis. However, this methodology enables sensitivity analyses to be carried out for weaker data such as this. The sensitivity analysis is carried out in order to determine whether changes in the uncertain parameter that has been assumed to take a particular value, or has been estimated by expert opinion, lead to changes in the conclusions of the analysis. In the situation where the uncertain parameter has a dramatic effect on the analysis it is necessary to obtain more information about this parameter before reporting the results.

There are a variety of drawbacks to using decision trees in a health economic evaluation. Primarily, the problem arises with time dependent factors that are required in the model, such as discounting factors, as it is not possible to specify a time variable in this type of methodology. This problem is not exclusive to decision trees as sometimes it is not easy to include time dependency in Markov models due to the memoryless property of the Markov model. This property is discussed in the next section. A second drawback is that decision trees can become very complicated and can involve many possible branches depending on the illness or disease that is under consideration. In particular, it is difficult to specify all of the potential events for persistent conditions that can occur over the length of the life of the patient.

Returning to the Butler et al. decision tree, the costs and benefits of a four dose vaccination regime are compared with those when no vaccination is adopted. A section of the decision tree is shown in Figure 4.5. In this figure, estimates for the expected costs and event probabilities are shown for the OM strand of the tree. These figures can be used to calculate total expected costs. For example, the expected cost of a severe case of OM is $(0.39 \times \$1,465) + (0.61 \times \$2,292) = \$1,969.47$. Expected costs and benefits for all of the pathways in the decision tree were calculated in order to decide whether to adopt PCV-7 in Australia. On the basis of the analysis carried out, Butler et al. concluded that adopting the four

dose vaccination schedule in Australia, at a cost of $90 per dose, would not prove to be cost-saving, regardless of the reduction in hospitalisations and in mortality.

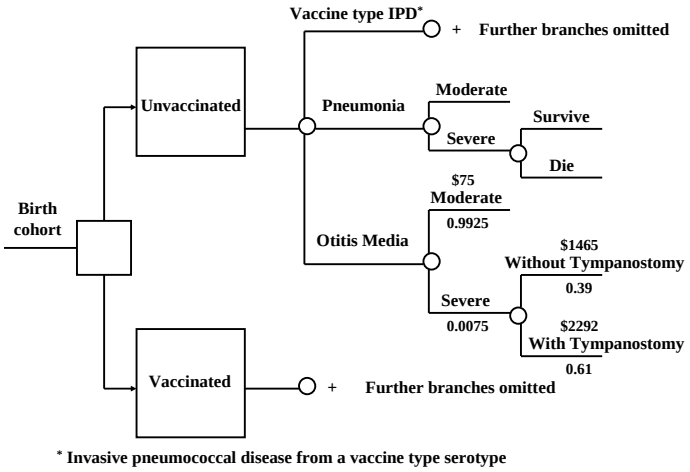

Figure 4.5: Decision tree model structure (adapted from Butler et al., 2004).

4.6.2 Markov chains

Markov chains may be used when the timing of the events being considered is important in the economic analysis. This type of model is a stochastic process defined in discrete time. As Markov chains consider time, they can be employed in situations unsuitable for the use of decision trees. Instead of branches that describe the progression of a patient dependent on the intervention adopted, Markov chains are based on the states a patient may take during defined cycles. A cycle in a Markov chain is a discrete period of time in which the patients may move between the various states. The duration of the cycle is determined by the health condition and the types of health care technology that are being considered in the CEA. Cycles are typically months of a year or years. A Markov chain may consist of several cycles and within each cycle there is a cost associated with each state.

The principal property, or Markov property, of this type of model is that it has

no memory. This means that to determine what state a patient may take in the future, only the present state need be considered. It is unimportant how the patient arrived at the current state. Once a patient has arrived at a certain state in the Markov chain, the patient can either move to another state or remain in that same state according to a set of given fixed probabilities. These probabilities are termed transition probabilities as they are used to define transitions the patient makes between states in the chain. In a Markov chain, the transition probabilities are fixed with respect to time. However, in another type of Markov model, a time-dependent Markov process, transition probabilities are allowed to vary over time (Briggs and Sculpher 1998). Time-dependent Markov processes are useful for models of continuing disease as, unlike the simple Markov chain, it is not assumed, for example, that there is the same probability of death for young individuals as for the elderly (Sloan 1996).

As mentioned previously, the memoryless property may cause problems with time dependency in models. For example, Drummond et al. (2005) discuss an example relating to AIDS patients. Patients can enter the AIDS state from either a state where the CD4 count is between 200 and 500 cells/mm^3 or from a state where the count is less than 200 cells/mm^3 but once in the AIDS state the model does not distinguish between patients who came from one CD4 state or the other. Drummond et al. state that this is a problem if evidence suggests that the mortality risk is higher for those who entered into the AIDS state from the higher CD4 count state than for those from the lower CD4 count state. However, this problem may be avoided by creating two AIDS states: one state for those who entered with lower CD4 counts and one for those who entered with higher CD4 counts.

Figure 4.6 shows a basic Markov chain consisting of only three possible states: 'Infected', 'Recovered' and 'Dead'. It can be noted that the state 'Dead' is an absorbing state. This means that once an individual has entered this state, he or she cannot depart from it. Therefore, it has an associated transition probability of remaining in this state, P_7, of 1.

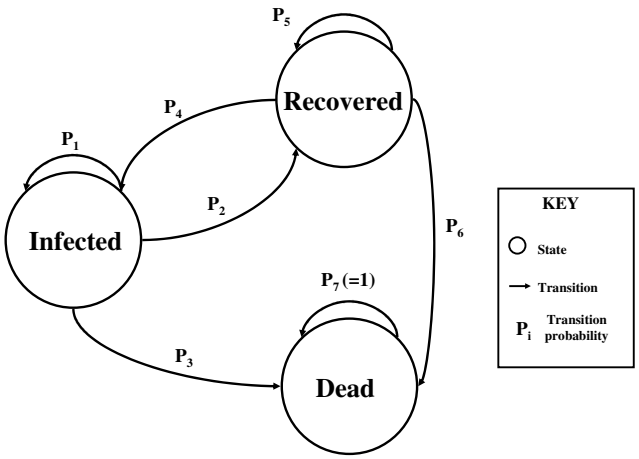

Figure 4.6: Markov model (adapted from TreeAge Software Inc., 2008).

	Infected	Recovered	Dead
Infected	P_1	P_2	P_3
Recovered	P_4	P_5	P_6
Dead	0	0	P_7

Figure 4.7: Matrix of transition probabilities.

Figure 4.6 may also be expressed as a square array called a transition matrix (Figure 4.7). In the transition matrix, the first row contains the one step transition probabilities of moving to an 'Infected', 'Recovered' or 'Dead' state if the current state is 'Infected'. The second row of the matrix contains the transition probabilities for each state if the patient is in a 'Recovered' state. Similarly, the third

row contains the probabilities of a patient making a transition from the 'Dead' state. As a patient must always be in one of the states defined in the Markov chain, the transition probabilities in each row must sum to one. Therefore, P_7 must equal one.

When using a Markov chain in a health economic analysis the expected costs for each intervention are calculated by first determining the length of time, on average, a patient spends in the different states. These times are then weighted by the costs associated with the state. To calculate these expected costs, the probabilities for a patient being in each state must be determined for every cycle in the model. To do this, the proportion of patients in each state is determined for every cycle by assessing the progression of a cohort of patients through the model. The important factors are the proportion of patients that were in each state in the previous cycle and the transition probabilities between the states. The costs for each state are multiplied by the proportion of the cohort of patients that are in the respective state for each cycle and these costs are then totalled for all cycles in the model to find the overall expected cost. The expected effect may be calculated in a similar manner (Drummond et al. 2005).

A Markov model was used in a Norwegian study to determine the cost-effectiveness of introducing PCV-7 to the routine infant immunisation schedule (Wisløff et al. 2006). The Markov process Wisløff et al. used involved cycles of 1 year from birth until death at age 100 years. A section of the Markov process is shown in Figure 4.8. This figure contains a section of the Markov model created to assess the various disease progressions from pneumococcal meningitis. The Markov model begins with newborn healthy children who can progress into one of four different pneumococcal disease or infection states. These diseases lead each patient into a mutually exclusive subsequent state such as hydrocephalus, epilepsy, neurological sequelae or death. As with the Australian study carried out by Butler et al. using a decision tree approach, the Norwegian study by Wisløff et al. concluded that the pneumococcal vaccination would not be cost-effective for routine use in a four dose infant immunisation schedule. The reasons behind this decision are discussed later in the chapter.

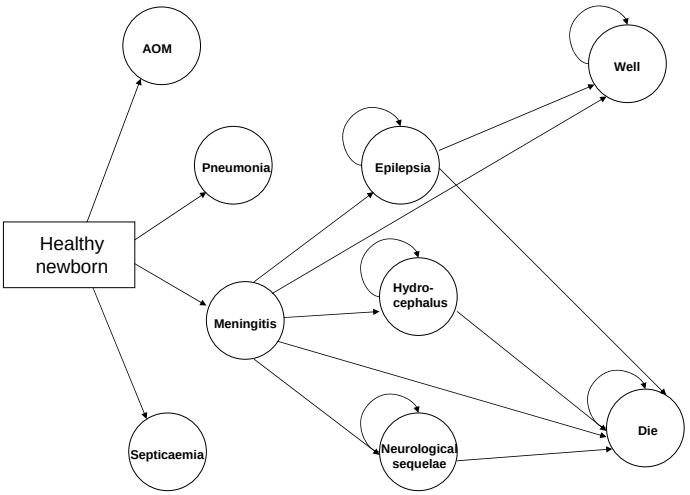

Figure 4.8: Section of Markov chain for the Norwegian PCV-7 cost-effectiveness analysis (adapted from Wisløff et al., 2006).

4.6.3 Monte Carlo simulations

Monte Carlo simulations are commonly used in probabilistic sensitivity analyses for cost-effectiveness models. In a probabilistic sensitivity analysis, a joint probability distribution is established which describes the uncertainty in the input parameters in the model. The Monte Carlo method involves repeated simulation of random input parameters for the model. Each set of input parameters is then adopted in the model to determine the effect of the changing parameters on the outcome. Ideally, at least 10,000 different simulations are required to obtain a clear idea of the uncertainty in the model.

Monte Carlo simulations have the benefit of allowing uncertainty and randomness to be taken into account in the cost-effectiveness model as probability distributions are adopted for the parameters rather than point estimates, thus enabling more informative predictions to be made.

Melegaro et al. (2004) made use of Monte Carlo simulations to carry out multivariate sensitivity analyses to determine the extent to which indirect vaccine

effects such as herd immunity affect the model results in a cost-effectiveness analysis of PCV-7 carried out for England and Wales. Latin Hypercube sampling was used to determine initial parameter values from probability distributions in the Monte Carlo simulations. In Latin Hypercube sampling, a set of n values are selected from j random variables by dividing the range of each of the j variables into n independent intervals which each have equal probability and randomly choosing a single value from each interval according to the probability density in the interval (Wyss and Jorgensen 1998). Melegaro et al. assumed different parameters, including age-specific incidence rates, case-fatality ratios, length of hospital stay and 35 different parameters related to costs of hospital care, followed a Uniform distribution.

age-specific incidence rates, casefatality ratios and length of stay in the hospital as well as for all the parameters related to the cost of care and treatment

The simulation was carried out 1,000 times in order to obtain a distribution for the outcome values.

Figure 4.9 shows the simulation results obtained from five different scenarios under consideration in the cost-effectiveness analysis. In the base case scenario, no indirect effects of vaccination such as herd immunity or serotype replacement were considered. Other scenarios were compared to this base case scenario to try to consider serotype replacement and herd immunity as these effects are known to occur in practice. Most alternative scenarios assumed a 5% reduction in pneumococcal disease for those not vaccinated with PCV-7. One scenario considered higher levels of herd immunity taken from levels observed in an American study by Whitney et al. (2003). As far as serotype replacement was concerned, two scenarios considered complete replacement of vaccine serotypes with non-vaccine serotypes (serotype replacement coefficient = 100%), and one scenario considered a varying level of serotype replacement between 0 and 1 (serotype replacement coefficient variable (0,1)). The results from this study showed that routine childhood immunisation with PCV-7 would not prove cost-effective in the base case scenario, where no herd immunity benefits are considered. If the herd immunity effects observed by Whitney et al. apply to England and Wales then the vaccine does prove cost-effective. In the scenarios where only a herd immunity effect of 5% is considered, the vaccine still proved to be cost-effective unless complete

serotype replacement occurs.

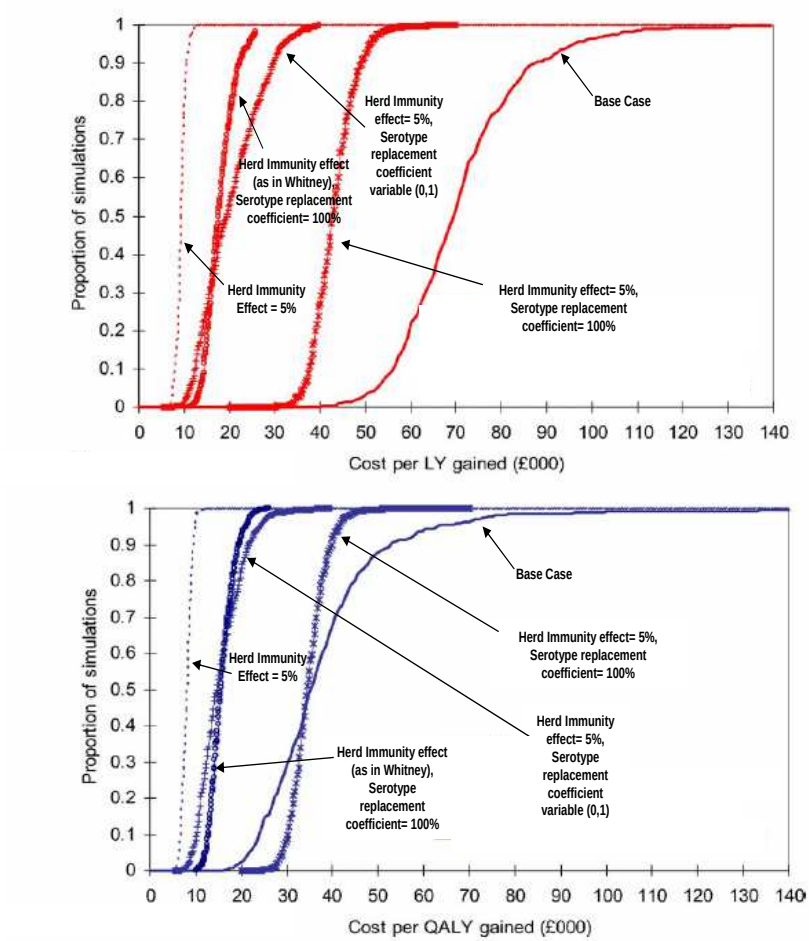

Figure 4.9: Graphs showing the cost per life year gained and cost per quality adjusted life year estimated using Monte Carlo simulations in a scenario sensitivity analysis (adapted from Melegaro et al., 2004).

4.6.4 Bayesian technique

To use Bayesian techniques in health economic analyses, prior probability distributions for the uncertain parameters in the model, such as the event rates and transition probabilities, must be specified. Thus, the Bayesian technique, unlike Monte Carlo simulations, requires data to update prior beliefs about these parameters. The prior distributions reflect the beliefs or knowledge of the modeller about the plausible values that may be taken by the parameters. However, it is perfectly acceptable to specify a prior distribution that indicates a complete lack of knowledge about the parameters. For example, a Uniform distribution may be adopted when there is no information available about the possible values a parameter may take. Such a prior is called a non-informative prior. In the case where valid information has been used as the basis for the prior distribution, the prior is called an informative prior. There are another two possible types of prior that may be relevant: skeptical priors and structural priors. Skeptical priors are used in situations where the null hypothesis is expected to be true whilst structural priors can be adopted when there is information available about how the model parameters are related to one another (O'Hagan and Luce 2003). In many situations there are a number of different distributions that may be selected for use as the prior distribution so there is no single correct distribution that must be adopted.

In addition to the prior distribution, the likelihood must be established. Information from clinical trials about the effect of the interventions under consideration is used to determine the likelihood function. The likelihood describes the support that exists for the different values of the effect of the interventions (Spiegelhalter et al. 2004).

Finally, in a Bayesian analysis, the prior distribution and the likelihood are combined to obtain the posterior distribution. Using a Bayesian approach, all inferences about the unknown parameter are made using the posterior distribution.

Using the notation of Gelman et al. (2004), let θ represent some unknown outcome parameter in the health economic analysis. A prior distribution must be specified for θ. This prior distribution is represented by $p(\theta)$. $p(y|\theta)$ is the likeli-

hood of obtaining the clinical trial data, y, given θ. The likelihood is effectively the conditional probability of observing the data, given a specific value for θ determined for each of the plausible values for θ (U.S. Department of Health & Human Services Food & Drug Administration 2008). Once this likelihood has been established, the posterior distribution, $p(\theta|y)$, can be determined using the following relationship:

$$p(\theta|y) \propto p(\theta)p(y|\theta). \tag{4.1}$$

One benefit of using Bayesian techniques in an economic analysis is the ease with which a model may be used to update the results when new information is available. That is, the posterior distribution may be adopted as the prior distribution and used to update the results about the unknown parameter θ when more data are available. As more information about the parameter is acquired, the more certain the modeller will be about the posterior distribution. The initial prior distribution chosen will have a small effect on the outcome distribution given the addition of more and more information for the unknown parameter.

Unfortunately, an example of the use of the Bayesian technique in a cost -effectiveness analysis of PCV-7 could not be identified. However, the Bayesian technique was adopted in a health economic analysis of an intervention to prevent HIV (Johnson-Masotti et al. 2001). The example focusses on a CUA carried out to assess HIV interventions for seriously mentally ill patients. To use the Bayesian approach, Johnson-Masotti et al. specified prior distributions for the key parameters for sexual risk such as the prevalence of HIV infection and condom effectiveness. In addition, prior distributions were specified for parameters such as the QALYs and intervention costs. 5,000 estimates for the parameters were obtained by sampling from the prior distribution and combining with empirical data available on parameters such as the number of partners and the number of unprotected sex acts. Johnson-Masotti et al. compare the results obtained from the Bayesian analysis to those from a univariate sensitivity analysis and state that they obtain comparable results. However, they state that using the Bayesian approach in a probabilistic cost-effectiveness analysis provides a greater degree of

certainty about the cost-effectiveness of the intervention under consideration.

4.7 Economic evaluations of routine childhood pneumococcal conjugate vaccination

Economic evaluations regarding health care can prove to be complex regarding issues of how best to measure the benefit outcome obtained. The evaluation of vaccine interventions can ultimately become even more difficult due to the fact that some benefits may be hard to measure or may not be observed immediately following use of the intervention. For example, herd immunity is a potential benefit occurring from the use of vaccines such as PCV-7 but it is difficult to obtain estimates of the scale of this immunity as the true impact may not be observed for many years.

In a review by McIntosh of various cost-effectiveness analyses of PCV, problems regarding the economic evaluation of this vaccine are discussed (McIntosh 2004). In this review, McIntosh explains that one of the key problems in analysing the PCV is in identifying the health burden of this bacterium. Pneumococcus is accountable for a variety of diseases such as meningitis, septicaemia, pneumonia and sinusitis, as well as common infections such as OM.

With certain pneumococcal diseases it is difficult to obtain a picture of the true burden in the population due to problems with the reporting of these diseases. It is difficult to obtain accurate estimates of the number of cases of childhood OM in a population due to differences in the way in which this condition may be treated. In the UK, some children will be seen by a GP where the incidence of OM may be recorded but the information may never be collected for analysis if there are no routine reporting systems for this type of condition. For example, in Scotland, the Scottish Meningococcus and Pneumococcus Reference Laboratory routinely collect information and samples from cases of invasive disease but not for non-invasive diseases or infections.

In various economic analyses of the PCV, the issue of assessing OM is addressed. In the Norwegian study of the cost-effectiveness of implementing PCV-7 discussed

previously (Wisløff et al. 2006), after carrying out a sensitivity analyses of various components in the Markov model used, it was noted that one of the important factors in determining the cost-effectiveness of the vaccination is the incidence of OM. This is the case since OM is the most common pneumococcal disease. OM was also found to be important in the sensitivity analysis due to the potential problems that can occur in the development of OM, such as the development of acute coalescent mastoiditis. However, Wisløff et al. pointed out that there was difficulty obtaining data on the occurrence of acute OM in Norway and that Danish data had to be used in the analysis. Similarly, in a Swedish cost-effectiveness analysis of PCV-7 (Bergman et al. 2008), the incidence of acute OM, as well as the ability of the vaccine to prevent it were identified as important factors in the sensitivity analysis of the variables in the Markov model used. With this analysis, the issue of the uncertainties surrounding true incidence rates of acute OM was raised and addressed through extensive sensitivity analyses varying the rates. Bergman et al. raised the point that a substantial component of cost savings in the use of PCV-7 would be due to lowered rates of complications attributable to the development of acute OM and that this was indeed the most influential variable in the Swedish Markov model. In an Italian cost-effectiveness study (Marchetti and Colombo 2005), limitations to the analysis regarding the treatment of OM were listed, such as the uncertainty in the rate of occurrence of OM which Marchetti et al. found difficult to take account of fully in the analysis. In addition, Marchetti et al. mentioned the exclusion of the potential benefit of reduction in the recurrence of OM in their Markov model. This exclusion is likely to be attributable to the difficulty in obtaining reliable data of this nature. These studies indicate the importance of accurately reflecting the true burden of OM in economic analyses of the pneumococcal conjugate vaccination programme as, even though OM is the least serious of the diseases attributable to *S. pneumoniae*, it is the most common of the pneumococcal diseases.

Other complications in the analysis of the PCV occur due to the fact there are at least 90 different types of the bacterium in circulation, with differing abilities to cause infection and disease. The conjugate vaccine currently on offer in the UK contains only seven of the 91 types of this bacterium and other vaccines that have been or are currently undergoing trials contain only ten and thirteen

types. Therefore, there may be problems which arise due to changing trends in the circulating pneumococcal types which could affect the impact of the vaccine.

Of the twenty-four economic analyses of the cost-effectiveness of pneumococcal conjugate vaccination published between 2000 and 2009 summarised in Table 4.1, only the UK study by Melegaro et al. (2004) directly tackles the issue of serotype replacement in depth in an economic evaluation. However, Ray et al. comment that their model results do take serotype replacement into account due to the fact that the rates of IPD for unvaccinated hosts were taken from observed cases and were irrespective of the serotype of the disease. Although, it is stated that if this replacement were to become greater then the future vaccine efficacy may be lessened.

In the economic analysis by Melegaro et al. a situation where, after implementation of PCV-7, there was a complete replacement of VT serotypes with NVT serotypes was considered. It was assumed that the NVT serotypes had the same ability to cause community acquired pneumonia and acute OM as the VT serotypes whilst estimates for the ability of non-VT serotypes to cause IPD were acquired from a study by Brueggemann et al. (2003). The conclusions from Melegaro et al. were mentioned earlier in the discussion of Monte Carlo simulations, with the vaccine proving cost-effective if herd immunity is taken into account, even when partial serotype replacement is present.

In a review of fifteen economic analyses of pneumococcal conjugate vaccination published between 2002 and 2006, the importance of considering herd immunity as one of the potential benefits of adopting PCV-7 in a childhood vaccination schedule was discussed (Beutels et al. 2007). Beutels et al. stated that herd immunity effects are greatly influential in a cost-effectiveness analysis but that herd immunity was only considered in 2 of the 15 papers they reviewed (McIntosh et al. 2005; Melegaro and Edmunds 2004a). However, following the publication of the USA study showing herd immunity effects (Whitney et al. 2003), more health economists are taking herd immunity into account.

Ten of the twenty-four economic evaluations mentioned previously were published between 2006 and 2009 and were thus not included in the Beutels et al. review.

Author	Country	CEA	CUA	CBA	DTree	Markov	Monte Carlo
Lieu et al. 2000	USA	✓			✓		
McIntosh et al. 2003	UK	✓			✓		
De Wals et al. 2003	Canada	✓	✓				✓
Lebel et al. 2003	Canada	✓			✓		
Ess et al. 2003	Switzerland		✓		✓		
Ruedin et al. 2003	Switzerland		✓		✓		
Bos et al. 2003	Netherlands		✓		✓		
Butler et al. 2004	Australian	✓	✓		✓		
Asensi et al. 2004	Spain	✓				✓	
Melegaro et al. 2004	UK	✓	✓		✓		✓
Salo et al. 2005	Finland	✓	✓			✓	
Marchetti and Colombo 2005	Italy	✓				✓	✓
Navas et al. 2005	Spain	✓	✓	✓	✓		
McIntosh et al. 2005	UK	✓			✓		
Ray et al. 2006	USA	✓			✓		
Wisløff et al. 2006	Norway	✓				✓	✓
Hubben et al. 2007	Netherlands	✓	✓		✓		✓
Sinha et al. 2007	72 developing		✓		✓		✓
Lloyd et al. 2008	Germany	✓			✓		
Tilson et al. 2008	Ireland	✓			✓		
Bergman et al. 2008	Sweden	✓	✓			✓	
Claes et al. 2009	Germany	✓			✓		
Silfverdal et al. 2009	Sweden	✓			✓		
Giorgi-Rossi et al. 2009	Italy		✓		✓		

Table 4.1: Key features of health economic models of PCV.

Of these ten analyses, eight addressed herd immunity in the evaluation. The two analyses that did not consider herd immunity were the international cost-effectiveness analysis involving 72 different countries (Sinha et al. 2007) and an Italian analysis (Giorgi-Rossi et al. 2009). Sinha et al. justified the exclusion of herd effects observed in the USA in their economic analysis by explaining that differences in population interactions and exposures between the USA and the developing countries may cause differences in herd immunity and, thus, assumptions involving US herd effects would not necessarily be relevant in their model. Giorgi-Rossi et al. state that they do not consider herd immunity in their model due to the fact they only consider a ten year period in the analysis. Thus, they felt that this reduces the relevance of herd immunity. In addition, they felt that this positive benefit of vaccination may be offset by other negative effects such as serotype replacement and antibiotic resistance.

In all but one of the eight studies which included herd immunity effects, favourable results were obtained regarding the cost-effectiveness of routine infant vaccination after inclusion of this indirect benefit. However, Wisløff et al. (2006) reported that in Norway, a four dose vaccination schedule would not prove to be cost-effective, even after inclusion of herd immunity and other indirect benefits of vaccination. If, however, a three dose schedule were adopted, where three doses elicit the same efficacy as four, the vaccine would prove cost-effective in Norway but only if other indirect effects, such as lost work time of parents with children suffering from pneumococcal disease, were included in the analysis in addition to herd immunity effects. In the study by Bergman et al. (2008), it was noted that even though herd immunity had been included in their analysis, some herd immunity effects are difficult to quantify so there is still an underestimation of the benefits of vaccine to unvaccinated individuals. Bergman et al. point out that their study includes only herd immunity in adults in the population and does not consider the potential impact of vaccinated children eliciting protection to siblings. Hubben et al. clearly show the benefit of including herd immunity in a cost-effectiveness analysis. They extend the model by Bos et al. (2003), in which a conclusion was reached that the inclusion of PCV-7 in a childhood vaccination schedule would not prove cost-effective in comparison with other health care interventions in use in the Netherlands, to include indirect benefits of vaccination,

such as herd immunity, and conclude that PCV-7 would prove cost-effective in the Netherlands.

One of the key issues in the economic analyses of PCV-7 is the price of the vaccine. PCV-7 is currently the most expensive infant vaccination available (Beutels et al. 2007) and, as such, the price of vaccination was highlighted as the most sensitive variable in economic analyses of PCV-7 (De Wals et al. 2003). With the vaccination price in mind, certain countries, such as the UK and Norway, have adopted a three dose immunisation schedule (two doses plus a booster dose) instead of the recommended four dose schedule (three doses plus a booster) on which the vaccine efficacy trials were based. In a Swedish economic analysis of the three dose schedule, one of the crucial assumptions in proving the cost-effectiveness of a three dose schedule was that three doses of PCV-7 elicit the same protection as four doses (Bergman et al. 2008).

Further to the issues discussed, McIntosh raises the point that there are greatly different rates of disease for different age groups within the population that will affect the ability to assess the cost-effectiveness of the vaccine (McIntosh 2004). In addition, various limitations were mentioned in the cost-effectiveness analyses studied which could affect the conclusions reached such as the unknown duration for which the vaccine would remain effective (Ruedin et al. 2003), the failure to acknowledge vaccination of high-risk groups with PCV-7 in a cost-effectiveness analysis (Butler et al. 2004; Marchetti and Colombo 2005) and the potential overestimation of the effect of vaccination in children due to disease and mortality occurring in children aged under 3 months of age who are too young to have received the PCV-7 vaccination (Bergman et al. 2008; Butler et al. 2004). Furthermore, other problems with cost-effectiveness analyses of PCV-7 occur due to the failure to quantify the beneficial effect the vaccine may have in preventing antibiotic resistant strains of pneumococcus (Navas et al. 2005; Bergman et al. 2008).

4.8 Summary

In this chapter, the different types of analysis and methodology which can be adopted in a health economic assessment have been discussed, with examples from published studies. In addition, published health economic analyses of PCV-7 have been summarised. In the next chapter, a statistical analysis of hospital episodes of the pneumococcal diseases meningitis, septicaemia and pneumonia is carried out. The intention of this analysis is to provide updated information to input into existing UK cost-effectiveness analysis models for PCV-7.

Chapter 5

Trend analysis of hospital episodes of septicaemia, meningitis and pneumonia

5.1 Introduction

In this chapter, a statistical analysis of cases of pneumococcal disease in England and Wales will be carried out with an aim to improve parameter estimates of PCV-7 efficacy to use in the CEA model created by Wyeth pharmaceuticals, discussed in Chapter 4 (McIntosh et al. 2003; McIntosh et al. 2005).

Cases of both pneumococcal and unspecified septicaemia, meningitis and pneumonia identified in England and Wales are considered in this chapter, with separate models fitted to each of these six classifications of disease to identify any trends in the occurrence of each disease prior to PCV-7 use. In addition, models are fitted to assess trends in two comparator groups, femur and forearm fractures, which are independent of pneumococcal infection or disease and should remain relatively stable in number over time. These models are fitted to the comparator groups in order to determine whether or not there is evidence of changes in the practice of recording hospital episodes in England and Wales.

5.2 Background

Invasive and respiratory pneumococcal diseases, such as septicaemia, meningitis and pneumonia, annually cause approximately 3 million deaths worldwide; 1 million of which are young children (Adrian et al. 2004). Both PPV-23 and PCV-7 were introduced to reduce the burden of pneumococcal disease in the age groups most at risk of developing pneumococcal disease, those aged 65 years and over and those aged under 2 years. However, it is believed that PCV-7 should reduce the burden of pneumococcal disease for all age groups due to the potential for herd immunity. This is not the case with PPV-23 which does not prevent carriage of the VT serotypes. It is important to assess the impact of these vaccines on the incidence of pneumococcal disease since introduction.

In this chapter, Hospital Episode Statistics (HES) (Hospital Episode Statistics 2007) from the NHS Information Centre for Health and Social Care of the number of cases of septicaemia, meningitis and pneumonia in England and Wales from 1993/94 to 2005/06 were used to identify trends for each disease over that time period. This data includes information gathered from over 300 Trusts and Primary Care Trusts in England and Wales and includes only information on in-patient and day cases. Out-patient information is not included.

The HES dataset contains information about the number of cases of each disease but, as an individual is able to have more than one episode of pneumococcal disease a year, the data are not representative of the number of patients that have hospital episodes of pneumococcal disease in England and Wales. The dataset includes information about the number of hospital episodes for ten different age groups: 0-3 months, 3-11 months (where those aged $>$ 3 months are included in this category), 1 year, 2 years, 3 years, 4-9 years, 10-17 years, 18-39 years, 40-64 years and 65 years and over.

The hospital episodes are categorised according to the International Classification of Diseases (ICD) codes. The dataset spans two different classification codes as ICD-10 codes were introduced in 1995/96 to replace the previously used ICD-9 codes. On examination of the disease descriptions associated with each code in the HES dataset, it was decided to combine the categories to create six categories

of health episode, as shown in Table 5.1, according to their clinical relevance.

Table 5.1: Hospital episode categories.

Disease	**ICD-9**	**ICD-10**
Pneumococcal Septicaemia	0382	A403
Unspecified Septicaemia	0389	A419
Pneumococcal Meningitis	3201	G001
Unspecified Meningitis	3209, 3229	G009, G039
Pneumococcal Pneumonia	481, 481-, 4819	J13X
Unspecified Pneumonia	485, 485-, 4850,	J180,
(includes unspecified bronchopneumonia	4851, 4856, 486,	J181, J188,
and lobar pneumonia)	486-, 4860, 4869	J189

These categories were combined as such based on the description of the episode provided with the HES data. For example, considering the ICD-9 codes for unspecified meningitis in Table 5.1, the description for code 3209 was 'bacterial meningitis, meningitis due to unspecified bacterium' whilst the description for code 3229 was 'meningitis of unspecified cause, meningitis unspecified'. Similarly, considering the ICD-10 codes for unspecified meningitis, G009 is described as 'bacterial meningitis, not elsewhere classified, bacterial' and G039 is 'meningitis due to other and unspecified causes, meningitis'.

PPV-23 was introduced for routine vaccination of the elderly in 2003 in the UK, with a staggered introduction in England and Wales with those aged over 85 years receiving the vaccine in 2003, those aged over 75 years in 2004 and those aged 65 years and over from 2005. Thus, this dataset may be used to assess whether or not PPV-23 had an impact on the number of hospital cases of septicaemia, meningitis and pneumonia. As PCV-7 was not introduced for routine use in the UK until 2006, the number of episodes of each disease can be predicted from the models for earlier years and these numbers may be compared to the observed number when new data are available. This will assist in determining whether or not the vaccine has reduced the number of hospital episodes of pneumococcal diseases septicaemia, meningitis and pneumonia. The primary aim of this analysis was to find suitable models to describe the trend for each of the six disease categories described in Table 5.1. The secondary aim was to identify whether or not there

were differences in the trends within each disease for the different age groups.

All statistical analysis in this chapter was carried out using R Version 2.9.1.

5.3 Initial analysis

To assess the trends of each category of disease from 1993/94 to 2005/06, plots of the number of cases of each disease against year were created for each age group.

From assessment of the plot of the number of cases of pneumococcal pneumonia for those aged between 0 and 3 months, shown in Figure 5.1, a substantial drop in the number of hospital episodes from 1994/95 to 1995/96 is observed. The same pattern was observed in each of the other age groups. It is likely that this difference is not due to a drop in episodes of pneumococcal pneumonia admitted to hospital but that the difference may be attributed to the change in the diagnosis coding from ICD-9 in 1994/95 to ICD-10 in 1995/96. On examination of the plot of unspecified pneumonia for the youngest age group, the opposite trend can be seen as there is an increase in the number of cases of unspecified pneumonia in 1995/96. Once again this pattern appears for all age groups. This could suggest that a greater number of cases of pneumonia remain undiagnosed as pneumococcal from 1995/96 onwards. As this will have an impact on the trend analysis, data from 1993/94 and 1994/95 is omitted from further analysis.

5.4 Trend analysis

Poisson and Negative Binomial regression were used to model the number of cases of each disease against time, in years from 1995/96 to 2005/06. Before carrying out the trend analysis, the population sizes for each year in England and Wales had to be determined to prevent results being influenced by increasing population sizes. In addition, as the age groups are not equal in size, the analysis was carried out to compare the rates of each disease across the age groups and not the number of cases. Thus, it was necessary to obtain these population estimates.

The mid-year population estimates were recovered from the Office for National

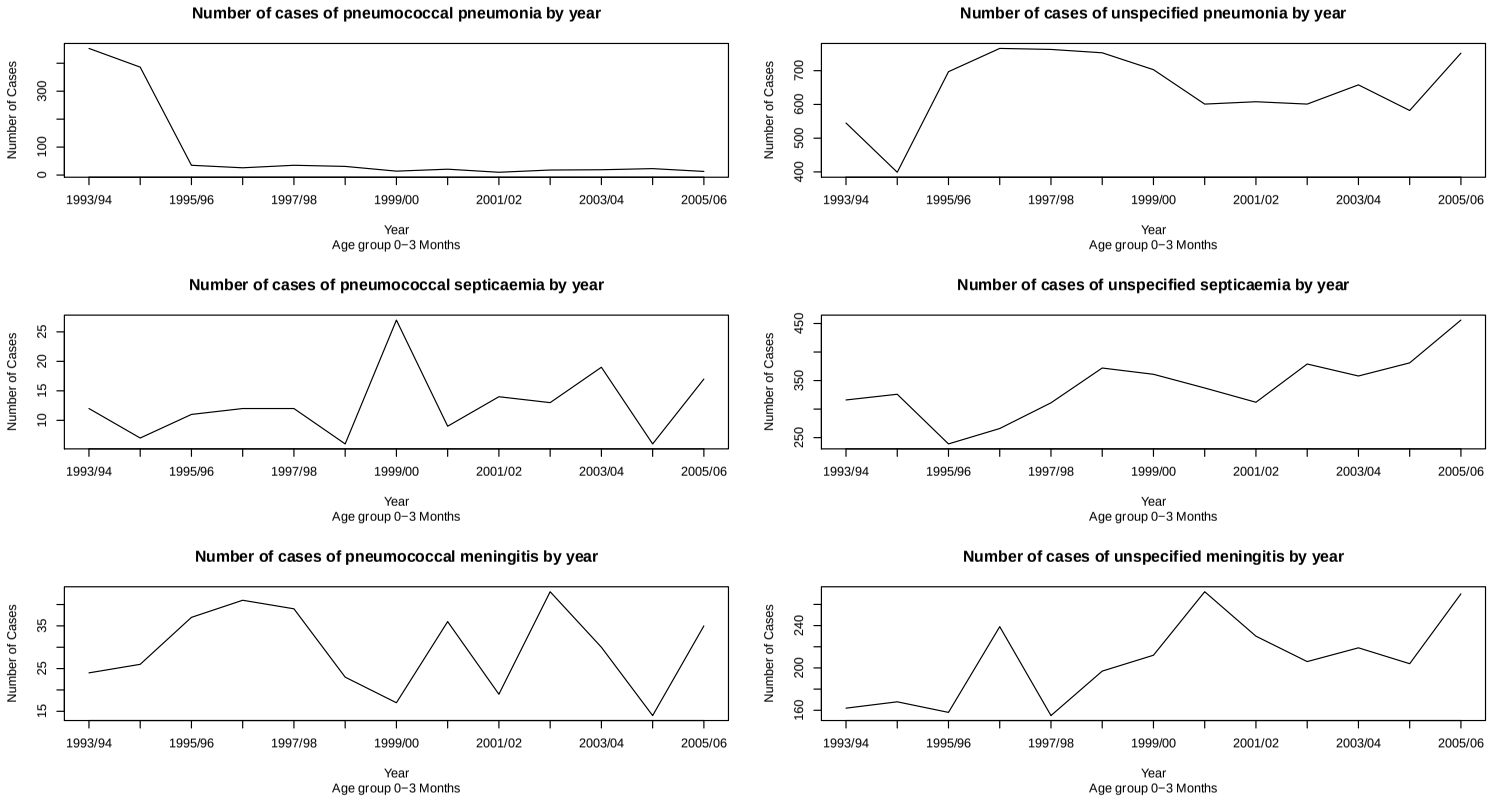

Figure 5.1: Trend plots for the 0 to 3 months age group.

Statistics (ONS) (Office for National Statistics 2008). The ONS only gives population estimates for age groups in years. Therefore, to obtain population sizes for the 0 to 3 months age group and the 3 to 11 months age group, a quarter and three quarters of the 1 year age group were taken for the 0 to 3 months age group and 3 to 11 months age group respectively to get a rough approximation of the population sizes. These population sizes were used as the basis for offsets in the Poisson and Negative Binomial regression models. The population sizes by year and age group are shown in Table 5.2.

5.4.1 Pneumococcal septicaemia

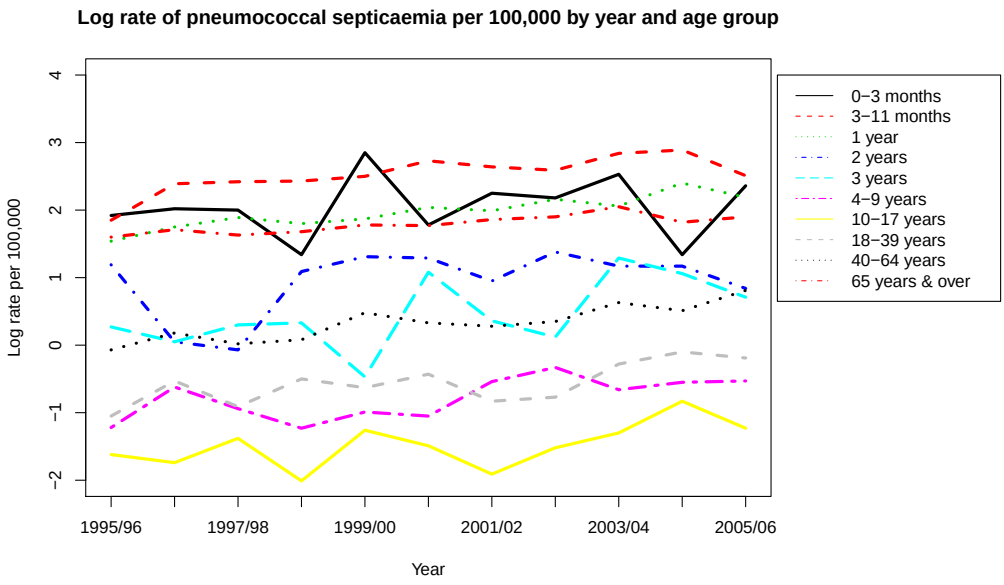

Figure 5.2: Log rate of episodes of pneumococcal septicaemia per 100,000 against year by age group.

The plot of the rate of hospital episodes of pneumococcal septicaemia per 100,000 against year by age group (Figure 5.2) shows that the highest rate of episodes generally occurred in the 3 to 11 months age group. However, the 0 to 3 months age group showed a higher rate in 1995/96 and 1999/00. No clear trend in the rates can be observed for any of the age groups.

Year	0-3M	3-11M	1Y	2Y	3Y	4-9Y	10-17Y	18-39Y	40-64Y	65Y+
1995/96	161,925	485,775	666,700	666,400	689,500	4,017,800	5,041,800	16,258,400	15,030,700	8,198,800
1996/97	159,150	477,450	646,100	665,500	666,800	4,108,100	5,133,400	16,183,600	15,148,800	8,221,400
1997/98	162,300	486,900	636,700	645,300	664,700	4,107,500	5,184,000	16,137,000	15,298,700	8,236,600
1998/99	157,850	473,550	647,700	637,800	645,100	4,086,500	5,229,100	16,101,700	15,482,800	8,258,100
1999/00	156,250	468,750	632,200	648,500	639,800	4,056,400	5,275,700	16,099,100	15,694,400	8,262,200
2000/01	151,675	455,025	622,600	631,500	648,300	4,006,600	5,326,300	16,095,400	15,915,700	8,287,000
2001/02	147,275	441,825	604,200	619,900	630,700	3,953,200	5,411,300	16,078,400	16,131,800	8,341,600
2002/03	146,950	440,850	588,000	602,600	618,700	3,888,300	5,466,300	15,996,000	16,343,400	8,389,500
2003/04	151,550	454,650	588,300	588,700	604,400	3,855,700	5,493,900	15,997,900	16,597,900	8,460,700
2004/05	157,275	471,825	605,500	588,200	589,000	3,802,500	5,502,900	15,973,800	16,834,700	8,519,900
2005/06	159,725	479,175	628,900	606,100	588,800	3,746,400	5,486,100	16,017,000	17,099,000	8,579,300

Table 5.2: Population sizes by year and age group.

To aid comparison of hospital episodes across the age groups, Table 5.3 shows the overall rate and 95% confidence interval for the proportion of cases with respect to the population size in the given age group. The confidence intervals were calculated for the proportions based on the Normal approximation to the Binomial distribution since $n\hat{p} > 5$, where n is the cumulative number of cases of pneumococcal septicaemia within an age group over multiple years and $\hat{p}$ is the estimated proportion of cases. The confidence intervals were calculated using the following equation:

$$\hat{p} \pm 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Table 5.3: Overall rate of hospital cases of pneumococcal septicaemia per 100,000 for each age group.

Age group	**Rate**	**95% C.I.**	
0-3 months	8.53	(7.15,	9.91)
3-11 months	12.89	(11.91,	13.87)
1 year	7.31	(6.67,	7.95)
2 years	2.80	(2.40,	3.19)
3 years	1.78	(1.46,	2.09)
4-9 years	0.47	(0.41,	0.54)
10-17 years	0.24	(0.20,	0.28)
18-39 years	0.59	(0.56,	0.63)
40-64 years	1.45	(1.39,	1.50)
65 years & over	6.05	(5.89,	6.21)

Table 5.3 shows that the overall rate of hospital episodes is greatest for the youngest age groups, with the highest rate of 12.89 per 100,000 for the 3 to 11 months age group, followed by 8.53 and 7.31 for the 0 to 3 months age group and the 1 year age group respectively. The 65 years and over age group also has a fairly high proportion of hospital episodes of pneumococcal septicaemia.

Poisson regression was used to model the number of episodes of pneumococcal septicaemia. In this model, year was included as a continuous explanatory variable and age group as a categorical variable. The age groups were combined into five age groups by grouping together the age groups from 2 years up to 40 to 64

years in the one category. Age groups 0 to 3 months, 3 to 11 months, 1 year and 65 years and over were retained as these groups are targeted by either PCV-7 or PPV-23. The categories used in the modelling are: 0 to 3 months, 3 to 11 months, 1 year, 2 to 64 years and 65 years and over. The reasoning behind these new age categories is that PCV-7 is administered to, and directly protects, those aged less than 2 years of age. In addition, PPV-23 is recommended for routine use in the elderly. By combining the age categories, 55 observations were used in the pneumococcal septicaemia model. A trend plot of pneumococcal septicaemia for the new age categories is shown in Figure 5.3.

On examination of the plot of the rate of episodes of pneumococcal septicaemia per 100,000 against year by age group for the five age group categories (Figure 5.3), it can be noted that the rate of cases of pneumococcal septicaemia in the 2 to 64 years age group is very low compared to the other age groups.

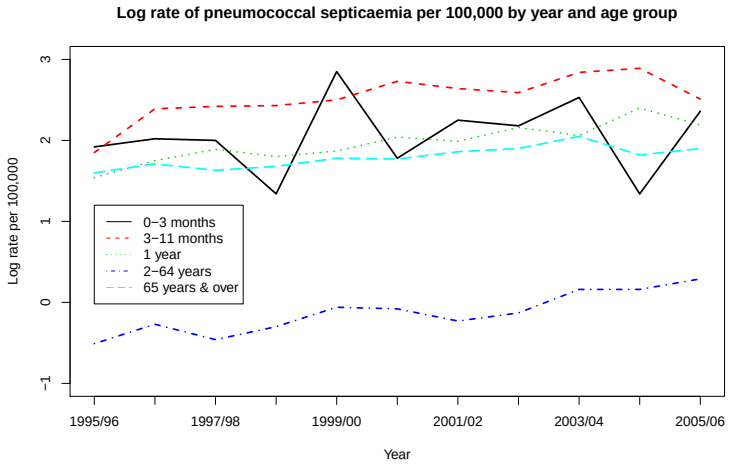

Figure 5.3: Log rate of episodes of pneumococcal septicaemia per 100,000 against year by age group with five age categories.

To attempt to account for the introduction of PPV-23 for the elderly in 2003 in the pneumococcal septicaemia model of all years, a change-point model was fitted to the data, similar to that carried out elsewhere (Bartsch et al. 2008). This complex model is used to test if there is any evidence of a change in slope in the 65 years and over age group following the introduction of PPV-23. In this model

an indicator term is used which allows the slope of the model to change in 2003. The year was centered at 2003 to ease interpretation of the parameter estimates. As PPV-23 is only administered to those aged 65 years and over routinely, an interaction between this age group and the indicator variable was fitted in the model.

In order to model the trend in the rate of pneumococcal septicaemia cases a Poisson regression model was fitted to the data as is appropriate when dealing with a response variable recording counts. This model included the variables year, age group, the two-way interaction between year and the 65 years and over age group, the two-way interaction between the indicator variable and this age group and the three-way interaction between year, the oldest age group and the year indicator. The model used an offset of the natural logarithm of the population size per 100,000 for each of the age groups. The model can be described using the following equation:

$$\begin{aligned}
\log(\text{rate}_i) &= \alpha + \beta_1(\text{Year} - 2003) + \beta_2\text{AgeGp} \\
&+ \beta_3(\text{Year} \geqslant 2003) \\
&+ \beta_4(\text{Year} - 2003)(\text{AgeGp} = 65\text{yrs \& over}) \\
&+ \beta_5(\text{Year} - 2003)(\text{Year} \geqslant 2003) \\
&+ \beta_6(\text{Year} \geqslant 2003)(\text{AgeGp} = 65\text{yrs \& over}) \\
&+ \beta_7(\text{Year} - 2003)(\text{AgeGp} = 65\text{yrs \& over})(\text{Year} \geqslant 2003),
\end{aligned}$$

where Year$\geqslant$2003 is the indicator variable. The three-way interaction term in the model is the variable required if there is evidence of a change in slope of the trend for the 65 years and over age group following the introduction of PPV-23. A significant two-way interaction between indicator and the 65 years and over age group is not indicative of a change in slope but rather a displacement from the estimated trend on the y-axis. Thus, it is of interest to assess the significance of this parameter as well as that of the three-way interaction.

A problem concerning the use of this change-point model is that there may be low power to detect any change in the slope due to the fact that there are only data for two observations following the year at which a change-point is thought to

have taken place. In addition, there may be problems due to correlations between the intercept and slope parameter as in a change point model the intercept at the change point and the slope before and after the change point are correlated.

A residual deviance of 107.65 was obtained on 44 degrees of freedom. As the residual deviance should be roughly equal to the residual degrees of freedom, it appears that this model is over-dispersed as the residual deviance is more than twice the residual degrees of freedom. If a Poisson regression model is the most appropriate model to use then the residual deviance should follow a χ^2 distribution with degrees of freedom equal to the residual degrees of freedom of the model (Vogt and Bared 1998). The p-value obtained for 107.65 following the χ^2 distribution with 44 degrees of freedom is less than 0.001 thus the null hypothesis that the Poisson model is appropriate to use to model the number of episodes of pneumococcal septicaemia should be rejected. The data are over-dispersed and Negative Binomial regression should be used. A Negative Binomial regression model is more appropriate to use as, unlike the Poisson distribution, the mean is not equal to the variance in this distribution. The Negative Binomial distribution is a generalisation of the Poisson distribution; the variance is equal to the mean plus some non-zero constant, k, multiplied by the mean squared.

A Negative Binomial regression model was fitted to the data. In order to determine whether or not each of the variables is required in the model, stepwise model selection procedures were used to identify the final model. The function stepAIC in R was used to identify the significant variables (Ripley and Venables 2002). This function carries out backward stepwise selection based on AIC where AIC is defined to be $-2L + 2k$ with L representing the maximum log-likelihood of the model being investigated and k is the number of parameters in the model. The optimal model is deemed to be the one with the smallest AIC value. The modelling results for the model including all possible variables is shown in Table 5.4.

Using backward stepwise selection, the only variables found to be significant in the model of pneumococcal septicaemia were year, age group and the interaction between the indicator variable and the 65 years and over age group. The results for this model are shown in Table 5.5, where Indicator represents the variable

Table 5.4: Parameter estimates and confidence intervals for the Negative Binomial regression model of the number of cases of pneumococcal septicaemia.

Parameter	**Estimate**	95% **C.I.**
Intercept	−9.181	(−9.359,−9.003)
Year	0.067	(0.051, 0.083)
3 to 11 months	0.413	(0.225, 0.601)
1 year	−0.153	(−0.345, 0.039)
2 to 64 years	−2.250	(−2.424,−2.076)
65 years & over	−0.356	(−0.558,−0.154)
Year:65 years & over	−0.018	(−0.043, 0.007)
Indicator	0.073	(−0.197, 0.343)
Year:indicator	−0.061	(−0.226, 0.104)
Indicator:65 years & over	−0.312	(−0.757, 0.133)
Year:indicator:65 years & over	0.092	(−0.180, 0.364)

Year $\geqslant$ 2003. The estimate of θ for the model shown in Table 5.5 is 223.5, 95% C.I. (17.8, 429.1), where θ is defined in the following expression for the variance of the Negative Binomial regression model, with μ representing the mean of the distribution:

$$\mu + \frac{\mu^2}{\theta}.$$

Thus, k described previously is equal to $\frac{1}{\theta}$. Therefore, in this model k is small at only 0.004.

As the three-way interaction term is not required in the model, there is no evidence that the slope changes for the 65 years and over age group from 2003. However, as mentioned previously, there are only two observations following the year at which a change-point is anticipated. Thus, there is little power to detect this change. The coefficient of year is positive. Therefore, the hospital episodes of pneumococcal septicaemia have been increasing between 1995 and 2005. The coefficient of the dummy variable for the 3 to 11 months age group is positive. Thus, this group has higher numbers of cases of pneumococcal septicaemia than the comparator, age group 0 to 3 months. All other age group coefficients are negative, with these groups having lower numbers of cases than the comparator. The interaction between the 65 years and over age group and the indicator vari-

able is negative. This indicates that, although there is no change in trend in cases in the over 65 years age group following 2003, a drop in the number of anticipated cases of pneumococcal septicaemia was observed in this group. This is shown in the plot of the predicted log rates of pneumococcal septicaemia, Figure 5.4.

Table 5.5: Parameter estimates and confidence intervals for the best fitting Negative Binomial regression model of the number of cases of pneumococcal septicaemia.

Parameter	**Estimate**	95% **C.I.**
Intercept	−9.208	(−9.384,−9.033)
Year	0.060	(0.047, 0.073)
3 to 11 months	0.413	(0.225, 0.601)
1 year	−0.153	(−0.346, 0.040)
2 to 64 years	−2.251	(−2.426,−2.077)
65 years & over	−0.289	(−0.466,−0.113)
Indicator	0.019	(−0.097, 0.135)
Indicator:65 years & over	−0.267	(−0.421,−0.113)

The plot of the deviance residuals for this model is shown in Figure 5.5. This plot shows a fairly even distribution of points around the zero line. Thus, the assumption that the residuals have zero mean appears reasonable.

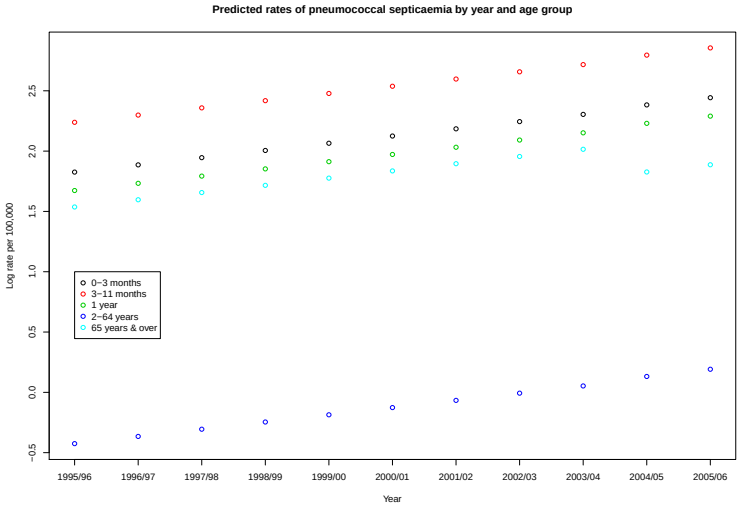

Figure 5.4: Plot of predicted log rates from the Negative Binomial regression model of pneumococcal septicaemia.

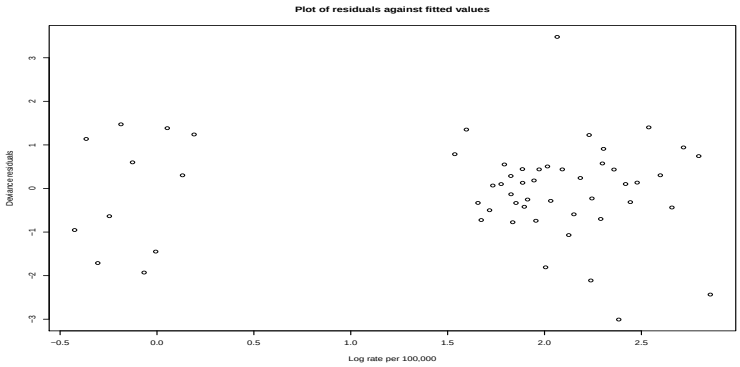

Figure 5.5: Plot of residuals against fitted values for Negative Binomial regression model of pneumococcal septicaemia.

5.4.2 Unspecified septicaemia

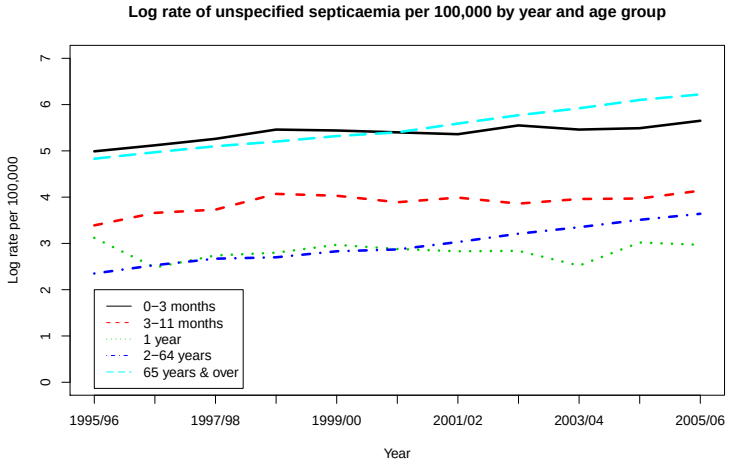

Figure 5.6: Log rate of episodes of unspecified septicaemia per 100,000 against year by age group with five categories.

The plot of the log rate of unspecified septicaemia per 100,000 against year by age group (Figure 5.6) shows the 0 to 3 months age group to have the highest rate of cases until 2000/01 when the 65 years and over age group shows the highest rate and continues to have the highest rate for all other years. The log rate of cases in the 65 years and over age group shows an increase over time. The 2 to 64 years age group had a similar rate of cases to the 1 year age group until 2000/01. However, the rate in the 2 to 64 years group showed an increasing trend over the period of study whilst the 1 year age group does not appear to have a distinct trend. Both of these groups, as well as the 3 to 11 months age group showed lower rate of cases than either the 0 to 3 months age group or the 65 years and over group. It appears that hospital episodes of unspecified septicaemia are most common amongst the elderly and the very young and that the rates for the elderly and for those aged 2 to 64 years of age have shown an increase over time.

The table containing summary statistics for the rate of unspecified septicaemia for each of the age groups (Table 5.6) shows that the 65 years and over age group and the 0 to 3 months age group have much higher rates, at over 200 cases per

Table 5.6: Overall rate of hospital cases of unspecified septicaemia per 100,000 for each age group.

Age group	**Rate**	**95% C.I.**
0-3 months	220.34	(213.31, 227.36)
3-11 months	49.26	(47.34, 51.18)
1 year	17.33	(16.34, 18.31)
2 years	8.71	(8.01, 9.41)
3 years	6.58	(5.98, 7.19)
4-9 years	3.14	(2.98, 3.31)
10-17 years	3.64	(3.49, 3.80)
18-39 years	11.75	(11.59, 11.91)
40-64 years	42.23	(41.93, 42.53)
65 years & over	270.00	(268.94, 271.07)

100,000 in the population, than the other age groups.

The analysis of the unspecified septicaemia was carried out in a similar manner to that of the pneumococcal septicaemia. A Poisson regression model involving the same variables as those considered in the pneumococcal septicaemia model was fitted for the number of episodes of unspecified septicaemia. However, this model was over-dispersed, with a residual deviance of more than twenty times the residual degrees of freedom. Therefore, a Negative Binomial regression model was fitted with the same variables. The parameter estimates for the full model and 95% confidence intervals are shown in Table 5.7.

Table 5.7 shows that there is no evidence of a change in slope in 2003 for the 65 years and over age group. Backward stepwise selection procedures were used to obtain the best fitting model to describe the trend in unspecified septicaemia. Table 5.8 shows the results for the best fitting model.

In the final model for unspecified septicaemia the interaction between year and the 65 years and over age group is significant with a positive coefficient. This indicates that the general trend for the 65 years and over age group was different to that for the other age groups. Therefore, a sharper increase is observed in cases of unspecified septicaemia in this age group than in all others. The coefficients

Table 5.7: Parameter estimates and confidence intervals for the full Negative Binomial regression model of the number of cases of unspecified septicaemia.

Parameter	**Estimate**	95% **C.I.**
Intercept	−5.974	(−6.099,−5.849)
Year	0.055	(0.035, 0.075)
3 to 11 months	−1.496	(−1.631,−1.361)
1 year	−2.528	(−2.669,−2.387)
2 to 64 years	−2.393	(−2.522,−2.264)
65 years & over	0.337	(0.117, 0.557)
Year:65 years & over	0.078	(0.035, 0.121)
Indicator	−0.019	(−0.380, 0.342)
Year:indicator	0.057	(−0.164, 0.278)
Indicator:65 years & over	0.113	(−0.655, 0.881)
Year:indicator:65 years & over	−0.062	(−0.532, 0.408)

of the dummy variables for the 3 to 11 months, 1 year and 2 to 64 years age group are all negative. Thus, the 0 to 3 months age group had higher rates of unspecified septicaemia than these groups. The highest rate was observed for the 65 years and over group.

Table 5.8: Parameter estimates and confidence intervals for the best fitting Negative Binomial regression model of the number of cases of unspecified septicaemia.

Parameter	**Estimate**	95% **C.I.**
Intercept	−5.945	(−6.049,−5.841)
Year	0.061	(0.045, 0.077)
3 to 11 months	−1.496	(−1.631,−1.361)
1 year	−2.527	(−2.668,−2.386)
2 to 64 years	−2.392	(−2.521,−2.263)
65 years & over	0.345	(0.184, 0.506)
Year:65 years & over	0.079	(0.048, 0.110)

In this model, θ was estimated to be 44.0, 95% C.I. (24.9, 63.1). The residuals plot for this model is shown in Figure 5.7. This plot shows that for the best fitting Negative Binomial regression model the assumption of zero mean is reasonable as most of the deviance residuals lie between -2 and 2 on the plot of residuals against fitted values. Figure 5.8 shows the predicted values of unspecified septicaemia for

each year and age group.

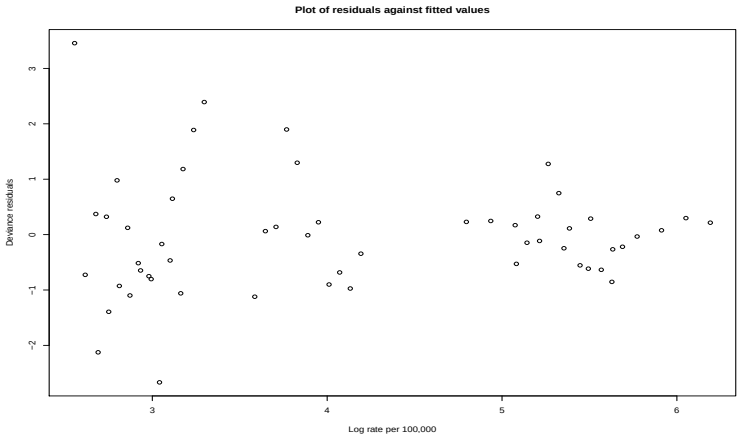

Figure 5.7: Residuals plot for the Negative Binomial regression model of unspecified septicaemia.

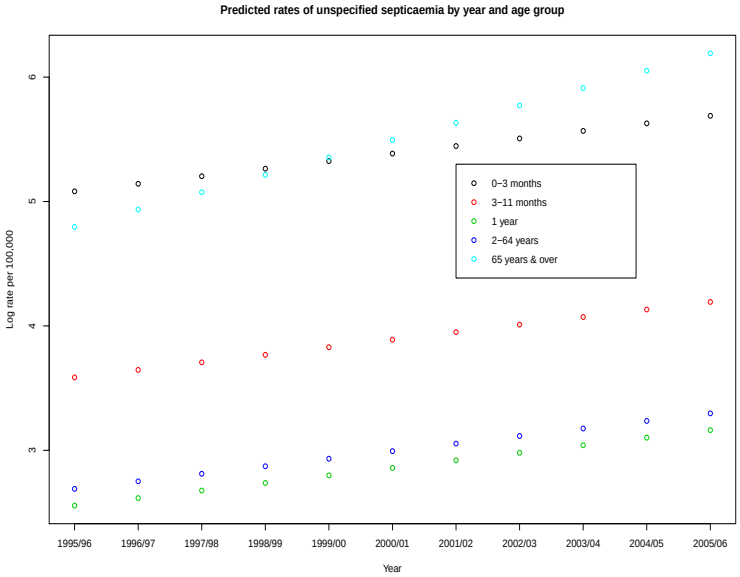

Figure 5.8: Plot of predicted log rates from the Negative Binomial regression model of unspecified septicaemia.

5.4.3 Pneumococcal meningitis

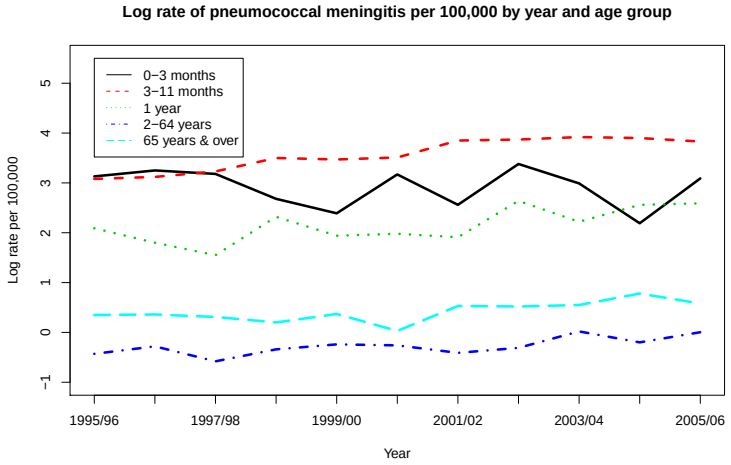

Figure 5.9: Log rate of episodes of pneumococcal meningitis per 100,000 against year by age group with 5 categories.

Figure 5.9 shows the 2 to 64 years age group to have the lowest rate of pneumococcal meningitis per 100,000 in the population of all age groups. In addition, the 65 years and over group has a low rate. Therefore, it appears that pneumococcal meningitis is most common amongst the very young, particularly those aged 3 to 11 months old. The general trend for most age groups appears to be an increasing trend, albeit only a slight increase. However, no clear trend is observed for the 0 to 3 months age group.

The table of summary statistics of the rate of pneumococcal meningitis for each group (Table 5.9) shows both the 0 to 3 months and the 3 to 11 months age groups to have a much higher overall rate than the other age groups.

As with the other models fitted for septicaemia, a Poisson regression model was fitted for pneumococcal meningitis in order to identify the best possible model to describe the change in number of hospital episodes with time. Once again, the Poisson regression model was over-dispersed with a residual deviance of 210.50 on 44 degrees of freedom and so a Negative Binomial regression model was fitted

Table 5.9: Overall rate of hospital cases of pneumococcal meningitis per 100,000 for each age group.

Age group	**Rate**	**95% C.I.**
0-3 months	19.51	(17.42, 21.60)
3-11 months	36.94	(35.28, 38.60)
1 year	8.99	(8.28, 9.69)
2 years	2.84	(2.44, 3.24)
3 years	1.75	(1.44, 2.06)
4-9 years	0.63	(0.56, 0.71)
10-17 years	0.35	(0.30, 0.39)
18-39 years	0.47	(0.43, 0.50)
40-64 years	1.14	(11.34, 11.44)
65 years & over	1.55	(1.47, 1.63)

to the data. The parameter estimates, standard errors and confidence intervals for the parameter estimates for the full model are shown in Table 5.10.

Table 5.10: Parameter estimates and confidence intervals for the full Negative Binomial regression model of the number of cases of pneumococcal meningitis.

Parameter	**Estimate**	95% **C.I.**
Intercept	−8.366	(−8.540,−8.192)
Year	0.054	(0.029, 0.079)
3 to 11 months	0.612	(0.434, 0.790)
1 year	−0.793	(−0.983,−0.603)
2 to 64 years	−3.243	(−3.417,−3.069)
65 years & over	−2.671	(−2.953,−2.389)
Year:65 years & over	−0.027	(−0.080, 0.026)
Indicator	−0.185	(−0.620, 0.250)
Year:indicator	0.087	(−0.178, 0.352)
Indicator:65 years & over	0.682	(−0.237, 1.601)
Year:indicator:65 years & over	−0.311	(−0.877, 0.255)

The best fitting model was identified using the backwards stepwise selection procedure as before. The only variables found to be significant in this model are year and age group, both with p-values of <0.001 based on F-tests with one and four degrees of freedom respectively. The results for this model are shown in Table

5.11. The estimate of θ for the model shown in Table 5.11 is 34.9, 95% C.I. (14.1, 55.7).

Table 5.11: Parameter estimates and confidence intervals for the best fitting Negative Binomial regression model of the number of cases of pneumococcal meningitis.

Parameter	**Estimate**	95% **C.I.**
Intercept	−8.393	(−8.546,−8.240)
Year	0.048	(0.030, 0.066)
3 to 11 months	0.612	(0.439, 0.794)
1 year	−0.796	(−0.990,−0.602)
2 to 64 years	−3.245	(−3.425,−3.065)
65 years & over	−2.552	(−2.736,−2.368)

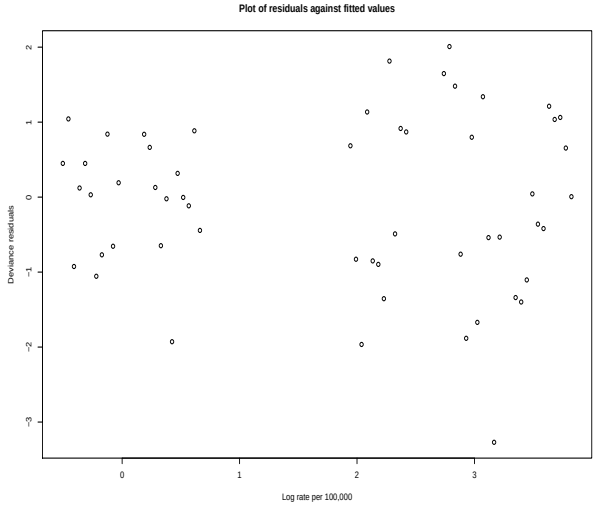

Figure 5.10: Plot of residuals against fitted values for the Negative Binomial regression model of pneumococcal meningitis.

The coefficient of year in the model of pneumococcal meningitis is positive. Thus, there is evidence that pneumococcal meningitis has been increasing with year. There is no evidence of a change in trend following the introduction of PPV-23. The 3 to 11 months age group has the highest rates of pneumococcal meningitis.

The 1 year, 2 to 64 years and the 65 years and over age group all have lower rates than the comparator, 0 to 3 months of age.

The residuals plot for the model described in Table 5.11 is shown in Figure 5.10. Figure 5.10 shows the points to be fairly evenly scattered about the zero line between -2 and 2. Therefore, the assumptions that the residuals have zero mean appears reasonable.

Figure 5.11 shows the predicted log rates of pneumococcal meningitis for each year and age group. The youngest age groups have the greatest predicted rates, with the 3 to 11 months group the highest.

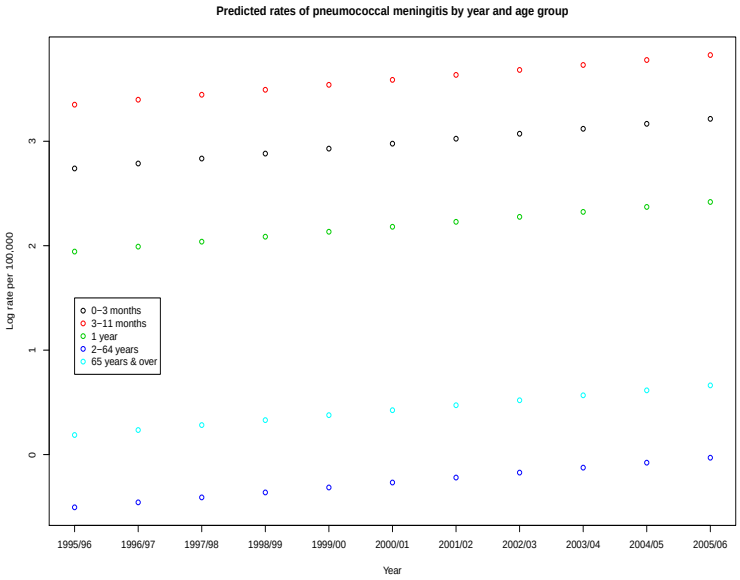

Figure 5.11: Plot of predicted log rates from the Negative Binomial regression model of pneumococcal meningitis.

5.4.4 Unspecified meningitis

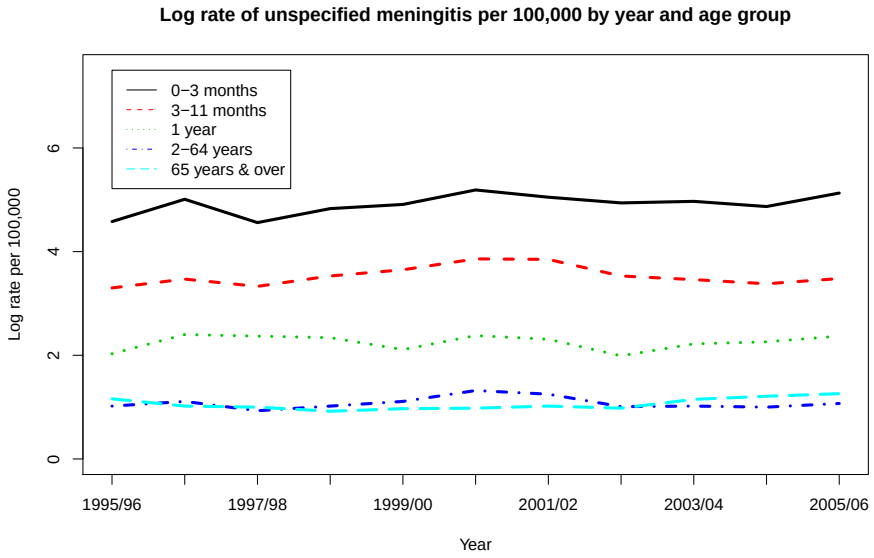

Figure 5.12: Log rate of episodes of unspecified meningitis per 100,000 against year by age group with five categories.

The trend plot of unspecified meningitis for the 5 age categories (Figure 5.12) shows the 2 to 64 years age group to have low log rates of cases, similar to those observed for the 65 years and over age group. Thus, it appears that unspecified meningitis is most common amongst the very young, in particular those under 3 months of age. The 0 to 3 months age group has higher rates per 100,000 in the population than all other age groups. However, the rate does not appear to show a great increase over time for this group. The next highest rate appears in the 3 to 11 months age group. This group shows a fairly even rate of cases across all years and does not appear to have an increasing or decreasing trend. All other age groups also appear to have fairly stable rates of cases across all years.

The table of summary statistics for the rate of unspecified meningitis (Table 5.12) shows that the 0 to 3 months age group did indeed have a much higher rate of cases than all other age groups, with a rate of more than 100 greater than the next highest average observed. The 40 to 64 years age group shows the lowest

Table 5.12: Overall rate of hospital cases of unspecified meningitis per 100,000 for each age group.

Age group	**Rate**	**95% C.I.**
0-3 months	137.97	(132.41, 143.53)
3-11 months	34.66	(33.05, 36.27)
1 year	9.60	(8.86, 10.33)
2 years	5.43	(4.88, 5.98)
3 years	5.08	(4.55, 5.61)
4-9 years	3.24	(3.07, 3.41)
10-17 years	3.53	(3.38, 3.68)
18-39 years	2.93	(2.85, 3.01)
40-64 years	2.55	(2.47, 2.62)
65 years & over	2.91	(2.80, 3.02)

average rate.

Following the same modelling procedure as described for pneumococcal septicaemia, a Poisson regression model for unspecified meningitis was found to be over-dispersed with a residual deviance of 302.34 on 44 degrees of freedom for the model. The results for the full Negative Binomial regression model are shown in Table 5.13.

Table 5.13: Parameter estimates and confidence intervals for the full Negative Binomial regression model of the number of cases of unspecified meningitis.

Parameter	**Estimate**	95% **C.I.**
Intercept	−6.493	(−6.597,−6.389)
Year	0.023	(0.005, 0.041)
3 to 11 months	−1.380	(−1.490,−1.270)
1 year	−2.664	(−2.789,−2.539)
2 to 64 years	−3.841	(−3.943,−3.739)
65 years & over	−4.001	(−4.185,−3.817)
Year:65 years & over	−0.024	(−0.061, 0.013)
Indicator	−0.317	(−0.621,−0.013)
Year:indicator	0.117	(−0.067, 0.301)
Indicator:65 years & over	0.465	(−0.178, 1.108)
Year:indicator:65 years & over	−0.068	(−0.460, 0.324)

The results for the best fitting model obtained using stepwise selection methods are shown in Table 5.14.

Table 5.14: Parameter estimates and confidence intervals for the best fitting Negative Binomial regression model of the number of cases of unspecified meningitis.

Parameter	**Estimate**	95% **C.I.**
Intercept	−6.547	(−6.637,−6.457)
Year	0.013	(0.001, 0.025)
3 to 11 months	−1.379	(−1.499,−1.259)
1 year	−2.665	(−2.798,−2.532)
2 to 64 years	−3.842	(−3.954,−3.730)
65 years & over	−3.861	(−3.977,−3.745)

The only variables required in the model for unspecified meningitis are year and age group. The coefficient of year is positive. Therefore, the hospital episodes of unspecified meningitis have been increasing with year in England and Wales. The coefficients for each of the dummy variables for age group are all negative. Thus, the 0 to 3 months age group has the highest predicted episodes of unspecified meningitis.

In this model, θ is 67.0. The 95% confidence interval for θ is (33.7, 100.3). The plot of the deviance residuals against fitted values is shown in Figure 5.13. The assumption that the residuals have zero mean appears reasonable from assessment of this plot.

The plot of the predicted log rate for each year and age group is shown in Figure 5.14. Only four lines are apparent from examination of Figure 5.14 as the predicted log rates of unspecified meningitis are practically equal for the 2 to 64 years age group and the 65 years and over group.

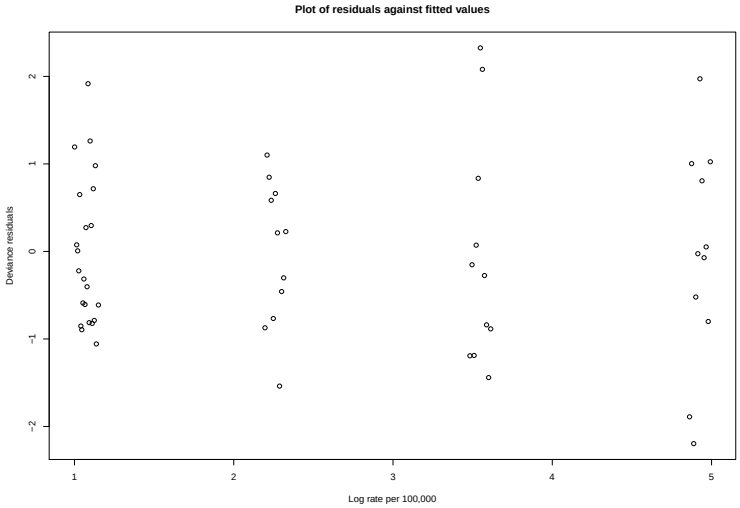

Figure 5.13: Plot of residuals against fitted values for the Negative Binomial regression model of unspecified meningitis.

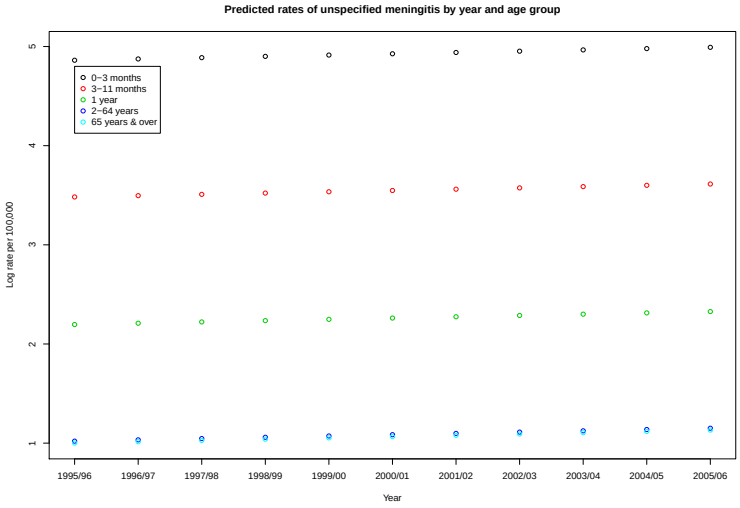

Figure 5.14: Plot of predicted log rates from the Negative Binomial regression model of unspecified meningitis.

5.4.5 Pneumococcal pneumonia

The trend plot of the rate of pneumococcal pneumonia for the five age categories (Figure 5.15) shows that the 2 to 64 years age group had the lowest rate of cases for every year from 1995/96 to 2005/06. The rate appears to remain relatively stable for this age group from 1995/96 to 2000/01, at which point an increase in the log rate of episodes is observed. A similar pattern is observed in the 65 years and over age group. The 0 to 3 months age group had the highest rate in 1997/98 but this was followed by a general decreasing trend in rate for this group. The 3 to 11 months age group also showed fairly high rates in comparison to all other age groups but, like the 0 to 3 months age group, showed a general decreasing trend.

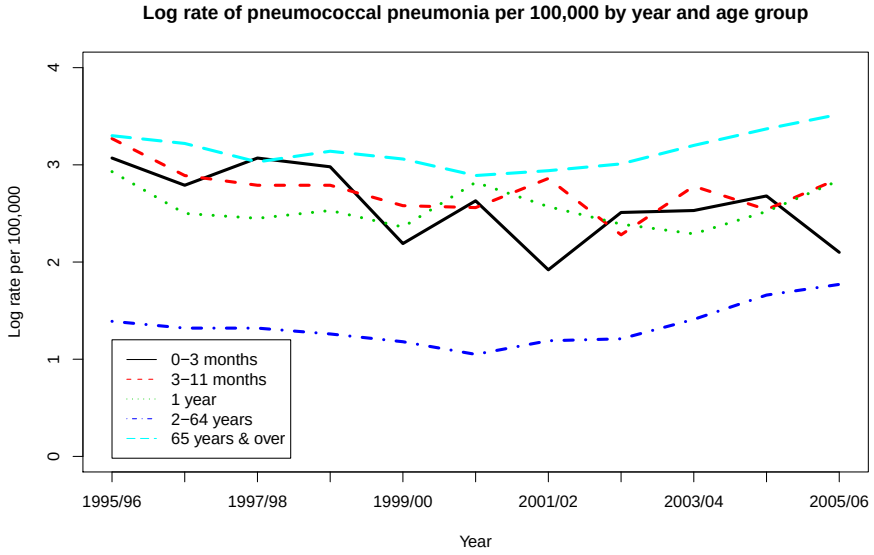

Figure 5.15: Log rate of episodes of pneumococcal pneumonia per 100,000 against year by age group with five categories.

Table 5.15 shows the highest overall rate of pneumococcal pneumonia of 23.82 per 100,000 occurred for the 65 years and over age group. The next highest overall of 16.12 per 100,000 was observed for the 3 to 11 months age group. The 0 to 3 months age group and the 1 year age group also showed high overall rates compared to other age groups of over 10 per 100,000.

Table 5.15: Overall rate of hospital cases of pneumococcal pneumonia per 100,000 for each age group.

Age group	**Rate**	**95% C.I.**
0-3 months	14.31	(12.52, 16.10)
3-11 months	16.12	(15.02, 17.22)
1 year	12.68	(11.84, 13.53)
2 years	6.51	(5.90, 7.11)
3 years	5.12	(4.59, 5.66)
4-9 years	2.20	(2.06, 2.34)
10-17 years	1.15	(1.06, 1.24)
18-39 years	3.00	(2.92, 3.08)
40-64 years	6.04	(5.92, 6.15)
65 years & over	23.82	(23.51, 24.14)

A Negative Binomial regression model was fitted to the pneumococcal pneumonia data as the Poisson regression model was deemed over-dispersed with a residual deviance of 494.12 on 44 degrees of freedom. The results from the full model are shown in Table 5.16.

Table 5.16: Parameter estimates and confidence intervals for the full Negative Binomial regression model of the number of cases of pneumococcal pneumonia.

Parameter	**Estimate**	95% **C.I.**
Intercept	−9.080	(−9.260,−8.900)
Year	−0.045	(−0.069,−0.021)
3 to 11 months	0.119	(−0.065, 0.303)
1 year	−0.124	(−0.308, 0.060)
2 to 64 years	−1.294	(−1.466,−1.122)
65 years & over	0.567	(0.320, 0.814)
Year:65 years & over	0.022	(−0.021, 0.065)
Indicator	0.166	(−0.232, 0.564)
Year:indicator	0.182	(−0.059, 0.423)
Indicator:65 years & over	0.062	(−0.689, 0.813)
Year:indicator:65 years & over	−0.015	(−0.474, 0.444)

Using stepwise selection procedures, the model shown in Table 5.17 is the best fit to the data. The interaction between year and the year indicator is significant,

Table 5.17: Parameter estimates and confidence intervals for the best fitting Negative Binomial regression model of the number of cases of pneumococcal pneumonia.

Parameter	**Estimate**	95% **C.I.**
Intercept	−9.062	(−9.235,−8.889)
Year	−0.039	(−0.059,−0.020)
3 to 11 months	0.119	(−0.067, 0.304)
1 year	−0.125	(−0.310, 0.060)
2 to 64 years	−1.300	(−1.472,−1.126)
65 years & over	0.508	(0.335, 0.681)
Indicator	0.184	(−0.156, 0.525)
Year:indicator	0.176	(−0.031, 0.384)

indicating that the mean slope from 2003 is different for all age groups. The coefficient of year is negative in this model and the coefficient for the interaction between the year and indicator variable is positive. Thus, the cases of pneumococcal pneumonia decrease between 1995 and 2003, after which they increase. The coefficients of the 3 to 11 months age group and the 65 years and over age groups are positive, indicating these two groups have higher rates than the 0 to 3 months age group. The oldest age group has the highest rates of pneumococcal pneumonia, the 2 to 64 years age group has the lowest.

θ in this model is estimated to be 50.1, 95% C.I. (22.9, 77.3). The residuals against fitted values plot for this model is shown in Figure 5.16. The assumption that the residuals have zero mean appears reasonable.

The plot of the predicted log rate for each year and age group is shown in Figure 5.17.

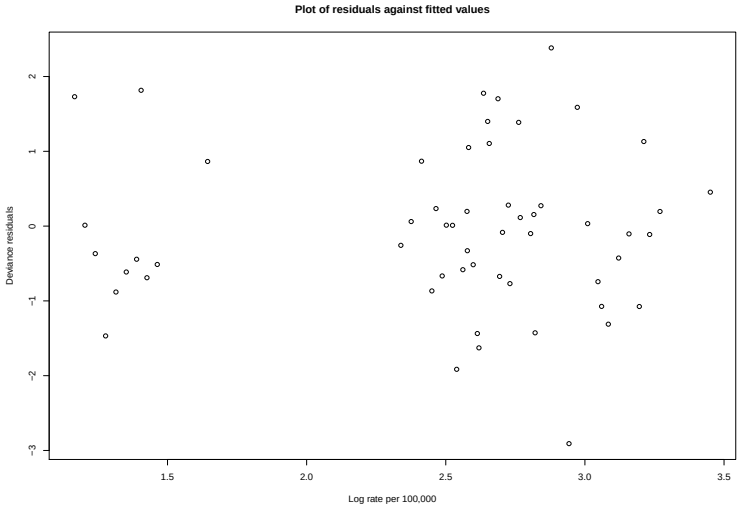

Figure 5.16: Plot of residuals against fitted values for the Negative Binomial regression model of pneumococcal pneumonia.

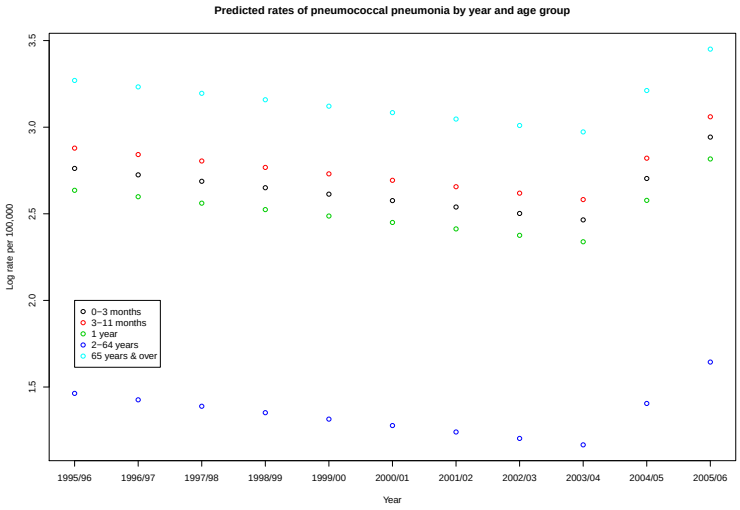

Figure 5.17: Plot of predicted log rates from the Negative Binomial regression model of pneumococcal pneumonia.

5.4.6 Unspecified pneumonia

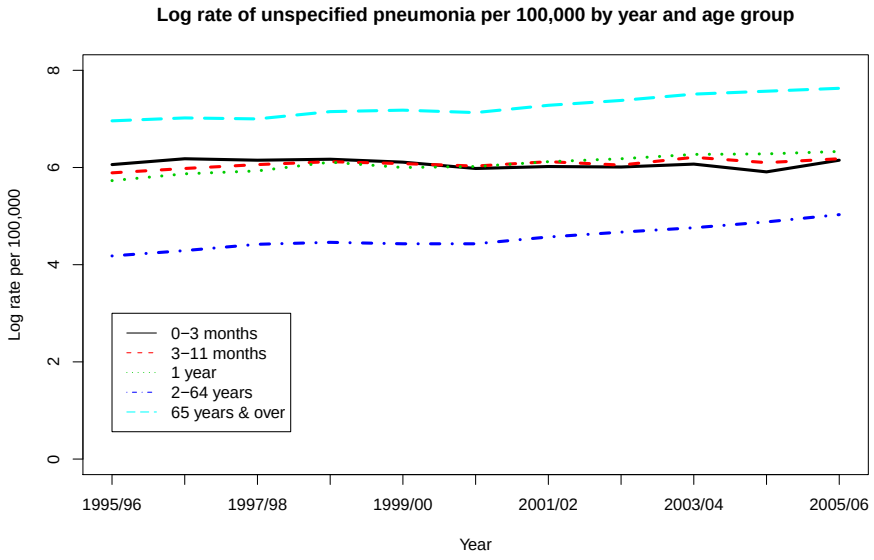

Figure 5.18: Log rate of episodes of unspecified pneumonia per 100,000 against year by age group with 5 categories.

The plot of the log rate of episodes of unspecified pneumonia for each age group (Figure 5.18) shows that the 65 years and over age group had the highest rate of episodes per 100,000. This age group shows a general increasing trend from 1995/96 to 2005/06. The 2 to 64 years age group also shows a slight increase in the log rates over time. No other age group shows a distinct trend from examination of this plot. The rate of unspecified pneumonia per 100,000 was lower for the 2 to 64 years age group than any other group.

Table 5.18 shows the summary statistics for the rate of hospital episodes of pneumococcal pneumonia for each age group. This table shows that the 65 years and over group had by far the highest rate per 100,000 than all other groups, at more than three times the next highest rate, which was observed for the 1 year age group. The lowest rate of hospital episodes of pneumococcal pneumonia was observed for the 10 to 17 years group, at only 32.22 per 100,000.

The Poisson regression model for the episodes of unspecified pneumonia was

Table 5.18: Overall rate of hospital cases of unspecified pneumonia per 100,000 for each age group.

Age group	**Rate**	**95% C.I.**
0-3 months	437.17	(427.29, 447.05)
3-11 months	435.77	(430.07, 441.46)
1 year	440.69	(435.74, 445.65)
2 years	277.52	(273.69, 281.44)
3 years	207.78	(204.40, 211.16)
4-9 years	81.62	(80.78, 82.47)
10-17 years	32.22	(31.76, 32.68)
18-39 years	57.20	(56.84, 57.55)
40-64 years	155.43	(154.58, 156.01)
65 years & over	1458.39	(1455.94,1460.84)

highly over-dispersed, with a residual deviance of 5,693 on 44 degrees of freedom. Therefore, a Negative Binomial regression model was fitted. The output from fitting this model is shown in Table 5.19.

Table 5.19: Parameter estimates and confidence intervals for the full Negative Binomial regression model of the number of cases of unspecified pneumonia.

Parameter	**Estimate**	95% **C.I.**
Intercept	−5.335	(−5.421,−5.249)
Year	0.033	(0.019, 0.047)
3 to 11 months	−0.010	(−0.102, 0.082)
1 year	−0.006	(−0.096, 0.084)
2 to 64 years	−1.520	(−1.610,−1.430)
65 years & over	1.259	(1.106, 1.412)
Year:65 years & over	0.031	(0.002, 0.060)
Indicator	−0.097	(−0.340, 0.146)
Year:indicator	0.089	(−0.060, 0.238)
Indicator:65 years & over	0.167	(−0.368, 0.702)
Year:indicator:65 years & over	−0.092	(−0.419, 0.235)

The model shown in Table 5.20 was obtained on performing stepwise model selection. The estimate of θ for this model is 87.7, 95% C.I. (53.2, 122.2).

Table 5.20: Parameter estimates and confidence intervals for the best fitting Negative Binomial regression model of the number of cases of unspecified pneumonia.

Parameter	**Estimate**	95% **C.I.**
Intercept	−5.317	(−5.391,−5.243)
Year	0.036	(0.026, 0.046)
3 to 11 months	−0.011	(−0.103, 0.081)
1 year	−0.007	(−0.099, 0.085)
2 to 64 years	−1.521	(−1.613,−1.429)
65 years & over	1.269	(1.155, 1.383)
Year:65 years & over	0.033	(0.011, 0.055)

The interaction between year and age group 65 years and over is significant. Thus, there is evidence that the 65 years and over age group has a different slope than the other age groups considered. The coefficient of year is positive. Therefore, the cases of unspecified pneumonia have been increasing with increasing year. The coefficient of the dummy variable for the 65 years and over age group is positive. This shows that this age group has higher rates of disease than the 0 to 3 months age group. In addition, this age group has a steeper slope relative to all other age groups. All other coefficients for the age groups are negative.

The plot of the deviance residuals against the fitted values for this model is shown in Figure 5.19 in which it can be observed that there are a couple of points which lie outwith the range -2 to 2. However, the assumption of zero mean appears reasonable.

The plot of the predicted log rate for each year and age group is shown in Figure 5.20. This shows no real difference in the predicted rates for the 0 to 3 months, 3 to 11 months and the 1 year age groups. The 65 years and over age group has the highest predicted rates and the 2 to 64 years age group has the lowest. The change in slope is minimal but can be observed in the plot.

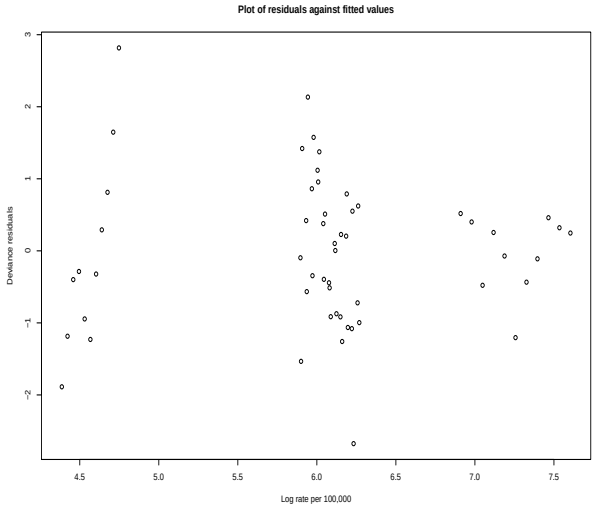

Figure 5.19: Plot of residuals against fitted values for the Negative Binomial regression model of unspecified pneumonia.

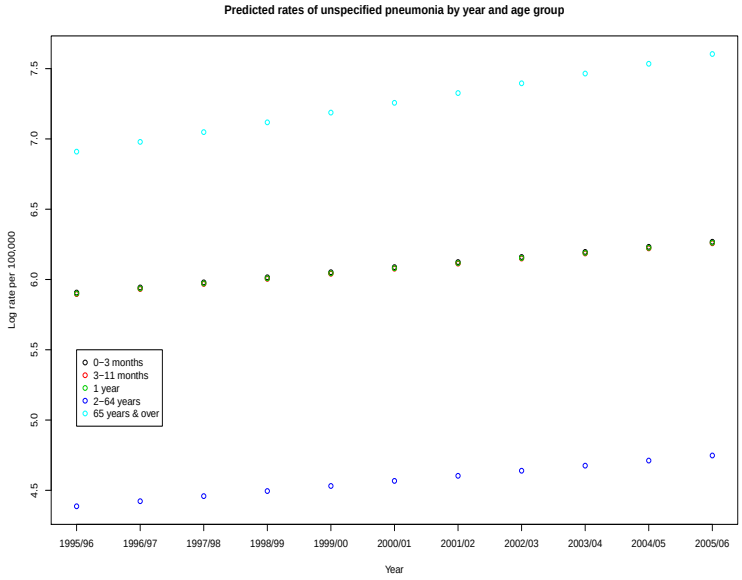

Figure 5.20: Plot of predicted log rates from the Negative Binomial regression model of unspecified pneumonia.

5.4.7 Comparator

To determine whether the trends observed in the HES of the pneumococcal and unspecified diseases described in this chapter are attributable to true changes in the number of cases of disease or whether they reflect changes in the reporting system of hospitals in England and Wales over time, a comparator group has to be considered. The HES comparator selected for this analysis is fractures as it can be assumed that the number of fractures observed annually should remain relatively stable within each age group. Data on the hospital episodes of both forearm fractures (ISD-10 code S52) and femur fractures (ISD-10 code S72) were downloaded from the HES website (Hospital Episode Statistics 2008). Unfortunately, the data involves only four age groups 0 to 14 years, 15 to 59 years, 60 to 74 years and 75 years and over and was only available for the years 2002/03 to 2005/06. For the analysis, the number of episodes was the number of finished consultant episodes as was the case with the data used in the other models in this chapter. Information was available on the gender of the patients. The appropriate estimates for the population sizes were obtained for these age groups (Office for National Statistics 2008).

Figures 5.21 and 5.22 show that for most age groups there are no clear increasing or decreasing trends for the log rate of forearm or femur fractures per 100,000. However, there appears to be an increasing trend in forearm fractures for the 75 years and over age group between 2002/03 and 2005/06.

Poisson regression models were fitted to the data and found to be over-dispersed. Thus, Negative Binomial regression models were fitted to the data for forearm and femur fractures separately and included the variables gender, age group, year and the interaction between year and age group. Using backward stepwise model selection based on AIC, as used in the other models in this chapter, the final model for forearm fractures involved only age group and gender; femur fractures only age group. Year was not significant. Thus, the null hypothesis that the number of forearm or femur fractures remains constant over time may not be rejected. This means that any trends observed in other hospital episodes during this period are likely to be attributable to changes in the number of cases of disease rather than differences in the reporting systems over time. However, a limitation of this

analysis is that data for the comparator groups were not available for the same duration and age groups that were considered in the analysis of pneumococcal and unspecified septicaemia, meningitis and pneumonia. Thus, it is difficult to determine whether or not the changes observed for these diseases are attributable to true increases or decreases in disease. In addition, as the analysis of these two comparator groups is limited to a shorter period than the analyses of the pneumococcal and unspecified diseases considered previously, the modelling has lower power to detect any trend effect. Thus, these results should be interpreted with caution. In order to have a suitable comparison of trends, an analysis of hospital episode statistics for femur and forearm fractures over the same period of time as the pneumococcal and unspecified diseases is required.

5.5 Conclusions

In this chapter, models were fitted to describe the trends in hospital episodes of pneumococcal and unspecified septicaemia, meningitis and pneumonia between 1995/96 and 2005/06. In each model a three-way interaction term was included to describe any change of slope following 2003/04 in the 65 years and over age group which may be attributable to the introduction of PPV-23. However, this variable was not found to be significant in any of the models fitted. Thus, no evidence was found of a change in slope for the 65 years and over age group.

All diseases considered in this chapter were observed to have an increasing trend between 1995/96 and 2005/06, other than pneumococcal pneumonia which showed a decreasing trend from 1995/96, followed by an increasing trend. The highest overall rates in disease were observed for the oldest and youngest age groups. The 65 years and over age group had the highest overall rates of unspecified septicaemia, pneumococcal pneumonia and unspecified pneumonia. The 3 to 11 months age group had the highest rate of pneumococcal septicaemia and pneumococcal meningitis whilst the 0 to 3 months age group had the highest rate of unspecified meningitis.

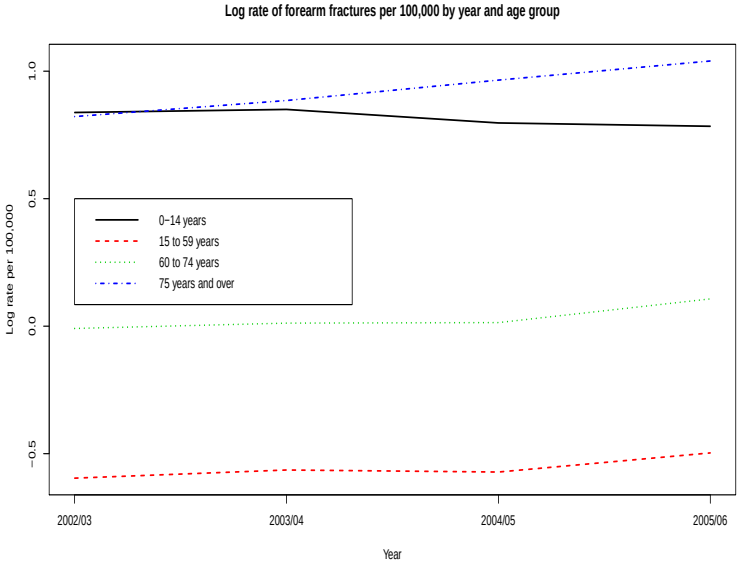

Figure 5.21: Log rate of episodes of forearm fractures per 100,000 against year by age group.

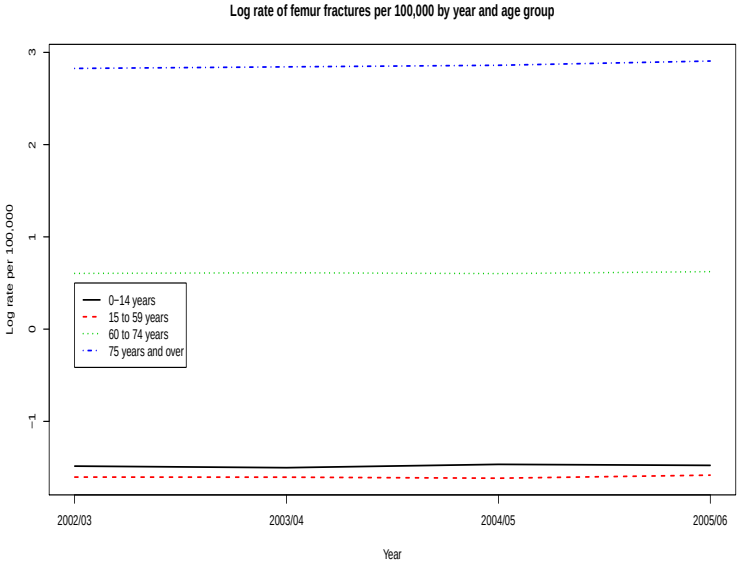

Figure 5.22: Log rate of episodes of femur fractures per 100,000 against year by age group.

The analysis carried out in this chapter adds to a previously published analysis of hospital episodes of pneumonia between 1997 and 2005, prior to the introduction of PCV-7 (Trotter et al. 2008). In this analysis, Trotter et al. consider only one episode of pneumonia per patient unlike the analysis of pneumonia carried out in this chapter in which the data are not representative of individual cases of disease but instead allow for the possibility of multiple hospital admissions. Trotter et al. observe a rise in the age-standardised incidence of pneumonia, particularly in the older age groups. The analysis carried out in this chapter for pneumococcal pneumonia contradicts these results as a decreasing trend was observed for all age groups between 1995 and 2003. The trend in the rates of disease is increasing from 2003 to 2005. However, the rate of unspecified pneumonia was observed to increase during the period 1995/96 to 2005/06. Thus, these results are in correspondence with the observations of Trotter et al. considering all cases of pneumonia together. The analysis carried out in this chapter separated the pneumococcal classifications of each disease from all others in order to determine any vaccine effect and also due to the fact that these models were created with the future impact of PCV-7 in mind.

Unlike the analysis of Trotter et al., the modelling undertaken in this chapter investigated whether or not PPV-23 had an impact on the cases of disease in the elderly. No significant evidence of a vaccine effect was found, although there was evidence of a deviation from the model predictions from 2003/04 for pneumococcal septicaemia. However, no change in slope was identified. The cases of pneumococcal septicaemia were lower than anticipated from 2003/04 for the 65 years and over age group. This could be attributable to the introduction of PPV-23 but the anticipated decreasing trend due to vaccine use was not observed.

Significant changes in slope from 2003/04 were identified for cases of pneumococcal pneumonia and unspecified pneumonia for all age groups. However, the change in slope for unspecified pneumonia was minimal. Considering pneumococcal pneumonia, an increase in the slope was observed from 2003/04, not a decrease due to any vaccine impact. As PPV-23 is only administered to those over the age of 65 years and does not provide any herd immunity as it does not prevent carriage, any effect observed in disease in the unvaccinated groups from

2003/04 cannot be attributable to vaccine use. In the analysis carried out in this chapter there was limited power to detect any changes in disease trends in the vaccine targeted age group due to the fact that only two years of data was available following the introduction of the vaccine.

It is unsurprising that no clear vaccine effect was found as the introduction of PPV-23 was staggered in England and Wales, with only those aged over 85 years of age receiving the vaccine in 2003/04 and those aged over 75 years in 2004/05. PPV-23 was not introduced for routine use in those aged 65 years and over until 2005/06. Thus, to identify whether or not PPV-23 has an impact in preventing pneumococcal disease more data are required for the years following 2005/06.

There was evidence of an interaction between year and age group for the trend in cases of unspecified septicaemia. The 65 years and over age group has a steeper slope than the other age groups, with lower predicted cases of disease than the 0 to 3 months age group in the early years of the analysis but the highest rates from 2000/01.

A limitation of the analysis carried out in this chapter is that data for the comparator group, fractures, were not available for the same duration and age groups that were considered in the analysis of pneumococcal and unspecified septicaemia, meningitis and pneumonia. Thus, it is difficult to determine whether or not the changes observed for these diseases are attributable to true increases or decreases in disease.

In future work, data for the same age groups and period for the comparator group should be obtained in order to determine whether or not there is evidence of a trend for this group that could affect the interpretation of the results obtained for the pneumococcal and unspecified diseases. In addition, it is intended that the models created in this chapter be used to predict the cases of pneumococcal and unspecified septicaemia, meningitis and pneumonia for the years following the introduction of PCV-7. These predictions assuming no PCV-7 effect will be compared to the observed hospital episode rates for these years to identify the effect of the conjugate vaccine. The estimates for the vaccine effect can then be used to update the cost-effectiveness model created by Wyeth pharmaceuticals

to assess PCV-7.

The next two chapters continue the investigation of pneumococcal disease in the UK by examining both the trends in serogroups, serotypes and MLSTs found in IPD in Scotland and the associations between these pneumococcal types and mortality.

Chapter 6

Analysing trends in serogroups/serotypes and MLSTs in Scottish IPD cases

6.1 Introduction

In addition to the assessment of factors influencing the long term efficacy and cost-effectiveness of PCV-7, one of the central themes in the thesis is exploring the importance of MLSTs in the ability of pneumococci to cause disease and their potential contribution to the loss of long term efficacy of PCV-7. The relationship between VT and NVT serotypes and MLSTs in carriage was explored previously in Chapters 2 and 3 through the use of mathematical modelling.

In this section of the thesis, the focus is on circulating pneumococcal serogroups, serotypes and MLSTs involved in IPD in Scotland. In addition, the relationship between serotypes and MLSTs involved in IPD in Scotland is explored. In this chapter and the subsequent chapter, through the use of existing statistical techniques, changes in serogroup, serotype and MLST distribution in IPD over recent years and the risks of fatality from IPD of different strains are analysed.

As mentioned in the introductory chapter of the thesis, the development of vaccines for the prevention of pneumococcal disease and infection is based upon the

serotypes responsible for the greatest burden of disease. Thus, it is important to continually assess the changing distribution of serotypes involved in IPD. It is necessary to additionally assess the MLSTs involved in IPD due to the potential for capsular switch to occur following the use of PCV-7.

This chapter begins with a discussion of serotype trends observed in IPD worldwide, both prior to and following the introduction of PCV-7. Following this, a trend analysis of the serogroups, serotype and MLSTs involved in IPD in Scotland prior to the introduction of PCV-7 is carried out.

6.2 Background

As discussed in the introductory chapter, pneumococcal vaccine formulations are based upon the serotype of the bacterium as this is the known virulence factor and currently 91 different pneumococcal serotypes within 46 serogroups have been identified (Henrichsen 1995; Park et al. 2007). A pneumococcal isolate may also be identified by its MLST. The MLST is defined according to 7 house-keeping genes identified within the genetic material of the pneumococcus (Enright and Spratt 1998). Hundreds of different MLSTs have been identified and some MLSTs have displayed an association with more than one serotype (Clarke et al. 2004).

Great interest lies in assessing the involvement of NVT serotypes in IPD cases and the burden of disease attributable to these serotypes. If NVTs are observed to increase in prevalence in IPD then vaccine formulations could require revisions to include NVTs responsible for increasing proportions of disease and infection.

6.2.1 Serogroup/serotype 1 IPD

Of particular interest for this study is serotype 1 as in previous published studies in Scotland, and throughout Europe, changes in the number of cases of IPD attributable to serotype 1, a serotype which is not included in PCV-7, have been observed.

In a Swedish study of serotypes involved in IPD between 1987 and 1997, a large

increase in the rate of pneumococcal bacteraemia was reported over the years, with the increase being related to a rise in the number of cases of both serotype 1 and serotype 14 bacteraemia, with a 10-fold increase in the number of cases of serotype 1 bacteraemia observed between 1992 and 1997 (Normark et al. 2001). In Scotland, the observed rate of serotype 1 IPD increased from 0.67 cases per 100,000 population in 2000 to 1.25 cases per 100,000 population in 2004 (Kirkham et al. 2006). In addition, in a study of cases of bacteraemia in the UK and Ireland, a significant change in serotype distribution was observed with increasing year of study, with serotype 1 increasing in prevalence from 4% of IPD in 2001 to 15.6% in 2006 (Farrell et al. 2008).

In contrast, a Danish study of IPD between 1996 and 1999, in which serotype 1 was identified as the most common disease-causing serotype overall, showed a reduction in cases attributable to serotype 1 from 26.5% of all invasive disease in 1996 to only 5.4% in 1999 (Konradsen and Kaltoft 2002). However, Konradsen and Kaltoff discuss a decrease in the overall incidence of IPD in Denmark in the years 1996 to 1999 which coincides with the observed decrease in serotype 1 IPD. A further contradictory study involves data on IPD from Oxfordshire, UK in which a reduction in disease attributable to serotype 1 was observed between 1996 and 2005 (Foster et al. 2008).

Considering the association between serotype 1 and IPD outwith Europe, a study of trends in pneumococcal serogroups (serotype specific information was not available for this analysis) involved in IPD between 1929 and 1998 in the USA was carried out (Feikin and Klugman 2002). This analysis showed a decrease in serogroup/serotype 1 IPD over this period. The authors state that the reduction of this serogroup and others such as 3 and 5 is likely to be attributable to the introduction of antibiotic treatment between the 1930s and 1940s. Fiekin and Klugman report an increase in the PCV-7 serogroups between 1929 and 1998.

A study of IPD in New Zealand reported serotype 1 as decreasing in IPD cases between 1998 and 2005, prior to the routine use of PCV-7 (Heffernan et al. 2008). Heffernan et al. also report reductions in serotype 9V, 7F and 12F disease and increases in 14, 6B and 4 disease.

6.2.2 Age associations with serogroup/serotype 1

Differing serotype prevalence amongst cases of IPD in different age groups has been documented (Kaplan et al. 2002; Konradsen and Kaltoft 2002; Kyaw et al. 2000). Focussing on serotype 1 once again, Kaplan et al. (2002) report increased proportions of serotype 1 IPD with increasing age in their study of children who attended 8 different hospitals in the USA. 16.2% of IPD isolates were identified as serotype 1 amongst all cases in children over the age of 10 years and only 2% in children under 2 years of age. Kaplan et al. also report serogroups/serotypes 3 and 23 had increased proportions in IPD for older children, whilst serotype 14 was involved in less IPD amongst older children.

In the Konradsen et al. (2002) study, although serotype 1 was reported to be the most common cause of IPD between 1995 and 1999 overall, it was not the most common cause of IPD in those aged under 2 years, where serotype 14 accounts for most disease. Serotype 1 is only the eleventh most common serotype in this age group but appears as the most common for the 2 to 59 years age group and the over 65 years age group.

Similarly, in a study of Scottish invasive isolates collected during the period 1988 to 1999, serotype 1 was the most common disease-causing serotype for the age group 5 to 64 years but not for those under 5 years old, where serotype 14 was the most prevalent. Serotype 14 was also reported as the most prevalent in disease of those aged 65 years and over (Kyaw et al. 2000).

6.2.3 MLST 306 and serotype 1

MLST 306 has been documented to be associated with serogroup/serotype 1 IPD (Jefferies et al. 2004; Serrano et al. 2005; Munõz-Almagro et al. 2008), with some countries documenting an increase in MLST 306 disease corresponding to an increase in serotype 1 disease (Normark et al. 2001).

In addition to an observed increase in rate of serogroup/serotype 1 IPD in Scotland, Kirkham et al. (2006) report an increase in MLST 306 serogroup/serotype

1 IPD from 0.04 cases per 100,000 in the population in 2001 to 0.81 cases per 100,000 in the population in 2004. However, they document a fairly constant rate in MLST 227 serogroup/serotype 1 IPD over the same period, the other MLST commonly found associated with serogroup/serotype 1. Thus, Kirkham et al. conclude that the increase in serogroup/serotype 1 IPD in Scotland is attributable to the increase in MLST 306 serogroup/serotype 1 IPD. However, their study focusses on the genetic composition of serotype 1 isolates in IPD and the rates of serotype 1 IPD are reported but no statistical analysis appears to have been carried out. Thus, the analysis carried out in this chapter should add to the work carried out by Kirkham et al. as statistical techniques are adopted to determine whether or not there is evidence of a changing trend in serogroup/serotype 1 disease or MLST 306 disease or both.

6.2.4 Effect of PCV-7

Many of the existing studies of serotypes involved in IPD concern the comparison of a pre-conjugate vaccination period to a post vaccination period to examine any changes in serotype distribution likely to be related to the use of PCV-7. In the USA, great reductions in IPD were documented which were not limited to the vaccine targeted age group (Whitney et al. 2003). However, recent studies have shown increases in IPD due to the NVT 19A in the USA following PCV-7 use (Pelton et al. 2007, Albrich et al. 2007, Beall et al. 2006, Huang et al. 2005b). Increases in this serotype have been observed following PCV-7 use outwith the USA (Aguiar et al. 2008; Munõz-Almagro et al. 2008). Aguiar et al. (2008) also report increases in NVT 7F disease in adults in Portugal but state that decreases in the VT serotypes 4, 6B, 14 and 23F were observed in IPD; Munõz-Almagro et al. document increases in the NVT serotypes 1, 5 and 6A disease in addition to 19A. Although it is speculated that the increase in 19A disease is attributable to vaccine use, in Belgium 19A was documented to increase in IPD but the authors state that this cannot be attributable to any vaccine effect as it was prior to routine PCV-7 implementation (Amrine-Madsen et al. 2008).

One of the primary concerns regarding the increase in serotype 19A IPD in particular is that the emergence of this serotype in the post-vaccine era is likely to

be attributable to a capsular switch event with serotype 4 (Brueggemann et al. 2007). Thus, it is becoming evermore important to assess the sequence types involved in IPD as well as the serotypes.

A study comparing circulating pneumococcal strains in Barcelona pre- and post-PCV-7 introduction reported increases in rates of IPD due to the NVT serotypes 1, 5, 7F, 12F, 19A, 22F and 24 (Ardanuy et al. 2009). Ardanuy et al. consider the MLSTs in the analysis of IPD in Barcelona and report the most common to be MLST 156, MLST 260, MLST 306, MLST 191, MLST 289, MLST 180 and MLST 81. It is reported that significant increases are observed in MLST 306 (associated with serotype 1), MLST 191 (serotype 7F), MLST 989 (serotype 12F) and MLST 433 (serotype 22F and 19A) from the pre-PCV-7 period (years 1997 to 2001) to the late-PCV-7 period (years 2005 to 2007).

6.2.5 Serotype and MLST associations

Associations between serotypes and MLSTs involved in IPD have been explored in the literature. In a study in the USA of cases of IPD in 1999, 2001 and 2002, years prior to and following the introduction of PCV-7, the serotype and MLST associations are described (Beall et al. 2006). Beall et al. state that it is likely that MLSTs which are associated with more than one serotype occur due to capsular switch events and that the data presented in their study indicate that such an event is rare since only 11 MLSTs amongst 177 identified from 2,100 collected isolates were associated with more than one serotype. Only 1 of these MLSTs, 199, is associated with both VT and NVT serotypes. However, a Scottish study of IPD prior to the introduction of PCV-7 showed MLSTs associated with more than 1 serotype to occur more frequently (Clarke et al. 2004). This study considered 368 IPD isolates collected in the first 6 months of 2003. Among 97 different MLSTs identified, 14 MLSTs were found associated with more than 1 serotype. Of these 14 MLSTs, 12 were associated with both VT and NVT serotypes.

6.3 Methods

In the analysis carried out in this chapter, the data were obtained from the Scottish Invasive Pneumococcal Disease Enhanced Surveillance (SPIDER) on all cases of IPD in Scotland between 1999 and 2006. Serogroup information for the cases of IPD was available for all years. Serotype and MLST information was available from 2003.

The SPIDER records contain information about the year and week the case of IPD was recorded. The years were grouped from week 40 of one year to week 39 of the next year to ensure winter seasons were grouped together as most cases of IPD occur during winter. The information used in the analysis was all cases of IPD in SPIDER which were identified from blood or cerebrospinal fluid (CSF) samples. The analysis was carried out on this data from 1999/00 to 2005/06. Within the SPIDER records information was also available on the age of the patient. These ages were grouped into six categories: 0 to 4 years, 5 to 34 years, 35 to 49 years, 50 to 64 years, 65 to 74 years and 75 years and over, as these age groups have been adopted in literature elsewhere looking at IPD in Scotland (Mooney et al. 2008). However, the Mooney et al. analysis combined the age groups 65 to 74 years and 75 years and over whereas this analysis makes a distinction between these age groups. This is due to the fact that the elderly have one of the highest burdens of IPD, particularly amongst the most elderly, and so these groups are separated as there may be differences in the serogroups found in disease.

Logistic regression models were used to investigate whether or not there is evidence of a trend for each of the serogroups, serotypes and MLSTs found in IPD in Scotland. The model can be described by the following equation:

$$\log\left(\theta/(1-\theta)\right) = \alpha + \beta \times x. \tag{6.1}$$

For the serogroup trend analysis, in the logistic regression model (6.1) θ is the proportion of serogroup i IPD in Scotland and x is the value of the continuous explanatory variable year from 1999/00 to 2005/06. This model was fitted for each of the serogroups found in IPD. Similar models were fitted for serotypes and MLSTs. However, in these models, data were only available for the years

2003/04 to 2005/06. To identify significant trends, either increasing or decreasing, for serogroups/types and MLSTs, the p-value for the variable year in the logistic regression model was examined.

Based on the observed increasing rates of serotype 1 IPD in Scotland between 2000 and 2003 (Kirkham et al. 2006), the primary hypothesis of this analysis of the Scottish IPD data was to determine whether or not there is significant evidence of an increasing trend in serogroup/serotype 1 IPD in the years prior to the introduction of PCV-7. The secondary hypothesis was to determine whether or not there is evidence of an increasing or decreasing trend for any other serogroup/serotype found in IPD in Scotland. As MLST 306 is a potentially influential factor for any increase in serotype 1 IPD observed, the primary hypothesis in the MLST analysis is that MLST 306 has an increasing trend.

As hypotheses are specified regarding serogroup/serotype 1 and MLST 306 trends, the p-values for year in each of these logistic regression models were compared to a significance level of 0.05. All other models were adjusted for multiple testing using the Bonferroni correction factor described below:

$$\alpha/n. \tag{6.2}$$

In the Bonferroni correction, α was fixed at a level of 0.05 and n is the total number of serogroups observed in at least 1% of all IPD from 1999/00 to 2005/06. Similarly, for serotypes and MLSTs, n is the total number observed in at least 1% of IPD between 2003/04 to 2005/06.

The analysis in this chapter was carried out using R version 2.8.0.

6.4 Results

6.4.1 Cases of IPD in Scotland

On average, approximately 651 cases of IPD per year were reported in Scotland between 1999/00 to 2005/06. Table 6.1 shows the results for each year.
Table 6.1 shows that the minimum number of cases of IPD occurred in 1999/00

Table 6.1: Number of cases of IPD reported in Scotland each year.

Year	99/00	00/01	01/02	02/03	03/04	04/05	05/06
Number	538	594	566	743	672	710	737

and the maximum in 2002/03, with a difference of 205 cases.

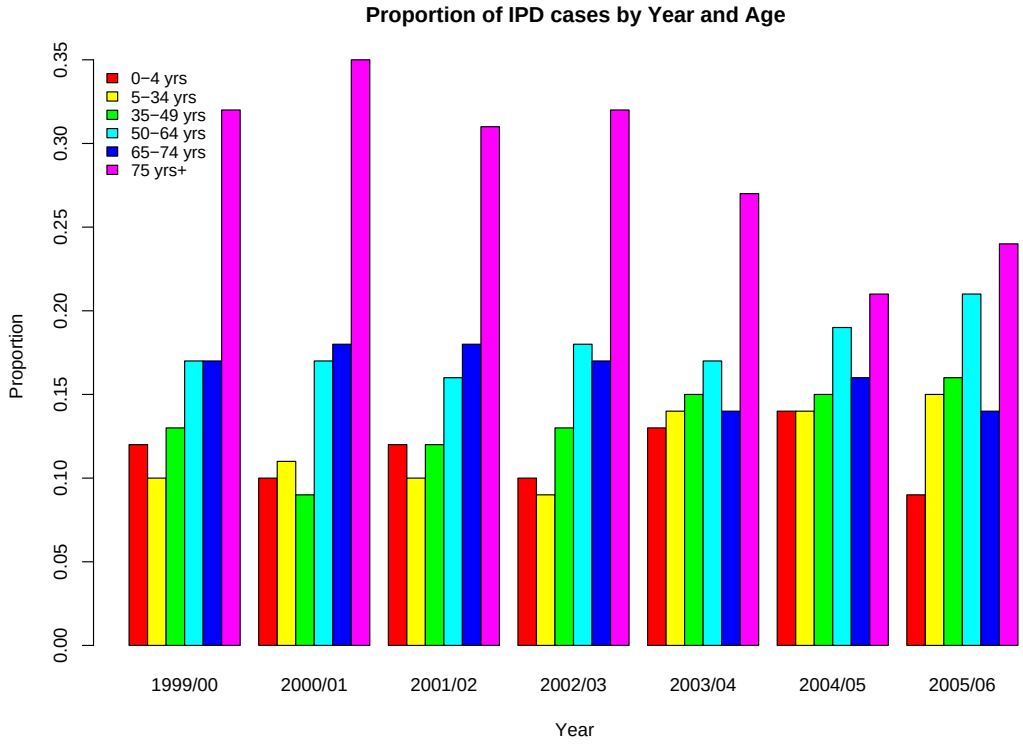

Figure 6.1: Proportion of IPD cases for each age group from 1999/00 to 2005/06.

From examination of Figure 6.1, it can be observed that IPD is most common among the elderly; in particular, those aged 75 years and over, with approximately 28% of all observed IPD occurring in this age group. The proportion of IPD in the 65 to 74 years age group and the 75 years and over age group shows a slight

decline after 2002/03. This is likely to be attributable to the introduction of PPV-23 in 2003 to prevent IPD in those aged 65 years and over. The introduction of the vaccine is not taken into account in the trend analysis as only serogroup information is available from 1999/00 and the vaccine is serotype specific. As serotype information is available from 2003/04, a trend analysis was carried out for the older age groups to identify which, if any, of the PPV-23 VT serotypes had a decreasing trend following the introduction of the vaccine.

6.4.2 Serogroup, serotype and MLST distribution in IPD

In Scotland, from 1999/00 to 2005/06, 35 different serogroups were identified in cases of IPD. Figure 6.2 shows the serogroups responsible for more than 1% of all cases of IPD during the period of study.

It can clearly be observed from examination of Figure 6.2 that serogroup 14 is the most common serogroup found in IPD in Scotland, accounting for more than 17% of all cases. Serogroups 9 and 1 are the next two most common serogroups, accounting for 9.28% and 8.44% of IPD respectively. Serogroups 14 and 9 appear to have been commonly associated with IPD in Scotland for many years as a study of Scottish IPD between 1988 and 1999 showed serogroups 14 and 9 to be the two most prevalent of all serogroups in IPD during that period (Kyaw et al. 2000). However, serogroup 1 was only the sixth most common serogroup involved in IPD between 1988 and 1999.

A total of 46 different serotypes were identified in IPD in Scotland between 2003/04 and 2005/06. Figure 6.3 shows the serotypes which each were responsible for more than 1% of all cases of IPD between 2003/04 and 2005/06, with vaccine type serotypes highlighted.

Figure 6.3 shows that, of the 20 serotypes that each account for more than 1% of IPD in Scotland, 17 are found in at least one of the two pneumococcal vaccines, PCV-7 and PPV-23. The other 3 serotypes, 19A, 6A and 9N, have been identified as PCV-7 related serotypes (Whitney et al. 2003). Thus, PCV-7 may prove effective in preventing disease from these 3 serotypes.

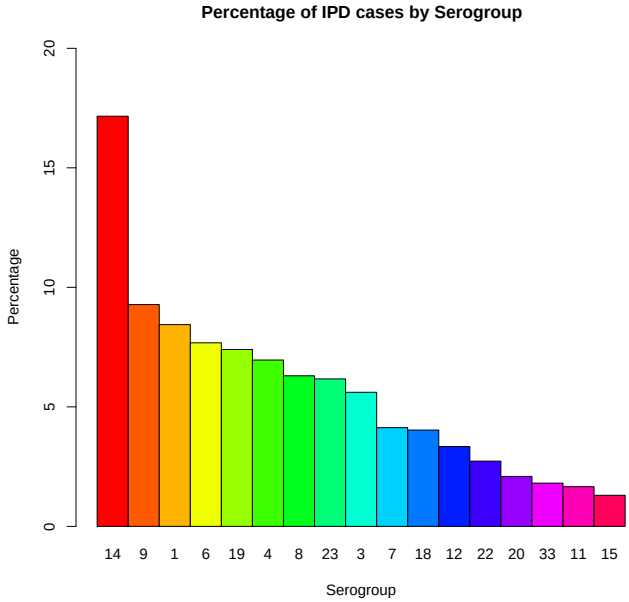

Figure 6.2: Percentage of IPD cases attributable to each serogroup from 1999/00 to 2005/06 (for serogroups that account for more than 1% of IPD).

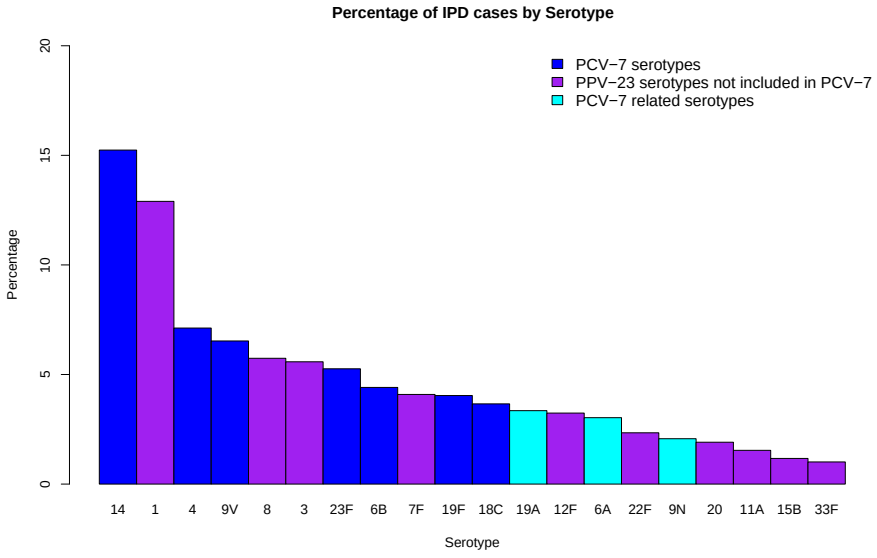

Figure 6.3: Percentage of IPD cases attributable to each serotype from 2003/04 to 2005/06 (for serotypes that account for more than 1% of IPD).

Together, the 10 most common disease-causing serotypes in Scotland account for approximately 71% of all IPD between 2003/04 and 2005/06. The serotypes found in PCV-7 (4, 6B, 9V, 14, 18C, 19F, 23F) account for approximately 47% of all those IPD cases serotyped during this period and the PCV-7 related serotypes (6A, 9A, 9L, 9N, 18A, 18B, 18F, 19A, 19B, 19C, 23A, 23B) together account for a further 9% of all IPD. Kyaw et al. (2000) report overall coverage of 61% of all IPD by PCV-7 between 1988 and 1999.

The serotypes found in PPV-23 (1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19A, 19F, 20, 22F, 23F, 33F) account for approximately 90% of all IPD for which the specific serotype was identified in Scotland between 2003/04 and 2005/06. Kyaw et al. (2000) report a coverage of 96% by PPV-23 of cases of Scottish IPD between 1988 and 1999. In total, 851 (18.66%) isolates were not specifically typed from 1999/2000 to 2005/06. However, serogroup information was available for these cases.

PCV-7 has a high efficacy in preventing vaccine type IPD and not only prevents disease and infection in the vaccinated population but also prevents vaccinated hosts from carrying these serotypes. PPV-23 does cover a high proportion of IPD in Scotland but this vaccine only has a moderate efficacy in preventing vaccine type disease. A meta-analysis carried out looking at the efficacy of the polysaccharide vaccine has shown the vaccine to have an estimated efficacy of 65% (95% C.I. (-49%, 92%)) in preventing vaccine type IPD (Melegaro and Edmunds 2004b). In addition, this vaccine has no impact on the carriage of the 23 VT serotypes. Thus, there is no potential for herd immunity through the use of this vaccine.

Figure 6.4 shows the MLSTs responsible for at least 1% of all IPD in Scotland between 2003/04 and 2005/06. MLST 9 is the most common MLST found in IPD, responsible for 9% of all IPD during the three year period. MLST 9 is commonly associated with serogroup/serotype 14. Approximately 60% of the serogroup/serotype 14 isolates found in IPD between 2003/04 and 2005/06 were MLST 9. The second most common MLST found in IPD was 306. Over 65% of serogroup/serotype 1 isolates were associated with MLST 306.

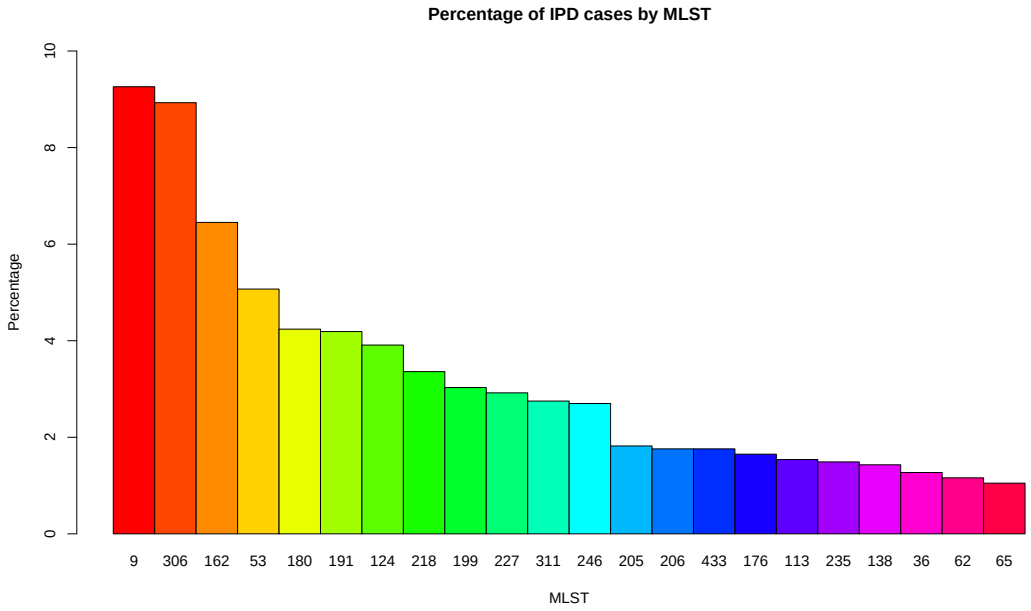

Figure 6.4: Percentage of IPD cases attributable to each MLST from 2003/04 to 2005/06 (for MLSTs that account for more than 1% of IPD).

There were 22 MLSTs each responsible for more than 1% of IPD in Scotland. The 10 most common disease-causing MLSTs together account for approximately 44% of all IPD in Scotland between 2003/04 and 2005/06. Of the 22 MLSTs shown in Figure 6.4, 16 were associated with serotypes found in PCV-7. The serotype associations are shown in Table 6.2. Only 2 MLSTs are associated with just the one serotype amongst the 22 most common. These are 227 and 36, associated with serotype 1 and 23F respectively. 5 isolates were non-typeable (NT), with only MLST information available.

Figure 6.5 shows the distribution of MLSTs associated with serogroup/serotype 1 IPD in Scotland. In total, serogroup/serotype 1 was associated with 23 different MLSTs during this three year period. From examination of this figure, it can be observed that MLST 306 is associated with the greatest percentage of serogroup/serotype 1 IPD each year and that the proportion increased from approximately 57% in 2003/04 to about 73% in 2005/06. This corresponds to a general increase in the number of serogroup/serotype 1 IPD, with only approxi-

MLST	Serotype (Number of isolates)
9	14 (167), 19F (1)
306	1 (154), 14 (1), 18C (1), 19F (1), 3 (1), 4 (1), 6B (1), NT (1)
162	9V (84), 19F (29), 1 (1), 20 (1), 6A (1), 9N (1)
53	8 (88), 3 (2), 1 (1)
180	3 (75), 1 (1), 6A (1)
191	7F (71), 7A (2), 19A (1), 4 (1), 6A (1)
124	14 (70), 9V (1)
218	12F (57), 12B (1), NT (3)
199	19A (38), 15B (13), 15C(2), 11A (1), NT (1)
227	1 (53)
311	23F (49), 14 (1)
246	4 (47), 19A (2)
205	4 (32), 1 (1)
206	4 (29), 1 (1), 9N (1), 9V (1)
433	22F (31), 22A (1)
176	6B (26), 1 (1), 18C (1), 19A (1), 23F (1)
113	18C (26), 18F (2)
235	20 (23), 7C (3), 4 (1)
138	6B (22), 6A (2), 1 (1), 7F (1)
36	23F (23)
62	11A (20), 9V (1)
65	6A (18), 6B (1)

Table 6.2: Serotypes associated with the 22 most common MLSTs found in IPD.

mately 11% of all IPD cases attributable to serogroup/serotype 1 IPD in Scotland in 2003/04 and 2004/05, increasing to over 16% in 2005/06. The number of different MLSTs associated with serogroup/serotype 1 IPD in Scotland decreased over time. There were 12 different MLSTs associated with serogroup/serotype 1 IPD in 2003/04, ten in 2004/05 and only six in 2005/06.

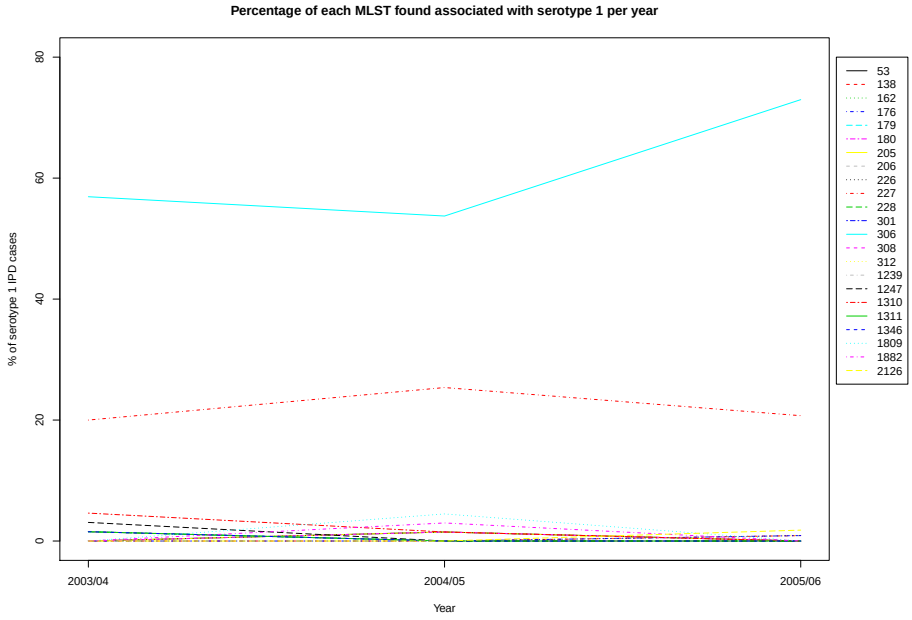

Figure 6.5: Percentage of serogroup/serotype 1 IPD associated with each MLST each year.

MLST 227 is the second most common MLST found associated with serogroup/serotype 1 IPD, with between 20 and 25% of this type of IPD attributable to MLST 227. MLST 227 serogroup/serotype 1 IPD did not vary greatly with year.

Figure 6.6 shows the distribution of MLSTs in serogroup/serotype 14 IPD in Scotland from 2003/04 to 2005/06. Overall, serogroup/serotype 14 IPD was associated with 31 different MLSTs. In this figure, it can be seen that MLST 9 was the most common MLST found in serogroup/serotype 14 IPD in Scotland in each of the years from 2003/04 to 2005/06. The percentage of serogroup/serotype 14 MLST 9 IPD cases increased slightly each year, with approximately 52% of serogroup/serotype 14 IPD attributable to this MLST in 2003/04 and 62% in

2005/06.

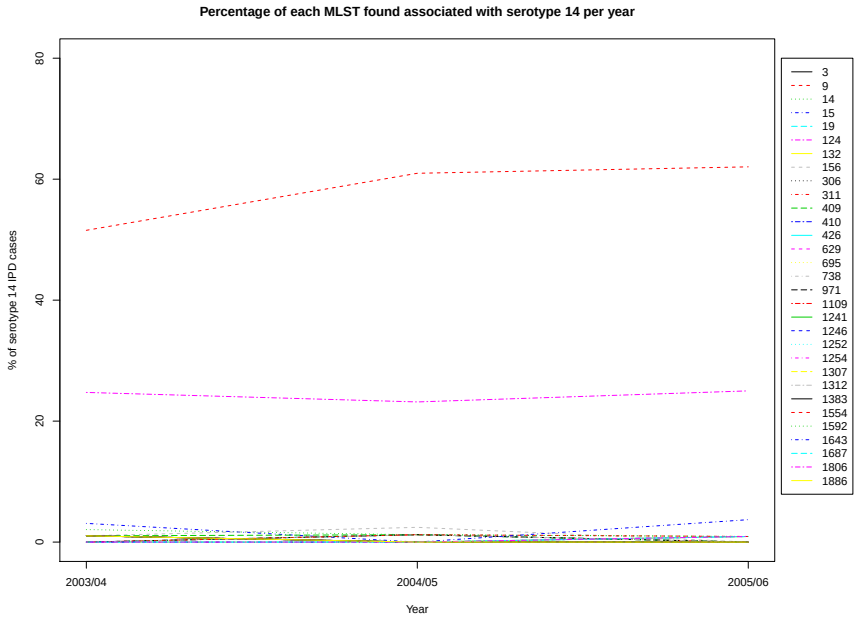

Figure 6.6: Percentage of serogroup/serotype 14 IPD associated with each MLST each year.

The second most common MLST found in serogroup/serotype 14 IPD is MLST 124. The proportion of serogroup/serotype 14 IPD attributable to this MLST remained fairly constant each year, averaging around 24%. The number of MLSTs associated with serogroup/serotype 14 IPD in Scotland decreased from 19 different MLSTs in 2003/04 to 11 in 2004/05 and 12 in 2005/06.

6.4.3 Trend Analysis

In Figure 6.7, the trends in IPD cases for the ten most common serogroups are shown. In addition, the trend for a group of serogroups consisting of those responsible for all other cases of IPD in Scotland is shown. Figure 6.7 supports the hypothesis that serogroup 1 is becoming more common in cases of IPD in Scotland over time. In 2005/06, serogroup 1 replaced serogroup 14 as the most common serogroup found in IPD. Serogroup 1 accounted for only approximately

5% of all IPD cases in 1999/00 but over 16% of all cases in 2005/06 suggesting a potentially significant increase over time. Serogroup 14 shows a decreasing trend in cases of IPD from 2001/02 to 2005/06. No other serogroup appeared to show a clear increasing or decreasing trend from 1999/00 to 2005/06.

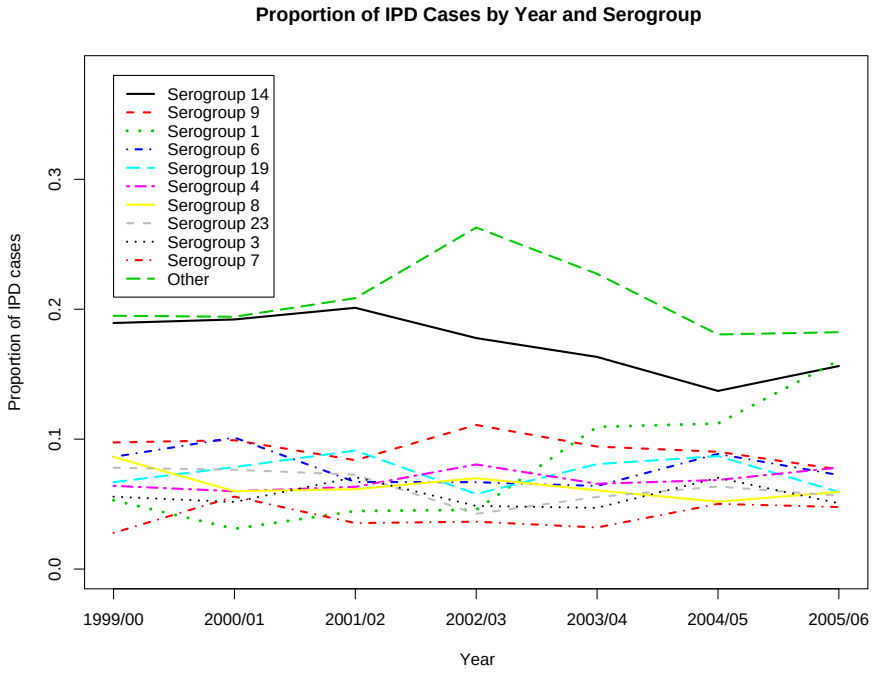

Figure 6.7: Trend plot of IPD cases by serogroup for the ten most common serogroups from 1999/00 to 2005/06.

Figures 6.8 to 6.13 show the trends in the ten most common serogroups found in IPD for each of the age groups. Figure 6.8 shows that for those aged 0 to 4 years, serogroup 14 was the most common serogroup of the 10 most common found in IPD for all years. There is no apparent increasing or decreasing trend for any of the 10 most common serogroups for this age group.

For the 5 to 34 years age group, it can be observed from examination of Figure 6.9 that serogroup 1 shows an increasing trend in cases of IPD and from 2001/02 was the serogroup accountable for the greatest proportion of cases of IPD in this group, with a marked increase from 23% of all cases of IPD in 2004/05 to

around 40% of cases in 2005/06. No other serogroup shows a clear increasing or decreasing trend during this period.

In the 35 to 49 years age group, serogroup 1 again appears to show an increasing trend in Figure 6.10, only 8% of all cases of IPD in 1999/00 and 21% of all cases in 2005/06 were attributable to serogroup 1. Between 6% and 10% of all cases of IPD each year were attributable to serogroup 14. Serogroup 8 was the most common cause of IPD in this age group in 1999/00, approximately 13% of all cases were attributable to this serogroup. However, by 2004/05 serogroup 8 was only responsible for about 4% of all IPD and was only responsible for 6% in 2005/06. Other than serogroup 1, none of the 10 most common serogroups show a clear increasing or decreasing trend over the period of study for this age group. The trend plot of serogroups for the 50 to 64 years age group (Figure 6.11) shows serogroup 14 to be the most common serogroup in IPD every year apart from 2004/05 where serogroup 4 was accountable for most invasive disease. There is no apparent increasing or decreasing trend for any of the ten most common serogroups in this age group.

Serogroup 14 was the most common cause of IPD in the 65 to 74 years age group for all years between 1999/00 and 2005/06 apart from 2002/03 when serogroup 9 was most common. There is no clear increasing or decreasing trend for any of the serogroups shown in Figure 6.12.

Serogroup 14 was the most common cause of IPD in the 75 years and over group for all years apart from 2004/05 when serogroup 3 was accountable for a slightly higher proportion of cases. The trend plot for this age group (Figure 6.13) shows no apparent trend, either increasing or decreasing, for any of the common disease-causing serogroups.

Examination of a trend plot of the 10 most common MLSTs in the period 2003/04 to 2005/06 (Figure 6.14) shows that most MLSTs remain fairly constant in proportion over the 3 year period. MLST 306 shows an increase from 2004/05 to 2005/06, replacing MLST 9 as the most common MLST in IPD in 2005/06. This corresponds with serogroup 1 replacing serogroup 14 to become the most common serogroup found in IPD as MLST 306 is the most common MLST associated with

serogroup 1 IPD in Scotland and MLST 9 is the most common MLST associated with serogroup 14.

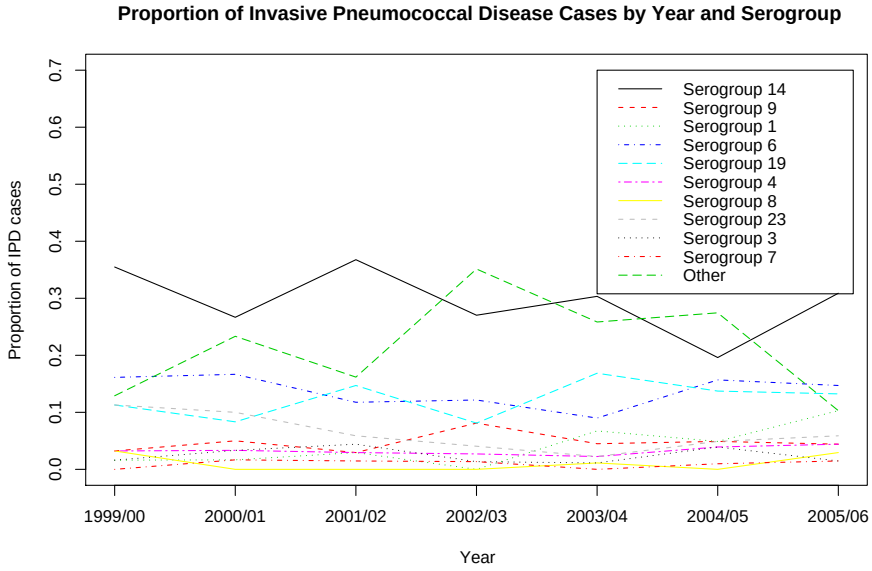

Figure 6.8: Trend plot of IPD cases by serogroup for those aged 0-4 years.

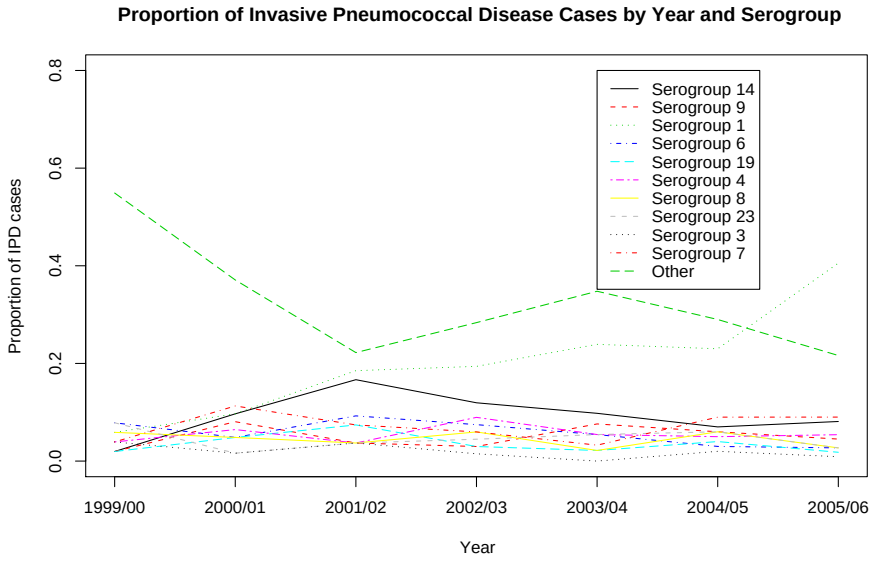

Figure 6.9: Trend plot of IPD cases by serogroup for those aged 5-34 years.

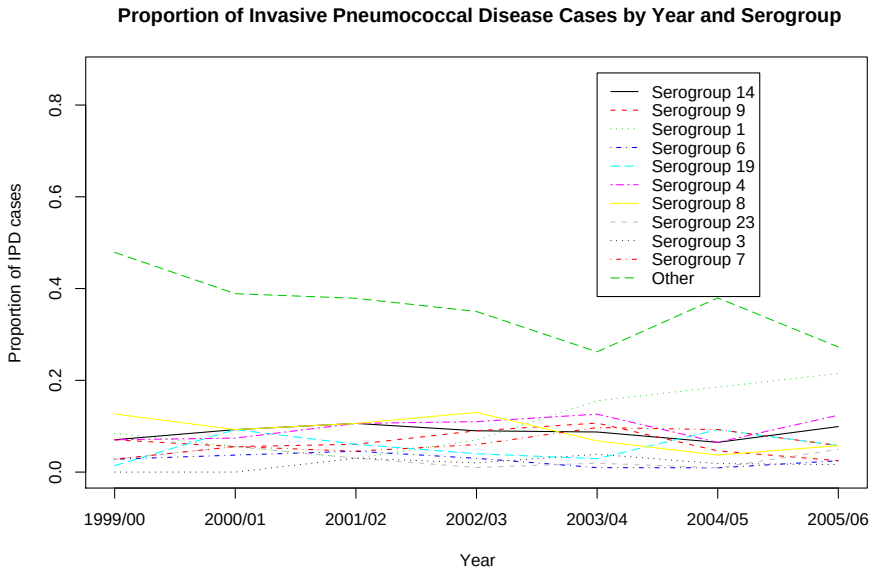

Figure 6.10: Trend plot of IPD cases by serogroup for those aged 35-49 years.

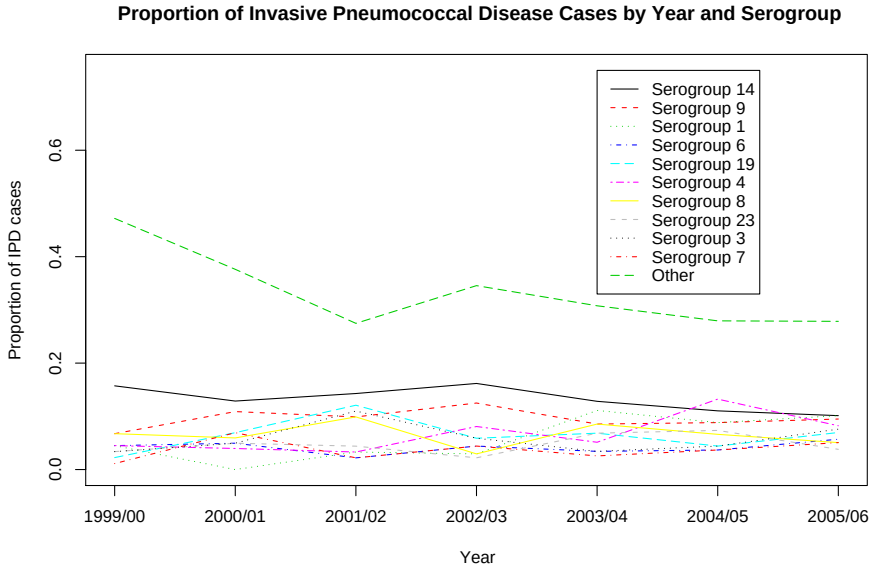

Figure 6.11: Trend plot of IPD cases by serogroup for those aged 50-64 years.

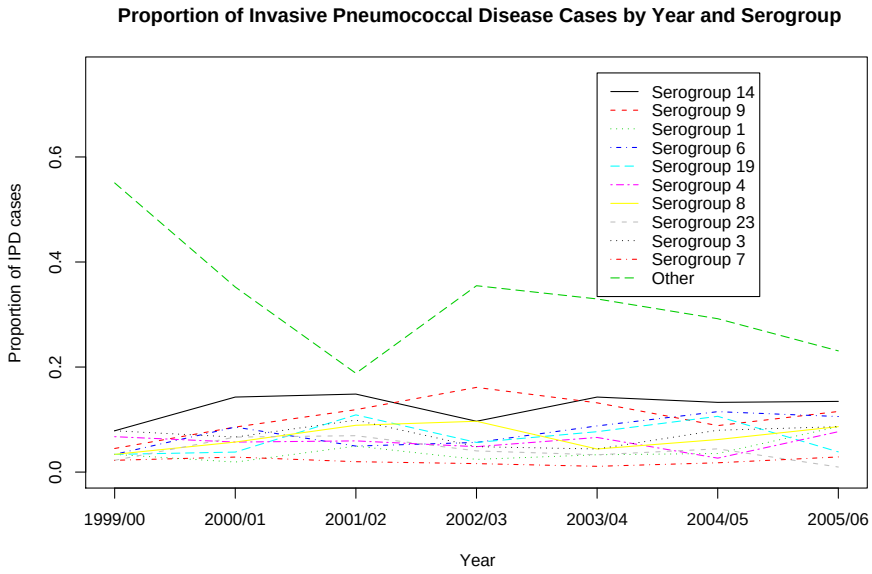

Figure 6.12: Trend plot of IPD cases by serogroup for those aged 65-74 years.

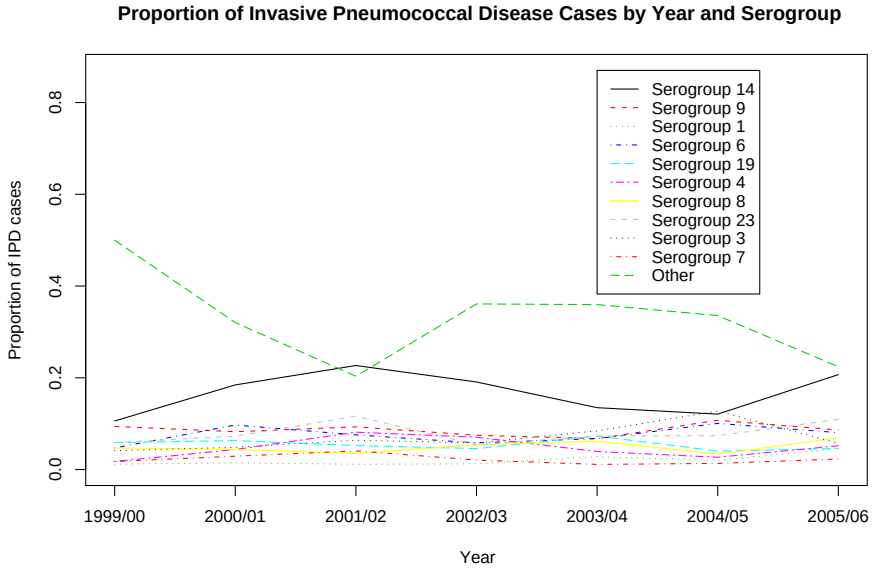

Figure 6.13: Trend plot of IPD cases by serogroup for those aged 75 years & over.

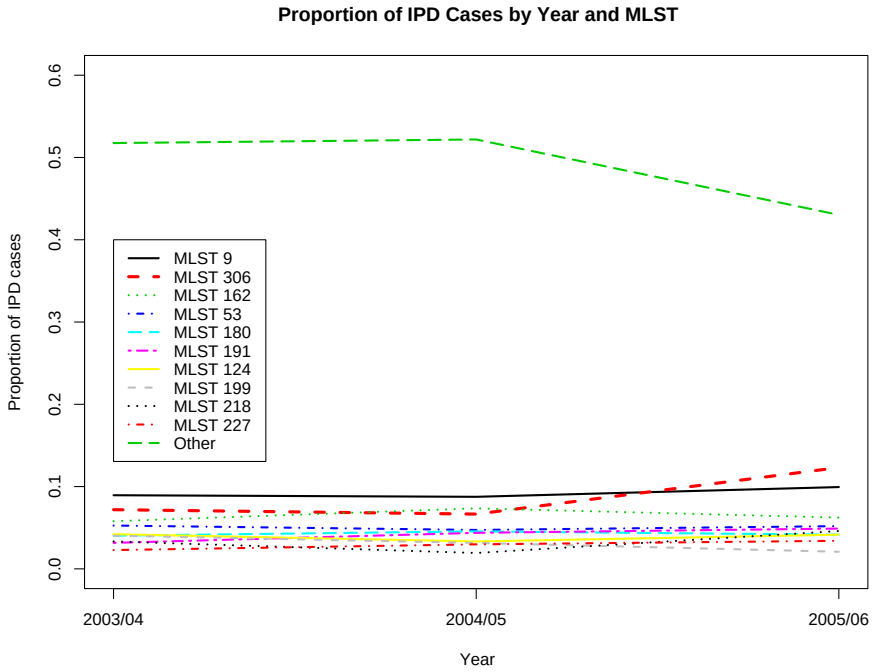

Figure 6.14: Trend plot of IPD cases by MLST for the ten most common MLSTs from 1999/00 to 2005/06.

6.4.4 Logistic regression

Logistic regression models were used to formally assess whether or not there is significant evidence of an increasing or decreasing trend for each of the serogroups and MLSTs involved in IPD in Scotland.

Serogroup trends

There were 35 serogroups in total observed in IPD in Scotland between 1999/00 and 2005/06. Table 6.3 shows the odds ratio (OR), Bonferroni adjusted 95% C.I. and p-value of year in the model for serogroups involved in at least 1% of IPD in Scotland. As a specific hypothesis was assumed about serogroup 1 having an increasing trend, no adjustment for multiple testing has to be made for the test for this serogroup and the p-value obtained for the serogroup 1 logistic regression may be compared to the significance level of 0.05.

Table 6.3: Results of the logistic regression model for the serogroups responsible for at least 1% of IPD.

Serogroup	**exp(β)**	**Adjusted 95% C.I.**	***p*-value**
14	0.938	(0.883, 0.997)	0.003
9	0.968	(0.892, 1.047)	0.230
1	1.359	(1.238, 1.493)	< 0.001
6	0.969	(0.891, 1.057)	0.328
19	0.984	(0.900, 1.074)	0.595
4	1.039	(0.946, 1.134)	0.284
8	0.948	(0.865, 1.044)	0.135
23	0.941	(0.858, 1.035)	0.084
3	0.995	(0.902, 1.100)	0.917
7	1.036	(0.926, 1.167)	0.357
18	0.983	(0.876, 1.104)	0.685
12	1.040	(0.914, 1.183)	0.400
22	0.977	(0.849, 1.125)	0.648
20	0.948	(0.808, 1.113)	0.360
33	0.936	(0.789, 1.110)	0.285
11	0.969	(0.811, 1.160)	0.638
15	1.013	(0.829, 1.239)	0.864

The p-value for year in the regression model for serogroup 1 is <0.001 and the OR ($\exp(\beta)$) is greater than 1. Therefore, there is significant evidence that serogroup 1 has been increasing in IPD in Scotland during the period 1999/00 to 2005/06.

The p-values for year for each of the other logistic regression models have to be compared to a significance level of 0.003, using the Bonferroni correction factor shown previously (6.2) to adjust for the 17 tests carried out for each serogroup observed in at least 1% of all IPD cases. Serogroup 14 is borderline significant with a p-value of 0.003. The OR for this serogroup is less than 1. Thus, there is significant evidence to suggest that serogroup 14 has been decreasing in IPD in Scotland from 1999/00 to 2005/06. No other serogroup was found to have a significant increasing or decreasing trend.

Separate logistic regression models were fitted to the serogroups found in IPD in each of the 6 age groups defined previously to assess trends within each group.

There was no information about the age of the patient for 32 cases of IPD. No adjustment was made for multiple testing for the serogroup 1 models. The results are shown in Table 6.4.

Table 6.4: Results of the logistic regression model for the serogroups responsible for at least 1% of IPD for those aged 0 to 4 years.

Serogroup	**Number**	**exp(β)**	**Adjusted 95% C.I.**	***p*-value**
14	151	0.947	(0.832, 1.079)	0.284
6	71	0.994	(0.837, 1.181)	0.932
19	66	1.081	(0.903, 1.294)	0.258
18	39	0.934	(0.748, 1.166)	0.426
23	31	0.844	(0.657, 1.084)	0.076
9	25	1.050	(0.794, 1.388)	0.649
1	22	1.461	(1.022, 2.088)	0.006
4	17	1.064	(0.760, 1.490)	0.629
3	13	0.991	(0.681, 1.443)	0.950
15	12	0.890	(0.603, 1.312)	0.433
33	7	0.988	(0.594, 1.643)	0.952

In total, amongst the 523 cases of IPD observed in the 0 to 4 years age group, there were 25 serogroups observed; 11 of which were each responsible for at least 1% of IPD within this age group. The results from the logistic regression analysis for this age group are shown in Table 6.4. The only serogroup found to have a significant trend was serogroup 1 with a p-value of 0.006. This serogroup was shown to increase over time.

There were 537 cases of IPD between 1999/00 and 2005/06 in the 5 to 34 years age group. 26 different serogroups were identified; 15 of which were responsible for at least 1% of IPD. Thus, after adjustments for multiple testing, a p-value of < 0.003 is required for significance. The results for these 15 serogroups are shown in Table 6.5. Once again, serogroup 1 was the only serogroup to show a significant trend in IPD, with a p-value of < 0.001.

25 different serogroups were accountable for IPD in the 35 to 49 years age group in Scotland. Of these serogroups, there were 17 which each accounted for at least

Table 6.5: Results of the logistic regression model for the serogroups responsible for at least 1% of IPD for those aged 5 to 34 years.

Serogroup	**Number**	**exp(β)**	**Adjusted 95% C.I.**	***p*-value**
1	122	1.358	(1.145, 1.611)	< 0.001
14	49	0.962	(0.778, 1.188)	0.617
7	39	1.009	(0.783, 1.255)	0.915
4	30	0.973	(0.748, 1.266)	0.782
18	29	0.840	(0.646, 1.093)	0.074
9	28	1.005	(0.764, 1.322)	0.960
6	28	0.809	(0.618, 1.058)	0.032
23	24	0.934	(0.699, 1.249)	0.524
8	23	0.896	(0.668, 1.201)	0.309
12	21	0.815	(0.600, 1.107)	0.070
19	18	0.875	(0.628, 1.218)	0.270
3	9	0.770	(0.483, 1.227)	0.128
15	7	1.269	(0.683, 2.355)	0.297
22	7	1.015	(0.592, 1.742)	0.941
33	6	0.841	(0.480, 1.475)	0.403

1% of the IPD within this age group, shown in Table 6.6. In total, there were 623 cases of IPD observed in this age range. Serogroup 1 had a significant positive trend (p-value < 0.001). In the logistic regression model for serogroup 8 in this age group, year had a significant p-value of 0.002. A p-value of < 0.003 is required for significance using the Bonferroni correction factor. The estimated OR for year in the logistic regression model for serogroup 8 is less than 1 indicating a decreasing trend in serogroup 8 IPD from 1999/00 to 2005/06. No other serogroup had a significant increasing or decreasing trend for this age group.

From the 828 cases of IPD, 24 different serogroups were observed in the 50 to 64 years age group. A trend test was carried out for only the 16 responsible for at least 1% of IPD. The results are displayed in Table 6.7. Only serogroup 1 had a significant trend with a p-value of 0.001. In the logistic regression models for the 65 to 74 years age group, none of the 16 most common serogroups had a significant trend, see Table 6.8. In total, 28 serogroups were observed in the 727 cases of IPD in this age group.

Table 6.6: Results of the logistic regression model for the serogroups responsible for at least 1% of IPD for those aged 35 to 49 years.

Serogroup	**Number**	**exp(β)**	**Adjusted 95% C.I.**	***p*-value**
1	80	1.301	(1.067, 1.586)	< 0.001
4	62	1.017	(0.836, 1.237)	0.810
14	54	0.964	(0.784, 1.185)	0.617
8	52	0.795	(0.646, 0.977)	0.002
7	41	1.082	(0.849, 1.379)	0.372
9	40	0.866	(0.687, 1.091)	0.087
12	34	1.053	(0.811, 1.368)	0.588
19	34	1.053	(0.811, 1.368)	0.588
18	23	1.264	(0.891, 1.793)	0.066
20	23	0.914	(0.675, 1.237)	0.413
23	17	0.958	(0.673, 1.363)	0.737
6	15	0.829	(0.573, 1.199)	0.159
3	12	1.119	(0.716, 1.748)	0.489
22	10	1.194	(0.719, 1.981)	0.337
11	9	0.774	(0.482, 1.243)	0.137
10	7	1.026	(0.588, 1.790)	0.898
33	7	0.986	(0.570, 1.706)	0.943

Serogroup 1 has a significant trend for the 75 years and over group with a p-value of 0.044. After adjusting for multiple testing for the 18 most common disease-causing serogroups in this age group, no other serogroups were found to have an increasing or decreasing trend. The results are shown in Table 6.9. In total, 29 serogroups were observed in the 1,290 cases of IPD in this age group between 1999/00 and 2005/06.

Serotype trends

A trend analysis was carried out for the serotypes for the years 2003/04 to 2005/06 since specific serotype information was available for these years. The results are shown for the serotypes responsible for at least 1% of all IPD in this period in Table 6.10.

From examination of Table 6.10, no serotype appears to have a significant in-

Table 6.7: Results of the logistic regression model for the serogroups responsible for at least 1% of IPD for those aged 50 to 64 years.

Serogroup	**Number**	**exp(β)**	**Adjusted 95% C.I.**	***p*-value**
14	108	0.889	(0.767, 1.030)	0.029
9	80	0.968	(0.819, 1.143)	0.588
4	59	1.165	(0.949, 1.431)	0.040
19	53	0.980	(0.801, 1.200)	0.784
8	52	0.935	(0.764, 1.145)	0.362
1	52	1.320	(1.047, 1.666)	0.001
3	48	0.998	(0.806, 1.235)	0.983
23	39	1.030	(0.815, 1.304)	0.727
6	35	1.000	(0.782, 1.279)	0.998
12	34	1.013	(0.790, 1.299)	0.890
7	32	1.022	(0.791, 1.322)	0.816
22	25	0.839	(0.629, 1.118)	0.092
18	18	0.826	(0.590, 1.156)	0.120
20	16	1.207	(0.819, 1.779)	0.185
11	14	1.045	(0.709, 1.541)	0.757
33	10	0.951	(0.607, 1.489)	0.760

creasing or decreasing trend following Bonferroni adjustment for the 19 tests. However, as a hypothesis was specified for serotype 1, the p-value obtained in this test should be compared to the significance level 0.05. Thus, serotype 1 appears to have a significant trend during the period 2003/04 to 2004/05. The OR for this serotype is greater than 1. Therefore, there is significant evidence that serotype 1 is increasing in IPD with increasing year.

As PPV-23 was introduced in 2003 for routine administration of those aged 65 years and over in Scotland, logistic regression models were fitted to assess the serotype trends for the 65 to 74 years age group and the 75 years and over age group to determine whether or not the vaccine had an impact on the VT serotypes in these groups. Approximately 85% of typeable IPD in the 65 to 74 years age group is attributable to the 23 serotypes found in PPV-23, whilst in the 75 years and over group, 30% of the typeable IPD is attributable to the 23 serotypes found in PPV-23. Serotype information was missing for approximately 11% of

Table 6.8: Results of the logistic regression model for the serogroups responsible for at least 1% of IPD for those aged 65 to 74 years.

Serogroup	**Number**	**exp(β)**	**Adjusted 95% C.I.**	***p*-value**
14	91	0.988	(0.839, 1.164)	0.846
9	79	1.027	(0.862, 1.224)	0.670
6	56	1.125	(0.917, 1.381)	0.116
3	52	0.950	(0.772, 1.170)	0.506
8	50	1.017	(0.822, 1.259)	0.828
19	48	1.021	(0.823, 1.267)	0.795
4	41	0.930	(0.737, 1.173)	0.394
23	30	0.834	(0.633, 1.100)	0.073
1	29	1.142	(0.862, 1.514)	0.197
18	21	0.971	(0.705, 1.338)	0.803
22	18	1.026	(0.725, 1.452)	0.840
11	17	1.049	(0.735, 1.497)	0.712
7	15	0.932	(0.639, 1.360)	0.614
12	14	1.314	(0.858, 2.013)	0.080
33	12	0.868	(0.567, 1.331)	0.365
20	10	0.882	(0.554, 1.403)	0.457

IPD cases for the 65 to 74 years age group and the same percentage for those aged 75 years and over. From the logistic regression modelling, no significant trends were identified for either age group.

MLST trends

There were 158 MLSTs involved in IPD in Scotland in 2003/04, 140 in 2004/05 and only 115 in 2005/06 showing a reduction in the diversity of MLSTs over time. In total, 273 different MLSTs were identified in IPD in Scotland. The ten most common MLSTs together account for 51% of all IPD in Scotland between 2003/04 and 2005/06.

The MLSTs associated with the PPV-23 serotypes during the period from 2003/04 to 2005/06, which includes those serotypes found in PCV-7, are shown in Tables 6.11 and 6.12.

Table 6.9: Results of the logistic regression model for the serogroups responsible for at least 1% of IPD for those aged 75 years or over.

Serogroup	**Number**	**exp(β)**	**Adjusted 95% C.I.**	***p*-value**
14	219	0.976	(0.874, 1.091)	0.548
9	110	0.955	(0.822, 1.109)	0.392
23	101	1.017	(0.871, 1.188)	0.758
6	96	1.003	(0.856, 1.175)	0.961
3	86	1.095	(0.927, 1.294)	0.128
19	70	0.915	(0.760, 1.102)	0.180
8	64	1.023	(0.847, 1.236)	0.736
4	63	0.986	(0.814, 1.194)	0.835
22	43	1.002	(0.796, 1.261)	0.983
7	29	0.886	(0.668, 1.176)	0.237
18	28	1.124	(0.847, 1.492)	0.255
20	28	0.826	(0.614, 1.111)	0.073
33	28	1.013	(0.764, 1.344)	0.902
12	26	1.015	(0.757, 1.362)	0.887
1	26	1.246	(0.921, 1.686)	0.044
11	21	0.969	(0.698, 1.343)	0.784
15	19	1.257	(0.882, 1.793)	0.073
16	14	1.023	(0.688, 1.521)	0.874

The diversity index was calculated for the MLSTs, as described by Brueggemann et al. (2003), in which it was defined that the genetic diversity, λ, of a population of bacteria provides a measure of the probability that two randomly selected pneumococcal isolates will be of different types. As with the Brueggemann et al. analysis, the Simpson index (Simpson 1949), D, is used to obtain an unbiased estimate of λ where

$$\lambda = \sum_{i=1}^{n} \frac{n_i(n_i - 1)}{N(N-1)} \quad \text{and} \quad D = 1 - \lambda. \tag{6.3}$$

In (6.3), n_i is the number of isolates of the i^{th} type and N is the overall number of isolates in the sample. Here, $N = 1,815$ since MLST information was not available for 304 isolates between 2003/04 and 2005/06. $i = 1, ..., 273$ as there are 273 distinct MLSTs observed. λ was calculated to equal approximately 0.035.

Table 6.10: Results of the logistic regression model for the serotypes responsible for at least 1% of IPD between 2003/04 and 2005/06.

Serotype	**Number**	**exp(β)**	**Adjusted 95% C.I.**	***p*-value**
14	287	0.976	(0.785, 1.214)	0.759
1	243	1.265	(0.998, 1.604)	0.006
4	134	1.100	(0.811, 1.491)	0.384
9V	123	0.887	(0.647, 1.216)	0.288
8	108	1.006	(0.718, 1.410)	0.959
3	105	1.023	(0.728, 1.438)	0.852
23F	99	1.013	(0.715, 1.436)	0.916
6B	83	0.975	(0.667, 1.425)	0.136
7F	77	1.224	(0.819, 1.829)	0.160
19F	76	0.942	(0.634, 1.400)	0.674
18C	69	0.707	(0.464, 1.077)	0.022
19A	63	0.799	(0.517, 1.235)	0.151
12F	61	1.390	(0.879, 2.196)	0.045
6A	57	1.209	(0.761, 1.922)	0.253
22F	44	0.726	(0.431, 1.224)	0.088
9N	39	0.961	(0.556, 1.660)	0.840
20	36	0.851	(0.482, 1.504)	0.430
11A	29	0.925	(0.491, 1.743)	0.732
15B	22	0.990	(0.479, 2.045)	0.970

An approximate 95% confidence interval for the genetic diversity, D, may be calculated as follows (Grundmann et al. 2001):

$$D \pm 2\sqrt{\sigma^2},$$

where

$$\sigma^2 = \frac{4}{N}\left(\sum_{i=1}^{n}\left(\frac{n_i}{N}\right)^3 - \left(\sum_{i=1}^{n}\left(\frac{n_i}{N}\right)^2\right)^2\right).$$

Thus, the genetic diversity, D, of the MLSTs from IPD was 0.965 (95% C.I. (0.962, 0.968). Therefore, there is a high probability that two randomly selected

Serotype[a]	MLST (Number of isolates)
1	306 (154), 227 (53), 1310 (4), 1809 (3), 1247 (2), 1882(2), 2126 (2), 53 (1), 138 (1), 162 (1), 176 (1), 179 (1), 180 (1), 205 (1), 206 (1), 226 (1), 228 (1), 301 (1), 308 (1), 312 (1), 1239 (1), 1311 (1), 1346 (1)
2	-
3	180 (75), 260 (3), 1003 (3), 1486 (3), 53 (2), 1220 (2), 1253 (2), 232 (1), 233 (1), 306 (1), 312 (1), 1300 (1), 1344 (1), 1377 (1), 1682 (1), 1867 (1), 1887 (1), 2263 (1)
4	246 (47), 205 (32), 206 (29), 247 (4), 695 (4), 4(1), 24 (1), 191 (1), 208 (1), 235 (1), 244 (1), 306 (1), 393 (1), 409 (1), 1178 (1), 1309 (1), 1884 (1), 1951 (1), 2143 (1)
5	100 (1), 1400 (1)
6B	176 (26), 138 (22), 469 (5), 96 (2), 146 (2), 1256 (2), 65 (1), 90 (1), 135 (1), 156 (1), 170 (1), 273 (1), 306 (1), 315 (1), 402 (1), 473 (1), 1002 (1), 1235 (1), 1240 (1), 1251 (1), 1612 (1), 1865 (1), 1981 (1)
7F	191 (71), 1122 (2), 138 (1)
8	53 (88), 404 (6), 1110 (4), 33 (1), 54 (1), 944 (1), 1380 (1), 1801 (1)
9N	405 (14), 66 (13), 834 (4), 162 (1), 166 (1), 206 (1), 1197 (1), 1302 (1), 1305 (1), 1885 (1)
9V	162 (84), 156 (9), 163 (5), 609 (3), 405 (2), 44 (1), 62 (1), 124 (1), 165 (1), 206 (1), 312 (1), 466 (1), 608 (1), 644 (1), 838 (1), 1243 (1), 1306 (1), 1630 (1), 2025 (1)
10A	97 (6), 461 (1), 1282 (1), 1497 (1), 2068 (1)
11A	62 (20), 408 (2), 513 (2), 199 (1), 1180 (1), 1304 (1)

Table 6.11: MLSTs associated with 12 of the PPV-23 serotypes found in IPD between 2003/04 and 2005/06.

[a]PCV-7 serotypes highlighted in red.

Serotype[a]	MLST (Number of isolates)
12F	218 (57), 220 (2), 223 (1), 1952 (1)
14	9 (167), 124 (70), 14 (4), 15 (4), 1643 (4), 156 (3), 409 (2), 1554 (2), 1592 (2), 3 (1), 19 (1), 132 (1), 306 (1), 311 (1), 410 (1), 426 (1), 629 (1), 695 (1), 738 (1), 971 (1), 1109 (1), 1241 (1), 1246 (1), 1252 (1), 1254 (1), 1307 (1), 1312 (1), 1383 (1), 1687 (1), 1806 (1), 1886 (1)
15B	199 (13), 411 (1), 419 (1), 579 (1), 583 (1), 1262 (1), 1888 (1), 2264 (1)
17F	392 (5), 393 (1), 1793 (1)
18C	113 (26), 121 (5), 697 (4), 110 (3), 1255 (3), 1381 (3), 114 (2), 308 (2), 638 (2), 1343 (2), 1361 (2), 119 (1), 176 (1), 306 (1), 716 (1), 1073 (1), 1233 (1), 1242 (1), 1303 (1), 1308 (1), 1711 (1), 2134 (1)
19A	-
19F	162 (29), 426 (7), 177 (4), 309 (4), 179 (2), 420 (2), 1218 (2), 9 (1), 43 (1), 58 (1), 171 (1), 236 (1), 251 (1), 306 (1), 312 (1), 416 (1), 422 (1), 423 (1), 425 (1), 476 (1), 644 (1), 654 (1), 1002 (1), 1035 (1), 1233 (1), 1359 (1), 1718 (1)
20	235 (23), 1435 (3), 1312 (2), 20 (1), 162 (1), 568 (1), 1950 (1)
22F	433 (31), 698 (3), 819 (3), 455 (2), 868 (1), 1244 (1), 1525 (1)
23F	311 (49), 36 (23), 1682 (7), 33 (2), 1237 (2), 81 (1), 176 (1), 427 (1), 442 (1), 778 (1), 833 (1), 1179 (1), 1245 (1), 1499 (1), 1868 (1), 1869 (1)
33F	100 (11), 673 (7), 60 (1)

Table 6.12: MLSTs associated with 11 of the PPV-23 serotypes found in IPD between 2003/04 and 2005/06.

[a]PCV-7 serotypes highlighted in red.

isolates will be of different MLST type.

As 156 MLSTs only appeared once in IPD in Scotland, 41 only twice and a further 21 appeared only three times, for the purpose of the trend analysis of MLSTs involved in IPD, only the MLSTs responsible for at least 1% of IPD in Scotland (shown in Figure 6.14) were assessed. Table 6.13 shows the exponentiated coefficient, Bonferroni adjusted 95% confidence interval and p-value of year resulting from each of the logistic regression models for the 22 MLSTs involved in at least 1% of IPD in Scotland. The p-value for the explanatory variable year in the logistic regression model with the proportion of MLST 306 IPD in Scotland as response is 0.001, shown in Table 6.13, which is below the significance level of 0.05. Therefore, as the OR for year in this model is greater than 1, there is evidence of an increasing trend in MLST 306 IPD.

All other p-values for year in the regression models were compared to a significance level of 0.002, using the Bonferroni correction factor to correct for multiple testing. No other MLST shows a significant increasing or decreasing trend.

6.5 Conclusions

In Scotland, between 1999/00 and 2005/06 on average approximately 650 cases of IPD occurred per year, with the greatest burden of disease amongst the elderly, particularly those aged over 75 years. A slight decline in the proportion of IPD for this age group was observed from 2003/04 to 2005/06 which is potentially attributable to the introduction of PPV-23. However, on carrying out logistic regression modelling of the serotypes observed in IPD for the vaccine targeted age groups, 65 to 74 years and 75 years and over, between 2003/04 and 2005/06 none of the PPV-23 serotypes showed a statistically significant decreasing trend in IPD.

The VT serogroup/serotype 14 was responsible for the greatest overall burden of IPD in Scotland, with 17% of disease between 1999/00 and 2005/06 caused by this serogroup. The observation that serogroup 14 is the most common disease-causing serogroup corresponds with other IPD studies in Europe prior to the

Table 6.13: Results of the logistic regression model for the MLSTs responsible for at least 1% of IPD between 2003/04 and 2005/06.

MLST	**exp(β)**	**Adjusted 95% C.I.**	***p*-value**
9	1.062	(0.804, 1.402)	0.539
306	1.395	(1.042, 1.869)	0.001
162	1.030	(0.741, 1.432)	0.797
53	1.012	(0.700, 1.464)	0.924
180	1.011	(0.678, 1.508)	0.939
191	1.242	(0.823, 1.875)	0.134
124	0.998	(0.658, 1.515)	0.987
218	1.239	(0.784, 1.956)	0.183
199	0.715	(0.444, 1.151)	0.045
227	1.221	(0.750, 1.990)	0.244
311	0.917	(0.561, 1.498)	0.615
246	0.974	(0.593, 1.601)	0.883
205	1.207	(0.654, 2.228)	0.386
206	1.217	(0.652, 2.271)	0.372
433	0.596	(0.314, 1.132)	0.022
176	1.178	(0.620, 2.237)	0.468
113	1.020	(0.530, 1.965)	0.930
235	0.773	(0.390, 1.531)	0.284
138	1.155	(0.581, 2.295)	0.552
36	1.190	(0.572, 2.475)	0.499
62	1.135	(0.531, 2.429)	0.636
65	1.629	(0.691, 3.838)	0.106

introduction of PCV-7 (Serrano et al. 2005; Normark et al. 2001) and a previous Scottish IPD study reported this serogroup/serotype to be the most prevalent in disease in a period preceding that of this analysis, 1988 to 1999 (Kyaw et al. 2000). The burden of disease attributable to serogroup 14 is almost double that of the next most common serogroup found in IPD, serogroup 9, and is slightly more than double that of the third most common, serogroup 1. However, it is interesting to observe serogroup/serotype 1 replacing serogroup/serotype 14 as the most common cause of IPD in 2005/06 after a period in which serotype 14 appears to dominate in IPD.

Serogroup 14 was the most common cause of disease overall. However, in the 6 age groups considered in this chapter, the serogroup attributable to most disease varies. Thus, there appears to be an age association with the serogroups involved in IPD in Scotland as observed elsewhere. In a study of IPD in children, Kaplan et al. (2002) observe increasing proportions of serotype 1 IPD with increasing age and decreasing proportions of serotype 14. This appears similar to the observations in this study where serogroup 14 appears as the most common cause of disease in every year for the 0 to 4 years age group, whilst in the 5 to 34 years old group serogroup 1 was the most common cause of disease in all but one year when serogroup 7 was accountable for most IPD. Similar results were presented in the Kyaw et al. (2000) study of IPD isolates in Scotland in a period prior to the one considered in this analysis, where serogroup 14 was the most common cause of IPD in those aged under 5 years. Kyaw et al. report serogroup 1 to be the most common cause of disease in those aged 5 to 64 years of age. However, the analysis in this chapter considers the age groups 5 to 34 years, 35 to 49 years and 50 to 64 years separately. Amongst the disease isolates for those aged 35 to 49 years, serogroup 8 was the most common until 2002/03 following which serogroup 1 was the most common; in the 50 to 64 years age group serogroup 14 was the most common cause of IPD for every year but 2004/05 when serogroup 4 was attributable for the highest burden of disease.

Although IPD attributable to 35 different serogroups was observed in this study, the majority of disease was attributable to a much smaller number of serogroups, with only 17 serogroups each responsible for at least 1% of all cases of IPD. During the three years for which serotype specific information was available, 47% of all isolates collected were PCV-7. Thus, if serotype proportions in IPD remain constant beyond 2005/06, the introduction of the vaccine should have a substantial impact on the burden of disease due to herd immunity preventing disease in the unvaccinated groups.

There are many MLSTs found in cases of IPD in Scotland. The genetic diversity of the MLSTs found in IPD was high with a probability of around 0.97 that two randomly selected pneumococcal isolates will have different MLSTs as in only three years of study, 273 different MLSTs were identified but many occur rarely.

Only 22 MLSTs accounted for at least 1% of IPD between 2003/04 and 2005/06. The most common MLST in IPD was identified to be MLST 9, predominantly associated with serogroup/serotype 14, with only one MLST 9 isolate associated with serotype 19F. The second most prevalent disease-causing MLST was MLST 306, primarily associated with serogroup/serotype 1. However, MLST 306 was also identified to be associated with a single case of each of the serotypes 14, 18C, 19F, 3, 4 and 6B, in addition to a nontypeable isolate. In a study of IPD in the USA, Beall et al. (2006) documented that the association of MLSTs with more than one serotype was a rare event. However, in this study this occurrence appears far more commonly with all MLSTs amongst the 22 most common associated with more than one serotype, 12 of which are associated with both a VT and NVT serotype.

The results from the analysis of serogroup trends in Scotland show significant evidence that serogroup 1 has increased in IPD incidence with year from 1999/00 to 2005/06. This increasing trend is apparent in all age groups but is not statistically significant in the 65 to 74 years group where no serogroup was found to have a statistically significant trend. The highest increasing trends for serogroup 1 were observed amongst those aged 5 to 34 years old and those aged 35 to 49 years old. The results obtained provide statistical evidence to add to the previous speculation that serogroup 1 has been increasing in Scotland (Kirkham et al. 2006) and corresponds with results obtained for the UK and Ireland in which serogroup 1 bacteraemia was found to increase over time (Farrell et al. 2008). Contradictory results have been obtained regarding serogroup/serotype 1 IPD in the UK, as a study of serotypes involved in IPD in Oxfordshire between 1996 and 2005 showed serotype 1 to have significantly decreased in incidence over that period (Foster et al. 2008). However, Foster et al. state that it appears that the largest decline occurred around 1999 and that from 2000 serotype 1 IPD appeared relatively stable. Thus, although the increase in trend of serogroup 1 IPD was not observed from 1999 as the case in Scotland, it may not be the case that the opposite trend is true for this period in the Oxfordshire study.

Amongst all cases of IPD, serogroup 14 was found to have a statistically significant decreasing trend but this trend was not confirmed statistically in any of the

analyses of the different age groups separately, although this trend was observed in all but one of the age classes. Overall, no other serogroup was identified as having a statistically significant trend. However, in cases of IPD in the 5 to 34 years age group, the NVT serogroup/serotype 8 was found to have a significant decreasing trend.

The results of the MLST analysis show an increasing trend in MLST 306 IPD, as hypothesised. This corresponds to the increase in serogroup 1 IPD. No other MLST was identified to have significantly increased or decreased between 2003/04 and 2005/06. However, as there are many MLSTs observed in disease and there are only three years of data there is little power to detect trends in the MLSTs.

PCV-7 does not provide protection against IPD caused by serogroup/serotype 1. Thus, it is concerning to see this serotype increasing in IPD in Scotland prior to vaccine use. In 2006, PCV-7 was introduced in Scotland and, assuming similar results to those obtained in the USA, will greatly reduce the burden of disease attributable to the VT serotypes 4, 6B, 9V, 14, 18C, 19F and 23F. However, amongst the 10 most common serotypes causing disease in Scotland in the 3 years prior to PCV-7 use, 4 (serotypes 1, 8, 3 and 7F) are not included in this vaccine formulation. This is of interest as there is potential for these virulent NVT serotypes, and other IPD causing NVT serotypes, to become evermore prevalent in IPD following reductions in the VT serotypes, as has been observed elsewhere (Aguiar et al. 2008; Munõz-Almagro et al. 2008). In addition, as there are several MLSTs associated with more than one serotype in Scotland, there is the potential for capsular switch events to occur which could also affect the overall disease burden in Scotland.

The recently developed 10-valent PCV contains the same 7 serotypes found in PCV-7, as well as serotypes 1, 5 and 7F; two of which are serotypes found amongst the 10 most common in IPD between 2003/04 and 2005/06. Thus, the introduction of this vaccine could aid in the prevention of disease in Scotland, particularly if serotype 1 continues to increase in IPD following PCV-7 implementation. The 13-valent vaccine contains these 10 serotypes in addition to the other common disease-causing serogroups 3, 6A and 19A. The introduction of this vaccine would

perhaps be necessary should the burden of disease become greater due to increases in 19A disease, as observed in the USA.

In this chapter, logistic regression modelling was carried out to assess the trends. However, an alternative approach would have been to use multinomial logistic regression, a method which is adopted in Chapter 8 of this thesis. However, as this analysis was carried out in conjunction with microbiologists, the logistic regression approach was preferred due to the ease of explanation of the results from this method.

The main limitation of the analysis carried out in this chapter is that serotype specific information is not available prior to 2003/04 and so the only trend analysis possible from 1999/00 is on the serogroups. This means the trends for the VT serotypes prior to the introduction of PCV-7 cannot be examined from 1999/00. However, serogroup information has been adopted in previous trend analyses of pneumococci where serotype information is unavailable (Feikin and Klugman 2002). Feikin et al. state that there are biological justifications for carrying out an analysis of serogroups due to the fact that levels of antibiotic resistance are similar for serotypes within a serogroup and PCV-7 may offer some cross-protection.

In future work, information collected about Scottish IPD in the 3 years following the introduction of PCV-7 should be used in a similar analysis to that carried out in this chapter to assess the impact of vaccine introduction. Pre-vaccination trends could be compared to post-vaccination trends to identify any significant differences attributable to PCV-7 use. It will be of interest to examine whether or not the NVT serotype 1 continues to increase in cases of IPD in Scotland and if any other NVT serotypes become more prevalent.

Chapter 7

Analysis of the association between pneumococcal serogroups and MLSTs and mortality

7.1 Introduction

This chapter follows from the previous chapter in thatIt appears that in this situation the Cochran-Mantel-Haenszel test may not be the most appropriate analysis to use in this situation as this test has low power for detecting an association in which the patterns of association for some strata are in the opposite direction of those displayed by other strata. Thus, a nonsignificant result can also indicate that the pattern of association does not have enough strength or consistency to dominate any other pattern. statistical techniques are used to examine serogroups and MLSTs involved in IPD. In particular, this chapter assesses associations between serogroups and MLSTs and mortality. As there are many different pneumococcal serotypes which have the potential to cause IPD, it is important for the development of vaccinations to establish not only which are the most prevalent in disease but also which are associated with a greater risk of death from IPD as this can influence the serotypes which should be included in a pneumococcal vaccine.

7.2 Background

The association between mortality and serotype in Scotland is unknown. A study of 103 patients suffering from pneumococcal bacteraemia between 1993 and 1995 in the Grampian region of Scotland documented serotypes 6A and 19A as having higher fatality rates than the other serotypes observed in disease (McKenzie et al. 2000). However, the sample size in this study was too small to obtain statistically significant results.

In other countries attempts have been made to determine the potential for specific serotypes in IPD to result in fatality. However, some involve assessment of associations with mortality for groups of serotypes rather than for individual serotypes. For example, Sjöström et al. (2006), Alanee et al. (2007), and Jansen et al. (2009) group serotypes in slightly different ways but base the groupings according to their invasive disease potential as identified by a meta-analysis of carriage and disease rates of various serotypes (Brueggemann et al. 2004).

7.2.1 Associations between groups of serotypes and mortality

Sjöström et al. (2006) group the serotypes into high (serotypes 1 and 7F), medium (4, 9V, 14, 18C), and low (3, 6A, 6B, 8, 19F, 23F) invasive disease potential. They state that although having high invasive disease potential, serotypes 1 and 7F caused no fatalities in their study of 494 adults and that serotype 1 was only found among younger adults. Sjöström et al. believe that these invasive serotypes are most able to cause disease and thus are generally observed in disease in healthy individuals whilst the other serotypes found in disease are more opportunistic, causing disease in elderly and weaker individuals. Unfortunately, due to the small number of patients in the study, Sjöström et al. state that they were unable to determine whether variations in disease severity in previously healthy individuals are attributable to the capsular type, clonal type or both.

As in the Sjöström et al. study, Alanee et al. (2007) group serotypes 1 and 7F

together as invasive. However, Alanee et al. also include all other serotypes found in serogroup 7 in this invasive group, as well as all serotypes within serogroup 5. This group is compared to a group involving the PCV-7 serotypes and a third group of paediatric serotypes (6, 9, 14, 19 and 23). In a univariate analysis of the 796 adult patients in the study, examining associations between these groups of serotypes and mortality and disease severity, statistically significant results were obtained for an increased risk of severe disease or fatality when infected with paediatric serotypes. However, in a multivariate analysis involving other co-morbidities no significant associations were identified between any of these groups of serotypes and disease severity or mortality.

Jansen et al. (2009) consider two groups of serotypes: a reference group composed of types 1, 5 and 7F as in the Alanee et al. analysis but, in addition, 15B, 20 and 33F are included in this group. It is concluded that the group composed of serotypes 3, 19F, 23A, 16F, 6B, 9N and 18C are associated with increased case-fatality rates (CFRs) when compared to the invasive reference group. As with the Sjöström et al. study, Jansen et al. state that serogroups 1 and 7 affect relatively healthy adults, as does serogroup 5. Jansen et al. state that although they had a relatively large number of patients in the study for which isolates could be typed (1,142 patients) they were unable to assess serotype specific associations with mortality due to the small number of isolates observed for each serotype.

A fourth analysis of mortality associations involved the assessment of groupings of serotypes. In this analysis of 160 patients, the serotypes involved in IPD were grouped according to VT or NVT serotypes based on PCV-7 (Chen et al. 2009). In this study, neither group was associated with mortality outcome. However, Chen et al. identified associations between strains with high levels of antibiotic resistance and fatal outcome.

7.2.2 Associations between individual serotypes and mortality

In a recent much larger population based study of 18,858 cases of IPD in Denmark, a statistical analysis was carried out of serotype specific associations with

mortality (Harboe et al. 2009). In this analysis, serotype 1 was selected as the reference category for comparison as it was the most commonly isolated serotype in IPD. The results show serotypes 31, 11A, 35F, 17F, 3, 16F, 19F, 15B and 10A were significantly associated with highly increased mortality when compared to serotype 1 for those aged at least 5 years. There appeared to be different mortality associations for those under 5 years. As fatalities in this age group were low in number, statistical precision was low and no statistically significant associations were identified. However, it was stated that the case fatality for this age group was highest for serotype 1 and serotypes 14, 6A, 7F and 4 appeared to be associated with decreased mortality compared to serotype 1.

Other smaller studies have assessed serotype specific associations with mortality attributable to IPD (Henriques et al. 2000; Martens et al. 2004; Rückinger et al. 2009; Balakrishnan et al. 2000). Two of these studies document associations with mortality for serotype 3, with this serotype associated with increased risk of fatality (Henriques et al. 2000; Martens et al. 2004). In addition, Martens et al. (2004) report a reduced relative risk of fatality associated with serotype 1 IPD. Ruckinger et al. (2009) report that serotype 7F is associated with an increased risk of severe and fatal outcome due to IPD after adjusting for other co-morbidities. Ruckinger et al. state that due to the fact that serotype 1 may be associated with a reduced risk of IPD fatality, as observed in the study by Martens et al., those studies mentioned previously which group serotypes 1 and 7F together may be missing associations with 7F and increased risk of death. Balakrishnan et al. (2000) report in their study of 104 cases of pneumococcal bacteraemia that serotype 14 was common and significantly associated with higher case fatality. This is in contrast to the study by Henriques et al. in which serotype 14 was the most common serotype overall but seemed to be associated with lower risks of severe disease and death.

A further study investigating mortality associated with pneumococcal meningitis focussing on the three serotypes 1, 3 and 9V reports significantly lower CFRs for serotype 1 than for 3 and 9V (Østergaard et al. 2004). Another investigating IPD in those aged at least 50 years in the USA following PCV-7 use reports significantly higher CFRs for IPD attributable to serotypes 19F, 23F, 3 or 11A

and lower rates for serotype 12F when compared to serotype 14 (Lexau et al. 2005).

No studies investigating mortality associations for MLSTs could be identified.

7.3 Methods

In this chapter, data on all cases of IPD, identified from blood or CSF samples in Scotland recorded at the Scottish Meningococcal and Pneumococcal Reference Laboratory (SMPRL) at Stobhill Hospital in Glasgow, were assessed. This information was linked to death certification records by the General Register Office for Scotland. As with the work of Harboe et al. (2009), deaths from any cause within 30 days of the sample being submitted to the SMPRL were classed as a fatality attributable to IPD. The data were used to determine whether or not there is evidence of an association between serogroup and MLST and mortality.

Data were available on cases of IPD in Scotland from January 1992 to December 2007. Serogroup information was available for all years. Routine MLST determination was not carried out at SMPRL on all samples until 2003. However, MLST analysis was carried out on many invasive isolates collected in 2001 and 2002 as part of a study funded by Wyeth pharmaceuticals and this information was made available for this mortality analysis. Information on the age and sex of the patient was also available. No information was available on co-morbidities of the patients. 68 duplicate samples from patients were removed prior to analysis. These duplicates occurred as certain patients, those with pneumococcal meningitis, had both a CSF and a blood sample issued to the SMPRL.

As the data available for the analysis in this chapter are also population-based and a relatively large amount of information is available, each serogroup and MLST is considered separately for associations with mortality. The primary aim of this analysis was to identify whether or not there was significant evidence of an association between any of the serogroups identified in IPD in Scotland and mortality. The secondary aim was to determine if any MLSTs are associated with mortality. In addition, it was of interest to assess associations between serogroups

and MLSTs and mortality for different age groups.

A 2×2 contingency table was used to display mortality against serogroup. Table 7.1 shows a 2×2 contingency table for serogroup i. In Table 7.1, a is the number of deaths within 30 days of developing IPD attributable to serogroup or MLST i, b is the number who survived more than 30 days of serogroup or MLST i IPD, c is the number of deaths within 30 days of developing IPD attributable to all other serogroups or MLSTs (i.e. all non serogroup or MLST i IPD deaths), and d is the number of survivors of more than 30 days of disease from all other serogroups or MLSTs.

Table 7.1: Example of a contingency table of mortality by serogroup.

Serogroup	**Fatalities**	**Survivals**
Serogroup i	a	b
All other serogroups	c	d

ORs of dying within 30 days from IPD due to particular serogroups or MLSTs compared to dying within 30 days from IPD due to all other serogroups or sequence types were calculated. The OR is calculated as follows:

$$OR = ad/bc. \tag{7.1}$$

If an OR of 1 is obtained, an individual with serogroup or MLST i is as likely to die within 30 days as an individual with IPD and a serogroup or MLST other than i. An OR of greater than 1 can be interpreted as being indicative of an increased probability of death within 30 days of IPD due to the serogroup or MLST invasive disease, whilst an OR of less than 1 is indicative of a reduced probability for the serogroup or MLST to cause death within 30 days. This technique was used in a paper establishing the invasive disease potential of serotypes and MLSTs among children in Oxford, England (Brueggemann et al. 2003). It is possible to use the most prevalent disease-causing serotype as the comparator in analyses of mortality associations (Lexau et al. 2005; Harboe et al. 2009). In this analysis each serogroup or MLST was compared to all others. This approach was used in another serotype mortality association analysis (Martens et al. 2004). However, for the unstratified analyses of serogroup and MLST associations with mortality,

the most common disease-causing serogroup and MLST were considered as the baseline in order to determine whether or not the results from the two approaches differ.

To formally determine whether or not there is evidence that certain serogroups or MLSTs are significantly associated with a greater or reduced risk of fatality, the Fisher's Exact Test was employed. This test is used to establish whether or not there are associations between two categorical variables. The null hypothesis is that there is no association between the two variables and the alternative hypothesis is that there is an association. The Fisher's Exact Test is similar in purpose to the χ^2 Test of Association. The Fisher's Exact Test is more appropriate than the χ^2 Test when the expected count in some cells is small. The χ^2 Test is based on a large sample approximation whilst the Fisher's Exact Test uses exact probabilities from the hypergeometric distribution. The Fisher's Exact Test does not have a formal test statistic or critical value (Simon 2000). However, a p-value for the test may be obtained.

To determine the p-value for the Fisher's Exact Test, let the sum of all entries of Table 7.1 equal n and let the row sums of the table equal x_1 $(= a + b)$ and x_2 $(= c + d)$ respectively, and the column sums equal y_1 $(= a + c)$ and y_2 $(= b + d)$. The conditional probability of obtaining the observed 2×2 contingency table is calculated as follows:

$$P = \frac{x_1!x_2!y_1!y_2!}{n!a!b!c!d!}. \tag{7.2}$$

Following this, all possible values of a, b, c and d to obtain the fixed row and column values x_1, x_2, y_1 and y_2 must be determined, and for each the conditional probability specified in (7.2) must be calculated (Weisstein 1999). The sum of all of these conditional probabilities should equal 1.

The p-value may be calculated by summing all conditional probabilities which are less than or equal to the conditional probability calculated for the observed values of a, b, c and d (Weisstein 1999). The p-value should then be compared to a specified critical level such as 0.05. If the calculated p-value is less than

the critical level then the null hypothesis may be rejected and there is significant evidence to suggest an association between the two categorical variables.

As many serogroups and MLSTs may be rarely observed in IPD, or even if common may be little observed in fatalities, the Fisher's Exact Test is more appropriate to use than the χ^2 Test for this analysis. For the purpose of determining whether or not there is an association between mortality and serogroup or MLST, one of the categorical variables is whether or not the IPD is from serogroup or MLST i, the other is whether or not the patient survived beyond 30 days of the sample being submitted to the SMPRL. As many MLSTs appear infrequently in IPD in Scotland, only those responsible for at least five cases of invasive disease were assessed. The Bonferroni correction factor (6.2) mentioned in the previous chapter, was used to adjust for multiple testing.

The Cochran-Mantel-Haenszel Test was used to carry out age adjustment when testing the association between serogroups or MLSTs and mortality. For this analysis, the continuous variable age was grouped into the following categories: 0-4 years, 5-34 years, 35-49 years, 50-64 years, 65-74 years and 75 years and over. These age groups are the same as those adopted in Chapter 6.

The null hypothesis of the Cochran-Mantel-Haenszel Test is that there is no association between the two categorical variables across all strata whilst the alternative hypothesis is that there is an association between the two categorical variables in at least one of the strata. Here, the strata are the age categories and the two categorical variables are as described for the Fisher's Exact Test. In the Cochran-Mantel-Haenszel Test, there are k 2×2 contingency tables; one for each of the k strata. The Cochran-Mantel-Haenszel Test has a greater power than the Fisher's Exact Test to establish whether or not there are non-random associations between the two categorical variables as individuals are grouped in strata of similar individuals (Agresti 2007). It is expected that by examining age group in the analysis an association may be more effectively identified between serogroup or MLST or both and mortality if the serogroup or MLST (or both) involved in IPD are related to age also. However, the Cochran-Mantel-Haenszel test has low power for detecting associations when the patterns of association for

some strata are in the opposite direction of those displayed by other strata. Thus, a nonsignificant result does not necessarily mean that there is no association but can also indicate that the pattern of association does not have enough strength or consistency to dominate any other pattern. An alternative, potentially more suitable test, which could have been adopted is the Breslow-Day test. However, this test requires a large sample size within each stratum. Thus, limiting its usefulness.

The pooled OR when carrying out the stratified analysis is calculated as follows:

$$OR = \frac{\sum_j \left(\frac{a_j d_j}{n_j}\right)}{\sum_j \left(\frac{b_j c_j}{n_j}\right)}. \tag{7.3}$$

In (7.3), $j = 1, ..., k$ as there are k strata. a, b, c and d have the same definition as in (7.1). The test statistic, χ^2_{CMH}, for the Cochran-Mantel-Haenszel Test defined by Cochran (1954) and Mantel-Haenszel (1959) is found using the following calculation:

$$\chi^2_{\text{CMH}} = \frac{\left(\sum_{j=1}^k a_j - \sum_{j=1}^k x_{1j} y_{1j}/n_j\right)^2}{\left(\sum_{j=1}^k x_{1j} x_{2j} y_{1j} y_{2j}/n_j^2(n_j - 1)\right)}. \tag{7.4}$$

The test statistic, χ^2_{CMH}, defined in (7.4) should approximately follow the $\chi^2(1)$ distribution under the null hypothesis. All quantities defined in (7.4) are as described earlier in this chapter. The Cochran-Mantel-Haenszel Test was also used to investigate age and gender combinations as strata.

Alternative analysis approached which could have been adopted in this analysis include log-linear models. However, the tests described previously were adopted due to the ease with which the results can be interpreted and described to the microbiologists who supplied the data.

All statistical analysis in this chapter was carried out using R Version 2.8.0.

7.4 Results

During the period from January 1992 to December 2007, blood or CSF samples from 5,959 patients with IPD were submitted to the SMPRL for typing. Of these 5,959 patients, 5,119 (85.90%, 95% C.I. (85.02, 86.79)) survived beyond 30 days of the sample submission and 833 (13.95%, 95% C.I. (13.10, 14.80)) died. This fatality rate is comparable to that obtained in other studies of IPD (Ewig et al. 1999; Alanee et al. 2007). There were 7 (0.12%) patients that could not be matched to data from the General Register Office on death certification records.

Figure 7.1 shows the number of cases of IPD by year. The number of cases of IPD appears to be increasing over time, with a minimum of 29 cases of IPD observed in 1992 and a maximum of 697 observed in 2006. A very small number of cases of IPD were recorded in 1992. The number remained fairly low, appearing generally to increase each year, until 2000. The increasing number of observations is likely due to increased surveillance rather than real increases in the number of cases of IPD in Scotland. In 2000, PCV-7 was introduced for routine use in the USA. Therefore, it is likely that the numbers increased markedly due to increased awareness by scientists and medical professionals working with the bacterium which in turn led to increased reporting of results to the reference laboratory. This could potentially have implications on the results obtained in the analysis in this chapter if in the earlier years the more severe cases were reported. However, this is unlikely as 30 day mortality was identified by data linkage following the report of IPD. On average there were approximately 372 cases of IPD per year (median = 292). Omitting 1992, the mean number of cases of IPD per year was 395 (median = 346).

Figure 7.2 shows the proportion of deaths and survivals from IPD each year from 1992 to 2007. From this plot, it appears that the proportion of fatalities attributable to IPD has decreased over the period of study even though the number of cases has been increasing over time as displayed in Figure 7.1. In fact, 1992, although having the lowest number of cases of IPD observed, had the highest proportion of fatalities with 7 of the 29 cases (24%) of IPD resulting in death within 30 days of the blood or CSF sample being submitted to the SMPRL. In 2006, the highest number of cases of IPD was observed. In this year, only 12% of

cases resulted in a fatality attributable to IPD. To test the null hypothesis that the proportions of fatalities remained constant over time, a logistic regression was fitted to the data with the mortality outcome as the response variable and year as the explanatory variable. A p-value of $< 1 \times 10^{-5}$ was obtained for the variable year. Thus, there is significant evidence to reject the null hypothesis that the proportion of fatalities remains constant over time. The coefficient of year is negative (-0.04), indicating a decrease in the odds of fatality with increasing year.

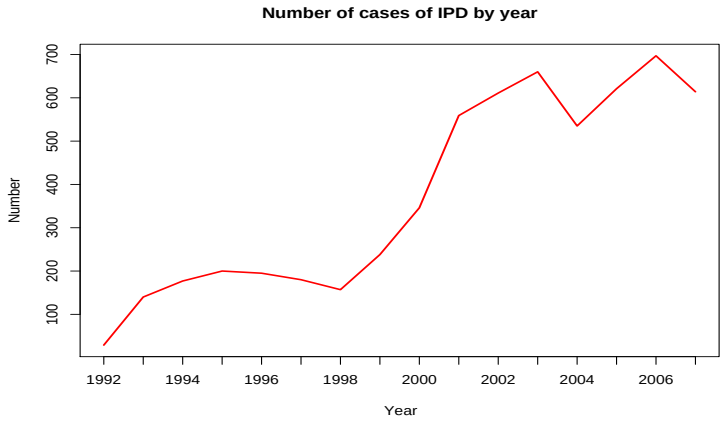

Figure 7.1: Plot of the number of cases of IPD observed in Scotland each year from 1992 to 2007.

The age of patients in the dataset ranged from 0 years to 99 years, with a mean age of approximately 53 years (median = 61 years). 301 patients had missing age information. 3,015 (50.60%) patients in the dataset were male and 2,913 (48.88%) were female. 31 patients had missing gender information. Figure 7.3 shows the proportion of fatalities and survivors of IPD within each gender by six different age groups. This graph shows that a marginally higher proportion of all females who acquired IPD had fatal outcomes than males who acquired IPD within the age groups 0 to 4 years and 35 to 49 years. For the other age groups, the opposite result is true.

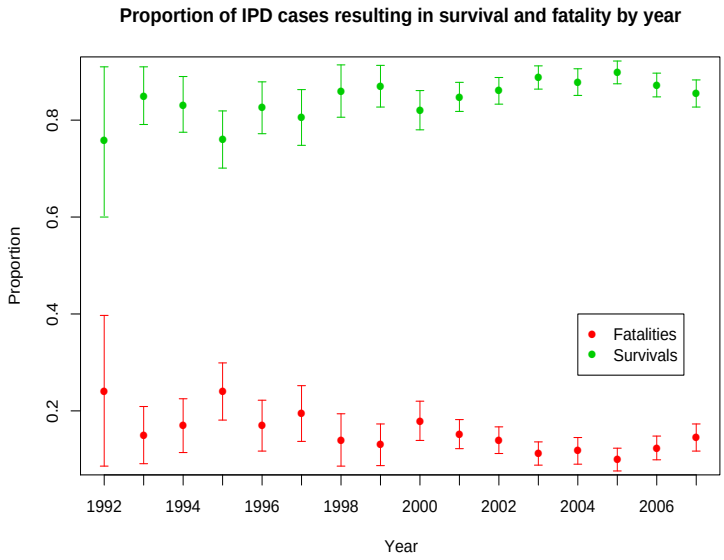

Figure 7.2: Plot of the proportion of deaths and survivals of IPD each year in Scotland from 1992 to 2007 (95% confidence intervals shown).

35 different serogroups were observed in IPD in Scotland between January 1992 and December 2007. Figure 7.4 shows the proportion of fatalities for each of the 10 most common disease-causing serogroups. Serogroup 3 has the highest proportion of deaths from IPD with 85 cases of IPD out of 349 (24%) resulting in death within 30 days of the specimen being submitted to the SMPRL. Serogroups 19 and 23 have the next highest rates of fatality at 18% and 15% respectively. Serogroup 1 has the lowest rate of fatality of the 10 most common IPD causing serogroups, with only 5% of cases of IPD resulting in death attributable to IPD. Serogroup 7 also has a fairly low percentage of IPD fatality at only 8%.

371 MLSTs were identified in cases of IPD between 2001 and 2007. Of the 10 most common disease-causing MLSTs, MLST 180 has the highest observed percentage of IPD attributable fatality at 22%, shown in Figure 7.5. MLST 306 has the lowest at only approximately 3%.

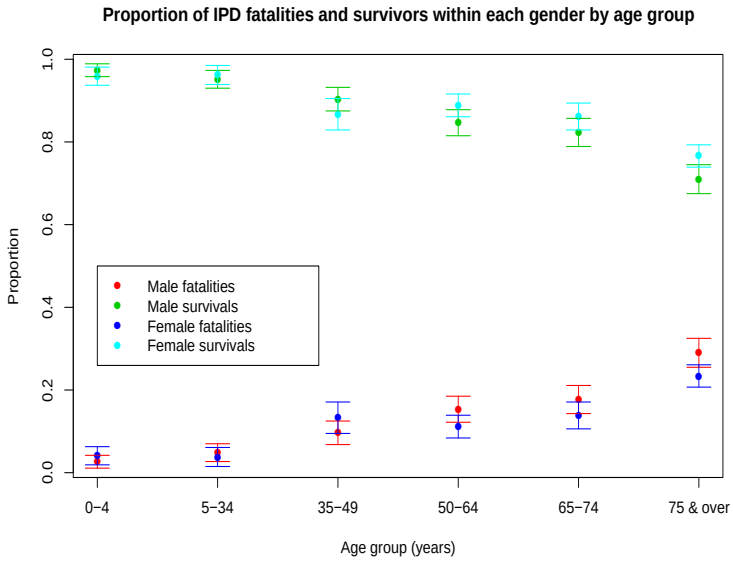

Figure 7.3: Plot of the proportion of deaths and survivals of IPD in Scotland for each sex by six age groups (95% confidence intervals shown).

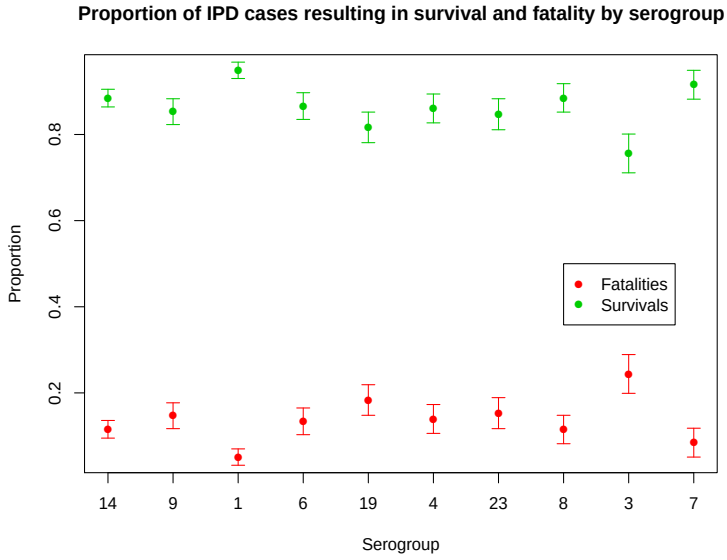

Figure 7.4: Plot of the proportion of deaths and survivals of IPD in Scotland for each of the 10 most common serogroups (95% confidence intervals shown).

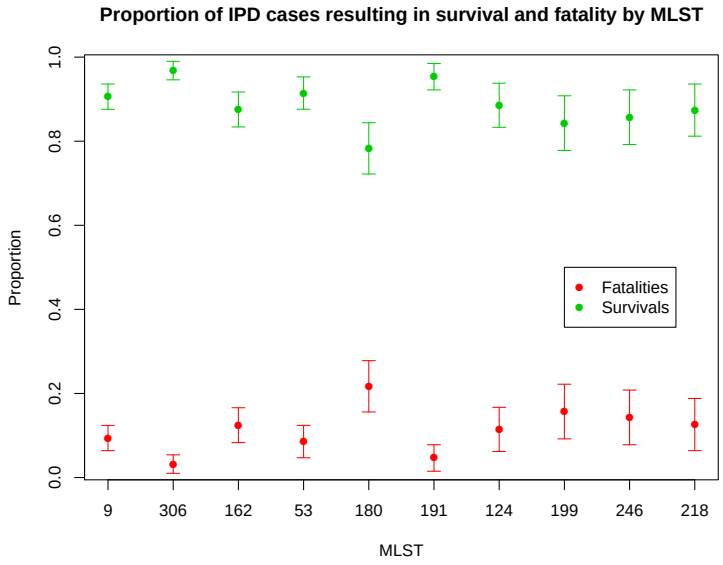

Figure 7.5: Plot of the proportion of deaths and survivals of IPD in Scotland for each of the 10 most common MLSTs (95% confidence intervals shown).

7.4.1 Tests of association

Serogroup analysis

Table 7.2 shows the results from the Fisher's Exact Test of association between each of the 20 most common serogroups found in IPD and mortality.

In Table 7.2, the OR and Bonferroni adjusted 95% confidence interval for the risk of fatality of each of the 20 most common serogroups found in IPD in Scotland are shown. The p-value for the Fisher's Exact Test of association between mortality and serogroup is also displayed. Using the Bonferroni correction factor to adjust for multiple testing, a p-value of less than 0.0018 was used (0.05/28 as 28 tests were carried out for each of the 28 serogroups observed in at least 5 cases of IPD to be assessed at the 5% significance level). From examination of Table 7.2, there is significant evidence of an association between mortality and the serogroups 1, 3, and 16. Serogroup 11 is borderline significant with a p-value slightly larger than the critical value 0.0018. Serogroups 3, 11 and 16 all display ORs greater than 1. Thus, there is an increased risk of a fatal outcome on obtaining IPD

Table 7.2: Results from the Fisher's Exact Test of association between mortality and serogroup.

Serogroup	**Fatalities**	**Total**	**OR**	**Adjusted 95% C.I.**	***p*-value**
14	106	919	0.78	$(0.54, 1.09)$	0.02
9	79	537	1.07	$(0.70, 1.59)$	0.60
1	26	513	0.31	$(0.15, 0.57)$	<0.002
6	62	463	0.95	$(0.59, 1.47)$	0.78
19	83	453	1.43	$(0.94, 2.12)$	0.01
4	57	410	1.00	$(0.61, 1.57)$	1.00
23	59	385	1.13	$(0.69, 1.77)$	0.40
8	41	357	0.79	$(0.44, 1.33)$	0.18
3	85	349	2.10	$(1.37, 3.16)$	<0.002
7	23	272	0.56	$(0.26, 1.08)$	0.01
18	17	213	0.53	$(0.21, 1.12)$	0.01
12	30	186	1.20	$(0.59, 2.21)$	0.39
22	26	170	1.12	$(0.53, 2.15)$	0.57
20	9	106	0.57	$(0.15, 1.56)$	0.12
11	27	106	2.15	$(0.99, 4.32)$	<0.002
33	11	76	1.05	$(0.31, 2.76)$	0.87
15	12	72	1.24	$(0.38, 3.21)$	0.49
10	7	44	1.17	$(0.23, 3.95)$	0.66
16	13	34	3.87	$(1.11, 12.15)$	<0.002
31	10	34	2.59	$(0.65, 8.45)$	0.02

attributable to any of those 3 serogroups. Serogroup 1 has an OR of less than 1. Therefore, there is a reduced risk of fatality on obtaining IPD attributable to this serogroup.

In addition, if tested singly at the 5% significance level, serogroups 7, 14, 18, 19 and 31 have a significant association with 30-day mortality. Serogroup 7, 14 and 18 are associated with a reduced risk of a fatal outcome whilst 19 and 31 are associated with an increased risk.

Using serogroup 14, the most common disease-causing serogroup in Scotland, as the baseline for comparison in the analysis serogroups 1, 3, 11 and 16 were identified to have significant associations with mortality, as before. Serogroup 1 was found to have a lower risk of fatality than serogroup 14 whilst 3, 11 and 16

had a higher risk. In addition to these four serogroups, serogroup 19 was found to be associated with mortality when considering serogroup 14 as the baseline, with serogroup 19 identified to have a higher risk of fatality than serogroup 14.

The Cochran-Mantel-Haenszel Test was used to determine whether or not there is an association between serogroup and mortality adjusting for the age strata. Once again, the Bonferroni correction factor was used to adjust for multiple testing for the 28 serogroups. Table 7.3 shows the results from the Cochran-Mantel-Haenszel Test for the 20 most common serogroups found in IPD.

Table 7.3: Results from the Cochran-Mantel-Haenszel Test of association between mortality and serogroup by age group.

Serogroup	**OR**	**Adjusted 95% C.I.**	***p*-value**
14	0.79	$(0.55, 1.13)$	0.05
9	1.01	$(0.67, 1.53)$	0.98
1	0.41	$(0.21, 0.80)$	$< 1.00 \times 10^{-4}$
6	0.96	$(0.60, 1.52)$	0.82
19	1.57	$(1.03, 2.39)$	$< 1.00 \times 10^{-3}$
4	0.95	$(0.58, 1.54)$	0.77
23	0.99	$(0.62, 1.58)$	0.99
8	0.72	$(0.42, 1.25)$	0.07
3	1.72	$(1.12, 2.62)$	$< 1.00 \times 10^{-4}$
7	0.65	$(0.32, 1.32)$	0.07
18	0.68	$(0.30, 1.53)$	0.16
12	1.25	$(0.65, 2.42)$	0.34
22	0.97	$(0.49, 1.92)$	0.96
20	0.45	$(0.15, 1.37)$	0.03
11	1.98	$(0.95, 4.12)$	< 0.01
33	0.87	$(0.31, 2.48)$	0.80
15	1.22	$(0.43, 3.45)$	0.68
10	1.23	$(0.33, 4.60)$	0.79
16	3.40	$(1.06, 10.89)$	$< 2.00 \times 10^{-3}$
31	2.64	$(0.76, 9.10)$	0.02

On examination of the p-values for the Cochran-Mantel-Haenszel Test shown in Table 7.3, there is evidence of an association between mortality and serogroups 1, 3, 19 (p-value ≈ 0.0009) and 16 (p-value ≈ 0.0011) across at least some of the age strata. Serogroup 1 has an OR lower than 1. Therefore, there is evidence to

suggest that patients acquiring serogroup 1 IPD have a reduced odds of fatality compared to IPD from all other serogroups. Serogroups 3, 16 and 19 have ORs greater than 1 and thus there is evidence to suggest that there is an increased risk of death from IPD if these serogroups are acquired.

On carrying out a Cochran-Mantel-Haenszel Test using combinations of age group and gender as the strata, significant associations were discovered between mortality and serogroup for serogroups 1, 3, 16 and 19. Serogroups 3, 16 and 19 are associated with an increased risk of fatality compared to all other serogroups across all age groups and both genders. Serogroup 1 has a reduced odds of fatality.

The results of the unadjusted and adjusted analyses are similar, with serogroups 1, 3 and 16 found to be significantly associated with mortality in all tests carried out, with serogroup 1 associated with a reduced risk of fatality and serogroups 3 and 16 with an increased risk. However, there are differences in the unadjusted and adjusted analyses. Serogroup 11 is borderline significant in an unadjusted test of association with mortality but is not significant after adjusting for age or age and gender. In addition, serogroup 19 which is not statistically significantly associated with mortality prior to adjustment when comparing serogroup 19 to all other serogroups is significant when age or both age and gender are taken into account. On further assessment of serogroup 19 case-fatality it can be observed that the CFR for serogroup 19 was comparable to the overall fatality rate for the 0 to 4 years and 5 to 34 years age groups. However, the CFR for serogroup 19 was higher than the overall CFR for all other age groups. This is shown in Table 7.4.

Serogroup 11 has an overall CFR of 25.5%. However, amongst those aged under 5 years only one case of serogroup 11 IPD is recorded and this does not result in fatality; only 8 cases of IPD amongst those aged 5 to 34 years are attributable to serogroup 11 and all of these patients survive. Amongst other age groups there is a relatively high CFR for serogroup 11 IPD. For example, for those aged 75 years and over, 46.4% of serogroup 11 IPD result in a fatal outcome. Thus, after adjustment for age group this serogroup does not appear significantly associated with mortality. It appears that in this situation the Cochran-Mantel-Haenszel

Age group	0 to 4 yrs	5 to 34 yrs	35 to 49 yrs	50 to 64 yrs	65 to 74 yrs
Serogroup 19 case-fatality rate (%)	3.3	4.3	23.8	18.9	25.8
Overall IPD case-fatality rate (%)	3.2	4.6	11.1	13.6	15.5

Age group	75 yrs and over
Serogroup 19 case-fatality rate	30.7
Overall IPD case-fatality rate	25.0

Table 7.4: IPD fatality rate and serogroup 19 case-fatality rate for all age groups.

test may not be the most appropriate analysis to use in this situation as this test has low power for detecting an association in which the patterns of association for some strata are in the opposite direction of those displayed by other strata. Thus, a nonsignificant result can also indicate that the pattern of association does not have enough strength or consistency to dominate any other pattern.

MLST analysis

Table 7.5 shows the results from the Fisher's Exact Test of association between each of the 20 most common IPD causing MLSTs in Scotland and mortality. A p-value of less than 0.0006 is required for significant evidence of an association between any MLST and mortality as an adjustment for multiple testing has to be made since tests were carried out on the 79 MLSTs observed in at least five cases of IPD in Scotland between 2001 and 2007. From examination of Table 7.5, there is only significant evidence of an association between MLSTs 306 and 180 and mortality. MLST 306 has an OR of less than 1. Therefore, there is evidence to suggest that there is a reduced risk of fatality within 30 days of a case of IPD from MLST 306 IPD than from all other MLST IPD. MLST 180 has an OR greater than 1. Therefore, this suggests that there is an increased risk of fatality within 30 days of IPD from MLST 180 than from all other MLSTs. If considered at the 5% significance level without adjustment for multiple comparisons, MLST 191 is significantly associated with a reduced risk of fatality within 30 days of IPD.

MLST 180 is commonly associated with serogroup/serotype 3 IPD. Of all the MLST 180 isolated identified in this study, 98.2% were associated with serogroup/ serotype 3. On the other hand, MLST 306 is commonly associated with serogroup/ serotype 1 with 96.4% of all MLST 306 isolates identified to be serogroup/serotype 1.

The unstratified analysis assessing associations with MLST and mortality was also carried out with MLST 9, the most common MLST found in disease in Scotland, as the baseline group for comparisons. MLST 180 was the only MLST found to have an association with mortality in this analysis, with a higher risk of fatality attributable to MLST 180 disease than to MLST 9 disease.

Table 7.5: Results from the Fisher's Exact Test of association between mortality and MLST.

MLST	**Fatalities**	**Total**	**OR**	**Adjusted 95% C.I.**	***p*-value**
9	34	361	0.72	$(0.35, 1.33)$	0.08
306	8	250	0.22	$(0.04, 0.67)$	$< 1.00\times10^{-6}$
162	30	241	1.01	$(0.47, 1.96)$	0.92
53	17	199	0.65	$(0.23, 1.49)$	0.10
180	38	175	2.05	$(1.01, 3.90)$	$< 1.00\times10^{-4}$
191	8	171	0.34	$(0.07, 1.03)$	$< 8.00\times10^{-4}$
124	16	140	0.91	$(0.31, 2.19)$	0.90
199	19	121	1.34	$(0.49, 3.08)$	0.26
246	16	112	1.19	$(0.40, 2.91)$	0.56
218	14	111	1.03	$(0.32, 2.61)$	0.88
311	15	99	1.28	$(0.41, 3.22)$	0.36
227	6	97	0.46	$(0.07, 1.66)$	0.06
433	10	80	1.01	$(0.24, 3.01)$	1.00
205	7	68	0.81	$(0.14, 2.83)$	0.71
176	11	64	1.48	$(0.37, 4.41)$	0.25
206	8	63	1.03	$(0.20, 3.45)$	0.85
113	6	60	0.79	$(0.11, 2.98)$	0.70
62	9	53	1.46	$(0.31, 4.81)$	0.29
36	9	52	1.49	$(0.31, 4.94)$	0.29
235	7	49	1.19	$(0.19, 4.35)$	0.66

The Cochran-Mantel-Haenszel Test was used to adjust for age groups in the analysis of the association between MLST and mortality. Table 7.6 shows the results for the 20 most common disease-causing MLSTs. This table shows that none of the 20 most common disease-causing MLSTs have a significant association with mortality across all age strata.

7.4.2 Mortality analysis prior to and following PCV-7 use

To determine whether or not the introduction of PCV-7 had an impact on the serogroups, serotypes and MLSTs associated with mortality, a pre and post vaccine analysis was carried out. As no information was available regarding the week of the case of disease, it was not possible to separate the cases of disease in 2006 into pre-vaccine and post-vaccine periods. 2006 was included in the pre-vaccine

Table 7.6: Results from the Cochran-Mantel-Haenszel Test of association between mortality and MLST by age group.

MLST	**OR**	**Adjusted 95% C.I.**	***p*-value**
9	0.75	$(0.39, 1.46)$	0.17
306	0.35	$(0.10, 1.26)$	< 0.01
162	0.95	$(0.47, 1.95)$	0.90
53	0.56	$(0.22, 1.41)$	0.04
180	1.67	$(0.85, 3.28)$	0.01
191	0.40	$(0.11, 1.47)$	0.02
124	0.76	$(0.29, 2.00)$	0.41
199	1.39	$(0.56, 3.44)$	0.27
246	1.27	$(0.48, 3.32)$	0.49
218	1.05	$(0.38, 2.91)$	0.88
311	1.12	$(0.41, 3.07)$	0.81
227	0.69	$(0.15, 3.08)$	0.50
433	0.81	$(0.25, 2.64)$	0.64
205	0.86	$(0.21, 3.61)$	0.88
176	1.72	$(0.49, 5.96)$	0.20
206	1.04	$(0.27, 4.05)$	0.92
113	0.98	$(0.21, 4.55)$	0.96
62	1.33	$(0.35, 4.97)$	0.61
36	1.14	$(0.31, 4.22)$	0.88
235	1.02	$(0.25, 4.21)$	0.97

analysis as it is unlikely that PCV-7 would have a great impact on disease in the first three months following routine implementation. The post-vaccine analysis only includes the year 2007.

The results for the serogroup associations in the pre-vaccine period were the same as those for the whole period of data with serogroups 1, 3, 11 and 16 found to be significantly associated with mortality when compared to all other serogroups. As before, serogroups 3 (OR 2.10, Bonferroni adjusted 95% C.I. (1.33, 2.04)), 11 (OR 2.56, C.I. (1.16, 2.56)) and 16 (OR 4.82, 95% C.I. (1.33, 16.34)) are associated with an increased risk of fatality, serogroup 1 (OR 0.34, C.I. (0.16, 0.64)) with a reduced risk. However, in the post-vaccine period, only serogroup 1 was found to have a significant association with mortality (OR 0.08, Bonferroni adjusted 95% C.I. (0, 0.84)). Serogroup 3 and 11 are not found to be associated with an

increased risk of fatality in the post-vaccine period and serogroup 16 is not found amongst the most common disease-causing serogroups in 2007. However, the post-vaccine period consists of only one year of IPD cases. Thus, there is much less power to detect an association with 30-day mortality in the post-vaccine period than in the pre-vaccine period.

In both the pre and post-vaccine periods, only MLST 306 was found to be significantly associated with mortality with an OR of 0.29 (Bonferroni adjusted 95% C.I. (0.06, 0.88)) in the pre-vaccine period and an OR of 0 (95% C.I. (0, 0.85)) in the post-vaccine period. Therefore, there are no differences observed in the MLST associations in the pre and post-vaccine periods.

7.5 Conclusions

In the analysis carried out in this chapter, it has been shown that the proportion of fatalities within 30 days of report of IPD has decreased with increasing year from 1992 to 2007. However, the number of cases of IPD reported increased substantially between 1992 and 2001. This is likely to be attributable to increased reporting of IPD cases in Scotland rather than a true increase in IPD as PCV-7 was introduced in the USA in 2000. Thus, this is likely to have resulted in increased interest in IPD leading to increased surveillance.

The highest IPD CFR is observed in those aged 75 years and over and the lowest amongst those aged 0 to 4 years and 5 to 34 years of age. Serogroup 14 was the most common serogroup found in cases of IPD. This has been observed in other studies (Balakrishnan et al. 2000; Henriques et al. 2000; Sjöström et al. 2006), whilst another study documented higher numbers of serogroup 1 IPD (Harboe et al. 2009). Serogroup 3 has the highest CFR, as observed elsewhere (Henriques et al. 2000). Other studies document this serotype amongst the three with the highest CFRs (Rückinger et al. 2009; Lexau et al. 2005). Serogroups 19 and 23 had the next highest CFRs. These serogroups have also been linked with high CFRs in other studies (Henriques et al. 2000; Lexau et al. 2005). The lowest overall CFR observed in this study was for serogroup 1 IPD. This concurs with low CFRs for serotype 1 observed elsewhere (Jansen et al. 2009). However,

serotype 1 was observed to have among the highest CFRs for those aged under 5 years in another study (Harboe et al. 2009). However, this result was not found to be statistically significant. MLST 180, identified to be commonly associated with serogroup/serotype 3 IPD in Scotland, had the highest CFR of all MLSTs, whilst MLST 306, commonly associated with serogroup/serotype 1 IPD, had the lowest.

The analysis carried out in this chapter has shown significant associations between serogroup 1 and mortality, with those with serogroup 1 IPD at a reduced risk of fatality. This concurs with another study of serotype mortality association (Martens et al. 2004). Serogroups 3 and 16 were identified to be significantly associated with an increased risk of fatality compared to all other serogroups. The result for serogroup 3 is in agreement with other studies (Henriques et al. 2000; Martens et al. 2004), whilst the result for serogroup 16 concurs with results obtained in another study (Harboe et al. 2009). Serogroup 11 was found to be borderline significant, with results suggesting that this serogroup is associated with an increased risk of fatality. This is also in agreement with the Harboe et al. study in which 11A was identified to be significantly associated with an increased risk of mortality. Lexau et al. (2005) also document 11A to have significantly higher CFRs than the comparator serotype 14.

After age-adjustment, serogroup 19 was found to be significantly associated with an increased risk of 30-day mortality compared to all other serogroups. 19F was also identified to be associated with highly increased mortality in the large Danish study carried out by Harboe et al. (2009). McKenzie et al. (2000) reported 19A to be associated with high mortality in their small study in Grampian, Scotland. However, due to the small sample size, statistically significant results could not be obtained. In contrast, Henriques et al. (2000) report that 19A seems to be associated with lower risks of severe disease and death but report 19F to have increased CFRs.

Regarding the MLST associations with mortality, both MLST 180 and 306 were identified to be associated with overall 30-day mortality from IPD in Scotland. MLST 180, which is primarily associated with serogroup 3 in this study, was

found to have an increased risk of fatality, whilst MLST 306, associated with serogroup 1, has a reduced risk of fatality.

No other study of pneumococcal mortality identified assessed the association between MLSTs and fatal outcome. Thus, the analysis carried out in this chapter contributes to the understanding of the importance of MLST in the disease outcome of pneumococci. From the analysis carried out in this chapter it cannot be ascertained whether or not the serogroup or MLST is more important in determining disease outcome. A simultaneous evaluation of the involvement of both serogroup and MLST in disease outcome would require the use of log-linear modelling. However, as MLST information is not available for all years this would reduce the power of the analysis and would mean that the associations between serogroup and 30-day mortality identified in this chapter may not be uncovered. Thus, the analysis approach adopted in this chapter is reasonable.

The main limitation of this analysis is that data were not available on co-morbidities of those with IPD in Scotland other than age. In an international study of associations between serotypes and IPD outcome (Alanee et al. 2007), which controlled for co-morbidities, it was concluded that factors specific to the individual patient rather than the serotype of the bacterium are of greater importance in determining disease outcome. No association was identified between particular serotypes and outcome after controlling for co-morbidities. In a study of IPD fatality in the Netherlands in which co-morbidities were also studied (Jansen et al. 2009), the serogroups 1, 5 and 7 which are known to have high IPD potential in children (Brueggemann et al. 2004) were observed to affect adults who were fairly healthy whilst other serogroups with low to moderate IPD potential were more common amongst those with underlying health conditions or are of an older age. By examining age in this study, one of the potential confounders has been assessed. However, future studies of IPD mortality should consider other potential health confounders to determine whether or not serogroup 1 is consistently significantly associated with lower odds of fatality. This may be the case, as in another study which did include co-morbidities (Martens et al. 2004).

Chapter 8

Analysis of factors affecting the uptake and timing of PCV-7

8.1 Introduction

In this chapter, single-level and multi-level modelling techniques are adopted to assess the uptake of PCV-7 in Scotland since its introduction in the childhood immunisation schedule in September 2006. Models are used to explore not only the factors affecting whether or not the third dose, or booster dose, is received by children but also whether or not the vaccine is administered to children later than scheduled.

PCV-7 is currently administered in a three dose schedule in the UK. The first dose is administered at 2 months of age, the second dose at 4 months of age and the third, booster vaccination, at 13 months of age. The third dose coincides with the timing of the MMR vaccination and is administered one month later than the Hib/Men C booster vaccination (NHS Immunisation 2008). It is of interest to assess what the effect of receiving the Hib/Men C booster late has on the timing of the PCV-7 booster. It is believed that delays in vaccine uptake have occurred in Scotland due to the addition of the PCV-7 and the Hib/Men C booster to the immunisation schedule as many more vaccines now need to be administered to children within the first two years of life.

Initially, univariate response models are created to look at three different response variables. The first model considers whether or not the PCV-7 booster is received and this variable involves all children for whom information is available. The second model addresses whether or not the vaccine is administered late. This variable only considers a subset of the data: those children who received the vaccine. Although, those who have not received the vaccine could potentially be administered the vaccine at a later date and thus could be considered as 'late' for this response variable, it is uncertain whether or not they will ever receive the vaccine. Finally, the third variable considered is the number of months late the PCV-7 booster is administered.

Further models will then be presented which combine the response variables considered in the univariate analyses. The first of these involves combining the two binary response variables which give information about whether or not the vaccine has been administered and whether or not it has been administered late. The second model looks at the addition of categories based on the continuous response variable for the number of months late the vaccine has been administered to this multivariate response model.

Data are available on vaccine uptake in Scotland from quarter 3 of 2004 to quarter 3 of 2008. The analysis in this chapter involves data on Scottish children born in quarters 3 and 4 of 2006 as PCV-7 was only introduced for routine vaccination in September 2006 and the analysis focuses only on those eligible to receive the vaccine according to the three dose schedule, excluding all involved in any catch-up campaign, and who have reached the age at which the PCV-7 booster and Hib/Men C booster should be administered. No information collected in 2007 was included in the analysis in order to have cohorts with sufficiently long follow-up time to have an idea of how late the vaccines are being administered. The dataset includes information on all vaccinations administered in the childhood vaccination schedule in Scotland which includes MMR, diphtheria (three dose schedule), tetanus (three dose schedule), pertussis (three dose schedule), polio (three dose schedule), Hib B (three dose schedule), PCV-7 (three dose schedule), Meningitis C (three dose schedule) and the Hib/Men C booster. Information on whether or not each child has received each vaccine is included, as well as the date

at which the vaccine was administered. *Date of birth*, *Health board of residence*, *Gender* and *Postcode sector* are also included where the postcode sector is the first five entries of a postcode, where the fourth entry could potentially be a space (e.g. G12 8). In addition, *Deprivation quintiles* (1=affluent to 5=deprived) and *Deprivation deciles* (1=affluent to 10=deprived) are recorded in the dataset. Only the Hib/Men C booster and the PCV-7 booster are considered in the modelling in this chapter as it is important that children receive the PCV-7 booster in order to obtain full protection against pneumococcal disease from the vaccine serotypes. The Hib/Men C booster is considered since it is administered one month prior to the PCV-7 booster. Thus, it is likely that any delays in uptake of this vaccine will result in delays in uptake of the PCV-7 booster.

In order to obtain postcode level variables to be considered in the models, 2001 Census data were downloaded from CASWEB, the web interface to census aggregate outputs and digital boundary data (Census Dissemination Unit 2001) and linked to the vaccine uptake data by postcode sector. The postcode sector variables selected for consideration are *Percentage of individuals aged 0 to 4 years* ('Aged0-4'), *Percentage of households with no car* ('NoCar'), *Percentage born other EU* ('OtherEU') (i.e. the percentage born in European Union countries other than the United Kingdom and Republic of Ireland), *Percentage of individuals who are not of white race* ('NotWhite'), *Percentage aged 16-74 years with no qualifications* ('NoQual'), *Percentage aged 16-74 years with level 4 qualifications* ('Level4'), *Percentage aged 16-74 years who work as large employers* ('LargeEmp'), *Percentage aged 16-74 years who work in routine employment* ('Routine'), *Percentage aged 16-74 years who are unemployed* ('Unemployed') and, finally, *Percentage aged 16-74 years who work in agriculture* ('Agriculture'). The variables 'NoCar', 'HighEmp', 'Routine' and 'Unemployed' were selected as measures of income to determine whether or not income has an effect on vaccine uptake; 'OtherEU' and 'NotWhite' were selected to examine whether or not nationality and race have an impact on the timing of the vaccine. The variables 'NoQual' and 'Level4' provide information about the educational status of individuals in each postcode sector as this may have an impact on vaccine uptake. These two particular variables were selected as they reflect the two extremes for which data are available. Those with no qualifications have the lowest educa-

tional attainment with no high school qualifications such as 'O' Grades, Standard Grades or Scottish Vocational Qualifications. Those with 'Level 4' qualifications have the highest level of attainment with either a first degree, higher degree or a professional qualification. The variable 'Agriculture' is included to identify those who work in rural areas as opposed to urban areas which may affect the timing of vaccinations.

8.2 Uptake of PCV-7 and other vaccines worldwide

8.2.1 Uptake of infant immunisations

In this section, published studies of childhood vaccination uptake are considered. As discussed in the introductory chapter of this thesis, the heptavalent PCV was first introduced for routine use in children in the USA in 2000. Thus, there are not many published studies involving factors affecting routine uptake of PCV-7. Available studies focus on associations with the vaccine impact on invasive disease and physician opinions on administering the new vaccine (Abuelreish et al. 2007; Davis et al. 2003), whilst for the analysis in this chapter interest lies in examining factors which affect uptake and timing. Unlike Scotland, the USA does not have a national health service and thus published studies of PCV-7 uptake focus on financial restrictions associated with vaccine administration (Stokley et al. 2006) which are not relevant to the analysis carried out in this chapter. In addition, unlike Scotland, problems arose in vaccine administration in the USA due to vaccine shortages in 2001 and 2004 which have been taken into account in studies of routine PCV-7 uptake (Smith et al. 2007).

A 2008 USA study of routine PCV-7 uptake is relevant to the analysis undertaken in this chapter, acknowledging that the vaccine shortages were likely to have affected the uptake of PCV-7 but focussing on child, mother and household characteristics affecting uptake (Nuorti et al. 2008). Most notably, the results

from the logistic regression analysis of up-to-date (UTD) and age-appropriate vaccination identified significant effects associated with race, ethnicity, vaccine provider and household income, with lower age-appropriate vaccinations linked to black race, Hispanic ethnicity, public health immunisation providers and low household incomes. Other variables considered in the analysis were mother's educational and marital status, number of children in the household, number of health care providers and place of residence (metropolitan central city, metropolitan non-central city and non-metropolitan area).

As few published studies of routine uptake of PCV-7 exist, published studies of uptake of other childhood immunisations were assessed to identify factors associated with non-receipt or delay in vaccine uptake. Table 8.1 shows a summary of the studies considered. These represent a sample of vaccine uptake studies to demonstrate the type of analyses carried out. The studies consider a variety of factors impacting on childhood immunisations, ranging from problems associated with adverse publicity for vaccines (Friederichs et al. 2006), changes to schedules (Cameron et al. 2007) to financial issues (Stokley et al. 2006; Zimmerman et al. 1999) and socioeconomic factors affecting uptake (Nuorti et al. 2008).

Examination of Table 8.1 shows that most published childhood immunisation studies considered involved the use of logistic regression in the statistical analysis. In general, logistic regression was adopted to assess UTD status of vaccines. This is defined to be the proportion of children who have received appropriate vaccinations by a specified age or during a certain age range (Rodewald et al. 1999). UTD analysis has been criticised as it does not provide information about delays in vaccination and delays may help explain persistence of infections (Akmatov et al. 2008). Thus, survival analysis has been used in various studies to assess vaccine uptake as this approach allows time-to-event to be assessed and therefore delays in vaccinations can be examined (Dayan et al. 2006; Akmatov et al. 2008; Clark and Sanderson 2009). Other studies look at age-appropriate vaccination; Dombkowski et al. (2002) state that the "distinction between age-appropriate vaccination and up-to-date vaccination status is important, since children with lengthier vaccination delays experience longer periods of increased susceptibility than those with shorter or no vaccination delays".

Year	Author	Country[a]	Vaccine[b]	Analysis[c]
1993	Bobo et al.	USA	DTP, OPV, MMR	Logistic regression
1994	Lieu et al.	USA	MMR	Logistic regression
1999	Zimmerman et al.	USA	DTP, MMR	Hierarchical linear modelling
2004	Reading et al.	UK	MMR, Pertussis	Logistic regression
2004	Dombkowski et al.	USA	DTP, Polio, MMR	Logistic regression
2004	Guttman et al.	Canada	DTP, HIB, MMR	Multi-level logistic regression
2004	Bardenheier et al.	USA	DTP, Polio, HIB, Hep C	Logistic regression
2005	Fiks et al.	USA	DTP, Polio, MMR, HIB, Hep B	Logistic regression
2005	Steyer et al.	USA	DTP, Polio, Varicella, Hep B, MMR, HIB	Logistic regression
2006	Friederichs et al.	UK	MMR	Logistic regression
2006	Dayan et al.	Argentina	DTP, MCV, Hep B	Survival analysis
2006	Ozcirpici et al.	Turkey	BCG, DTP, Polio, Hep B, Tetanus	Logistic regression
2006	Stokley et al.	USA	PCV-7	Logistic regression
2006	Torun et al.	Turkey	BCG, Hep B, OPV, DTP, MCV	Logistic regression
2007	Datar et al.	India	Polio, non-polio	Multinomial regression
2007	Cameron et al.	UK	MMR, DTP, HIB, Polio	KDE, linear regression
2008	Akmatov et al.	CIS	DTP	Kaplan-Meier
2008	Nuorti et al.	USA	PCV-7	Logistic regression
2009	Clark et al.	45 countries	BCG, DTP, MCV	Survival analysis

Table 8.1: Studies of infant immunisation uptake.

[a]CIS= Commonwealth of Independent States countries: Armenia, Kazakhstan, Kyrgyzstan and Uzbekistan.

[b]DTP=diptheria, tetanus and pertussis; OPV=oral poliovirus; MMR=measles, mumps and rubella; HIB=haemophilus influenzae type B; Hep B=hepatitis type B; BCG=bacille Calmette-Guérin; MCV= measles containing vaccine.

[c]KDE=Kernel density estimate.

Another statistical technique, a technique used in the analysis in this chapter, used in the studies assessed is multinomial logistic regression (Akmatov et al. 2008; Datar et al. 2007). Akmatov et al. consider whether the vaccine was not received, received on time or delayed. Results from this analysis suggest that differences between countries are of greater importance in the risk of delayed vaccination than any individual risk factor amongst those considered: place of residence (city, town, countryside), marital status of mother (currently married, currently unmarried), education of mother ($<$ 10 years, 10-11 years, $>$ 11 years), birth order/family size (1, 2, $>$2). Akmatov et al. (2008) found differences in vaccination delays for urban and rural areas, with children living in cities at a higher risk of having delayed vaccinations. This is stated to be attributable to low socioeconomic status of city inhabitants. Datar et al. (2007) use a multinomial model to determine the factors affecting whether a child has no cover, some cover or full age-appropriate vaccination cover. The authors identified the significant variables of their multinomial models to be the sex of the child, the maternal literacy and whether or not the child belonged to a tribe household. The analysis showed that urban India had a higher vaccine coverage than rural India but that this higher coverage was not universal across the urban areas.

As this chapter contains an analysis of vaccine uptake in Scotland, it is of note to mention that two of the UK studies displayed in Table 8.1 are Scottish studies (Friederichs et al. 2006; Cameron et al. 2007). Cameron et al. consider the uptake of the MMR immunisation and the changes to the DTP, HIB and polio vaccines in the childhood immunisation schedule in Scotland. This analysis differs from other uptake studies assessed as, for MMR, the focus is on visually depicting the age distribution of uptake in children using Kernel Density Estimates (KDEs), described in the Methods section of this chapter, to assess delays in vaccine administration. In addition, regression techniques are used to predict final vaccine uptake figures. KDEs were used to assess differences in the distributions for the health boards (HBs) and for different levels of deprivation. The study showed that uptake of the vaccine differs for the HBs and that there are greater delays for children in the most deprived category than for all other categories. The study by Friederichs et al. is based on assessing the impact of adverse publicity about MMR from 1998 and shows evidence of a slight rise in late uptake, with

late vaccination associated with deprivation. The results suggest that the most affluent either have the vaccine on time or not at all whilst delays are shown for the most deprived.

To briefly summarise the findings in the studies considered, one of the key variables identified in many of the studies shown in Table 8.1 in determining whether or not vaccines are UTD or delayed relates to family size. The larger the number of siblings the greater the risk of delayed vaccination (Lieu et al. 2000; Reading et al. 2004; Dombkowski et al. 2004; Bardenheier et al. 2004; Ozcirpici et al. 2006). In addition, the birth order appears to be important, with first born children having an increased odds of receiving the vaccine without delay (Bobo et al. 1993; Dayan et al. 2006).

Educational levels of the care giver, most often the mother, also features in some of the models. Mothers or fathers with higher levels of education are more likely to have children with UTD vaccines (Bobo et al. 1993; Dombkowski et al. 2004; Ozcirpici et al. 2006; Torun and Bakırcı 2006; Datar et al. 2007). Mother's marital status is also found to be significant, with married mothers less likely to have children with delayed vaccinations (Bardenheier et al. 2004; Akmatov et al. 2008). In addition, older mothers were found to be associated with increased odds of delay (Reading et al. 2004).

Concerning income and financial issues, households with lower income are identified as less likely to have UTD vaccinations (Guttmann et al. 2006), as are those receiving assistance with medical expenses (Dombkowski et al. 2004). Health insurance features as a significant variable in various studies (Bardenheier et al. 2004; Steyer et al. 2005; Dayan et al. 2006; Dombkowski et al. 2004), with no insurance associated with delays.

Finally, considering area level factors, Ozcirpici et al. (2006) show UTD vaccinations and vaccination coverage is lower for rural over urban areas whilst Akmatov et al. state that higher delays are found in cities.

8.2.2 Multi-level modelling of vaccine uptake

Multi-level models have been used in published analyses of vaccine uptake. Only one multi-level analysis identified involved the uptake of a pneumococcal vaccine. However, this analysis did not involve PCV-7 but PPV-23, assessing administration to elderly nursing home residents. The emphasis in this study was comparing single-level logistic regression to multi-level logistic regression to show how results may differ (Bardenheier et al. 2005). Similar, but not identical results were obtained on carrying out single-level and multi-level analyses and Bardenheier et al. state that it is important to consider multi-level analyses where there is some correlation amongst observations as those variables found to be statistically significant using the single-level approach but not in the multi-level approach are likely to be exhibiting effects attributable to residual correlation. In the analysis considered by Bardenheier et al., the authors state that the multi-level modelling approach is preferred as there is apparent between nursing home variation in the outcome measure. However, the authors state that using multi-level analysis has some drawbacks in that it is difficult to interpret the parameter estimates obtained in the model. ORs cannot be interpreted in the same manner as in a single-level logistic regression due to the model structure and thus all that can be stated meaningfully is that certain variables increase or decrease the log odds of the response.

In Vietnam, a multi-level approach was taken by Ali et al. in analysing vaccine uptake data as social and ecological factors which are associated with spatial variation were deemed important to assess (Ali et al. 2007). The study looked at individual-level factors such as age, gender, educational attainment, and household-level factors such as literacy, age and gender of the head of the household, affecting uptake of the typhoid vaccine or the hepatitis A vaccine administered to children of school age in Hue, Vietnam.

A spatial structure was considered in the Ali et al. model as households were grouped together according to their distance from selected spatial reference points. Ali et al. use the statistical method kriging to define the spatial pattern of vaccine uptake for the area considered. In the multi-level analysis, both school and

neighbourhood were considered as levels in the model. As the data are not nested, since students may not necessarily attend a school within the neighbourhood in which they reside, a cross-classified model was used in the analysis. Thus, students are the level one components and both neighbourhood and school are level two components. Ali et al. consider two types of area; low-vaccine coverage areas and high-vaccine coverage areas.

The results from the analysis show significant student level and ecological level factors. Vaccine uptake was higher amongst females and those students with a younger, male or non-literate head of household. Income also affected the uptake, with those students from lower income households displaying higher uptake than those from higher income households. In addition, greater uptake was observed for those at a greater distance to a hospital. The Ali et al. study provides interesting results. However, the authors state that as the study was created to assess vaccine effectiveness and not factors affecting vaccine uptake there are potential drawbacks which include the fact that the uptake was not monitored under normal public health vaccination conditions. Results showing greater uptake by low-income and lower educational status households may be due to the greater levels of promotion of the vaccine and the free availability of the vaccine.

Multi-level modelling techniques were used in two of the childhood immunisation studies listed in Table 8.1. Both Guttman et al. and Zimmerman et al. consider patient and provider levels in their UTD analysis as patients who are under the care of the same physician may have correlated outcomes.

8.3 Methods

The primary aim of this analysis is to determine factors which affect the timing of the uptake of the PCV-7 booster. The hypothesis is that the timing of the subject level variable Hib/Men C booster, scheduled to be administered 1 month prior to the PCV-7 booster, will have an impact on the timing of the PCV-7 booster.

KDEs are used to provide a picture of the age at which the PCV-7 booster and the Hib/Men C booster are received, as used by Cameron et al. in their analysis, described earlier in this chapter, to provide a picture of the uptake of the MMR vaccine. As with the Cameron et al. analysis, the smoothing parameter used for the final graphical display presented in the KDEs is chosen through visual examination, varying the parameter until a smooth distribution can be seen.

KDEs can be used in a similar manner to histograms to determine distributional shapes. However, unlike the histogram, KDEs display the distribution in the form of a smooth curve. A kernel function, w, is used to obtain the smooth shape and the kernel estimator is described as follows:

$$\hat{f}(y) = \frac{1}{nh} \sum_{i=1}^{n} w \left(\frac{y - y_i}{h} \right) \tag{8.1}$$

where y_i are assumed to be independent and identically distributed random variables. In (8.1), h controls the variance of the kernel function w and is known as the bandwidth. By varying h, the smoothness of the estimated distribution is altered. w may be assumed to be any type of symmetric distribution with zero mean (Bowman and Azzalini 1997). Commonly the Gaussian distribution is adopted and in the KDEs shown in this chapter the Gaussian distribution was used.

Single and multi-level modelling techniques are adopted to determine the significance of various factors which may have an effect on the uptake of the PCV-7 booster. Three levels are considered in this modelling. These are the individual level, postcode district level and HB level. The postcode district is the first four entries of the postcode, where the fourth entry could be a space, e.g. G12 or AB10. Three response variables are considered in this chapter. These are the binary response variables *PCV-7 booster received* and *PCV-7 booster administered late*, and the continuous response variable *Number of months late PCV-7 booster is administered*. All individuals who received the PCV-7 booster prior to the age of 13 months are classed as 'not late' in the binary variable *PCV-7 booster administered late*, all others are classed as 'late'. For the continuous response variable, only vaccinations classed as 'late' are considered. Clearly, these

response variables are interdependent. The variable *PCV-7 booster received* is a censored response as for those who have not received the vaccine it cannot be determined whether or not these children will receive the vaccine late or will never receive the vaccine. Children who have not received the PCV-7 booster will have missing observations in both the *PCV-7 booster administered late* and the *Number of months late PCV-7 booster is administered* response variables and all those classed as 'not late' in the *PCV-7 booster administered late* variable will be classed as missing in the *Number of months late PCV-7 booster is administered* variable.

Each of the response variables described will be considered separately in single-level and multi-level models in this chapter. Multi-level modelling techniques combining these response variables in a multivariate response model will then be explored.

Multi-level modelling is a technique commonly adopted in the analysis of social science data as it is able to model complex variability structures involving nested observations (Snijders and Bosker 1999). Examples of possible uses of multi-level modelling include educational studies where pupils are nested within classes or schools, or household studies where individuals are nested within households within areas (Rasbash et al. 2008). In these examples, pupils or individuals are considered as the level 1 units, classes or households as level 2 and schools or areas as the level 3 unit in a multi-level model, where each unit is assumed to have its own component of variation. Modelling in this manner allows for the fact that the correlations of pupils within a class, or individuals within a household, will be greater than those of pupils or individuals between classes or households. In other words, it takes account of the fact that pupils within a class will have greater similarity to one another than to pupils in a different class. Ignoring the variance structure in such data can lead to incorrect conclusions, with insignificant variables identified as significant and vice versa, although it is unusual for significant variables to be identified as insignificant.

Multi-level analysis is adopted in this chapter as it is hypothesised that children living within the same area, postcode district or HB, are likely to have more

similar vaccination uptake patterns than those in different areas. Another way in which this analysis could have been carried out is using a spatial approach, similar to that described by Ali et al. using a statistical technique such as kriging to model spatial patterns in uptake.

All analysis in this chapter was carried out using MLwiN version 2.11 and R version 2.9.1.

8.4 Results

8.4.1 The data

The dataset contains information on 28,672 children who all met the criteria to have been administered the three doses of PCV-7 according to the routine immunisation schedule, having reached the age of 2 months when PCV-7 was introduced for routine use. Data were available up to the end of March 2008. The earliest date of birth for this subset of the data was 1st July 2006 and the latest was 31st December 2006. As the children have different birth dates they will have different lengths of follow up. Thus, in order to analyse vaccine uptake a cut-off of age 22 months was taken as the follow-up time since all children were followed to at least 22 months of age.

13,974 (48.74%) of the children were females, 14,698 (51.26%) males. Table 8.2 shows the number of children eligible for inclusion in this analysis for each of the HBs in Scotland. It can be noted that there are 126 (0.44%) individuals with missing HB information. HBs G (Greater Glasgow) and S (Lothian) contain the largest number of children to be included in the analysis with over 30% of children in the data residing in either of these two locations. HB R (Orkney) contains the fewest with only 0.33% of Scottish children born between 1st July 2006 and 31st December 2006 residing in this region. Shetland (HB Z) and the Western Isles (HB W) also have low percentages at only 0.45% and 0.50% respectively.

In the data considered in this analysis there are 932 postcode sector levels. To reduce the number of levels to be considered in the multi-level modelling, a new postcode district variable was created by selecting the first four elements of the

HB [a]	A	B	C	F	G	H	L	N	R	S	T	V	W
Number	2055	547	2154	2042	4829	1104	3406	2919	95	4540	2144	1638	144
% of total	7.17	1.91	7.51	7.12	16.84	3.85	11.88	10.18	0.33	15.83	7.48	5.71	0.50

HB	Y	Z
Number	801	128
% of total	2.79	0.45

Table 8.2: Number of children born between 1st July 2006 and 31st December 2006 in each Scottish HB.

[a] A=Ayrshire & Arran, B=Borders, C=Argyll & Clyde, F=Fife, G=Greater Glasgow, H=Highland, L=Lanarkshire, N=Grampian, R=Orkney, S=Lothian, T=Tayside, V=Forth Valley, W=Western Isles, Y=Dumfries & Galloway, Z=Shetland.

postcode sector character, where the fourth element could be a space. These postcode districts were matched to a list of possible postcode districts to check for misclassifications. 30 mismatches were found. Of these, 28 were missing entries. The other 2 misclassifications were districts "ML55" and "L8" which were recoded as missing. 409 postcode district levels remained for inclusion in the analysis. One of the postcode districts was for Cumbria. Thus, the individual with this postcode district was excluded from the dataset as only Scottish data are considered in this analysis. Postcode district information was used to identify some of the missing HB levels.

As mentioned previously, the dataset contains information about two deprivation categories: *Deprivation quintiles* and *Deprivation deciles*. Only *Deprivation quintiles* is considered in this chapter. The deprivation quintiles are taken from the Scottish Index of Multiple Deprivation (SIMD). The SIMD was created in order to have an area measure of relative deprivation for different data zones in Scotland (The Scottish Government 2009).

Table 8.3: Number of children in each deprivation quintile.

Deprivation quintile	1	2	3	4	5
Number of children	5050	5428	5540	5805	6723
% of total	17.61	18.93	19.32	20.25	23.45

Table 8.3 shows the number and percentage of children that fall into each of the deprivation quintiles. The percentage increases slightly with increasing deprivation, with fewest children found in the affluent category and highest found in the most deprived category. As with the HB information, there are 126 children with missing deprivation quintiles information. Unfortunately, as deprivation is not specific to postcode sector, this information cannot be recovered.

Considering the vaccine information for all children, amongst those children born in quarters 3 and 4 of 2006, 28,281 (98.64%) had received the first dose of PCV-7 by March 2008, 27,881 (97.24%) had received the second dose of PCV-7 and 26,414 (92.12%) had received the third (booster) dose of PCV-7. The earliest

date at which the first dose of PCV-7 is administered is the 4th September 2006 which coincides with the introduction of the vaccine to the routine immunisation schedule (Cullen and Gillooly 2006).

On assessment of the dates at which the booster dose of PCV-7 was administered to children, unusual dates appear. The booster is supposed to be administered to those aged 13 months. This means that if 2 month old children are being routinely administered the first dose of PCV-7 from 4th September 2006 then the routine booster vaccinations should take place from around the 4th August 2007. The minimum date observed for the booster is 7th September 2006. 222 of the booster vaccinations were received prior to 4th August 2006.

On examination of the ages of the children who received these 222 early vaccinations, it can be noted that the minimum age at which the vaccine was administered was 0 months of age and the mean age was 6.67 months (median = 5.97 months). These early vaccinations may perhaps be attributable to children in high risk groups, such as those who are immunocompromised, receiving the PCV-7 booster prior to the age of 13 months to ensure that they are protected against pneumococcal disease. However, all vaccinations administered prior to the age of 2 months are not considered in the analysis as the routine childhood immunisation schedule in the UK is in place for children from the age of 2 months. In total, 13 children received the PCV-7 booster and 28 received the Hib/Men C booster prior to the age of 2 months. These cases were omitted. All other early vaccines are retained as it is feasible that some children may receive the vaccines early if they are at an increased risk of developing infections or disease so it is unclear whether or not these other discrepancies are errors.

In the final cleaned dataset, 26,067 (90.91%) of children born in quarters 3 and 4 of 2006 received the PCV-7 booster prior to 22 months of age. The mean age at which the PCV-7 booster was received was 14.76 (standard deviation (S.D.) 1.88) months. The minimum was 2.73 months of age, after omitting those vaccinated prior to two months of age, and the maximum was 21.97 months. 25,823 (90.06%) of children received the Hib/Men C booster by 22 months of age. The mean age at which Hib/Men C is received is 13.16 months (S.D. 1.81 months). Thus,

the PCV-7 booster appears to be administered, on average, almost two months later than the routine vaccination schedule recommendations and the Hib/Men C booster is administered over a month later than scheduled.

8.4.2 Hib/Men C booster and the PCV-7 booster

The Hib/Men C booster is administered to children routinely at age 12 months. Interest lies in determining the effect of late administration of the Hib/Men C booster on the timing of the PCV-7 booster.

Initially, uptake of the PCV-7 booster is assessed and is compared to uptake of the Hib/Men C booster. Table 8.4 shows a cross-tabulation of PCV-7 booster receipt by age 22 months by Hib/Men C receipt by age 22 months. Table 8.4 shows that the majority of children (85.66%) received both vaccines. Of those that received the Hib/Men C booster, only 4.89% did not receive the PCV-7 booster and, amongst those who received the PCV-7 booster, only a slightly higher percentage, 5.78%, did not receive the Hib/Men C booster. Overall, 4.68% of children received neither vaccine by 22 months of age. A χ^2 Test of Association resulted in a significant p-value of less than 0.001. Therefore, there is evidence to reject the null hypothesis of no association between these two variables. Table 8.4 shows lower observed numbers than expected for children receiving the PCV-7 booster and not the Hib/Men C booster and for children receiving the Hib/Men C booster and not the PCV-7 booster. Higher observed numbers are shown for children receiving either both or neither of the vaccines than expected.

Table 8.4: Observed counts for PCV-7 booster received against Hib/Men C booster received (expected counts in brackets).

		PCV-7 booster	
		Not received	**Received**
Hib/Men C booster	**Not received**	1343 (258.85)	1506 (2590.15)
	Received	1262 (2346.15)	24561 (23476.85)

In considering late uptake of PCV-7 up to the age of 22 months, those who had

not received the vaccination were not included in this analysis. Thus, only 26,067 children are considered. The reasoning behind the exclusion was mentioned in the introductory section of this chapter. All children who received vaccination after the age of 395.2 days (roughly 13 months) are classed as having received the vaccination late. All other children are classed as having received the vaccine on time. By this classification, 676 (2.59%) of children who received the PCV-7 booster received it on time. For those who received the vaccine late, the mean number of months late is 1.86 (S.D. 1.73 months); the maximum number of months late is 8.97. The cut-off of 13 months seems strict as the vaccine has to be administered at age 13 months and anyone aged between 13 and 14 months is classified as late. Thus, only 2.59% of children receive the vaccine on time according to this cut-off point. A more appropriate cut-off to use for late vaccinations may be those vaccinations administered later than one month after the recommended age. This cut-off has been adopted in a previous vaccine uptake analysis (Akmatov et al. 2008). Considering a classification of late uptake as any vaccinations administered after the age of 14 months, 10,493 (40.25%) children receive the vaccine on time. Therefore, clearly many children receive the vaccine between the ages of 13 and 14 months. Histograms and kernel density estimates[1] of the number of months late may be assessed. These plots are of months late at the 13 month cut-off. In both plots of the timing of the PCV-7 booster, Figure 8.1, peaks in the number of vaccinations can be observed to take place between approximately 0 and 0.3 months late followed by a gradual decline tailing off at around 6 months late.

For the Hib/Men C booster vaccination, all children who received the booster after the age of 364.8 days (roughly 12 months) are classed as having received the vaccine late whilst all other children received the vaccine on time. Once again, the cut-off for follow-up is 22 months of age. 2,849 (9.94%) of those who received the Hib/Men C booster were administered the vaccine on time. Amongst children who received the vaccine late, the mean number of months late is 1.28 (S.D. 1.81 months); the maximum is 9.97 months. As with the PCV-7 booster, plots of the number of months late the Hib/Men C booster was administered are examined.

[1]KDE has a Gaussian kernel and bandwidth = 0.15 months.

The KDE of the number of months late receiving the Hib/Men C booster[2], Figure 8.2, shows a much sharper peak at 0 months late than the KDE of months late receiving the PCV-7 booster, Figure 8.1. In addition, there is a much steeper downwards slope shown in Figure 8.2, with the majority of those children who receive the Hib/Men C vaccine late receiving the vaccine within 2 months of the date at which it should have been administered.

A cross-tabulation of the timing of the PCV-7 booster uptake, where late is classified as greater than 13 months of age, against the timing of Hib/Men C booster uptake is shown in Table 8.5. A χ^2 Test of Association carried out to test the relationship between the binary variables *PCV-7 booster administered late* and *Hib/Men C booster administered late* provides strong evidence at the 5% significance level (p-value < 0.001) of an association. The expected counts calculated in the χ^2 Test are also shown in Table 8.5.

Table 8.5: Observed counts for late PCV-7 booster uptake (late if administered after 13 months of age) against late Hib/Men C booster (expected counts in brackets).

		PCV-7 booster	
		Not Late	**Late**
Hib/Men C booster	**Not Late**	192 (48.31)	1792 (1935.69)
	Late	406 (549.69)	22171 (22027.31)

In Table 8.5, it can be noted that higher numbers of cases are expected than observed for one vaccine administered on time, the other late; lower numbers are expected than observed for both administered on time or both administered late.

The correlation between *Number of months late PCV-7 booster is administered* and *Number of months late Hib/Men C booster is administered* is calculated to be 0.35. Thus, there is a positive relationship between these two variables but it is fairly weak.

[2]KDE has a Gaussian kernel and bandwidth = 0.15 months.

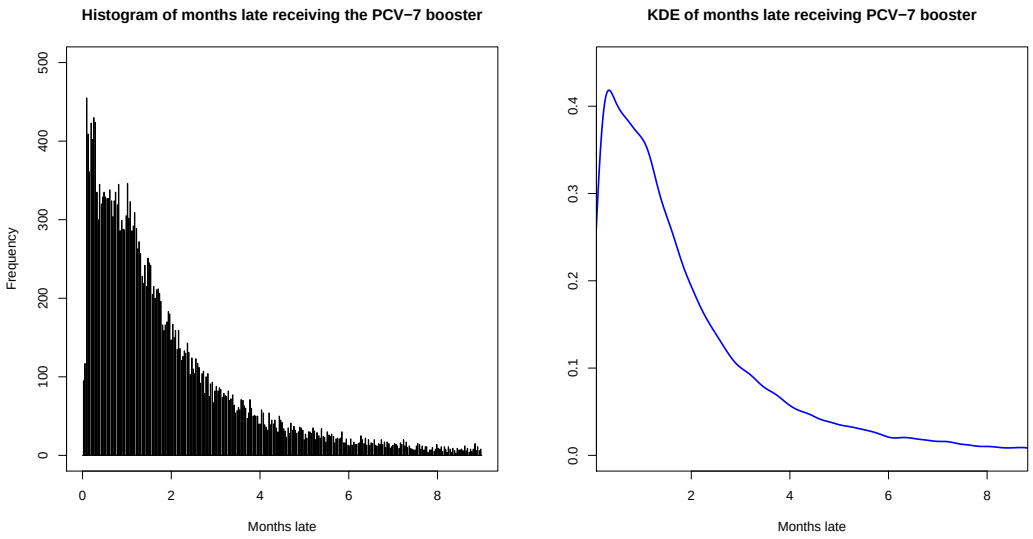

Figure 8.1: Histogram and KDE of months late receiving the PCV-7 booster for children born in quarters 3 and 4 of 2006.

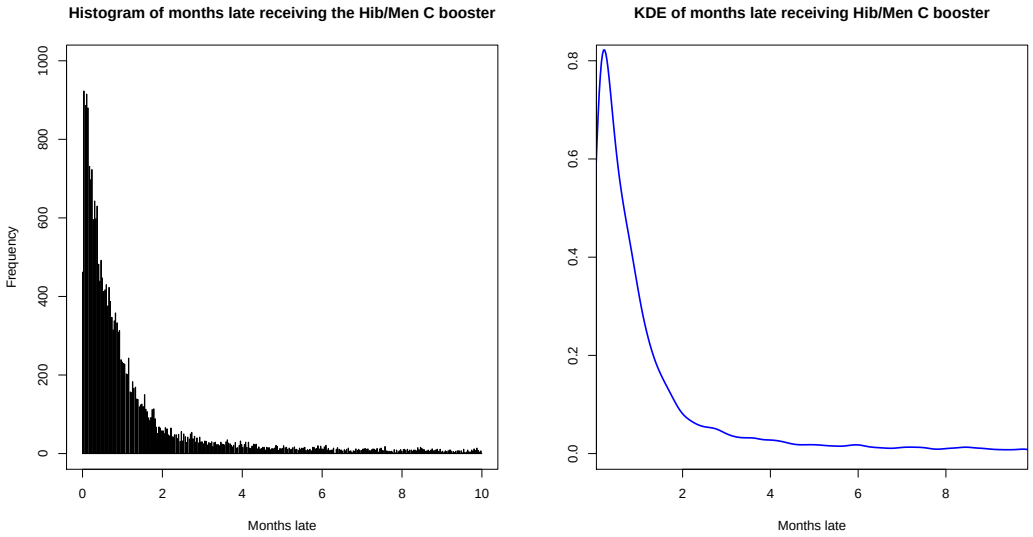

Figure 8.2: Histogram and KDE of months late receiving the Hib/Men C booster for children born in quarters 3 and 4 of 2006.

8.4.3 Modelling the binary response PCV-7 booster received

Descriptive analysis

The first univariate response model considers the variables important in predicting the binary response *PCV-7 booster received* ('PCVReceived'). This model, and the subsequent models, consider the possibility that the variation in the data can be explained by three levels which are the individual level, postcode district level and HB level. Postcode district was selected for use in the analysis rather than postcode sector as this considerably reduces the number of different areas within level 2 adopted in the analysis. The postcode sector level variables were aggregated in order to obtain the variables at district level. The individual level variable 'Gender' is considered in the model, as well as the area level deprivation measure 'SCSIMD5', as are all of the postcode district level variables from CASWEB described earlier. Although 'SCSIMD5' is an area level measure, it appears in the model as an individual level variable due to the fact the area at which this variable is measured is smaller than the postcode sector or district. The binary variable *Hib/Men C booster received* ('HibReceived') is also considered. The variables from CASWEB, unlike 'Gender' and 'HibReceived', are not individual level variables but have been aggregated to postcode district before inclusion in the modelling.

Figure 8.3 shows the proportion of children who received the PCV-7 booster within each HB. Orkney has the lowest uptake of all the HBs; only 58.95% of children born in quarters 3 and 4 of 2006 received the PCV-7 booster by age 22 months in this HB. The next lowest uptake is observed in the Shetlands. However, the percentage of children who received the PCV-7 booster is much higher at 80.47%. HBs H, W and F, the Highlands, the Western Isles and Fife, also display uptake of less than 90%. There is strong evidence of an association between HB and 'PCVReceived' with a p-value of less than 0.001 obtained in a χ^2 Test of Association.

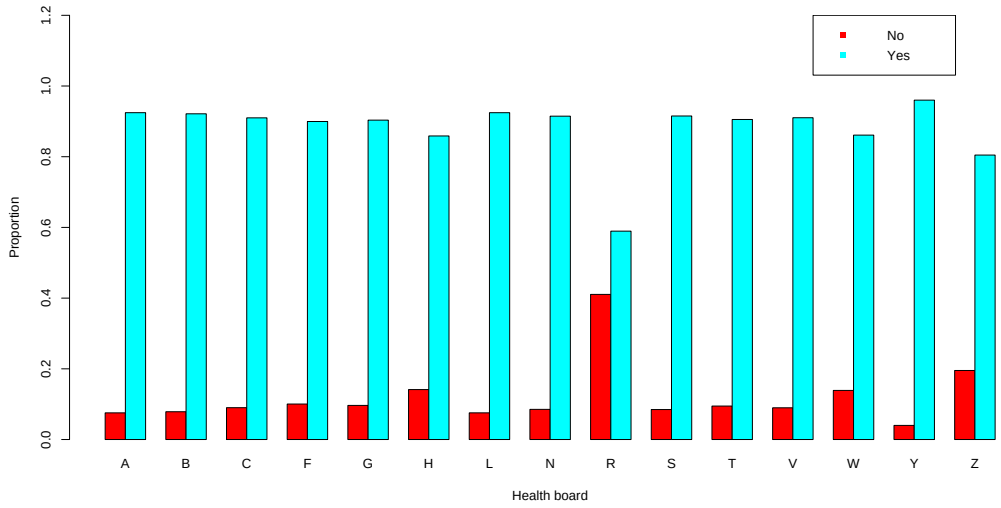

Figure 8.3: Barchart of proportion administered the PCV-7 booster by HB (A=Ayrshire & Arran, B=Borders, C=Argyll & Clyde, F=Fife, G=Greater Glasgow, H=Highland, L=Lanarkshire, N=Grampian, R=Orkney, S=Lothian, T=Tayside, V=Forth Valley, W=Western Isles, Y=Dumfries & Galloway, Z=Shetland).

Table 8.6: Observed counts for PCV-7 booster uptake by gender.

		PCV-7 booster	
		Not received	**Received**
Gender	**Female**	1224	12750
	Male	1381	13317

The cross-tabulation of 'PCVReceived' by 'Gender', Table 8.6, shows no great difference in PCV-7 booster uptake by sex. 91.24% of females received the vaccine and a slightly lower percentage of males, 90.60%, received the vaccine. As expected, there is no evidence of an association between 'PCVReceived' and 'Gender' at the 5% significance level (p-value = 0.06) on carrying out a χ^2 Test of Association.

From examination of Figure 8.4, it appears that the proportion of children who do not receive the PCV-7 booster increases slightly with increasing deprivation.

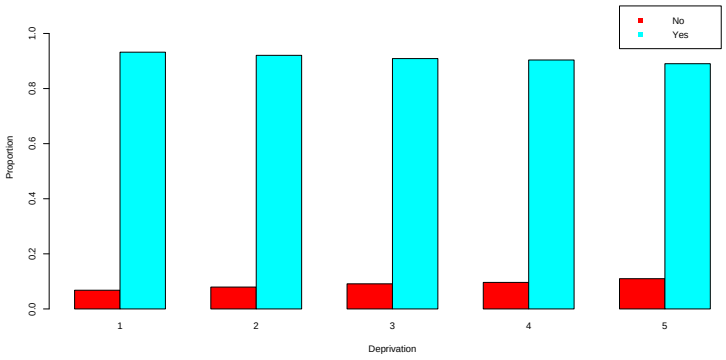

Figure 8.4: Barchart of proportion administered the PCV-7 booster by deprivation.

Considering the effect of postcode district on vaccine uptake, there are 48 postcode districts with a vaccine uptake of less than 80%. Of these 48 districts, 5 districts had no vaccinations; IV28, PA66, PA76, PH23 and PH5. Amongst the other 43 districts, there is only one child present in each of the postcode districts IV28, PA66, PH23 and PH5 and only three children in PA76. On average, there were roughly 70 children per district, with numbers ranging from only one child in a district to a maximum of 433.

Plots of the postcode district level variables by the proportion of children vaccinated in each district are examined to obtain an impression of which variables may be important to include in a model of 'PCVReceived'. On examination of the plots shown in Figures 8.5 and 8.6, none appear to show great differences in the proportion vaccinated with the PCV-7 booster for differing percentages of each of the postcode district level variables. Therefore, it does not appear that any of the postcode district level variables are particularly important in determining whether or not the vaccine is received. This may be due to the fact that the uptake is so high there is little scope for predicting areas with low uptake.

In summary, it appears that whether or not a child receives the Hib/Men C booster is important in determining whether or not the PCV-7 booster is administered. In addition, HB appears to play an important part in determining whether or not a child receives the vaccine, with the lowest observed uptake rates for the island HBs. Deprivation is also perhaps important as it seems that the

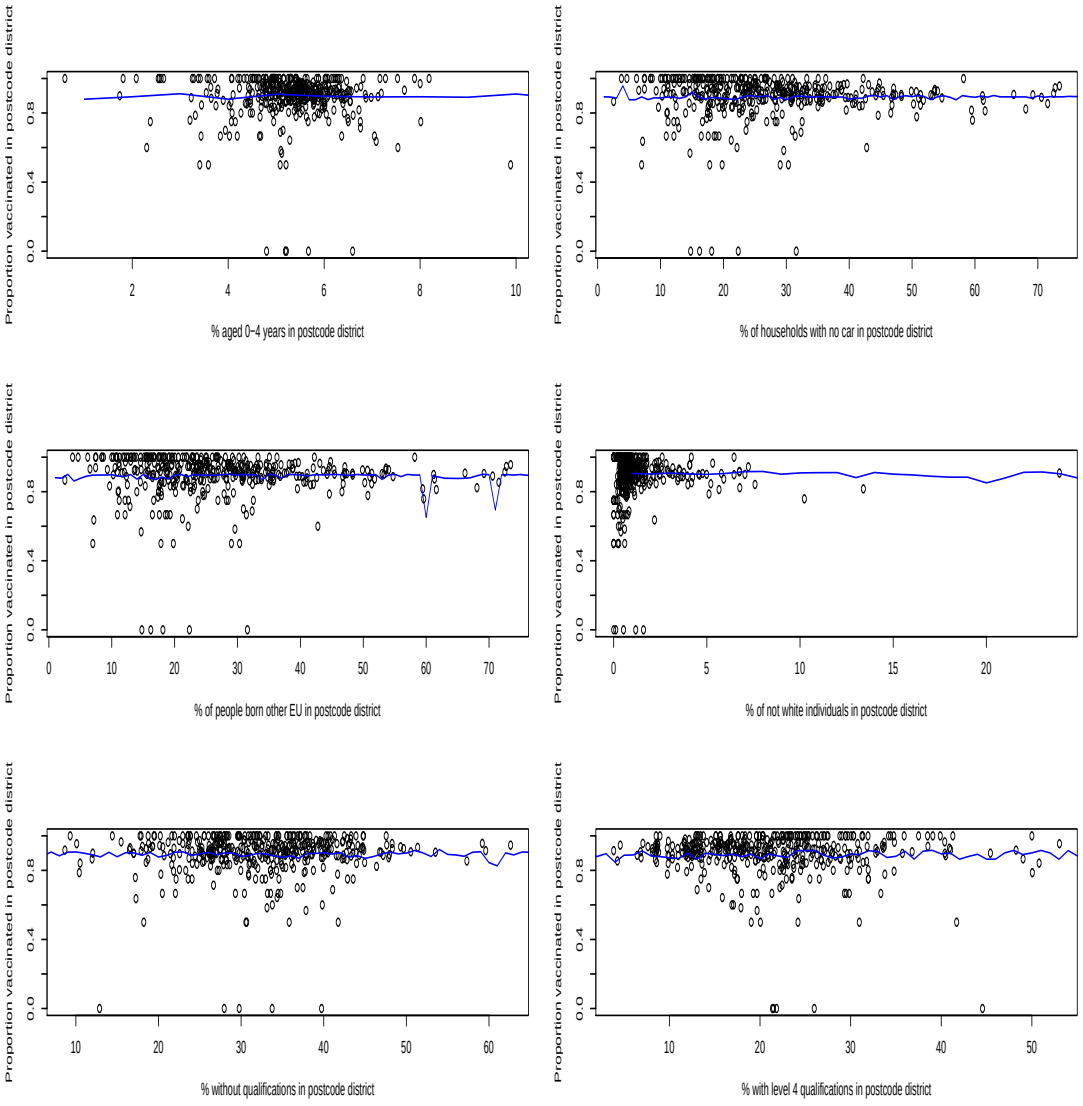

Figure 8.5: Scatterplots of the postcode district level variables 'Aged0-4', 'NoCar', 'OtherEU', 'NotWhite', 'NoQual' and 'Level4' by the proportion vaccinated in the district.

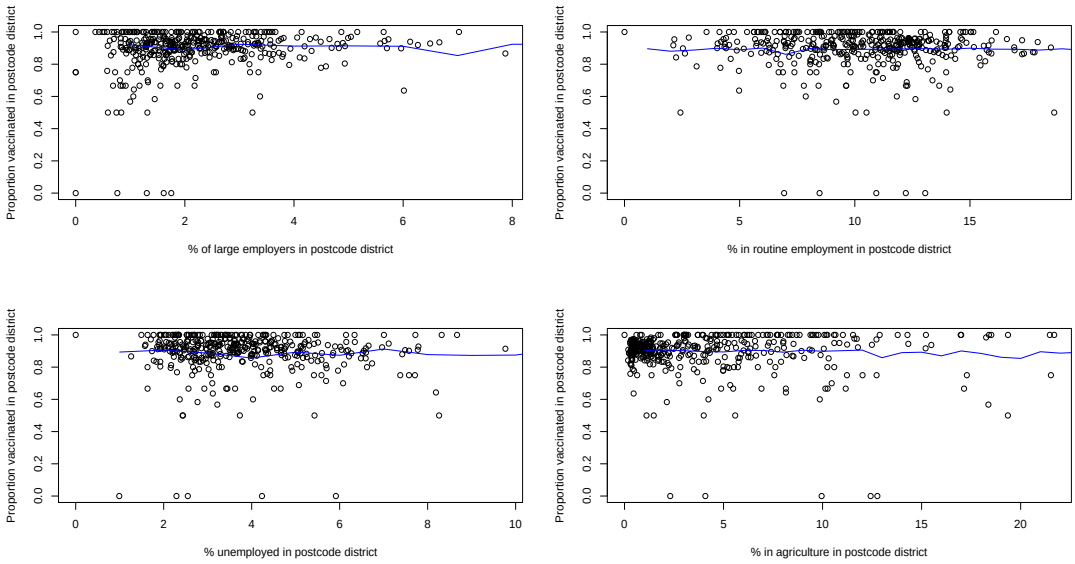

Figure 8.6: Scatterplots of the postcode district level variables 'LargeEmp', 'Routine', 'Unemployed' and 'Agriculture' by the proportion vaccinated in the district.

uptake decreases with increasing deprivation. From assessment of the postcode district level variables, it does not appear that any are particularly useful in determining the outcome.

Modelling

Single-level models

Initially, single-level models are considered to examine the importance of each of the explanatory variables independently in models of 'PCVReceived' without considering variability explained by either postcode district or HB. As this response variable is binary, single-level logit models are fitted. The single-level

logit model can be expressed as follows:

$$\begin{aligned} y_i &\sim \text{Binomial}(n_i, \pi_i), \\ \text{logit}(\pi_i) &= \alpha + \beta x_i, \\ \text{var}(y_i|\pi_i) &= \pi_i(1-\pi_1). \end{aligned} \tag{8.2}$$

For the models considered in this section, y_i in (8.2) is 'PCVReceived', where each child is followed to the age of 22 months, which is the binary response for the i^{th} unit, i.e. the i^{th} individual, where $i = 1, ..., 28,640$ as there is information on vaccine uptake for all children in the data. 32 of the observations did not have either postcode district or HB information and thus could not be linked to postcode district level variables. Since this is such a small number of the total number of observations, these 32 children were not considered in the analysis. π_i is the probability that $y_i = 1$, i.e. this is the probability that child i received the PCV-7 booster vaccination. $\pi_i/(1-\pi_i)$ is the odds that $y_i = 1$ and $\exp(\beta)$ is the OR. x_i represents the individual value of the variable chosen for inclusion in the model. Table 8.7 shows the results of Wald Tests of $H_0 : \beta = 0$ for each of the single-level models where each explanatory variable was considered independently as a predictor of 'PCVReceived'. The Wald Test is based on the result that the Wald statistic should approximately follow a Normal distribution with zero mean and variance equal to one and the Wald statistic is defined to be the estimated coefficient of a variable divided by its estimated standard error (S.E.). All modelling carried out in MLwiN used the default modelling approach of first order marginal quasi-likelihood.

The single-level models for 'PCVReceived' for each of the explanatory variables independently show significant non-zero coefficients for all but four of the possible explanatory variables. The null hypothesis that $\beta = 0$ cannot be rejected for 'Gender', 'NoQual', 'Level4', 'Routine' and 'Agriculture' at the 5% significance level. 'HibReceived' appears to be the most significant of all the explanatory variables in determining 'PCVReceived' as it has the highest Wald Test statistic. 'SCSIMD5' has the next highest Wald Test statistic and results in a significant

effect testing on four degrees of freedom. This variable included as a categorical variable, with deprivation category 1 as the baseline for comparison, in a single-level model of 'PCVReceived' gives the coefficients -0.166 (S.E. 0.075), -0.319 (S.E. 0.073), -0.380 (S.E. 0.071) and -0.526 (S.E. 0.068) for the four dummy variables for categories 2 to 5 respectively. The intercept in this model is 2.619 (S.E. 0.056). As the coefficient gets smaller with increasing deprivation there is a decreased probability of receiving the PCV-7 booster for increasing deprivation.

Table 8.7: Wald Tests for single-level logit models of 'PCVReceived'.

Predictor	**Test statistic**	***p*-value**
HibReceived	3626.066	< 0.001
Gender	3.358	0.067
Aged0-4	4.014	0.032
NoCar	23.662	< 0.001
OtherEU	22.219	< 0.001
NotWhite	10.781	0.001
NoQual	0.069	0.793
Level4	1.257	0.262
LargeEmp	18.089	< 0.001
Routine	0.707	0.400
Unemployed	19.195	< 0.001
Agriculture	0.092	0.762
SCSIMD5	60.955	< 0.001

From assessment of Table 8.7, it is surprising that so many of the postcode district level variables appear to have non-zero coefficients as there did not appear to be great associations between any of these variables and PCV-7 booster uptake shown in Figures 8.5 and 8.6. However, this is attributable to the fact that the models being fitted are single-level so all observations are considered in one estimate whilst the figures are based on over 400 postcode districts. These postcode district level variables may not be significant on fitting a multi-level model as ignoring variance structures can lead to insignificant variables being identified as significant, as mentioned earlier. Therefore, these postcode district variable

effects will be explored further on consideration of two-level models.

Correlations

The correlations between each of the postcode district variables are calculated to determine whether or not certain variables may not be required in the model if other variables are included.

Examination of Table 8.8 shows the highest correlation of -0.89 exists between 'NoQual' and 'Level4'. This high negative correlation is expected as one measures the lowest group of qualifications in a district and the other the highest in the census. It is unsurprising that there is a high negative correlation between the percentage with no qualifications and the percentage who work as large employers as these positions presumably are appointed on the basis of relevant qualifications. Similarly, it is hardly unexpected that the correlations between the percentage with no qualifications and the percentage working in routine employment or who are unemployed are highly positive. The opposite results are true for the percentage with level 4 qualifications, as expected. Moderately high positive correlations are observed between 'OtherEU' and 'Level4' and 'NoCar' and 'NoQual', respectively. The first of these correlations indicates that the higher the proportion of individuals born in other EU countries, i.e countries outwith the UK and Republic of Ireland, the higher the percentage in the district with higher qualifications. The second correlation makes intuitive sense as the higher the proportion of unemployed individuals within a postcode district, the higher the proportion of households with no car. As there are some high correlations between these variables it may not be necessary to include both variables involved in high correlations in a model of 'PCVReceived'. For example, there is evidence that both 'LargeEmp' and 'Unemployed' have non-zero coefficients when modelling 'PCVReceived' individually. However, it may not be necessary to include both variables together in a model as both provide measures of income in a district and there is a relatively high negative correlation of -0.67 between these variables. Thus, the addition of 'LargeEmp' to a model of 'PCVReceived' which already involves 'Unemployed' may not provide any significant additional information.

Table 8.8: Correlations between each of the postcode district level variables.

	0-4[a]	**NoC**	**OEU**	**NW**	**NQ**	**L4**	**LEmp**	**R**	**UnE**	**Agr**
0-4	1.00	-0.06	-0.45	-0.24	0.22	-0.35	0.05	0.27	0.11	-0.17
NoC		1.00	-0.08	0.26	0.62	-0.42	-0.58	0.30	0.73	-0.43
OEU			1.00	0.34	-0.64	0.69	0.39	-0.52	-0.38	-0.02
NW				1.00	-0.24	0.40	0.18	-0.43	0.01	-0.29
NQ					1.00	-0.89	-0.81	0.79	0.76	0.01
L4						1.00	0.78	-0.87	-0.63	0.04
LEmp							1.00	-0.75	-0.67	-0.11
R								1.00	0.54	0.09
UnE									1.00	-0.15
Agr										1.00

[a]0-4='Aged0-4', NoC='NoCar', OEU='OtherEU', NW='NotWhite', NQ='NoQual', L4='Level4', LEmp='LargeEmp', R='Routine', UnE='Unemployed', Agr='Agriculture'.

In consideration of these postcode district variables, it is important to acknowledge the possibility of the presence of ecological fallacy (Robinson 1950). This is where incorrect interpretations are made about associations between variables due to the fact that aggregated data are used rather than individual level data. This means that the assumption is being made that individuals within a postcode district have the average characteristics of that district which can lead to incorrect interpretations. Thus, care must be taken when assessing the results from the area level variables in the models considered in this chapter. In addition, the correlations described above for the aggregated data may not be present if information was available on the individual level.

In further individual level models, HB is considered as a categorical explanatory variable of 'PCVReceived'. HB G is taken as the baseline for comparison as this is the HB in which the largest number of children in the dataset are found. Table 8.9 shows the coefficients and S.E.s of each of the HBs in the model of 'PCVReceived' on comparison with HB G, Greater Glasgow.

In Table 8.9 it can be observed that five of the HB coefficients are negative indicating that these HBs have a lower probability of receiving the PCV-7 booster than the baseline comparator HB G. These HBs with negative coefficients are Fife, Highlands, Orkney, Western Isles and Shetland. The predicted probability of receiving the PCV-7 booster in Greater Glasgow, HB G, is approximately 0.91. Table 8.9 shows that most HBs have a high predicted probability of receiving the vaccine with ten HBs displaying probabilities of greater than 0.90. Orkney has the lowest predicted probability of receiving the vaccine at only 0.59, whilst the other islands and the Highlands have much higher predicted probabilities of at least 0.80.

Table 8.9: HB coefficients and S.E.s in model of 'PCVReceived' and predicted probabilities of receiving the PCV-7 booster (A=Ayrshire & Arran, B=Borders, C=Argyll & Clyde, F=Fife, G=Greater Glasgow, H=Highland, L=Lanarkshire, N=Grampian, R=Orkney, S=Lothian, T=Tayside, V=Forth Valley, W=Western Isles, Y=Dumfries & Galloway, Z=Shetland).

HB	**Coefficient**	**S.E.**	**Probability**
A	0.263	0.096	0.92
B	0.230	0.166	0.92
C	0.071	0.089	0.91
F	-0.049	0.088	0.90
H	-0.444	0.099	0.86
L	0.272	0.081	0.92
N	0.134	0.082	0.91
R	-1.875	0.214	0.59
S	0.142	0.072	0.92
T	0.024	0.088	0.91
V	0.082	0.099	0.91
W	-0.413	0.246	0.86
Y	0.946	0.187	0.96
Z	-0.821	0.228	0.80

Two-level models

Next, a two-level model with postcode district as the higher level is considered. In MLwiN, the parameter estimates for a two-level discrete response model are found by initially carrying out first order marginal quasi-likelihood, a crude approximation which can lead to biased estimates, then adopting these estimates as starting values for the more reliable method of penalised quasi-likelihood (PQL). The starting values from the crude approach are often required as PQL is a less stable method which often has convergence problems (Rasbash et al. 2008).

A two-level random intercept model is fitted. In this model the intercept α_j is

split into a fixed component, α, and a random component, υ_j, which is assumed to be Normally distributed with zero mean and constant variance σ_υ^2. In this model j represents the postcode district level. Thus, υ_j is the random effect attributable to each postcode district level. The parameter estimate for this model is shown:

$$\text{logit}(\pi_{ij}) = \alpha_j,$$

where

$$\alpha_j = 2.270 + \upsilon_j. \tag{8.3}$$

The estimated S.E. for α in 8.3 is 0.037 and the estimate for σ_υ^2 is 0.246 with S.E. 0.033. An approximate test that may be used to determine whether or not there is evidence that the postcode district level variation is required in the model is the Wald Test of $(0.246/0.033)^2$. A p-value of less than 0.001 is obtained on comparison with the χ^2 statistic with one degree of freedom. Thus, there is significant evidence to reject the null hypothesis that $\sigma_\upsilon^2 = 0$. Therefore, there is evidence of a significant postcode district effect in determining 'PCVReceived'.

Three-level models

Next, it is necessary to determine whether or not a third level should be included to incorporate HB level variation in the model of 'PCVReceived'. In order to use MLwiN to perform multi-level modelling the data must be arranged according to the specified hierarchy which, in this series of models, has to be HB first, then postcode district within HB, followed by individual within postcode district within HB. The associations between postcode district and HBs were examined to determine whether or not postcode district is nested within HB.

21 of the postcode districts were found to be associated with more than one HB so, as this is only a small number of all the districts, a cross-classified multi-level model should not be necessary for the modelling. To ensure the hierarchy is as required for the analysis, each of the postcode districts linked to more than one HB were grouped into the largest of the two or three HBs with which the district

was associated. As the majority of the postcode districts associated with more than one HB are predominantly found in one of the HBs this should not greatly affect the modelling results.

Initially, a three-level random intercept model is considered. An additional random parameter, ν_k, representing the random effect attributable to each HB is entered in the model for the random intercept shown in (8.3) and is assumed to follow a $N(0, \sigma_\nu^2)$ distribution. In this model, k represents the HB level. All residual parameters are assumed to be uncorrelated. A p-value of 0.017 is obtained on testing the null hypothesis that $\sigma_\nu^2 = 0$ in this random intercept model. Thus, there is evidence of a significant HB effect at the 5% significance level.

As both postcode district and HB have been deemed significant in explaining variability in 'PCVReceived', the next stage in the modelling process is to add potentially important explanatory variables to the model. Each of the variables which were found to have non-zero coefficients in the previous individual single-level models of 'PCVReceived' are added to this three-level model, assuming each of the coefficients do not vary randomly with HB. Table 8.10 shows the coefficients and S.E.s obtained.

It can be noted from examination of Table 8.10 that 'NoCar', 'Aged0-4', 'NotWhite' and 'LargeEmp' are not required in the final model as the p-values obtained from the Wald Tests for these variables are all greater than 0.05. After removal of these variables, the final random-intercept model contains 'HibReceived', 'SCSIMD5' and 'OtherEU'.

The estimate of α in the final model is 0.335 (S.E. 0.185) and the parameter estimates for σ_v^2 and σ_ν^2 are 0.087 and 0.381. Therefore, there is more variation between HBs than between postcode districts.

Table 8.10: Results from the three-level model of 'PCVReceived'.

Variable	**Coefficient**	**S.E.**	***p*-value.**
HibReceived	2.784	0.047	< 0.001
$\text{SCSIMD5}_{\text{cat 2}}$	-0.161	0.079	
$\text{SCSIMD5}_{\text{cat 3}}$	-0.258	0.079	
$\text{SCSIMD5}_{\text{cat 4}}$	-0.282	0.080	
$\text{SCSIMD5}_{\text{cat 5}}$	-0.344	0.082	< 0.001
Aged0-4	-0.025	0.049	0.626
NoCar	0.002	0.004	0.508
OtherEU	-0.234	0.065	0.019
NotWhite	-0.026	0.014	0.063
LargeEmp	0.009	0.040	0.827

The next stage in the modelling process is the introduction of a random coefficient for 'HibReceived' to allow for the possibility that 'HibReceived' could vary by postcode district and HB. It is reasonable to consider that the uptake of the Hib/Men C booster, and hence the relation to PCV-7, may vary randomly by postcode district and health board as evidence had been found that the uptake of the PCV-7 booster varies randomly by these two area variables. This model introduces four new parameters to the three-level random intercept model. The random variance and covariance parameters are:

$$\begin{bmatrix} \nu_{0k} \\ \nu_{1k} \end{bmatrix} \sim N(0, \Omega_\nu) : \Omega_\nu = \begin{bmatrix} \sigma^2_{\nu 0} & \sigma_{\nu 01} \\ \sigma_{\nu 01} & \sigma^2_{\nu 1} \end{bmatrix}$$

and

$$\begin{bmatrix} \upsilon_{0jk} \\ \upsilon_{1jk} \end{bmatrix} \sim N(0, \Omega_\upsilon) : \Omega_\upsilon = \begin{bmatrix} \sigma^2_{\upsilon 0} & \sigma_{\upsilon 01} \\ \sigma_{\upsilon 01} & \sigma^2_{\upsilon 1} \end{bmatrix}. \tag{8.4}$$

In (8.4), $\sigma^2_{\nu 0}$ is the variation in the intercepts across the HB summary lines and $\sigma^2_{\nu 1}$

is the variation in the slopes across the HB lines. $\sigma_{\nu 01}$ is the covariance between the HB intercept and slope. The parameters $\sigma^2_{\upsilon 0}$, $\sigma^2_{\upsilon 1}$ and $\sigma_{\upsilon 01}$ can be interpreted in a similar fashion for the postcode district within HB variation. A Wald Test is used to test the significance of these new parameters and a test statistic of 9.942 is obtained on 4 degrees of freedom, giving a p-value of 0.041. Therefore, the null hypothesis that these parameters equal zero may be rejected at the 5% significance level. Thus, it appears that the random coefficient for 'HibReceived' is required in the model. The final model of 'PCVReceived', including the variance parameters and S.E.s is shown in Table 8.11.

As HB has only 15 levels and these levels represent all possible categories of HB, it is possible to fit HB as a fixed effect in a two-level model to examine the differences in uptake of the HBs. Table 8.12 shows the modelling results. In the model of 'PCVReceived' with a random effect attributable to postcode district variability and with HB included as a fixed effect, shown in Table 8.12, the Greater Glasgow HB, HB G, is taken as the baseline for comparison as it is the largest HB. The coefficients for HBs L (Lanarkshire), S (Lothian) and Y (Dumfries and Galloway) are positive. Thus, given the other variables included in the model, children found in these HBs have a greater odds of receiving the PCV-7 booster than those in Greater Glasgow. HBs R and Z, Orkney and Shetland, have high negative coefficients. Thus, there are lower odds of receiving the booster for children in these two HBs. The next highest negative coefficient amongst the HBs is for HB W, the Western Isles. The odds of receiving the booster are lower for the Western Isles than Greater Glasgow. Thus, clearly there is a lower probability of receiving the PCV-7 booster for children living in island HBs than for those living in mainland Scotland. It can be observed that the coefficients and S.E.s of the other explanatory variables do not differ greatly to those obtained in the three level model shown in Table 8.11.

Table 8.11: Parameter estimates and variance components for the final model of 'PCVReceived'.

Variable	**Parameter**	**Estimate**	**S.E.**
Intercept	α	0.323	0.186
HibReceived	β_1	3.070	0.118
$SCSIMD5_2$	β_2	-0.154	0.090
$SCSIMD5_3$	β_3	-0.276	0.090
$SCSIMD5_4$	β_4	-0.315	0.090
$SCSIMD5_5$	β_5	-0.385	0.090
OtherEU	β_6	-0.281	0.058

Variance component	**Parameter**	**Estimate**	**S.E.**
HB intercept	$\sigma^2_{\nu 0}$	0.328	0.145
HB intercept and slope covariance	$\sigma_{\nu 01}$	-0.093	0.080
HB slope	$\sigma^2_{\nu 1}$	0.120	0.070
PCDIST within HB intercept	$\sigma^2_{\upsilon 0}$	0.208	0.056
PCDIST intercept and slope covariance	$\sigma_{\upsilon 01}$	-0.094	0.055
PCDIST within HB slope	$\sigma^2_{\upsilon 1}$	0.183	0.074

To summarise, in the final model of 'PCVReceived', there are significant random effects corresponding to postcode district and HB. Thus, there is evidence of significant area variation in the uptake of the PCV-7 booster. A random coefficient for the variable 'HibReceived' was also found to be significant in the model. The other variables included in the final multi-level logistic regression model of 'PCVReceived' are 'SCSIMD5' and 'OtherEU'. Given the other variables in the model, the coefficients of each of the categories of 'SCSIMD5' are negative and decrease in size with increasing category from -0.154 on comparing category 2 to the baseline category 1 to -0.385 comparing category 5 to category 1. Thus, increasing levels of deprivation decrease the log odds of receiving the PCV-7 vaccine.

The coefficient of the continuous postcode district level variable 'OtherEU' is negative. Thus, the log odds of receiving the PCV-7 booster decrease as the percentage of individuals born in EU countries other than the UK and Republic of Ireland within a postcode district increases.

Finally, the fixed part of the binary covariate for 'HibReceived' is positive, taking a value of 3.07. Thus, the log odds of receiving the PCV-7 booster for a child who receives the Hib/Men C booster are higher than that for a child who does not receive the Hib/Men C booster. However, as the coefficient of 'HibReceived' varies randomly with postcode district and HB this result may not be true for all areas. However, the variance components for the slope are estimated at 0.12 and 0.18. Thus, the effect of HB is small in comparison to the fixed estimate of 3.07 and so the effect of the Hib/Men C booster receipt on the PCV-7 booster uptake is generally in the same direction for all HBs and postcode districts.

8.4.4 Modelling the binary response PCV-7 booster administered late

Descriptive analysis

In assessing the variable *PCV-7 booster administered late* ('PCVLate'), as mentioned earlier, only those who received the vaccine during the first 22 months of life are considered in the analysis. Thus, information on 26,067 children is used. Earlier in this chapter, significant associations were identified between this binary response and the binary explanatory variable *Hib/Men C booster administered late* ('HibLate'), see Table 8.5. Other explanatory variables are considered to determine their importance in predicting 'PCVLate'.

The percentage of vaccinations received on time varies from a very low 0.54% in Lanarkshire to the highest 8.93% in Orkney where only 56 children were vaccinated with the PCV-7 booster. Considering postcode district, of the 408 postcode districts included in the data, 6 districts (AB1, IV28, PA66, PA76, PH23, PH5) have no observations for 'PCVLate' for children born in quarters 3 and 4 of 2006. For the 402 districts with PCV-7 booster observations, in 187 (45.83%) there were no vaccinations administered prior to 13 months old.

Table 8.12: Results from the two-level model of 'PCVReceived' with random slope for 'HibReceived' and HB as a fixed effect (A=Ayrshire & Arran, B=Borders, C=Argyll & Clyde, F=Fife, G=Greater Glasgow, H=Highland, L=Lanarkshire, N=Grampian, R=Orkney, S=Lothian, T=Tayside, V=Forth Valley, W=Western Isles, Y=Dumfries & Galloway, Z=Shetland).

Variable	**Coefficient**	**S.E.**
Intercept	0.652	0.124
HibReceived	3.007	0.061
SCSIMD5_2	-0.162	0.090
SCSIMD5_3	-0.275	0.090
SCSIMD5_4	-0.315	0.089
SCSIMD5_5	-0.397	0.089
OtherEU	-0.283	0.056
HB_A	-0.080	0.152
HB_B	-0.177	0.236
HB_C	-0.120	0.147
HB_F	-0.239	0.164
HB_H	-0.467	0.162
HB_L	0.217	0.151
HB_N	-0.253	0.144
HB_R	-2.256	0.362
HB_S	0.003	0.135
HB_T	-0.146	0.157
HB_V	-0.222	0.167
HB_W	-0.949	0.348
HB_Y	0.635	0.255
HB_Z	-1.296	0.396

Associations between 'PCVLate' and 'Gender' and 'SCSIMD5' are assessed using χ^2 Tests of Association.

HB [a]	A	B	C	F	G	H	L
No. received PCV-7 booster on time	66	26	44	49	79	52	17
No. received PCV-7 booster late	1834	478	1916	1788	4284	896	3132
% **per HB received PCV-7 booster late**	96.53	94.84	97.76	97.33	98.19	94.51	99.46

HB	N	R	S	T	V	W	Y	Z
No. received PCV-7 booster on time	157	5	85	16	38	9	28	3
No. received PCV-7 booster late	2513	51	4070	1925	1453	115	741	100
% **per HB received PCV-7 booster late**	94.12	91.07	97.95	99.18	97.45	92.74	96.36	97.09

Table 8.13: Number of children who received the PCV-7 booster on time or late in each Scottish HB.

[a]A=Ayrshire and Arran, B=Borders, C=Argyll and Clyde, F=Fife, G=Greater Glasgow, H=Highland, L=Lanarkshire, N=Grampian, R=Orkney, S=Lothian, T=Tayside, V=Forth Valley, W=Western Isles, Y=Dumfries and Galloway, Z=Shetland.

Table 8.14: Observed counts for late PCV-7 booster uptake by gender.

		PCV-7 booster	
		Not Late	**Late**
	Female	320	12430
Gender	**Male**	356	12961

From examination of Table 8.14, it appears unlikely that there is an association between 'Gender' and 'PCVLate'. Similar figures were obtained for late PCV-7 uptake amongst males and females, with 97.33% of all vaccinated males and 97.49% of all females recorded as having received the PCV-7 booster late. A χ^2 Test of Association confirms there is no evidence of an association between these two variables (p-value 0.429).

On assessment of Table 8.15, it does not appear that there is a great difference in the timing of PCV-7 administration to children for each category. A regression of the odds of receiving the vaccine late by deprivation, treating deprivation as a continuous variable, gives a p-value of 0.164 which is not significant. Thus, there is no significant evidence to suggest that the proportion of children vaccinated late changes with increasing deprivation.

Considering late uptake in each postcode district, only one of the postcode districts, KW6, has a proportion of children less than 0.6 who were vaccinated late. In this district, all children were vaccinated on time. However, this postcode district contained only one child. 3 postcode districts had a proportion vaccinated on time of between 0.3 and 0.5. These districts are PA67, IV21 and FK20, each with 33.33% vaccinated on time.

Table 8.15: Observed counts for late PCV-7 booster uptake by deprivation quintile.

Deprivation Category	PCV-7 not late	PCV-7 late (%)
1	143	4564 (96.96%)
2	122	4875 (97.56%)
3	134	4901 (97.34%)
4	120	5126 (97.71%)
5	155	5830 (97.41%)

Considering a cut-off of 14 months for the age at which the vaccine should be received to be considered on time, 92 districts have greater than 50% of vaccinations received on time.

Plots of the postcode district level variables by the proportion of children vaccinated late in each district are examined to obtain an impression of which variables may be important to include in a model of 'PCVLate'. In these plots, the cut-off for the PCV-7 booster to have been received on time is 13 months. On examination of the plots shown in Figures 8.7 and 8.8, there does not appear to be any substantial difference in the proportion vaccinated late with the PCV-7 booster for differing percentages of each of the postcode district level variables. Thus, it does not appear that any of the postcode district level variables are particularly important in determining whether or not the vaccine is received late.

In conclusion, it appears that the individual level measurement 'HibLate' is associated with 'PCVLate'. It also appears that HB may be important in determining 'PCVLate'. No postcode district level measurements appear to have a particularly large effect on whether or not the PCV-7 booster is administered late.

Modelling

Single-level models

As with the modelling procedure adopted for 'PCVReceived', the hierarchical

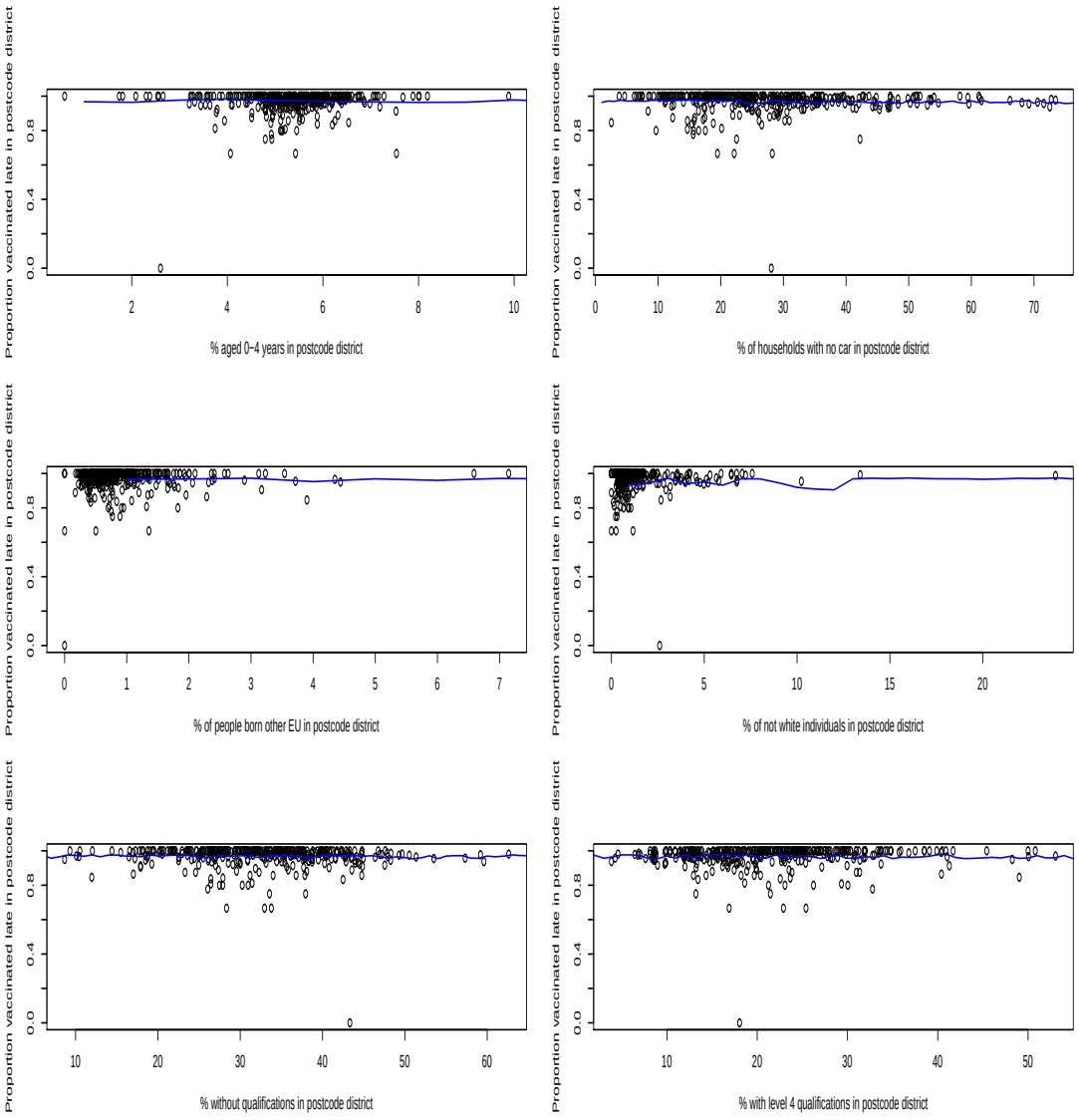

Figure 8.7: Scatterplots of the postcode district level variables 'Aged0-4', 'NoCar', 'OtherEU', 'NotWhite', 'NoQual' and 'Level4' by the proportion vaccinated late in the district.

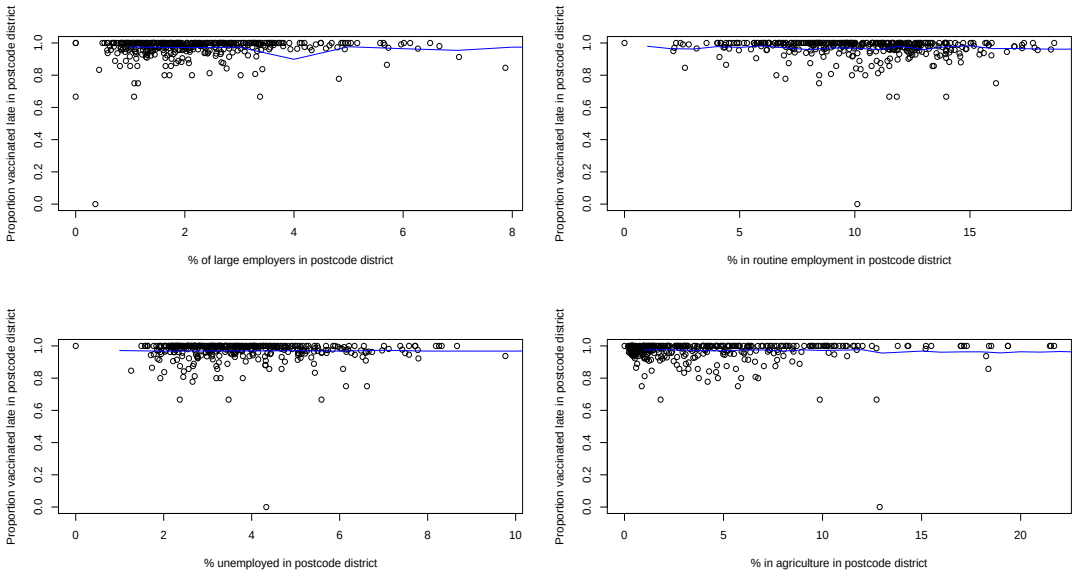

Figure 8.8: Scatterplots of the postcode district level variables 'LargeEmp', 'Routine', 'Unemployed' and 'Agriculture' by the proportion vaccinated late in the district.

nature of the data is initially ignored to determine which variables are potentially important in determining 'PCVLate' when the postcode district and HB level variation are not taken into account. Since 'PCVLate' is a binary response variable, the modelling is the same as that for 'PCVReceived'. Table 8.16 shows the results of Wald Tests of $H_0 : \beta = 0$ for each of the single-level models.

Considering Table 8.16, the null hypothesis that $\beta = 0$ may be rejected at the 5% significance level for the models containing the explanatory variables 'HibLate', 'Aged0-4', 'NoCar', 'OtherEU', 'NoQual', 'Level4', 'Unemployed' and 'Agriculture'. 'Gender', 'NotWhite', 'LargeEmp' and 'Routine' do not appear significant in determining 'PCVLate'. Assessing the categorical variable 'SCSIMD5' in a model of 'PCVLate', 'SCSIMD5' does not appear useful in determining 'PCVLate' as a p-value of > 0.05 was obtained comparing this model to the null model.

From examination of Table 8.13 it can be observed that the percentage of late vaccinations is extremely high for each HB. Thus, a less strict cut-off of 14 months is examined. The predicted probabilities of receiving the vaccine late for this cut-off

are shown in Table 8.17. Differences between the HBs are much more noticeable for the 14 month cut-off for the vaccine to have been received late. Examination of Table 8.17 shows that most HBs have a probability of between 0.54 and 0.62 of receiving the PCV-7 booster late. The lowest predicted probability is 0.49 for HB Y, Dumfries and Galloway, whilst the highest probability for HB T, Tayside, is much greater than that for all other HBs at 0.89.

Table 8.16: Wald Tests for single-level logit models of 'PCVLate'.

Predictor	**Test statistic**	***p*-value**
HibLate	370.250	< 0.001
Gender	0.579	0.447
Aged0-4	31.050	< 0.001
NoCar	6.394	0.011
OtherEU	44.364	< 0.001
NotWhite	1.714	0.190
NoQual	7.105	0.008
Level4	13.870	< 0.001
LargeEmp	0.002	0.964
Routine	0.171	0.679
Unemployed	4.767	0.029
Agriculture	21.902	< 0.001
SCSIMD5	6.174	0.187

Two-level models

To determine whether or not a second level should be added to the model of 'PCVLate' to account for variability between postcode districts, a random intercept model is fitted for the two-level model with i indexing the individuals and j the postcode districts. A Wald Test was carried out to test $H_0 : \sigma_v^2 = 0$, i.e. to test whether or not there is significant evidence that the postcode district level variance is required in the model. A test statistic of 36.678 is obtained, giving a p-value of less than 0.001. Therefore, the null hypothesis may be rejected. This

means there is significant evidence to retain the postcode district level in the model of 'PCVLate'.

Table 8.17: Predicted probabilities of receiving the PCV-7 booster late at 14 month cut-off for each HB (A=Ayrshire & Arran, B=Borders, C=Argyll & Clyde, F=Fife, G=Greater Glasgow, H=Highland, L=Lanarkshire, N=Grampian, R=Orkney, S=Lothian, T=Tayside, V=Forth Valley, W=Western Isles, Y=Dumfries & Galloway, Z=Shetland).

HB	**Probability**
A	0.63
B	0.61
C	0.54
F	0.61
H	0.58
L	0.59
N	0.55
R	0.61
S	0.62
T	0.89
V	0.55
W	0.52
Y	0.49
Z	0.59

Three-level models

The random effect of HB is considered in a three-level random intercept model. A p-value of 0.024 is obtained on testing the null hypothesis that $\sigma_\nu^2 = 0$ in this random intercept model. Thus, there is evidence of a significant HB effect.

Next, the individual level and postcode district level variables found to have non-zero coefficients in the single-level models are all entered into a three-level model.

As with the procedure adopted for the three-level model of 'PCVReceived', variables are removed one by one after carrying out a Wald Test of $H_0 : \beta_i = 0$ for the variable with the highest S.E. comparative to its coefficient in the model. The final three-level random intercept model is shown in Table 8.18.

Table 8.18: Results from the three-level model of 'PCVLate'.

Variable	**Coefficient**	**S.E.**
HibLate	1.904	0.101
NoCar	-0.021	0.007
Level4	-0.045	0.017

The final random intercept model of 'PCVLate' includes the variables 'HibLate', 'NoCar' and 'Level4'.

There is no evidence to suggest that it is necessary to allow 'HibLate' to vary randomly by postcode district and HB, with a p-value of 0.47 obtained on carrying out a Wald Test on four degrees of freedom for the additional four parameters included for this random effect.

The final model of 'PCVLate' adopting a 13 month cut-off for late vaccinations is shown in Table 8.19. In addition, the same model was fitted to the data with late classified as greater than 14 months. 'NoCar' is no longer significant in the model when the cut-off for late PCV-7 booster administration is 14 months. The random effects of HB and postcode district have greater significance in this model than in the model with the 13 month cut-off. 'Level4' was found to be significant in both the 13 month and the 14 month model. However, the direction of the effect of this variable differs in the models. In the 13 month cut-off model, the odds of receiving the vaccine late decreases as the percentage of individuals with level 4 qualifications increases. In the 14 month cut-off model, the odds of receiving PCV-7 late increases as the percentage of individuals with level 4 qualifications in a district increases.

As with the modelling for 'PCVReceived', a two-level model is fitted to 'PCVLate'

with HB entered as a fixed effect. The results from the modelling are shown in Table 8.20.

Table 8.19: Results from the three-level model of late PCV-7 booster, with late classified as greater than 13 or 14 months of age.

	13 months		**14 months**	
Variable	**Coefficient**	**S.E.**	**Coefficient**	**S.E.**
Intercept	3.349	0.370	-0.677	0.188
HibLate	1.901	0.101	1.118	0.052
NoCar	-0.027	0.006	-0.001	0.003
Level4	-0.016	0.008	0.006	0.004

In the two-level logistic regression model of 'PCVLate' with HB, shown in Table 8.20, HB L, Lanarkshire, is the only HB with a positive coefficient. Thus, the odds of receiving the vaccine late are highest for this HB compared to all others. On comparison with Greater Glasgow, the odds of receiving the vaccine late are lowest for Orkney at 0.123 times that of Greater Glasgow.

To summarise the modelling of 'PCVLate' with the 13 month cut-off, significant random effects attributable to HB and postcode district are included. The other variables in the final three-level logistic regression model are the individual level variable 'HibLate' and the postcode district level variables 'NoCar' and 'Level4'. The binary variable 'HibLate' has a positive coefficient. Thus, the log odds of receiving the PCV-7 booster late are higher for those children who receive the Hib/Men C booster late than for those who receive the booster on time, which makes intuitive sense. The coefficients of 'NoCar' and 'Level4' are both negative. Thus, the first result suggests that as the percentage of households with no car increases in a postcode district, the log odds of receiving the vaccine late decreases. The second result suggests that as the percentage of households with the highest level of qualifications increases, the probability of receiving the vaccine late decreases. The second result appears logical as it implies that the higher the educational attainment of a district, the more likely it is for children in that district to be vaccinated according to schedule. 'NoCar' is a measure

of deprivation and the results indicate that districts with low car ownership are more likely to receive the vaccine on time than late. With the 13 month cut-off, the vast majority of children, 90.91%, are classed as receiving the vaccine late. Thus, with such a small proportion of children receiving the vaccine on time, it appears that it may be more appropriate to use the less stringent cut-off of 14 months for late vaccinations as discussed previously.

In the three-level model for the 14 month cut-off for the PCV-7 booster to have been classed as late, only the variables 'HibLate' and 'Level4' are found to be significant. 'HibLate' in this model was defined as in the 13 month model with those aged over 12 months of age classed as late. 'NoCar' is not required in the model. The coefficient of 'HibLate' for this model is 1.118. Thus, the log odds of receiving the PCV-7 booster late for a child who receives the Hib/Men C booster late are higher for a child who receives the Hib/Men C booster on time. The coefficient of 'Level4' is positive, 0.006, in this model suggesting that as the percentage with the highest level of qualifications increases within a postcode district, the odds of receiving the vaccine late increase. This contradicts the result obtained for the 13 month cut-off.

8.4.5 Modelling the number of months late PCV-7 booster is administered

Descriptive analysis

In this final univariate response variable model, the variables important in predicting the continuous response *Number of months late PCV-7 booster is administered* ('PCVMonthsLate') are identified. In this analysis, only those children who receive the vaccine after the age of 13 months, i.e. those classed late in the response 'PCVLate', are considered. Thus, in total, 25,378 children are considered. The individual level variables 'Gender' and 'SCSIMD5' are considered in this model, as are all of the postcode district level variables assessed in the previous two models. The continuous individual level variable *Number of months late Hib/Men C booster is administered* ('HibMonthsLate') is also considered.

Table 8.20: Results for the fixed effect estimates of the two-level model of 'PCVLate' with HB as a fixed effect (A=Ayrshire & Arran, B=Borders, C=Argyll & Clyde, F=Fife, G=Greater Glasgow, H=Highland, L=Lanarkshire, N=Grampian, R=Orkney, S=Lothian, T=Tayside, V=Forth Valley, W=Western Isles, Y=Dumfries & Galloway, Z=Shetland).

Variable	**Coefficient**	**S.E.**
Intercept	4.299	0.399
HibLate	1.896	0.100
NoCar	-0.028	0.006
Level4	-0.017	0.007
HB_A	-1.443	0.248
HB_B	-1.634	0.334
HB_C	-0.508	0.263
HB_F	-1.012	0.287
HB_H	-1.833	0.289
HB_L	1.025	0.377
HB_N	-1.862	0.246
HB_R	-2.097	0.727
HB_S	-0.257	0.235
HB_T	-0.123	0.356
HB_V	-0.918	0.299
HB_W	-1.999	0.492
HB_Y	-1.509	0.332
HB_Z	-1.035	0.729

First of all the individual level variables are considered. The correlation between 'PCVMonthsLate' and 'HibMonthsLate' is 0.35. Thus, 'HibMonthsLate' does not appear to be particularly useful in determining 'PCVMonthsLate'. The mean number of months late the PCV-7 booster is administered does not appear to differ greatly by gender, with an average of 1.85 months for females and 1.88 months for males. A 95% confidence interval for the difference in the mean number of months late for males and females is (-0.012, 0.073). Since the interval

straddles zero, there is insufficient evidence to reject the null hypothesis that the mean number of months late differs by gender. The deprivation measure, 'SCSIMD5', also does not show great differences in 'PCVMonthsLate' between each quintile. The mean number of months late is 1.75 for category 1, 1.90 for category 2, 1.83 for category 3, 1.86 for category 4 and 1.95 for category 5. Therefore, it is not clear that 'SCSIMD5' is important in determining 'PCVMonthsLate'. However, there is a slight indication that the number of months late may perhaps increase with increasing deprivation as the lowest number of months late is 1.75 which is observed in the most affluent category and the highest number, 1.95, is seen in the most deprived category.

Next, the postcode district level variables are considered. Table 8.21 shows the correlations between each variable and 'PCVMonthsLate' at the individual level and at the spatial, postcode district, level. To consider the variables from the postcode district level perspective, 'PCVMonthsLate' is aggregated by postcode district and correlations between postcode district level 'PCVMonthsLate' and the postcode district level variables are assessed.

All individual level correlations between 'PCVMonthsLate' and the postcode district level variables are extremely small, see Table 8.21. The highest correlation is only 0.019 between 'OtherEU' and 'PCVMonthsLate'. The correlations over postcode district are also fairly small. The highest spatial correlation observed is for 'Agriculture' at 0.205, followed by 'Level4' at 0.165. It does not appear that the postcode district level variables are strongly associated with 'PCVMonthsLate'.

Table 8.24 shows the mean of 'PCVMonthsLate' for each of the HBs of residence. The lowest mean number of months late is 1.47 months, observed in Dumfries and Galloway; the highest is 2.44 in Tayside. A one-way ANOVA with response log(PCVMonthsLate) and HB as the explanatory variable shows a significant difference between at least two HBs (p-value < 0.001). The transformation of the response variable was used as the distribution of the number of months late was skewed.

Kernel density estimates are used to examine the differences in the distribution

Table 8.21: Individual level and spatial level correlations between postcode district level variables and the number of months late the PCV-7 booster is administered.

Variable	Aged0-4	NoCar	OtherEU	NotWhite	NoQual	Level4	LargeEmp
Individual level	-0.009	-0.005	0.019	0.003	-0.010	0.012	-0.004
Postcode district level	-0.101	-0.128	0.098	-0.060	-0.119	0.165	-0.040

Variable	Unemployed	Agriculture	Routine
Individual level	-0.007	0.006	0.014
Postcode district level	-0.138	-0.035	0.205

HB [a]	A	B	C	F	G	H	L	N	R	S	T	V
Mean	1.97	1.95	1.70	1.94	1.75	2.02	1.82	1.74	1.58	1.88	2.44	1.73
Standard deviation	1.76	1.80	1.72	1.80	1.76	1.91	1.73	1.67	1.27	1.67	1.65	1.70

HB	W	Y	Z
Mean	1.67	1.47	2.15
Standard deviation	1.65	1.37	1.82

Table 8.22: Mean number of months late the PCV-7 booster is administered for each Scottish HB.

[a]A=Ayrshire and Arran, B=Borders, C=Argyll and Clyde, F=Fife, G=Greater Glasgow, H=Highland, L=Lanarkshire, N=Grampian, R=Orkney, S=Lothian, T=Tayside, V=Forth Valley, W=Western Isles, Y=Dumfries and Galloway, Z=Shetland.

of the number of months late the PCV-7 booster is administered for each of the HBs. Figures 8.9 to 8.11 show these distributions. On examination of the KDE plots, most HBs show a peak shortly after 0 months late. However, Figure 8.10 show that Orkney has the greatest delay until the peak number of months late the vaccine is administered, followed by Tayside. However, Orkney displays a higher uptake around 0 months late than Tayside which shows a gradual incline to the peak at around 1 month late.

Assessment of Figure 8.9 shows that for HB A, Ayrshire and Arran, the majority of late vaccinations occur between the age of 13 months and 15 months. In the Borders, three peaks can be seen in the KDE. The first, largest peak occurs at around 0.5 months late, followed by a second smaller peak at around 1.5 months late and, finally, a much smaller peak can be observed around 3 months late. In HB C, Argyll and Clyde, the peak occurs at 0 months late and by 2 months late, the majority of late vaccinations have been administered. In Fife, HB F, a peak occurs at 0 months late followed by a second peak of roughly the same size at approximately 1 month late. Greater Glasgow and Highland, HBs G and H, show a similar picture to that of Argyll and Clyde with most late vaccinations administered by 2 months late.

The smaller HBs for the islands, such as Orkney, show more variability in the distributions due to the fact that a much smaller number of children are found in these HBs. Shetland, HB Z, shown in Figure 8.11, shows an initial peak just before 1 month late, followed by a subsequent much smaller peak at around 3.5 months late. The KDE for the Western Isles, again displayed in Figure 8.11, displays two peaks in the number of months late; the first large peak occurring just before 0.5 months late, the second at around 1.8 months late. Orkney, shown in Figure 8.10, shows the majority of late vaccinations are administered at around 1.2 months late, but all late vaccinations are administered within around 6 months of the age at which the PCV-7 booster should be received. Lanarkshire, HB L, in Figure 8.10 shows peak uptake around 0 months late which gradually tails off whilst the KDE for Grampian, HB N, shows a peak just before 1 month late. Lothian, HB S, and Forth Valley also have the highest uptake around 0 months late.

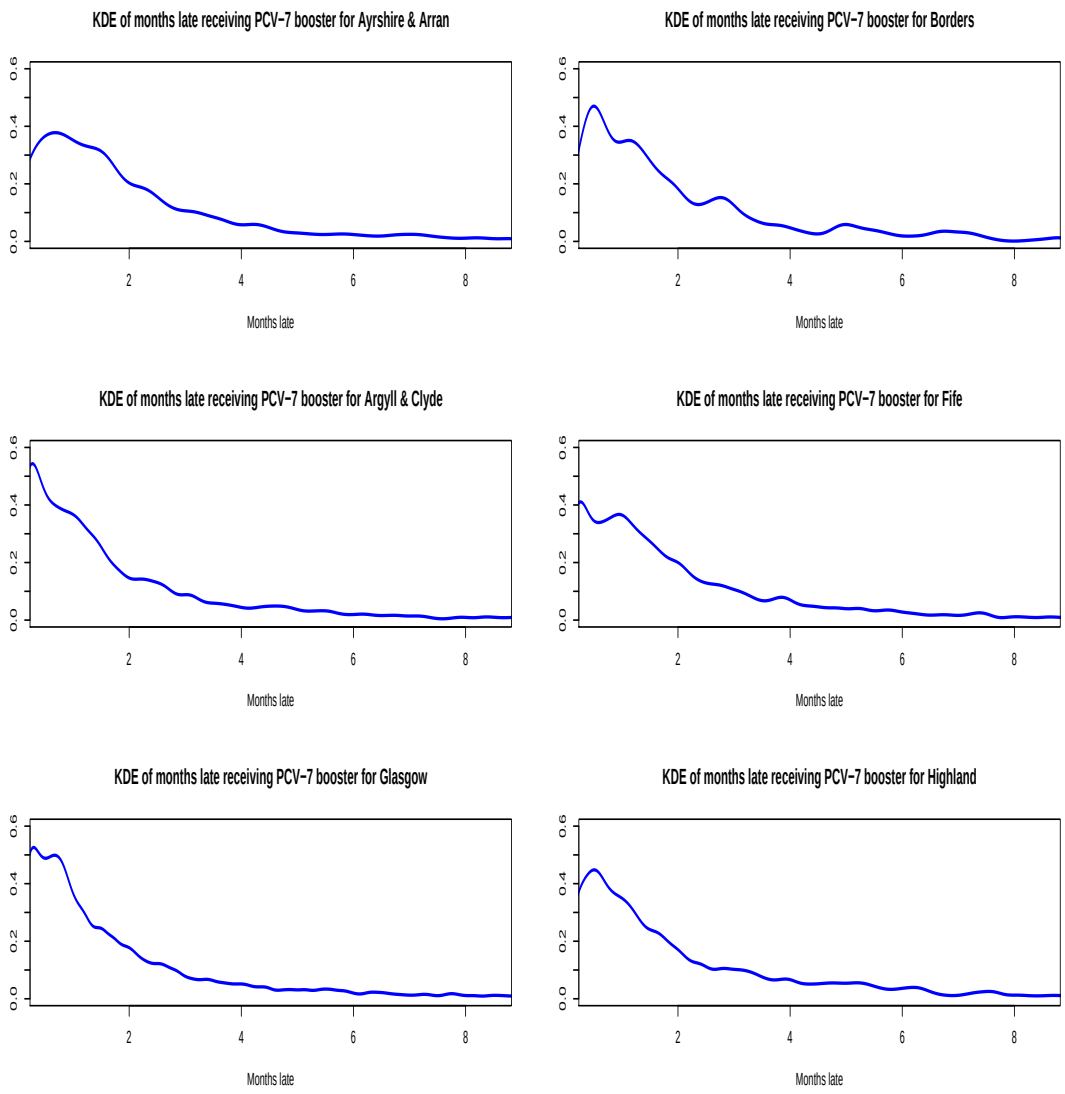

Figure 8.9: KDEs of number of months late the PCV-7 booster was administered in HBs A to H.

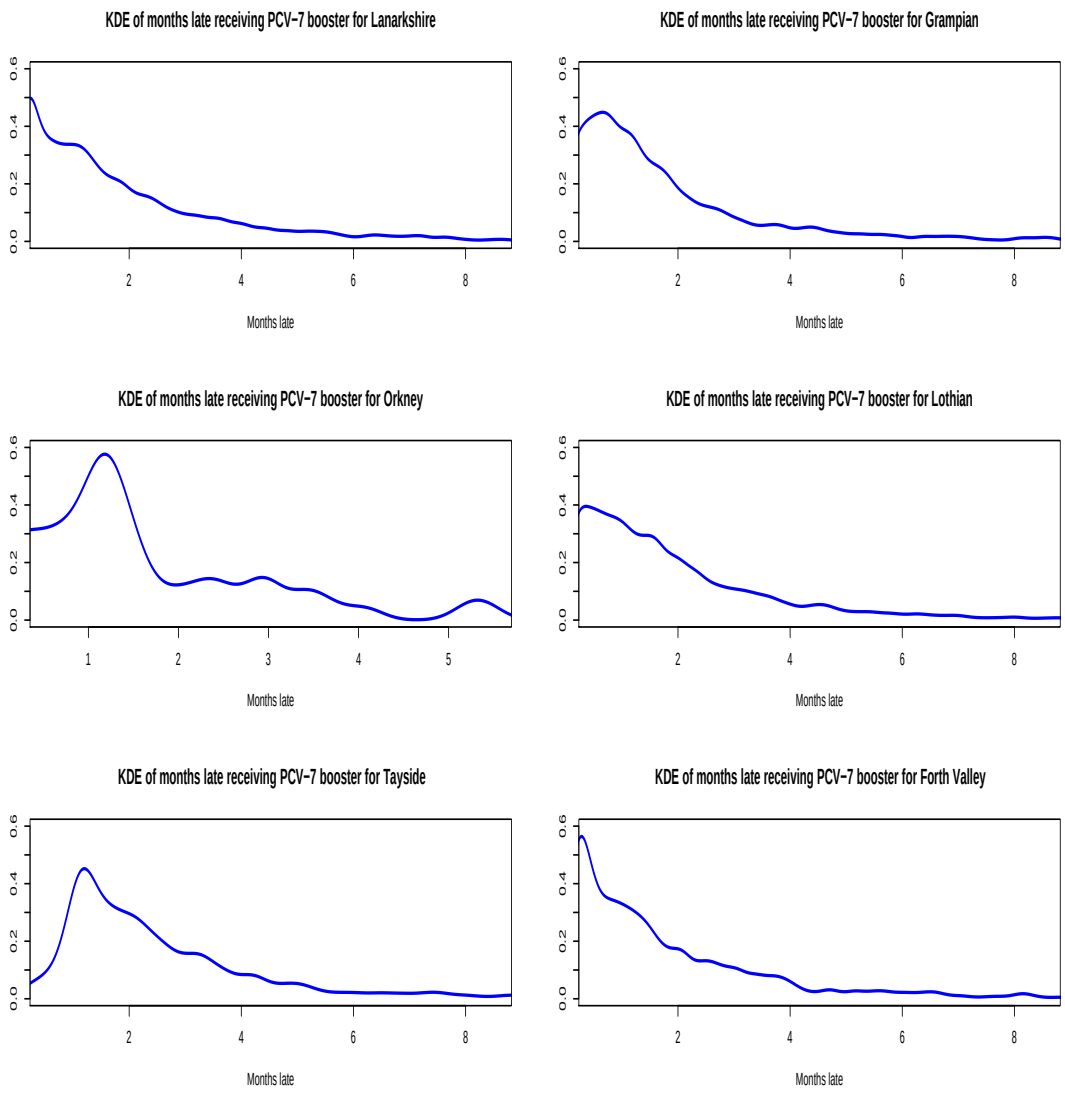

Figure 8.10: KDEs of number of months late the PCV-7 booster was administered in HBs L to V.

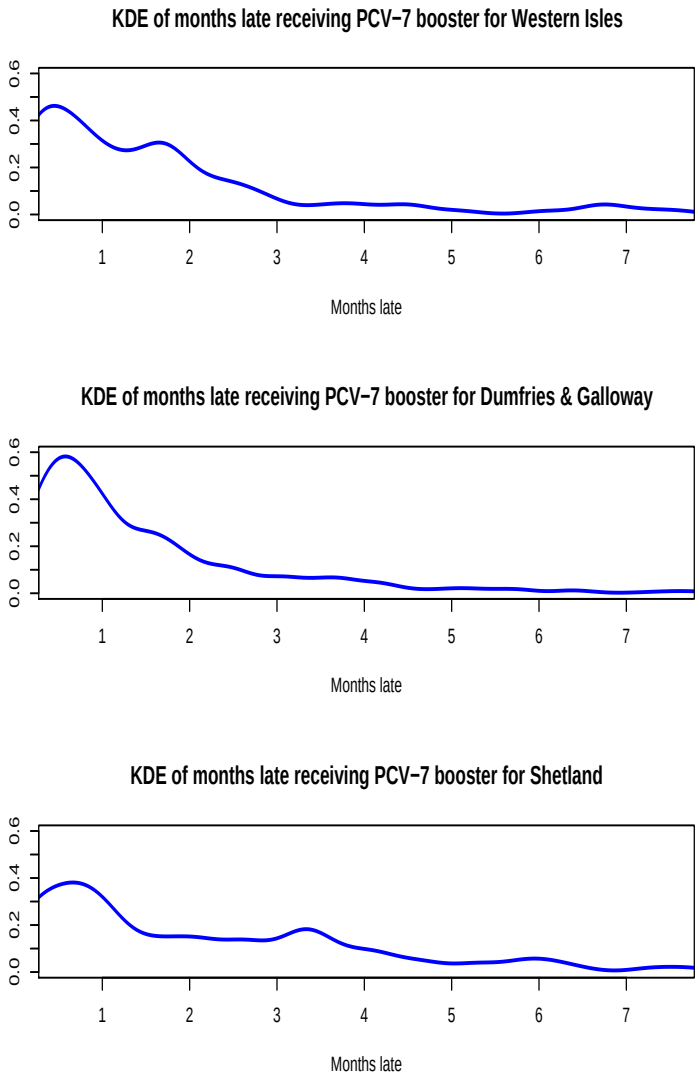

Figure 8.11: KDEs of number of months late the PCV-7 booster was administered in HBs W to Z.

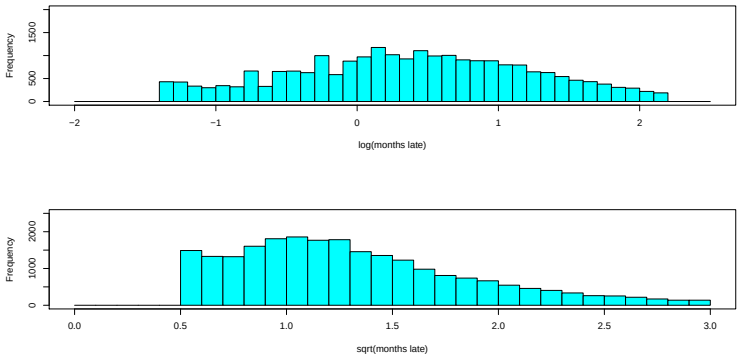

Figure 8.12: Histograms of log(PCVMonthsLate), and $\sqrt{\text{PCVMonthsLate}}$.

As the KDEs and histograms of 'PCVMonthsLate' shown in this section all appear skewed, it is likely that a transformation of 'PCVMonthsLate', such as the logarithm or square root of this variable, will be required to fulfill the necessary model assumptions. Figure 8.12 shows histograms of two possible transformations of 'PCVMonthsLate'. The transformations of 'PCVMonthsLate' shown in Figure 8.12 still appear to be a little skewed but are an improvement on the untransformed variables. The transformation log(PCVMonthsLate) appears to be the least skewed and thus will be considered in the modelling.

Modelling

Single-level models

As with the modelling approach used for the response variables *PCV-7 received* and *PCV-7 administered late*, single-level models involving each of the possible explanatory variables independently are fitted to discover which variables are significant in determining 'PCVMonthsLate'. A single-level model for this continuous variable is described using the following equation:

$$y_i = \alpha + \beta x_i + \epsilon_i. \tag{8.5}$$

In this model, y_i is log(PCVMonthsLate) for individual i, where $i = 1, ...,$

25,378 since 25,378 children in the data set received the PCV-7 booster by age 22 months and received it after 13 months of age; α is the intercept parameter and β is the slope parameter. x_i is the value of the explanatory variable x for individual i and ϵ_i is the residual error for individual i. ϵ_i is assumed to follow the Normal distribution with mean 0 and variance σ^2_ϵ. Table 8.23 shows the coefficients and S.E.s for each of the models of 'PCVMonthsLate' involving a single explanatory variable. In each of these models, and subsequent models with this response variable, iterative generalised least squares (IGLS) is used to obtain parameter estimates.

Table 8.23: Results from the single-level models of log(PCVMonthsLate).

Variable	**Coefficient**	**S.E.**	***p*-value**
HibMonthsLate	0.059	0.002	<0.001
Gender	0.017	0.011	0.127
Aged0-4	-0.005	0.008	0.553
NoCar	-0.001	0.0004	0.063
OtherEU	0.043	0.009	< 0.001
NotWhite	0.00001	0.002	0.996
NoQual	-0.002	0.0006	0.001
Level4	0.001	0.0006	0.024
LargeEmp	0.004	0.005	0.384
Routine	-0.001	0.002	0.416
Unemployed	-0.002	0.004	0.688
Agriculture	0.003	0.002	0.179
SCSIMD5_2	0.069	0.018	
SCSIMD5_3	0.025	0.018	
SCSIMD5_4	0.051	0.018	
SCSIMD5_5	0.080	0.017	< 0.001

Examination of Table 8.23 shows the significant variables for determining log(PCVMonthsLate) in the single explanatory variable models are 'HibMonthsLate', 'OtherEU', 'NoQual', 'Level4' and 'SCSIMD5'. The models of the sig-

nificant variables shown in Table 8.23 have R^2 values of 0.0432, 0.0009, 0.0005, 0.0002 and 0.0012. Thus, 'HibMonthsLate' appears to be the most important variable amongst those available in determining log(PCVMonthsLate), although this variable alone only explains 4.32% of the variability in the response variable. A model involving all of these variables gives an R^2 value of 0.0524. Thus, a total of 5.24% of the variability in log(PCVMonthsLate) is explained by these significant variables.

Table 8.24 shows the coefficient and S.E. for each HB using HB as a categorical variable in a model of log(PCVMonthsLate). As usual, HB G is the comparator.

Table 8.24: Results from the single-level model of log(PCVMonthsLate) by HB (A=Ayrshire & Arran, B=Borders, C=Argyll & Clyde, F=Fife, G=Greater Glasgow, H=Highland, L=Lanarkshire, N=Grampian, R=Orkney, S=Lothian, T=Tayside, V=Forth Valley, W=Western Isles, Y=Dumfries & Galloway, Z=Shetland).

HB	**Coefficient**	**S.E.**
A	0.155	0.024
B	0.084	0.041
C	0.009	0.024
F	0.142	0.024
H	0.133	0.032
L	0.099	0.021
N	0.030	0.022
R	0.011	0.122
S	0.128	0.019
T	0.427	0.023
V	0.025	0.027
W	-0.029	0.082
Y	-0.138	0.034
Z	0.223	0.087

In Table 8.24, two of the fourteen coefficients are negative, indicating that these HBs have a lower mean for log(PCVMonthsLate) than the baseline comparator

Greater Glasgow. These HBs with negative coefficients are the Western Isles and Dumfries and Galloway. All other HBs have a higher predicted mean than Greater Glasgow.

Two-level models

Next, two-level models are considered to examine the importance of each of the explanatory variables. The equations below describe the random intercept two-level model, or variance components model as it is also known:

$$\begin{aligned} y_{ij} &= \alpha_j + \beta x_{ij} + \epsilon_{ij}, \\ \alpha_j &= \alpha + \upsilon_j. \end{aligned} \tag{8.6}$$

In this model, y_{ij} in (8.6) is log(PCVMonthsLate) for individual i in postcode district j. x_{ij} corresponds to an individual level explanatory variable with fixed coefficient. As this is a random intercept model, α is indexed by postcode district. In (8.6), it can be seen that α_j can be expressed in terms of a fixed parameter, α, and a random parameter, υ_j. υ_j provides an estimate of the departure of each of the j postcode district intercepts from the fixed intercept. υ_j in this model is classed as a level two residual which is assumed to take the same value for all children in postcode sector j (Rasbash et al. 2008). Both υ_j and ϵ_{ij} are random components in this two-level model. Both are assumed to be Normally distributed with means equal to 0 and variances equal to σ_υ^2 and σ_ϵ^2 respectively. A further assumption is that these random residual parameters are uncorrelated. The parameter β in (8.6) remains fixed, as in (8.5).

To determine whether or not postcode district is important in explaining the variability in log(PCVMonthsLate) a random intercept model is fitted to the data. The fitted model is:

$$y_{ij} = \alpha_j + \epsilon_{ij},$$

$$\alpha_j = 0.379 + \upsilon_j. \tag{8.7}$$

The S.E. for the fixed part of α is estimated as 0.013. The estimate for σ_υ^2 is 0.038 and for σ_e^2 is 0.672. A Likelihood Ratio Test was carried out to test the importance of the random postcode district component. Using the approach described by Rasbash et al. (2008), this random-intercept model is compared to a model where $\sigma_\upsilon^2 = 0$, i.e. the single-level intercept only model. The difference in minus twice the log-likelihoods of these two models is equal to 591.12. This test statistic is compared to the χ^2 statistic with one degree of freedom as only one additional parameter is included in the variance components model (p-value < 0.001). Clearly, there is significant variation between postcode districts.

In the analysis, five of the postcode district levels are not considered in the model. From these five levels, four of the postcode districts have no children administered the PCV-7 booster (PA66, PA76, PH23 and PH5) and one has no children administered the vaccine late (KW6). Thus, these districts have no information on the number of months late the vaccine is administered.

The variance partition coefficient (VPC) is 0.054 (${\sigma^2}_\upsilon/(\sigma_\upsilon^2 + \sigma_\epsilon^2)$). Thus, only about 5.4% of the total variance in the mean of log(PCVMonthsLate) is attributable to differences between postcode districts which accounts for the low correlations observed between the postcode district level variables and 'PCV-MonthsLate'.

Three-level models

Next, a third level, HB, is considered. The three-level random intercept model can be expressed as follows:

$$\begin{aligned} y_{ijk} &= \alpha_{jk} + \beta x_{ijk} + \epsilon_{ijk}, \\ \alpha_{jk} &= \alpha + \nu_k + \upsilon_{jk}. \end{aligned} \tag{8.8}$$

It can be noted from comparison of (8.8) with the two-level random intercept model (8.6) that an additional random parameter has been included in the model, ν_k. This is the random departure from the fixed intercept, α of each of the HBs. Notice now that υ is indexed by both j and k to represent the random effect for postcode districts within HBs. As with ϵ_{ijk} (the individual level residual within postcode district within HB) and υ_{jk} (the postcode district level residual within HB), ν_k is assumed to follow a Normal distribution with 0 mean and constant variance, σ^2_ν. All residual parameters are assumed uncorrelated.

Fitting the three-level random intercept model to the vaccine uptake data gives the following parameter estimates:

$$\begin{aligned} y_{ijk} &= \alpha_{jk} + \epsilon_{ijk}, \\ \\ \alpha_{jk} &= 0.373 + \nu_k + \upsilon_j k. \end{aligned} \tag{8.9}$$

The S.E. of α in (8.9) is 0.033, more than twice the S.E. for the fixed intercept of (8.7). The estimates of the three residual components, σ^2_ν, σ^2_υ and σ^2_ϵ are 0.013 (S.E. 0.006), 0.022 (S.E. 0.003) and 0.673 (S.E. 0.006), respectively. The estimate for σ^2_υ has decreased in size by the addition of the random component for HB residuals. However, the estimate for σ^2_ϵ remains the same. A Wald Test of $H_0 : \sigma^2_\nu = 0$ gives a p-value of 0.025. Thus, there is evidence to suggest that there is significant variability in log(PCVMonthsLate) between HBs. Therefore, the random effect of HB should be retained in the model.

Considering the VPCs, the proportion of the variance which is attributable to differences between HBs, the intra-HB correlation, is 0.018 ($\sigma^2_\nu/(\sigma^2_\nu + \sigma^2_\upsilon + \sigma^2_\epsilon)$). This can be considered as a measure of the similarity of individuals within the same HB. Thus, 1.8% of the variation is at the HB level. The proportion of the variance which is due to differences between postcode districts, the intra-postcode district correlation, equals 0.049 ($(\sigma^2_\nu+\sigma^2_\upsilon)/(\sigma^2_\nu+\sigma^2_\upsilon+\sigma^2_\epsilon)$). This can be considered as a measure of the similarity of individuals within the same postcode district within the same HB. 4.9% of the variation is at the postcode district and HB level.

The similarity of postcode districts within the same HB can be measured using the formula shown in (8.10).

$$\frac{\sigma_\nu^2}{\sigma_\nu^2 + \sigma_\upsilon^2}. \tag{8.10}$$

A similarity measure of 37.14% was found for this model. Thus, knowing the number of months late the PCV-7 booster is received in one postcode district is relatively informative for determining the number of months late the PCV-7 booster is received in another postcode district within the same HB.

As both postcode district and HB have been deemed significant in explaining variability in log(PCVMonthsLate), the next stage in the modelling process is to add potentially important explanatory variables to the model. From the previous single-level modelling, 'HibMonthsLate' appears to be significant in determining log(PCVMonthsLate). This variable was added to the three-level variance components model. The parameter estimates obtained are shown in (8.11).

$$\begin{aligned} y_{ijk} &= \alpha_{jk} + 0.065 x_{1ijk} + \epsilon_{ijk}, \\ \\ \alpha_{jk} &= 0.486 + \nu_k + \upsilon_{jk}. \end{aligned} \tag{8.11}$$

As 'HibMonthsLate' is an individual level variable, the coefficient of 'HibMonthsLate' (x_{1ijk}) is similar to that obtained in the single-level model and the S.E. is the same. Thus, this variable is significant in determining log(PCVMonthsLate) in the three-level model. The estimate of the variance parameter for the random effect of HB, σ_ν^2, is 0.018 (S.E. 0.007) which is slightly larger than the parameter estimate obtained in the three-level intercept only model. The estimate for the variance parameter for the postcode district random effect is 0.012 (S.E. 0.002), which is lower than the 0.038 estimate for the random intercept model. This implies that some of the variability between postcode districts within HBs has been explained by the inclusion of the variable 'HibMonthsLate'. Similarly, the individual level residual variance estimate is reduced from 0.672 to 0.476 (S.E. 0.006) by the addition of this variable in the three-level model. Therefore, this variable also explains some of the variability between individuals within postcode

district in HB. The Likelihood Ratio Test can be employed to test the significance of the inclusion of this variable in the model. This model is compared to the three-level variance components model and a test statistic of 29698.54 is obtained which is clearly highly significant when compared to a χ^2 statistic with one degree of freedom. Thus, 'HibMonthsLate' should be retained in the model. A Wald Test of $H_0 : \sigma_\nu^2 = 0$ shows that there is still significant variability between the HBs even after adjusting for the number of months late the Hib/Men C booster is administered (p-value 0.019). Next, the possibility that the coefficient of 'HibMonthsLate' varies randomly with postcode district and HB is explored. This random coefficient model can be described as follows:

$$y_{ijk} = \alpha_{jk} + \beta_{jk} x_{1ijk} + \epsilon_{ijk},$$

where

$$\alpha_{jk} = \alpha + \nu_{0k} + \upsilon_{0jk},$$

and

$$\beta_{jk} = \beta + \nu_{1k} + \upsilon_{1jk}. \tag{8.12}$$

The random variance and covariance parameters are as described in (8.4). The parameters α and β are estimated to be 0.370 (S.E. 0.045) and 0.142 (S.E. 0.025) respectively in this model. The estimates of each of the variance and covariance parameters are shown:

$$\Omega_\nu = \begin{bmatrix} 0.025 \text{ (S.E. } 0.011) & -0.010 \text{ (S.E. } 0.005) \\ -0.010 \text{ (S.E. } 0.005) & 0.009 \text{ (S.E. } 0.003) \end{bmatrix}$$

and

$$\Omega_\upsilon = \begin{bmatrix} 0.018 \text{ (S.E. } 0.003) & -0.004 \text{ (S.E. } 0.001) \\ -0.004 \text{ (S.E. } 0.001) & 0.002 \text{ (S.E. } 0.0004) \end{bmatrix}.$$

The estimate for σ_e^2 is 0.446 (S.E. 0.006). The addition of the random slope to the model has reduced the residual variance from 0.476 to 0.446. Comparing this model with the three-level random intercept model with fixed slope for 'HibMonthsLate' gives a change in deviance of 578.67. The change in deviance should be compared to a χ^2 distribution with four degrees of freedom as there are four new parameters in (8.12). This is highly significant (p-value < 0.001). Thus, the random slope model is preferable to the random intercept model.

Further explanatory variables deemed significant in the single-level models of log(PCVMonthsLate) are added to the model. From the single-level models of each explanatory variable used independently to predict 'PCVMonthsLate', results shown in Table 8.23, 'SCSIMD5' is found to be significant. On adding this variable to the model, assuming the coefficient of 'SCSIMD5' does not vary by postcode district and HB, a change in deviance of 99.69 is obtained. This is significant when compared to the χ^2 statistic with four degrees of freedom for the four new dummy variables entered for 'SCSIMD5'. Therefore, 'SCSIMD5' should be retained in the model. Fitting the other variables in order of significance in the single-level model in the three-level model one by one and examining deviance statistics results in a final model involving 'HibMonthsLate', 'SCSIMD5' and 'NoQual' shown in Table 8.25.

Table 8.25: Results for the fixed effects of the three-level model of 'PCVReceived'.

Variable	**Coefficient**	**S.E.**
Intercept	0.458	0.026
HibMonthsLate	0.143	0.059
SCSIMD5_2	0.044	0.020
SCSIMD5_3	0.031	0.021
SCSIMD5_4	0.036	0.022
SCSIMD5_5	0.083	0.023
NoQual	-0.004	0.001

The estimates of the residual variance parameters are:

$$\Omega_\nu = \begin{bmatrix} 0.025 \text{ (S.E. } 0.011) & -0.011 \text{ (S.E. } 0.005) \\ -0.011 \text{ (S.E. } 0.005) & 0.009 \text{ (S.E. } 0.004) \end{bmatrix}$$

and

$$\Omega_\upsilon = \begin{bmatrix} 0.017 \text{ (S.E. } 0.003) & -0.004 \text{ (S.E. } 0.001) \\ -0.004 \text{ (S.E. } 0.001) & 0.002 \text{ (S.E. } 0.0004) \end{bmatrix}.$$

These estimates are almost identical to those obtained for the three-level model with random intercept term and 'HibMonthsLate' as the only explanatory variable, with a coefficient which varies randomly with HB. The fixed part of the coefficient of 'HibMonthsLate' in this final model is positive, indicating that log(PCVMonthsLate) increases as the number of months late the Hib/Men C booster is administered increases. The negative coefficient of 'NoQual' suggests that as the percentage aged 16-74 years with no qualifications in a postcode district increases, the number of months late the PCV-7 booster decreases.

The model assumptions must be verified. These are that the residuals for the three-levels each follow a Normal distribution with zero mean and constant variance.

Figures 8.13 and 8.14 can be used to assess the model assumptions for the individual level residuals, ϵ_{ijk}. The Normality plot, Figure 8.13, shows a slight curvature with some departures from a straight line. However, the assumption of normality appears reasonable. The plot of residuals against fitted values, Figure 8.14, shows the assumptions of zero mean and constant variance are questionable as there appears to be a slight impression of a negative slope to the plot. This suggests that the model is under-predicting at long delays in vaccine uptake and over-predicting at shorter delays.

The assumptions that the postcode district level residuals follow a Normal distribution with zero mean and constant variance all appear reasonable from examination of Figures 8.15 and 8.16. The Normality plot, Figure 8.15 shows a

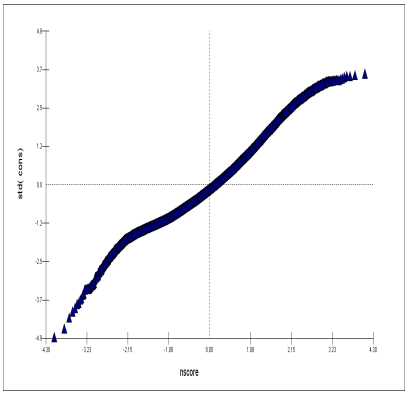

Figure 8.13: Normality plot of individual level residuals for final model of log(PCVMonthsLate).

relatively straight line and the plot of residuals against fitted values, Figure 8.16, shows a random scatter of points around zero.

From examination of the plots for the HB level residuals, Figures 8.17 and 8.18, the model assumptions appear reasonable for these residuals.

HB may be considered as a fixed effect in this model for the continuous variable log(PCVMonthsLate) with HB. The results are shown in Table 8.26 where it can be observed that all coefficients are positive. Therefore, all HBs have a higher estimated mean number of months late than the comparator Greater Glasgow. The HB with the highest mean number of months late is Tayside, HB T, with the largest estimated coefficient.

In conclusion, as with the other univariate models discussed, significant random effects are found for both HB and postcode district in the model of 'PCVMonthsLate'. The final three-level model involved the variables 'HibMonthsLate', 'SCSIMD5' and the postcode district level variable 'NoQual'. There is significant evidence that a random coefficient is required for 'HibMonthsLate'. Thus, the number of months late the Hib/Men C booster is received is allowed to vary by postcode district within HB. The fixed coefficient of 'HibMonthsLate' is positive. Thus, the mean number of months late the PCV-7 booster is administered increases as the number of months late the Hib/Men C booster is administered

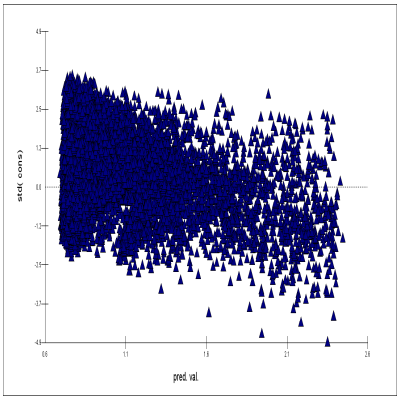

Figure 8.14: Individual level residuals by fitted values for final model of log(PCVMonthsLate).

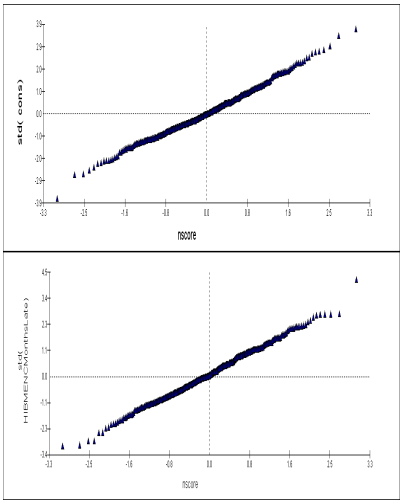

Figure 8.15: Normality plots for the random intercept & coefficient of 'HibMonthsLate' for PCDIST level residuals for final model of log(PCVMonthsLate).

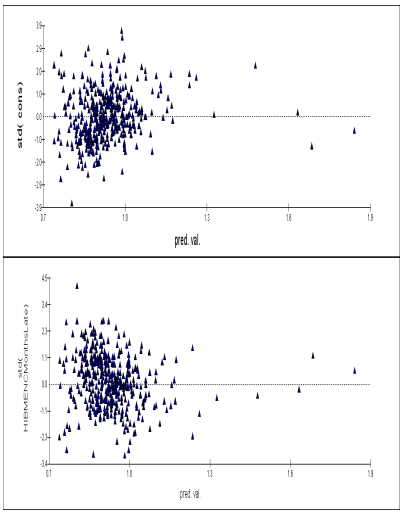

Figure 8.16: PCDIST level residuals by fitted values for the random intercept & random coefficient of 'HibMonthsLate' in final model of log(PCVMonthsLate).

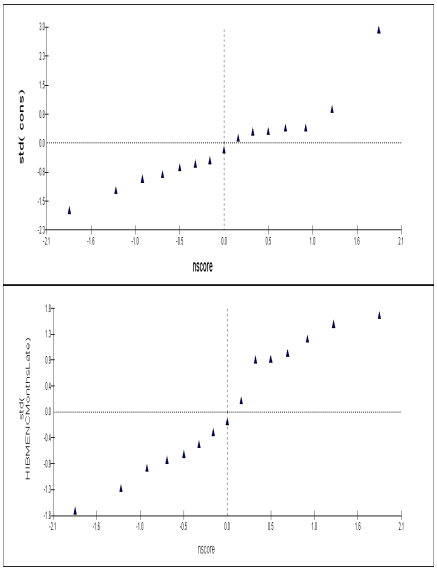

Figure 8.17: Normality plots for the random intercept & coefficient of 'HibMonthsLate' for HB level residuals for final model of log(PCVMonthsLate).

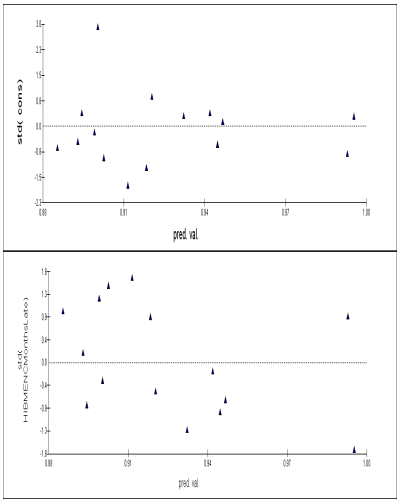

Figure 8.18: HB level residuals by fitted values for the random intercept & random coefficient of 'HibMonthsLate' in final model of log(PCVMonthsLate).

increases. The coefficients of each of the levels of 'SCSIMD5' are positive, with the coefficients of both category 2 and 3 found to be equal. Thus, given the other variables included in the model, there does not appear to be a difference in the mean number of months late the PCV-7 booster is administered for these two classes compared to the baseline category 1, least deprived. The coefficient of category 5, most deprived, is highest compared to the baseline category. Thus, the greatest difference in the mean number of months late appears to be between those least and most deprived, with those most deprived displaying a greater mean number of months late than those least deprived. The coefficient of 'NoQual' is negative. Thus, this suggests that the mean number of months late the PCV-7 booster is administered decreases with increasing percentages of individuals with no qualifications within postcode districts.

Conclusions from the univariate models of PCV-7 uptake and timing

To summarise the results from the univariate models of 'PCVReceived', 'PCVLate' and 'PCVMonthsLate', all three models had significant random effects attributable

Table 8.26: Results from the two-level model of log(PCVMonthsLate) with a random slope for 'HibMonthsLate' and HB as a fixed effect (A=Ayrshire & Arran, B=Borders, C=Argyll & Clyde, F=Fife, G=Greater Glasgow, H=Highland, L=Lanarkshire, N=Grampian, R=Orkney, S=Lothian, T=Tayside, V=Forth Valley, W=Western Isles, Y=Dumfries & Galloway, Z=Shetland).

Variable	**Coefficient**	**S.E.**
Intercept	0.293	0.050
HibMonthsLate	0.116	0.007
$SCSIMD5_2$	0.041	0.020
$SCSIMD5_3$	0.030	0.021
$SCSIMD5_4$	0.031	0.022
$SCSIMD5_5$	0.081	0.023
NoQual	-0.004	0.001
HB_A	0.205	0.038
HB_B	0.305	0.060
HB_C	0.113	0.041
HB_F	0.286	0.043
HB_H	0.082	0.046
HB_L	0.192	0.039
HB_N	0.130	0.037
HB_R	0.147	0.133
HB_S	0.282	0.035
HB_T	0.481	0.041
HB_V	0.302	0.046
HB_W	0.181	0.111
HB_Y	0.014	0.054
HB_Z	0.250	0.117

to variations between postcode district and HB. Thus, it appears that area is important in determining patterns in vaccine uptake and timing. In addition, the models of both 'PCVReceived' and 'PCVMonthsLate' involve random components in the coefficients of the variables associated with uptake of the Hib/Men

C booster; 'HibReceived' in the first case and 'HibMonthsLate' in the second. No evidence was found that the variable 'HibLate' varied randomly by postcode district and HB in the model of 'PCVLate'. However, Hib/Men C booster uptake and timing was found to be significant in each of the three models as expected since this vaccine is scheduled to be administered not long prior to the PCV-7 booster. Those who have received the Hib/Men C booster have a higher odds of receiving the PCV-7 booster; those who receive the Hib/Men C booster late are at an increased odds of receiving the PCV-7 booster late.

Deprivation was identified as a significant variable in both the model for 'PCVReceived' and the model for 'PCVMonthsLate'. The results suggest that the odds of receiving the vaccine decrease with increasing levels of deprivation and that the number of months late the vaccine is received is greatest for the most deprived. In contrast, no association was found between deprivation and the binary variable 'PCVLate'.

Educational level was found to be statistically significant in determining late vaccine uptake. However, contradictory results were obtained. The model of the binary response of late uptake with 13 month cut-off showed the odds of late uptake decrease as the percentage of individuals with the highest level of education increases within a postcode district. This result is in agreement with results obtained regarding parental education levels in studies of vaccine delay discussed previously (Bobo et al. 1993; Dombkowski et al. 2004; Ozcirpici et al. 2006; Torun and Bakırcı 2006; Datar et al. 2007). However, the logistic regression model with 14 month cut-off for late uptake resulted in a positive estimated coefficient for 'Level4', contradicting the result from the previous model. The model of log('PCVMonthsLate') had a negative coefficient for the no qualifications variable, indicating that the number of months late the vaccine is administered decreases with increasing proportions of low educational attainment within an area. It is worth noting once again the importance of considering the possibility of ecological fallacy in the analysis. Individual level information on parental educational status would have been of higher value in establishing whether or not a relationship exists between educational level and timing of vaccine.

Country of birth was identified as significant in determining whether or not the vaccine is administered with lower odds of receiving the PCV-7 booster as the proportion of individuals within a postcode district born in EU countries outwith the UK and Ireland increases.

'NoCar', a proxy for income within a district, was only found to be significant in one of the univariate models- the 13 month cut-off model for 'PCVLate'. The coefficient is negative indicating that the odds of receiving the PCV-7 booster late decrease as the proportion of households without a car within a postcode district increases.

8.4.6 Modelling the multivariate response combining PCV-7 received and PCV-7 administered late

In this section of the chapter, a model involving a combination of two of the univariate responses discussed previously will be considered. This model involves both the response 'PCVReceived' and 'PCVLate'. This multivariate response, 'PCVUptake', has three categories; the first category refers to those children who have not received the PCV-7 booster by 22 months of age, the second refers to children who receive the vaccine on time and the third refers to those who receive the vaccine late, i.e. after 13 months of age. Thus, a multinomial logistic regression analysis, such as that adopted by Akmatov et al., is appropriate to use to determine significant explanatory variables for this response. The cut-off of 14 months for late vaccinations is also considered.

The models discussed here take into account the variables found to be significant in determining the original binary responses separately. Both original models involve the random effects of postcode district and HB. The significant individual level and postcode district level variables considered are 'HibReceived', 'HibLate', 'SCSIMD5', 'OtherEU', 'NoCar' and 'Level4'. 'HibReceived' and 'HibLate' may be combined in the same manner as 'PCVReceived and 'PCVLate' for inclusion in this model. As 'Level4' is considered in this model, the correlated 'NoQual' is not considered.

The combined model is considered to examine the effects of several factors on vaccine uptake and timing. This type of modelling approach in a multi-level framework has been looked at previously with data involving survey participation (Durrant and Steele 2009). In this analysis, Durrant and Steele were interested in determining the variables that affect whether or not individuals respond to a survey, as well as the variables that determine why an individual may not respond. The two possibilities for non-response considered are non-contact, where an eligible household cannot be contacted, and refusal, where the household refused to be interviewed following contact. Thus, the response variable considered distinguishes between two possible mechanisms of non-response just as the analysis considered in this section distinguishes between two different mechanisms of vaccine receipt: on time or late. In the Durrant and Steele analysis, these two processes of non-response are contrasted with cooperation in the survey, whilst in the vaccine uptake analysis carried out in this section both mechanisms of vaccine receipt are contrasted with not received.

Durrant and Steele state that the benefit of this type of analysis, combining the different categories of non-response with the response category, over using separate logistic regression is that the effects of the variables on the probability of being in either of the non-response categories can be evaluated simultaneously and tested for equivalence. Thus, in this vaccine uptake analysis, the probabilities of receiving the vaccine on time and late may be considered in this manner using the approach adopted by Durrant and Steele.

Descriptive analysis

As both unvaccinated and vaccinated children are considered in the response variable, all data can be used in the analysis. As mentioned previously, 2,605 (9.09%) children did not receive the vaccine. Considering the 13 months of age cut-off for late PCV-7 booster uptake, 676 (2.36%) received the vaccine on time and 25,391 (88.56%) received the vaccine late. For the 14 months of age cut-off, 10,493 (36.60%) received the vaccine on time and 15,574 (54.32%) received it late.

Tables 8.27 and 8.28 provide information on each of the response categories for each of the potential explanatory variables for both the 13 months and 14 months

cut-offs for late uptake. Focussing on 'HibUptake', from examination of Tables 8.27 and 8.28 it can be observed that for the 13 month cut-off a high percentage, 93.22%, of children who received the Hib/Men C booster late received the PCV-7 booster late, as observed previously. However, the relative change in the timing of the vaccines is not so apparent for the 14 month cut-off. This is likely attributable to the fact that, although a less stringent cut-off is used for late classification of PCV-7 booster, the strict cut-off of 12 months for late classification of the Hib/Men C booster is used in both models. Concerning deprivation, as observed previously, the percentage who do not receive the vaccine increases with increasing deprivation. No clear trends can be seen for deprivation for the other two categories of response for the 14 month cut-off. However, for the 13 month cut-off, it appears that the percentage of late vaccinations decreases as the level of deprivation increases.

Table 8.27: Percentage distribution for each category of uptake for the PCV-7 booster using the 13 month cut-off for the potential explanatory variables.

Variable		**Not received**	**On time**	**Late**	**Total**
Individual level					*Number*
HibUptake (%)	Not received	47.14	2.74	50.12	2815
	On time	2.70	9.42	87.89	2015
	Late	5.07	1.71	93.22	23545
SCSIMD5 (%)	1	6.79	2.83	90.38	5048
	2	7.94	2.25	89.81	5413
	3	9.12	2.42	88.47	5513
	4	9.63	2.07	88.30	5777
	5	10.98	2.31	86.72	6624
Postcode level					
OtherEU (mean)		0.91	0.89	0.92	
NoCar (mean)		28.94	28.12	27.28	
Level4 (mean)		20.52	20.61	21.20	
Total		2553	670	25152	28375

Table 8.28: Percentage distribution for each category of uptake for the PCV-7 booster using the 14 month cut-off for the potential explanatory variables.

Variable		**Not received**	**On time**	**Late**	**Total**
Individual level					*Number*
HibUptake (%)	Not received	47.14	11.30	41.56	2815
	On time	2.70	65.91	31.39	2015
	Late	5.07	37.11	57.81	23545
SCSIMD5 (%)	1	6.79	38.87	54.34	5048
	2	7.94	35.11	56.95	5413
	3	9.12	37.44	53.45	5513
	4	9.63	36.81	53.56	5777
	5	10.98	35.36	53.67	6624
Postcode level					
OtherEU (mean)		0.91	0.89	0.92	
NoCar (mean)		28.94	28.12	27.28	
Level4 (mean)		20.52	20.61	21.20	
Total		2553	10399	15423	28375

Modelling

Single-level models

First, single-level models are examined. Let y_i be the outcome for individual i. Then y_i is coded as follows:

$$y_i = \begin{cases} 0 & \text{for not received,} \\ 1 & \text{for received on time,} \\ 2 & \text{for received late.} \end{cases}$$

The probability of obtaining each response may be denoted $\pi_i^{(s)}$ where $s = 0, 1, 2$. The baseline category for comparison is not received. Thus, the multinomial model for $s = 1, 2$, with $x_{ij}^{(s)}$ and $\beta^{(s)}$ representing the vectors of explanatory vari-

ables and coefficients respectively, may be expressed using the following equation:

$$\log\left(\frac{\pi_i^{(s)}}{\pi_i^{(0)}}\right) = \beta^{(s)^T} x_{ij}^{(s)}. \tag{8.13}$$

The model shown in (8.13) represents two simultaneous equations. When $s = 1$, the logarithm of the ratio of the probability of receiving the PCV-7 booster on time to the probability of not receiving the vaccine is represented and when $s = 2$, the logarithm of the ratio of the probability of receiving the vaccine late to that of not receiving the vaccine is represented. In the study by Akmatov et al., the baseline category in the multinomial logistic regression is vaccination received according to schedule. However, a different baseline category is chosen in this analysis as interest lies in the contrasts between those who do not receive the vaccine by age 22 months and those who receive the vaccine on time or late.

Using R version 2.9.1, a multinomial model was fitted to the response 'PCVUptake' using the multinom function from the nnet library. The variables deemed significant in the separate binary logistic regression models and a combined variable defining the Hib/Men C booster uptake as '0' for not received, '1' for received on time and '2' for received late, as adopted for 'PCVUptake' were included in the model. To determine whether or not each of the variables is required in the model, stepwise model selection procedures were used to identify the final model, similar to those used in the HES analysis in Chapter 5. In order to use stepAIC in the modelling, all missing data must first be omitted. Thus, 28,375 individuals are considered in the modelling in this section. The function dropterm in R is also used to examine the significance of each of the variables using the χ^2 Test. All explanatory variables considered are found to be significant and the results from the single-level multinomial model are displayed in Table 8.29. In Table 8.29, the subscript 1 denotes the category of 'PCVUptake' for PCV-7 booster received on time, 2 denotes that the booster was received late. Not received is the comparator for 'HibUptake' and category 1 is the comparator for 'SCSIMD5'.

The variable 'Level4' is not significant in a model of 'PCVUptake' with the 14

month cut-off for late uptake. As mentioned previously, the benefit of using a multinomial model combining the response categories is that the effects of each of the explanatory variables included in the model can be compared for each response.

Table 8.29: Results from the multinomial model of 'PCVUptake', with not received as the baseline category for comparison. The subscript 1 denotes on time and 2 denotes late.

	13 months		**14 months**	
Variable	**Coefficient**	**S.E.**	**Coefficient**	**S.E.**
Intercept_1	-1.849	0.271	-0.925	0.115
Intercept_2	0.666	0.136	0.342	0.100
$\text{HibUptake}_{\text{On time}_1}$	4.066	0.194	4.614	0.152
$\text{HibUptake}_{\text{On time}_2}$	3.392	0.142	2.553	0.146
$\text{HibUptake}_{\text{Late}_1}$	1.738	0.132	3.436	0.071
$\text{HibUptake}_{\text{Late}_2}$	2.850	0.049	2.553	0.051
SCSIMD5_{21}	-0.368	0.151	-0.336	0.088
SCSIMD5_{22}	-0.235	0.084	-0.166	0.084
SCSIMD5_{31}	-0.395	0.151	-0.385	0.087
SCSIMD5_{32}	-0.359	0.084	-0.312	0.084
SCSIMD5_{41}	-0.469	0.159	-0.396	0.088
SCSIMD5_{42}	-0.343	0.086	-0.280	0.085
SCSIMD5_{51}	-0.247	0.166	-0.471	0.093
SCSIMD5_{52}	-0.374	0.090	-0.281	0.089
OtherEU_1	0.265	0.877	-0.367	0.039
OtherEU_2	-0.253	0.051	-0.236	0.036
NoCar_1	-0.015	0.005	0.004	0.002
NoCar_2	-0.0002	0.002	-0.001	0.002
Level4_1	-0.023	0.008	-	-
Level4_2	-0.005	0.004	-	-

Formal tests to assess these effects have not been carried out in this chapter. However, informal tests may be carried out by simply examining and comparing

the coefficients of the explanatory variables. For example, in the 13 month cut-off model shown in Table 8.29, it can be observed that although the coefficients of the explanatory variables show differences when comparing received on time or received late to not received they generally have the same sign so the effect of each variable is the same on each response category compared to the baseline. The only variable that is different for each of the response categories compared to not received is 'OtherEU' where one coefficient is negative, the other positive. Similarly, for the 14 month cut-off it can be observed that the coefficients of the explanatory variables generally have the same sign for comparing received on time or received late to not at all. Only 'NoCar' displays differences when comparing received on time to not received and received late to not received. However, it is of interest to note that in both the 13 month and 14 month cut-off models the 'NoCar' explanatory variable, one of the coefficients is not statistically significantly different to zero.

The ORs for each of the levels of 'HibUptake' for the multinomial model with 13 month cut-off and that with 14 month cut-off can be observed in Table 8.30. The greatest OR of 100.89 is observed for the 14 month cut-off model, comparing the odds of receiving the Hib/Men C booster on time (OT) to being unvaccinated with the Hib/Men C booster (Unv) for the model of PCV-7 booster received on time compared to not being received. For the 13 month cut-off model, this OR is also very high at 58.32. This makes sense as it has already been shown that there are strong associations between receipt and timing of the PCV-7 booster with the receipt and timing of the Hib/Men C booster.

Table 8.30: ORs for each Hib/Men C booster uptake category for the single-level multinomial models of 'PCVUptake' with 13 and 14 month cut-offs for late uptake.

	13 month cut-off				**14 month cut-off**			
Hib	**OT vs. Unv**		**L vs. Unv**		**OT vs. Unv**		**L vs. Unv**	
	β	exp(β)	β	exp(β)	β	exp(β)	β	exp(β)
Unv	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
OT	4.066	58.323	3.392	29.725	4.614	100.887	2.553	12.846
L	1.738	5.686	2.850	17.288	3.436	31.062	2.553	12.846

Table 8.31 shows the ORs for each category of deprivation for the model with 13 month cut-off for late vaccination and the model with 14 month cut-off.

Table 8.31: ORs for each deprivation category for the single-level multinomial models of 'PCVUptake' with 13 and 14 month cut-offs for late uptake.

	13 month cut-off				**14 month cut-off**			
Dep.	**OT vs. Unv**		**L vs. Unv**		**OT vs. Unv**		**L vs. Unv**	
	β	exp(β)	β	exp(β)	β	exp(β)	β	exp(β)
1	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
2	-0.368	0.692	-0.235	0.791	-0.336	0.715	-0.166	0.847
3	-0.395	0.674	-0.359	0.698	-0.385	0.680	-0.312	0.732
4	-0.469	0.626	-0.343	0.710	-0.396	0.673	-0.280	0.756
5	-0.247	0.781	-0.374	0.688	-0.471	0.624	-0.281	0.755

From the ORs shown in Table 8.31, it can be observed that the probability of receiving the vaccine on time rather than not at all decreases as deprivation increases from category 1 to 4 for both the 13 month cut-off model and the 14 month cut-off model. However, this pattern continues to the lowest deprivation level, category 5, for the 14 month cut-off model but not for the 13 month cut-off

model, where the probability of receiving the vaccine on time rather than not at all increases. Comparing those who receive the vaccine late (L) to those who are unvaccinated, the probability of receiving the vaccine late decreases as deprivation increases from level 1 to level 2 for both models. However, the magnitude of the estimates of the other deprivation categories compared to deprivation level 1 do not show great differences.

On consideration of the other variables in the 13 month and 14 month cut-off models, for the 13 month cut-off, 'OtherEU' has a negative coefficient for the comparison of receiving the vaccine late and not receiving the vaccine. Thus, it appears that the odds of receiving the PCV-7 booster late decrease with increasing percentages of individuals born in other EU countries compared to not receiving the vaccine. However, 'OtherEU' is not significant for the 13 month cut-off model on comparison of on time to not at all. This is not the case for the 14 month cut-off, where 'OtherEU' is significant for both comparisons and both coefficients of 'OtherEU' found to be negative. The S.E.s for 'OtherEU' in the 14 month cut-off are similar due to the fact that the numbers of children falling into each of the PCV-7 response categories are more evenly split for this cut-off.

Considering 'NoCar', the coefficients are both negative in the 13 month cut-off model. However, 'NoCar' is not significant in the model comparing late uptake to not received. Thus, only the comparison of PCV-7 booster received on time and PCV-7 booster received late should be assessed for 'NoCar'. The results show that the odds of receiving the vaccine on time to not at all decrease as the percentage of households without a car increases within a postcode district. This seems reasonable since having a car is associated with higher income and lower deprivation and it has already been observed that children in the more deprived classes are less likely to receive the vaccine. In the 14 month cut-off, once again 'NoCar' does not appear to be required in the model comparing late uptake to not received. However, the coefficient of 'NoCar' is positive comparing vaccine received on time to vaccine not received. Thus, the odds of receiving the vaccine on time for this cut-off increase as the percentage of households within a postcode district with no car decreases.

'Level 4' does not appear in the 14 month cut-off model. In the 13 month cut-off model, the 'Level4' variable is not significant in the model comparing late uptake to not received. However, in the model comparing on time to not at all, this variable is significant and has a negative coefficient. Thus, the odds of receiving the vaccine on time decrease as the percentage of individuals with the highest level of qualifications increases in a postcode district. Thus, it appears that children in postcode districts with greater high levels of education are less likely to receive the vaccine.

To summarise, it appears that in both the 13 month and the 14 month models the effect of deprivation is similar for both comparisons. However, the effects of car ownership, education status and country of birth are more important for one comparison or the other.

To assess the ability of the final models shown in Table 8.29 in determining the correct classification of children in the dataset, a script was created in R to allocate each child into the class for which the child has the highest probability of response. Cross-tabulations of the predicted uptake against the true 'PCVUptake' can then be assessed to identify how strongly the model is able to determine the correct class. The probabilities of correct classification can be observed, and also the misclassification probabilities to the different classes, on assessing the cross-tabulation. The probability of correct classification for a certain class is calculated as the number of children in the dataset correctly identified in a class divided by the total number of children in that class. To establish the class of a new child entering the dataset, the values of the explanatory variables for that child may be substituted in the final models shown previously. The child is assigned to the class for which the highest probability is found. To assess how effectively the model can predict the class of a new child, cross-validation methods can be employed to discover whether or not the model is able to determine the correct outcome for the child. The cross-tabulation in Table 8.32 shows the proportion of children in each group correctly classified to each outcome, as well as the proportions misclassified for the final model with the 13 month cut-off.

Table 8.32: Cross-tabulation of the proportion of children classified to each of the PCV-7 booster uptake outcomes for the model using the 13 month cut-off.

	Predicted class			
True class	Not received	On time	Late	Number in class
Not received	0.134	0.000	0.865	2553
On time	0.042	0.000	0.958	670
Late	0.009	0.000	0.991	25152

From assessment of Table 8.32, it can be noted that the model of 'PCVUptake' displayed in Table 8.29 correctly classifies a high percentage, 99.1%, of children who receive the vaccine late but only a small percentage, 13.4%, of children who do not receive the vaccine. The model is unable to correctly classify any of the small number of children who receive the vaccine on time. The Kappa statistic was calculated to assess the agreement between the predicted and true classes, adjusting for the amount of agreement expected by chance (Cohen 1960). A statistic of 0.82 was obtained indicating strong agreement between the numbers in the predicted and true classes. Table 8.33 shows the proportions of correct and incorrect classifications for the model with the 14 month cut-off for late vaccinations. A Kappa statistic of 0.40 was obtained for this model indicating fair agreement.

Table 8.33: Cross-tabulation of the proportion of children classified to each of the PCV-7 booster uptake outcomes for the model using the 14 month cut-off.

	Predicted class			
True class	Not received	On time	Late	Number in class
Not received	0.449	0.021	0.530	2553
On time	0.024	0.128	0.847	10399
Late	0.065	0.041	0.894	15423

The model for the 14 month cut-off has a greater ability to determine those who do not receive the vaccine and those who receive the vaccine on time than the model

for the 13 month cut-off. 44.9% of children who did not receive the vaccine and 12.8% of children who received the vaccine on time are correctly classified. 89.4% of children who received the vaccine late are identified. The correct classification rate of those who received the vaccine on time, although improved upon for the 14 month cut-off rather than the 13 month cut-off, is still very poor. It appears that the model struggles to distinguish between those who receive the vaccine on time and those who receive it late as 84.7% of children who received the vaccine on time are incorrectly classed as late receivers.

Five-fold cross-validation is used to assess the predictive ability of each of the 'PCVReceived' models. To carry out this cross-validation the data are split into five groups of equal size, ensuring there are roughly equal numbers of children with the different classifications in each group. Thus, for example, within each group for the model with 13 month cut-off there are approximately 510 children who do not receive the vaccine, 134 who receive the vaccine on time and 5,030 who receive the vaccine late. The model is fitted to data from the first four of the five groups. This model is then used to predict the class of the fifth group. Then this method is repeated and four other groups are combined in the model and this model is used to predict the class of the group omitted. In total, five models are fitted and predictions for five different groups are made. The average percentage of children classified to each of the vaccine uptake categories can then be determined. The averages for the 13 month cut-off are shown in Table 8.34 and for the 14 month cut-off in Table 8.35.

Table 8.34: Average percentage of children classified to each of the PCV-7 booster uptake categories for the 13 month cut-off using five-fold cross-validation.

	Predicted class			
True class	Not received	On time	Late	Number in class
Not received	14.924	0.039	85.037	2553
On time	4.179	0.000	95.821	670
Late	1.069	0.000	98.931	25152

Table 8.35: Average percentage of children classified to each of the PCV-7 booster uptake categories for the 14 month cut-off using five-fold cross-validation.

	Predicted class			
True class	Not received	On time	Late	Number in class
Not received	44.810	2.115	53.075	2553
On time	2.547	13.033	84.419	10399
Late	6.035	5.005	88.960	15423

As expected from the cross-tabulations shown in Table 8.32 and Table 8.33, neither model has a high ability to correctly classify children who receive the PCV-7 booster on time. However, the cross-validation classifications are very similar to the raw percentages shown in these two tables suggesting the models are robust. The model for the 13 month cut-off is unable to correctly identify any of the children who fall into this category, with overall 95.8% of children who fall into this category predicted to be in the late vaccination class. The 14 month cut-off model performs a bit better, correctly determining the class of 13.0% of these children. However, almost 85% of these children are incorrectly classed as late vaccinations. The cross-validation shows that only 14.9% of children are correctly predicted to fall in the unvaccinated class for the 13 month cut-off, with 85.0% of this class of children incorrectly predicted to be vaccinated late. For the 14 month cut-off, a higher percentage are classed correctly as unvaccinated. However, still more than half of these children are incorrectly classed as vaccinated late.

On assessment of the histograms of estimated probabilities for each of the two models, Figure 8.19, it can be observed that the peak of the distribution for the 13 month cut-off occurs between 0.9 and 0.95. In the model a child is assigned to the class with the highest probability. In general, it is good to observe that the peak of the distribution occurs at such a high probability value as this means the classification for many children was strong. If the peak of the distribution were to occur at a lower value of 0.4 it would mean that it would be difficult to assign children to the correct class by simply using the model to find the highest probability. The number of children with estimated probabilities greater than

0.7 is calculated to be 23,725. Therefore, it appears that approximately 83.61% of the children in the dataset have decisive classifications based on the method of classification according to the highest probability. This appears odd as the model does not perform well in correctly classifying children who do not receive the vaccine and those who receive the vaccine on time. However, with this cut-off of 13 months, the vast majority of children, around 88%, in the dataset fall into the vaccinated late class and thus, since the model is able to correctly classify over 98% of these children, it makes sense that there is a high probability of overall successful predictions. For the 14 month cut-off, the peak probability is much lower at between 0.55 and 0.6. For this cut-off, the number of children with estimated probabilities of greater than 0.7 is much lower at only 22. Thus, in this model, only a tiny 0.08% of children have a decisive classification. A reduced classification of 0.6 still would only result in 11.75% correct classifications overall.

HB has been shown to be statistically significant as a fixed effect in previous models discussed in this chapter. Thus, HB is entered as a fixed effect in both the 13 month and 14 month cut-off multinomial models for 'PCVUptake'. HB is significant as a fixed effect in both models.

In the final 13 month single-level model including HB, in the comparison of received on time to not received, 'SCSIMD5' and 'NoCar' are the only significant variables. The coefficients of 'SCSIMD5' have changed slightly following the inclusion of HB in the model but all remain negative and show the same pattern as before. The coefficient of 'NoCar' is positive in this model, not negative as it was in the model without the HB effect. Comparing received late to not received, 'NoCar' and 'Level4' are not significant in this model. 'OtherEU' has a negative coefficient as in the model without the HB effect. The coefficients of the 'SCSIMD5' dummy variables are all negative as in the model without HB. However, the results are altered as, in the model without HB the coefficient of category 4 of 'SCSIMD5' is less negative than that of category 3. This is not the case in the model with HB where the coefficients are more negative with increasing deprivation category.

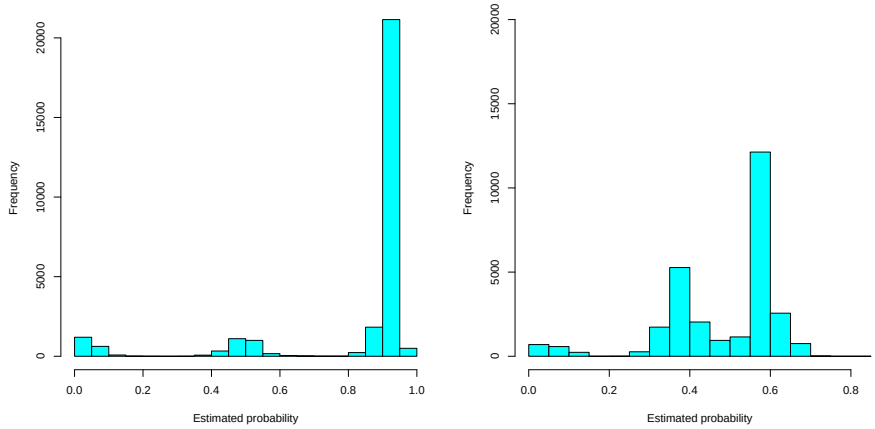

Figure 8.19: Histograms of the estimated probabilities for the correct class of PCV-7 booster uptake obtained from the models with 13 month cut-off and 14 month cut-off for late vaccination respectively.

Comparing received on time to not received in the 14 month cut-off model, only 'SCSIMD5' appears in addition to HB. The coefficients of 'SCSIMD5' changed in this model following the inclusion of HB but, as with the 13 month cut-off model, the general pattern observed remains the same. Similarly, 'SCSIMD5' is the only variable in addition to HB which appears in the model comparing late to not received. As with the 13 month cut-off model, comparing late to not received the coefficient of 'SCSIMD5' becomes more negative with increasing deprivation. This was not the case in the model without the HB effect.

The variable 'HibReceived' does not appear in either the 13 month or the 14 month cut-off model after inclusion of HB.

In the multinomial model, differences in the effects of each of the explanatory variables comparing vaccinated on time to not at all and vaccinated late to not at all can be examined. For the 13 month model it can be observed that the coefficients for the deprivation categories in each model are both negative. However, the patterns appear to differ for the two comparisons. The odds of receiving the vaccine late to not at all decrease with increasing deprivation whilst the odds of receiving the vaccine on time compared to not at all decrease with increasing deprivation to the fourth deprivation category. The odds of receiving the vaccine on

time are higher for those in the most deprived category than for those in the second most deprived category. This is not true of the 14 month model where both comparisons show reduced odds with increasing deprivation. In the 13 month model both 'NoCar' and 'Level4' have different effects in the two comparisons.

HB had the greatest variation in effect with health boards B (Borders), F (Fife) and V (Forth Valley) showing a different sign of coefficient when comparing received on time or received late to not at all in both the 13 month and the 14 month cut-off models. In addition, in the 13 month cut-off model, HB C (Argyll and Clyde), H (Highland), L (Lanarkshire), N (Grampian) and W (Western Isles) all displayed different effects for the two comparisons whilst for the 14 month cut-off the effects were most apparent in HBs A (Ayrshire and Arran), S (Lothian) and T (Tayside). In addition HB N (Grampian) in the 14 month cut-off model, although having the same sign of coefficient in each of the comparisons, had a much larger effect when comparing received late to not received than when comparing received on time to not received.

The parameter estimates for the 13 and 14 month cut-off models are shown in Tables 8.36 and 8.37. By using the category with the highest estimated probability to classify each child, all children in the 13 month cut-off model are classed as receiving the vaccine late. Thus, this model does not perform as well as the model without the HB effect in classifying children into the not received category, where 13.4% were correctly classified. Neither model was able to correctly classify any children in the received on time classification.

For the 14 month cut-off model, 1.4% of children who do not receive the vaccine are correctly classified, 3.5% of children who receive the vaccine up to the age of 14 months and 97.2% of children who receive the vaccine after the age of 14 months are correctly classified. Thus, the model does not perform well in predicting the children in two of the three categories. The model with the HB effect does not perform as well as the 14 month cut-off model without HB as using this model 44.9% of children were correctly classified as having not received the PCV-7 booster and 12.8% were classed as having received the vaccine on time.

Two-level models

In the next stage of modelling, a random effect for postcode district is entered in the model. MLwiN is used to carry out the multi-level modelling. As with the binomial multi-level models, the model is first fitted using first order MQL and then, following convergence, second order PQL is adopted as first order MQL produces severely biased estimates in multinomial logistic regression (Rasbash et al. 2008). In this multi-level multinomial model a random term, $\upsilon_j^{(s)}$, representing unobserved postcode district level characteristics is entered into the model described in (8.13). The random effects are assumed to follow a bivariate Normal distribution with mean $\mathbf{0}$ and variance $\mathbf{\Omega}$, where $\mathbf{\Omega}$ is defined to be

$$\begin{pmatrix} \sigma^{2(1)} & \sigma^{(12)} \\ \sigma^{(12)} & \sigma^{2(2)} \end{pmatrix}. \tag{8.14}$$

In (8.14), $\sigma^{2(1)}$ represents the residual between postcode district variance in the log odds of receiving the vaccine on time against not receiving the vaccine. Similarly, $\sigma^{2(2)}$ is the residual between postcode district variance in the log odds of receiving the vaccine late against not receiving the vaccine. Finally, $\sigma^{(12)}$ is the covariance between the postcode district effect on the probabilities of receiving the vaccine on time and receiving it late.

The variance components models for both the 13 month cut-off and the 14 month cut-off, i.e. the model with only a random intercept on postcode district, show zero S.E.s for the estimates of $\sigma^2_{\upsilon 0}$ and $\sigma_{\upsilon 01}$.

Table 8.36: Results from the multinomial model of 'PCVUptake' including fixed effect of HB with 13 and 14 month cut-off, comparing on time to not received (A=Ayrshire & Arran, B=Borders, C=Argyll & Clyde, F=Fife, G=Greater Glasgow, H=Highland, L=Lanarkshire, N=Grampian, R=Orkney, S=Lothian, T=Tayside, V=Forth Valley, W=Western Isles, Y=Dumfries & Galloway, Z=Shetland).

	13 month		**14 month**	
Variable	**Coefficient**	**S.E.**	**Coefficient**	**S.E.**
Intercept_1	-1.919	0.388	2.167	0.86
SCSIMD5_{21}	-0.397	0.148	-0.232	0.083
SCSIMD5_{31}	-0.464	0.150	-0.395	0.080
SCSIMD5_{41}	-0.623	0.158	-0.517	0.079
SCSIMD5_{51}	-0.522	0.164	-0.734	0.077
OtherEU_1	-0.092	0.116	-0.293	0.040
NoCar_1	0.010	0.006	-	-
Level4_1	0.007	0.009	-	-
HB_{A_1}	1.126	0.219	-0.026	0.104
HB_{B_1}	1.497	0.311	-0.071	0.179
HB_{C_1}	0.440	0.222	0.021	0.096
HB_{F_1}	0.558	0.242	-0.256	0.097
HB_{H_1}	0.942	0.252	-0.569	0.112
HB_{L_1}	-0.735	0.297	0.043	0.087
HB_{N_1}	1.492	0.223	0.044	0.093
HB_{R_1}	-0.071	0.510	-2.289	0.275
HB_{S_1}	0.351	0.208	-0.001	0.086
HB_{T_1}	-0.589	0.306	-1.439	0.113
HB_{V_1}	0.625	0.253	-0.036	0.107
HB_{W_1}	1.294	0.438	-0.455	0.267
HB_{Y_1}	1.925	0.323	0.973	0.193
HB_{Z_1}	-0.082	0.639	-1.102	0.260

Table 8.37: Results from the multinomial model of 'PCVUptake' including fixed effect of HB with 13 and 14 month cut-off, comparing late to not received (A=Ayrshire & Arran, B=Borders, C=Argyll & Clyde, F=Fife, G=Greater Glasgow, H=Highland, L=Lanarkshire, N=Grampian, R=Orkney, S=Lothian, T=Tayside, V=Forth Valley, W=Western Isles, Y=Dumfries & Galloway, Z=Shetland).

	13 month		**14 month**	
Variable	**Coefficient**	**S.E.**	**Coefficient**	**S.E.**
Intercept_2	3.202	0.185	2.172	0.084
SCSIMD5_{22}	-0.180	0.078	-0.141	0.079
SCSIMD5_{32}	-0.382	0.078	-0.356	0.078
SCSIMD5_{42}	-0.493	0.079	-0.473	0.076
SCSIMD5_{52}	-0.654	0.082	-0.608	0.674
OtherEU_2	-0.174	0.057	-0.233	0.038
NoCar_2	-0.005	0.003	-	-
Level4_2	-0.009	0.005	-	-
HB_{A_2}	0.109	0.109	0.906	0.101
HB_{B_2}	-0.042	0.179	0.248	0.173
HB_{C_2}	-0.035	0.097	0.060	0.094
HB_{F_2}	-0.205	0.109	0.072	0.094
HB_{H_2}	-0.637	0.122	-0.378	0.108
HB_{L_2}	0.114	0.094	0.312	0.086
HB_{N_2}	-0.098	0.111	9.125	0.092
HB_{R_2}	-2.282	0.228	-1.949	0.243
HB_{S_2}	0.111	0.091	0.342	0.084
HB_{T_2}	-0.063	0.102	0.536	0.093
HB_{V_2}	-0.114	0.116	0.049	0.105
HB_{W_2}	-0.582	0.254	-0.435	0.264
HB_{Y_2}	0.732	0.200	0.806	0.193
HB_{Z_2}	-1.073	0.239	-0.828	0.246

Thus, there is clearly some issue with convergence of this multi-level modelling in MLwiN. Under the belief that this problem is due to the possibility that the

random effect of postcode district is not significant in this multinomial model, a simulated response with postcode district effects was created to test this theory. In this simulated response variable, for half of the postcode district levels a third of the number of observations within the district were classed as 0, a third as 1 and a third as 2. For a further quarter of the postcode districts, 4/5 of the observations were classed as 0; 1/10 as 1 and 1/10 as 2. Finally, for the last quarter of the districts, 2/5 of the observations were classed as 0, 2/5 as 1 and 1/5 as 2. Thus, postcode district level variation should be observed in this response. However, on fitting the variance component model in MLwiN similar convergence issues were experienced. As a number of postcode districts have very few numbers of children within them, it is believed that this is the likely cause of the convergence problems. Thus, in the remaining analysis in this chapter only single-level models are considered and a fixed effect of HB is entered in the models to take into account area level variation.

8.4.7 Modelling the multivariate response PCV-7 received combined with PCV-7 received late and months late

In this final section of the vaccine uptake analysis chapter, a combination of all three of the univariate responses considered previously is analysed. In order to use a multinomial approach, as adopted for the modelling combining 'PCVReceived' and 'PCVLate', the number of months late at which the vaccine is administered is split into categories. The first categorical model looks at five categories where 0 represents children who have not received the vaccine by 22 months of age, 1 represents that the PCV-7 booster has been administered on time, i.e. by 13 months of age, 2 is the category for vaccinations administered up to 1 month late, 3 represents those vaccinated between 1 and 3 months late and 4 represents children vaccinated over three months late. Of the 28,672 children in the dataset, 2,605 (9.09%) fall into category 0; 676 (2.36%) are found in category 1; 9,964 (34.75%) in category 2; 10,527 (36.72%) in category 3 and 4,900 (17.09%) in category 4. Multinomial logistic regression was used for a similar purpose in an analysis of delays in uptake of the DTP vaccination in the USA, where uptake was split into three categories: 1 to 6 months delay, 7 or more months delay, no

vaccinations recorded (Dombkowski et al. 2004). Dombkowski et al. state that the benefit of this type of analysis is that it is possible to characterise children of all different states of vaccination uptake status. Thus, factors affecting both delays and whether or not the vaccine is UTD may be determined.

As mentioned in the previous section, the significant variables for determining either 'PCVReceived' or 'PCVLate' are 'HibReceived', 'HibLate', 'SCSIMD5', 'OtherEU', 'NoCar' and 'Level4'. For the continuous response 'PCVMonthsLate', which has been considered in categories to carry out the modelling in this section, the variables 'HibMonthsLate', 'SCSIMD5', and 'Level4' were found to be significant. As there are three different explanatory variables involving the uptake of the Hib/Men C booster, these variables are combined in a similar fashion to that of the response variable describing the PCV-7 booster uptake.

Modelling

Single-level models

Single-level multinomial models for this five category response, 'PCVTiming', may be fitted in the same manner as in for the previous model. Once again, the category for children who have not received the PCV-7 booster is taken as the baseline for comparison in the analysis. The model shown in (8.13) now represents four simultaneous equations. As before, when $s = 1$, the logarithm of the ratio of the probability of receiving the PCV-7 booster on time to the probability of not receiving the vaccine is represented. When $s = 2$, the logarithm of the ratio of the probability of receiving the vaccine up to one month late to that of not receiving the vaccine is represented, when $s = 3$, the vaccine received between 1 and 3 months late is compared to not at all and, finally, when $s = 4$, the vaccine received over 3 months late is compared to not at all.

Results from the multinomial regression of 'PCVTiming' using the same approach in R 2.9.1 as in the previous multinomial single-level model are shown in Table 8.39. The categorical explanatory variable involving the uptake and timing of the Hib/Men C booster, 'HibTiming', is not included in the final model as this term

is not deemed significant using stepAIC in R.

Table 8.38: Percentage distribution for each category of timing for the PCV-7 booster for each of the potential explanatory variables.

Variable		**Unv**	**OT**	**<1M**	**1-3M**	**>3M**	**Total**
Individual level							*Number*
HibTiming (%)	Unv	47.14	2.74	8.78	20.99	20.36	2815
	OT	2.70	9.42	56.60	24.38	6.92	2015
	<1M	2.75	1.86	49.28	35.77	10.35	15631
	1-3M	6.32	1.13	7.78	57.38	27.39	5235
	>3M	16.19	1.96	13.80	27.67	40.39	2679
SCSIMD5	1	6.79	2.83	36.46	38.97	14.95	5048
	2	7.94	2.25	33.51	38.50	17.80	5413
	3	9.11	2.42	35.49	36.75	16.23	5513
	4	9.63	2.07	35.23	35.66	17.42	5777
	5	10.98	2.31	33.60	34.52	18.59	6624
Postcode level							
OtherEU (mean)		0.91	0.85	0.90	0.89	0.90	
NoCar (mean)		28.94	31.85	28.09	27.71	28.11	
Level4 (mean)		20.52	19.06	20.65	20.86	20.89	
Total		2553	670	9875	10422	4855	28375

On comparing the effect of each of the explanatory variables in the four comparisons with the baseline category of not received in this multinomial model, it can be observed that 'Level4' appears to have the same effect in all comparisons. The negative coefficient of 'Level4' suggests that as the percentage of individuals within a postcode district increases the odds of receiving the vaccine on time, less than one month late, one to three months late or over three months late each decrease when compared to not received. Different effects are observed for the variable 'NoCar' with the comparisons between received on time, received between one and three months late or received more than three months late all showing negative coefficients for this variable whilst the coefficient for 'NoCar' comparing received up to one month late to not received is positive. Different

effects are also observed for the variable 'OtherEU' with the comparisons of received up to one month late, received between one and three months late and received more than three months late to not received all showing negative coefficients. The comparison of received on time to not received has a positive coefficient.

On assessment of 'SCSIMD5' in the four comparisons made in this model, it is of interest to assess whether the patterns of effect with increasing deprivation differ. All deprivation coefficients in the model are negative. On assessment of the coefficients in the comparison of received on time to not at all, the moduli of the coefficients generally become larger with increasing deprivation, indicating higher odds of receiving the vaccine on time to not at all for more deprived classes. However, the modulus of the coefficient for the most deprived class is lower than that of the second most deprived class. Similarly, on comparison of receiving the vaccine between one and three months late to not at all the negative coefficients become smaller with increasing deprivation. The same pattern is not observed for the other two comparisons. However, all four comparisons show the smallest negative coefficient for the most deprived classification.

Each child in the dataset is assigned to one of the five classes of 'PCVTiming' according to the highest probability determined by the model shown in Table 8.39. The cross-tabulation shown in Table 8.40 shows the proportions of predicted classifications for each of the true classes.

From examination of Table 8.40, it can be noted that the model performs poorly in distinguishing between the five classes, only classifying children into categories 3 and 4 of 'PCVTiming'. No children are correctly classified as not receiving the vaccine, receiving the vaccine on time or receiving the vaccine up to 1 month late. Each true class of 'PCVTiming' has between 24% and 35% of children incorrectly classified as receiving the vaccine between one to three months late and between 65% and 76% incorrectly classed as receiving the vaccine greater than three months late. A Kappa statistic of 0.35 shows fair agreement between the model predictions and the true observations. Retaining 'HibTiming' in the model results in little difference in the predictions.

By combining categories 1 and 2 in the response variable 'PCVTiming' to allow vaccinations received on time to be up to the age of 14 months, a new response variable is created with 0 representing not vaccinated, 1 representing vaccinated by 14 months of age, 2 representing vaccinated between 14 and 16 months of age, i.e. between 1 and 3 months later than scheduled, and 3 representing vaccinated greater than 16 months of age. As with the model with 5 categories, the categorical variable involving the Hib/Men C booster is not found to be significant. Unfortunately, this model also is unable to correctly identify the classification of the child on assessment of the highest probability. The results from this method are shown in Table 8.41. On assessment of the agreement, a Kappa statistic of 0.25 is obtained indicating fair agreement.

Table 8.39: Results from the multinomial model of 'PCVTiming', with not received as the baseline category for comparison.

Variable	**Coefficient**	**S.E.**	**Variable**	**Coefficient**	**S.E.**
Intercept_1	-0.234	0.240	Intercept_3	2.244	0.126
Intercept_2	2.087	0.127	Intercept_4	1.198	0.140
SCSIMD5_{21}	-0.399	0.147	SCSIMD5_{23}	-0.248	0.081
SCSIMD5_{22}	-0.341	0.081	SCSIMD5_{24}	-0.038	0.089
SCSIMD5_{31}	-0.462	0.147	SCSIMD5_{33}	-0.477	0.081
SCSIMD5_{32}	-0.486	0.081	SCSIMD5_{34}	-0.291	0.090
SCSIMD5_{41}	-0.606	0.155	SCSIMD5_{43}	-0.558	0.082
SCSIMD5_{42}	-0.578	0.083	SCSIMD5_{44}	-0.251	0.091
SCSIMD5_{51}	-0.521	0.161	SCSIMD5_{53}	-0.724	0.086
SCSIMD5_{52}	-0.812	0.086	SCSIMD5_{54}	-0.283	0.095
OtherEU_1	0.271	0.086	OtherEU_3	-0.103	0.049
OtherEU_2	-0.330	0.052	OtherEU_4	-0.155	0.056
NoCar_1	-0.016	0.004	NoCar_3	-0.004	0.002
NoCar_2	0.002	0.002	NoCar_4	-0.005	0.002
Level4_1	-0.021	0.008	Level4_3	-0.010	0.004
Level4_2	-0.003	0.004	Level4_4	-0.003	0.005

As previous models considered in this chapter show a significant HB effect when entering this variable as either a random or fixed effect, HB was entered as a fixed effect in the model of PCV-7 timing outcome with five categories. HB is found to be significant in the model, with both HB and deprivation category appearing as the most significant determinants of PCV-7 timing outcome. For brevity, the results for this modelling are not presented in the thesis. However, these are available on request. No great differences in the effect of HBs R (Orkney), Y (Dumfries and Galloway) and Z (Shetland) were observed for the four model comparisons. All other HBs displayed different effects in the four model comparisons.

Table 8.40: Cross-tabulation of true PCV-7 timing outcome with five categories against predicted outcome.

	Predicted class					
True class	Unv	OT	<1M	1-3M	>3M	Number in class
Unv	0.000	0.000	0.000	0.315	0.685	2553
OT	0.000	0.000	0.000	0.240	0.760	670
<1M	0.000	0.000	0.000	0.349	0.651	9875
1-3M	0.000	0.000	0.000	0.295	0.705	10422
>3M	0.000	0.000	0.000	0.301	0.699	4855

Table 8.41: Cross-tabulation of true PCV-7 timing outcome with four categories against predicted outcome.

	Predicted class				
True class	Unv	OT	1-3M	>3M	Number in class
Unv	0.000	0.000	0.590	0.410	2553
OT	0.000	0.000	0.625	0.375	10545
1-3M	0.000	0.000	0.592	0.408	10422
>3M	0.000	0.000	0.589	0.411	4855

In all multinomial models considered in this chapter, the results for the post-code district level variables should be interpreted with caution as the hierarchical

structure of the data has not been included in the model. Therefore, these significant variables may not have been found to be significant had this structure been incorporated.

Conclusions from the multivariate models of PCV-7 uptake and timing

In summary, the multivariate models of PCV-7 uptake and timing were created by combining the univariate responses considered in the previous section of the chapter. The benefit of combining the responses was the possibility of comparing the effects of each of the explanatory variables for the different response categories. The explanatory variables identified as significant for determining the univariate responses were included in the combined model. Unfortunately, the multi-level structure adopted in the creation of the univariate models was not possible for reasons discussed previously. Thus, the final models are individual level models only. However, as it is expected that there is variation between the different HBs, dummy variables representing this variable were entered into the model to assess area level variation.

The first multinomial model considered involved only three categories for vaccine not received, received on time and received late using both a 13 month and a 14 month cut-off. As with the univariate models, strong associations were observed between PCV-7 booster uptake and timing and the Hib/Men C booster uptake and timing. On consideration of deprivation category, the probability of receiving the vaccine on time rather than not at all decreases with increasing deprivation. On comparison of late uptake to non-receipt, the probability of receiving the vaccine late decreases as deprivation increases from level 1 to 2 but there does not appear to be a clear trend as the coefficients of categories 3 to 5 are all fairly similar.

Considering country of birth, the odds of receiving the PCV-7 booster on time compared to not at all decrease as the percentage of individuals born in other EU countries increases. Comparing late receipt to not receiving the vaccine, the odds of late receipt increase as the percentage born outwith the UK and Ireland increases. This is not true of the 14 month cut-off where the odds of late receipt decreases as the percentage born in the EU outwith the UK and Ireland increases.

As the proportion of households with no car increases, the odds of receiving the vaccine on time compared to not at all decreases and the odds of receiving the vaccine late compared to not at all decreases also. However, this is not true of the 14 month cut-off where the odds of receiving the PCV-7 booster on time compared to not at all increases with increasing proportions of households with no car.

Finally, on addition of HB to the model, differences were observed for the 13 and 14 month cut-offs. For the 13 month cut-off comparing on time to not received, HBs L, R, T and Z all have negative coefficients. Thus, these HBs have a lower odds of receiving the vaccine on time to not received than the comparator HB G. However, on assessment of the 14 month cut-off, HBs A, B, F, H, R, S, T, V, W and Z all have lower odds of receiving the vaccine on time to late than HB G. Comparing the category received late to not at all, HBs B, C, F, H, N, R, T, V, W and Z all have a lower odds than HB G of receiving the vaccine late for the 13 month cut-off whilst for the 14 month cut-off, only H, R, W and Z have lower odds.

The next multinomial model considered all three of the univariate responses considered in the previous section, with the number of months late grouped into three arbitrary categories: up to 1 month late, between 1 and 3 months late and more than 3 months late. In addition to these categories, not received and on time were considered as before. The same explanatory variables as considered in the first multinomial model were included in this model.

Considering deprivation quintiles in this model, the negative coefficients become greater in size with increasing deprivation comparing up to one month late to not received and between 1 and 3 months late to not received. Comparing received on time to not received, the negative coefficients of categories 1 to 4 increase in size but decrease from category 4 to 5. Comparing the more than 3 months late category to received on time, there is no observable pattern to the coefficients for deprivation.

The coefficient of the variable representing the proportion born in the EU outwith the UK and Ireland is positive comparing receiving the vaccine on time to not

at all. Therefore, as the proportion of individuals born in other EU countries increases in a district, the odds of receiving the vaccine on time compared to not at all increases. The coefficient of this variable comparing the other vaccine timing categories to not received are all negative indicating decreasing odds of late receipt with increasing proportions of individuals born in other EU countries.

All coefficients are negative for the educational attainment variable, indicating decreased odds are observed of receiving the vaccine on time, up to 1 month late, 1 to 3 months late or more than 3 months late compared to not at all as the percentage of individuals with the highest educational attainment increases.

On examination of the coefficients for the variable for the percentage of households with no car in a postcode district, negative values are observed comparing on time to not at all, between 1 and 3 months late to not at all and more than 3 months late to not at all. The coefficient for this variable comparing up to 1 month late to not at all is positive. Thus, in this case, the odds of receiving the vaccine up to 1 month late increases with increasing proportions of households with no car.

Finally, considering the area measure HB, lower odds of receiving the vaccine on time to not at all are observed for HBs L, R, T and Z compared to HB G. Comparing receiving the vaccine up to 1 month late to not at all, lower odds are observed for all HBs other than Y which has a higher odds than the comparator G. Examining the model for the next category, comparing between 1 and 3 months late to not at all, the HBs with positive coefficients and thus higher odds compared to HB G are A, B, L, N, S, T and Y. Finally, comparing more than 3 months late to not at all, HBs A, B, F, L, S, T and Y have higher odds of late uptake than HB G.

8.5 Conclusions

In this chapter, the uptake and timing of the newly introduced PCV-7 in Scotland was considered. Two 2006 birth cohorts with a 22 month follow-up period were assessed in this analysis. The uptake of PCV-7 was high for these cohorts, with over 90% of the children receiving each dose. The focus of the analysis in this

chapter was on the uptake of the booster dose at 13 months of age. 90.91% of the children received this dose by age 22 months. However, a substantial proportion of children did not receive the vaccine according to schedule. Using the stringent cut-off of 13 months, only 2.59% of children received the vaccine on time. The less strict cut-off of 14 months still results in a substantial percentage of late vaccinations, with only 40.25% receiving the vaccine by this age. On average, the PCV-7 booster was administered almost two months later than the routine vaccination recommendations.

The analysis carried out in this chapter involved the creation of separate univariate models to describe the PCV-7 booster uptake and timing, followed by the assessment of multivariate models combining the univariate response categories. The benefit of combining the response categories was that it became possible to make direct comparisons of the effects of each of the explanatory variables when comparing scheduled or late receipt to non-receipt. However, the benefit of using the univariate models was that it was possible to assess not only categorical variables but also the continuous response variable for the number of months late. Since a multinomial modelling approach is used, this response had to be split into categories. An approach based on the techniques adopted in survival analysis, such as those used by Dayan et al. (2006), Akmatov et al. (2008) or Clark et al. (2009), would perhaps have been more appropriate so that the number of months late could have been treated as a continuous response whilst allowing for those unvaccinated to also be dealt with.

In all of the three univariate response models considered, significant random variation attributable to both postcode district and HB was found. Therefore, it appears that area effects are significant in determining vaccine uptake and timing. This corresponds with another vaccine uptake analysis undertaken in Scotland in which varying patterns were observed in MMR uptake for the different HBs (Cameron et al. 2007). The smaller, island HBs show the greatest variability in the timing of vaccine administration. Orkney has the lowest uptake of the PCV-7 booster by age 22 months, with only approximately 59% of children in this HB receiving the vaccine. However, Orkney had the lowest proportion of late vaccinations of all HBs using the 13 month cut-off for late vaccinations. The

results are different for the 14 month cut-off, where Orkney still has amongst the lowest late vaccination percentages. Thus, it appears that in Orkney the vaccines are either received on time or not at all. For the 13 month cut-off, Dumfries and Galloway had the lowest proportion of late vaccinations. The odds of receiving the vaccine late were highest for Lanarkshire and Tayside for the 13 month cut-off. For the 14 month cut-off, Tayside has a much higher predicted probability of late vaccinations than all other HBs at 0.89.

The timing of the Hib/Men C booster is important in determining whether or not the subsequent PCV-7 booster is administered in a timely fashion, with delays in receipt of the Hib/Men C booster leading to delays in receipt of the PCV-7 booster. In addition, those who received the Hib/Men C booster had a higher odds of receiving the PCV-7 booster than those who did not.

As with previous Scottish analyses of vaccine uptake (Cameron et al. 2007; Friederichs et al. 2006), deprivation was identified to be an important determinant of vaccine uptake and timing. Cameron et al. observed greater delays in vaccine uptake for children in the most deprived category which corresponds with the analysis in this chapter in which the number of months late the vaccine was administered was highest for those most deprived. Friederichs et al. observe that the most affluent in Scotland either have the MMR vaccine on time or not at all. This may be the case with the PCV-7 booster where the greatest delays are observed amongst those most deprived. However, the odds of receiving the vaccine decrease with increasing deprivation. Thus, the least deprived are more likely to have the vaccine.

Educational level was identified to be significant in determining the timing of the vaccine administration. However, contradictory results were obtained on comparison of the model of the binary variable for PCV-7 booster timing and the continuous variable for the number of months late the booster is administered. The results for the binary model, where the odds of receiving the vaccine late increase as the percentage with high educational attainment within a postcode district decreases, correspond with observations from other studies (Bobo et al. 1993; Dombkowski et al. 2004; Ozcirpici et al. 2006; Torun and Bakırcı 2006;

Datar et al. 2007). However, the model for the continuous response contradicted these results. The correlations between area educational attainment and months late, both at the individual level and postcode district level, are fairly small. Thus, it is unclear how much information educational attainment contributes to the determination of the number of months late.

The country of birth and the percentage of households without a car were also identified as significant in one or other of the univariate models. The country of birth has an impact on whether or not the PCV-7 booster is received with lower odds of receipt associated with higher proportions of individuals born in EU countries outwith the UK and Ireland. This corresponds with the role of ethnicity in age-appropriate vaccinations observed in the USA (Nuorti et al. 2008). The results for the percentage of households with no cars, a proxy for the income of a district, are that the odds of receiving the PCV-7 booster late decrease with decreasing proportions of households without a car within a postcode district. This also corresponds with results from other vaccine analyses in which lower age-appropriate or UTD vaccinations are observed for households with lower incomes (Nuorti et al. 2008; Guttmann et al. 2006).

In the three-level models, no significant associations were identified between PCV-7 uptake and timing and the variables representing the different types of employment in a postcode district or the percentage who are not of white race within a district. The percentage of individuals who were not of white race within a postcode district is generally very low, with only up to around 8% of individuals in a district listed as not white. Similarly, the percentages of unemployed individuals and large employers was fairly low, ranging from around 0 to 10%. No significant effects were found for those working in agriculture. This variable was included in an attempt to have some variable representing urban and rural areas as higher percentages of individuals working in agriculture would imply rurality. This variable was highlighted as an important variable in some vaccine uptake analyses (Akmatov et al. 2008; Ozcirpici et al. 2006). However, contradictory results were obtained in these analyses, with Akmatov et al. observing higher delays in urban areas and Ozcirpici et al. observing higher delays in rural areas. This is likely to be due to the nature of the studies. Akmatov et al. state that

living in urban areas in CIS countries represents lower socioeconomic status and lower standards of living. However, in the Ozcirpici et al. study in Turkey, the opposite is true as there are inadequate health services and lower socioeconomic status in rural areas. Thus, the effect of how urban or rural an area is on the uptake and timing of vaccination appears to be confounded with other factors. Therefore, this may explain the reason this variable is not important in the analysis of the uptake of the PCV-7 booster in Scotland as these other socioeconomic factors, such as deprivation, have been included.

The other variable found not to have an association with the timing and uptake of the PCV-7 booster was the percentage of individuals aged 0 to 4 years living within a district. In other vaccine uptake analyses, the number of children within a household was found to be significant in determining vaccine status, with a larger number of siblings associated with a greater risk of delayed vaccination (Lieu et al. 2000; Reading et al. 2004; Dombkowski et al. 2004; Bardenheier et al. 2004; Ozcirpici et al. 2006). Thus, perhaps if information on the households of the children had been included in this analysis, then the number of children within a household may have had a significant effect. However, in contrast, the study involving PCV-7 uptake in the USA did not find significant associations between the number of children in a household and UTD vaccine status (Nuorti et al. 2008) so this variable may not be important in Scotland.

Multi-level modelling was not possible for the multivariate models combining the responses for the PCV-7 booster uptake and timing. However, HB was entered as a fixed effect in the models to attempt to describe the area variability in uptake. Using a similar approach to Akmatov et al. (2008), in which three categories of response are considered: not received, received on time and delayed, the importance of the cut-off chosen for late uptake was apparent as different results were obtained for the model with the 13 month and the model with the 14 month cut-off. For the country of birth, both cut-off points show decreasing odds of receiving the vaccine compared to not at all as the percentage of individuals born in other EU countries increases, as observed in the univariate models. However, contradictory results are compared for the different cut-off points comparing received late to not received. For the 14 month cut-off, the probability of late

receipt increases as the percentage born outwith the EU increases. This suggests that for the 14 month cut-off, there is an increased chance of receiving the vaccine late.

Contradictory results were obtained for car ownership, with the odds of receiving the vaccine on time increasing with increasing proportions of households with no car within a postcode district for the 14 month cut-off, whilst the opposite is true for the 13 month cut-off. The result for the 14 month cut-off is not intuitive as car ownership should aid in the receipt of vaccine. In addition, if car ownership provides a measure of higher income, then to correspond with other vaccine uptake analyses, higher income should correspond with increased odds of uptake. For both models, the odds of receiving the vaccine late decrease with as the percentage of households with no car increases.

Concerning HB, different results are obtained for the two cut-off points. For the 13 month cut-off, Lanarkshire, Orkney, Tayside and Shetland have lower odds than Greater Glasgow of receiving the vaccine on time to not at all, whilst for the 14 month cut-off, many more HBs have lower odds: Ayrshire and Arran, Borders, Fife, Highlands, Lothian, Forth Valley and the Western Isles in addition to the HBs with lower odds from the 13 month cut-off apart from Lanarkshire. Comparing late uptake to not received, the highland and island HBs appear with reduced odds of late uptake using the 14 month cut-off. This corresponds to the KDEs of the timing of the PCV-7 booster where the peak uptake is observed early, at around 1 month late, for these HBs.

The other multinomial modelling carried out in this chapter involved a combination of all three univariate responses. Once again, deprivation, HB, country of birth, educational attainment and car ownership feature in this model.

Limitations of the modelling, commented on earlier, include the fact that, although postcode district level variables were available to attempt to gauge factors such as levels of education and unemployment, no variables specific to the household or parent of the child are available. In other studies of uptake and timing of childhood immunisations, the educational status of the parent and variables relating to family size have been deemed statistically significant (Bobo

et al. 1993; Bardenheier et al. 2004; Dayan et al. 2006). Thus, this study perhaps could have been improved by having parental and household characteristics rather than postcode district characteristics.

In future work, other approaches to assess the spatial variability in vaccine uptake could be adopted, such as kriging as used in the study in Vietnam by Ali et al. (2007). Furthermore, approaches involving the continuous variable for months late may be more appropriate for the combined models. Techniques such as survival analysis could be adopted to reflect the censored observations for those who have not received the vaccine.

Chapter 9

Conclusions

9.1 Introduction

On the introduction of a new national immunisation schedule to prevent disease and infection it is important to examine the costs, benefits and potential limitations of such an intervention.

S. pneumoniae is responsible for a great burden of disease and infection worldwide, particularly amongst infants and the elderly so vaccines have been developed and introduced in recent years to combat these diseases. The recently introduced PCV-7 is routinely administered to children under the age of two years and is anticipated to not only prevent disease within this age group but throughout the population through herd immunity. In this thesis, factors which could potentially prevent the vaccine from having long term efficacy were explored, all of which will have an impact on the cost-effectiveness of this intervention. The aim of this thesis was to add to existing research on these problems through the examination of the importance of the genetic MLSTs in pneumococcal carriage and disease through the use of mathematical models and also statistical techniques to examine the types of pneumococci involved in disease. In addition, problems relating to limited vaccine effectiveness due to missed or delayed booster vaccinations were explored.

9.2 Main results

In order to theoretically investigate the importance of MLSTs and the relationship of these genes to the pneumococcal serotypes in carriage of the bacterium, differential equation models were created. These models explored the possibility of capsular switch occurring in the population by assuming that MLSTs could be associated with more than one serotype, only one of which was assumed to be a VT serotype. The conclusions from the modelling were that a vaccine, assumed to be completely effective in preventing carriage of VT serotypes, can result in elimination of a MLST only if the MLST is solely associated with the VT serotype.

The relationship between MLSTs and serotypes involved in pneumococcal carriage was modelled by adapting the 1997 Lipsitch serotype model of pneumococcal carriage. No existing mathematical models of pneumococcal carriage or disease identified considered MLSTs in the modelling. The fundamental conclusion reached from the modelling in this thesis was that in order for coexistence of MLSTs to occur within a population, it is essential for coexistence of MLSTs or serotypes to occur within an individual.

The importance of issues affecting vaccine effectiveness such as serotype replacement, occurring either through a capsular switch process or by NVT serotypes becoming more prevalent through the eradication of VT serotypes, and herd immunity were considered in the review of the literature on the assessment of the cost-effectiveness of PCV-7. As PCV-7 is a costly intervention, the impact of herd immunity on overall disease in the population was identified to be fundamental for the vaccine to be found to be cost-effective. However, the herd immunity effect may be offset by the impact of serotype replacement and antibiotic resistance. Thus, it is important to continually assess the effectiveness of the vaccine in preventing disease and the circulating bacterial types involved in disease to update the cost-effectiveness models.

Through the use of Negative Binomial models, it was found that hospital episodes of pneumococcal and unspecified meningitis and septicaemia, as well as unspecified pneumonia, increased between 1995/96 and 2005/06 in England and Wales.

Pneumococcal pneumonia was observed to decrease in England and Wales between 1995/96 and 2003/04, after which the cases began to increase. Thus, if the pneumococcal diseases were predominantly caused by VT serotypes then PCV-7 may prevent a substantial proportion of disease, particularly if the herd immunity effects within the UK match those observed in the USA.

In Scotland, the specific bacterial types involved in IPD prior to the use of PCV-7 were examined to identify any trends in disease-causing pneumococci. In addition, associations with 30 day mortality were identified. Through the use of logistic regression models, the VT serotype 14 was identified to be decreasing in cases of IPD in Scotland, whilst the NVT serotype 1 was found to significantly increase in disease prior to PCV-7 use. In correspondence with this result, there was significant evidence that MLST 306, commonly associated with serotype 1 in Scotland, increased in IPD. This could lead to problems with the long term effectiveness of PCV-7 as serotype 1, and other NVT serotypes, may cause greater levels of disease following the use of PCV-7, impacting on any conclusions reached on the cost-effectiveness of PCV-7.

On examination of the associations between serogroups and mortality, serogroup 3, a NVT serotype, was found to have the highest CFR in Scotland and the NVT serotype 1 the lowest. MLST 180, commonly associated with serotype 3 disease, had the highest CFR of all MLSTs, and MLST 306 the lowest. Using Fisher's Exact Tests, serogroup 1 was found to be significantly associated with a reduced risk of fatality, whilst serogroups 3, 11, 16 and 19 were associated with an increased risk of fatality.

The multi-level modelling undertaken in Chapter 8 of the thesis allowed the individual level and area level factors affecting the uptake of PCV-7 booster in Scotland to be determined. Generally, the uptake of the three doses of PCV-7 in Scotland is high but the uptake is lower for the booster dose than the other two doses. If children do not receive all doses of the vaccine or are subject to delays in vaccine administration this can impact on the effectiveness of the vaccine in preventing disease in the population. Area was found to be a significant factor in determining the uptake and timing of the PCV-7 booster, with island HBs in par-

ticular showing great variability in the timing of the vaccine. Timing and uptake of the Hib/Men C booster is also critical for determining the uptake and timing of the PCV-7 booster, with delays in uptake of the Hib/Men C booster leading to delays in the subsequent vaccination. Deprivation, education and country of birth were also identified to be significant.

9.3 Discussion

The work summarised in the previous section shows the diverse range of analyses involved in the determination of the effectiveness of a new vaccine intervention.

The mathematical modelling undertaken in Chapters 2 and 3 provide an insight into the relationship between MLSTs and serotypes found in pneumococcal carriage and show that should the MLST be identified as important in the ability of the bacterium to cause disease this could result in problems for the vaccine effectiveness if MLSTs are associated with a VT and a NVT serotype. The models assume the vaccine to be completely effective in preventing carriage of the pneumococcal serotype which may not be a realistic assumption. However, the assumption that a proportion of children remain unvaccinated allows for the possibility of VT serotypes to remain in the population at equilibrium. The critical conclusion reached through the modelling was that coexistence should be considered possible within the individual in order to allow two MLSTs to coincide in a population at equilibrium. Thus, it is recognised that future modelling is required incorporating coexistence. In addition, in order to fully examine the relationship between MLSTs and serotypes, information from large longitudinal carriage studies is required in which both the MLSTs and serotypes are recorded so that parameter estimates may be obtained to use with these theoretical mathematical models.

The statistical analyses of cases of disease in England and Wales show that hospital episodes of pneumococcal disease were rising prior to the use of PCV-7. Thus, there is the potential for PCV-7 to have an impact on reducing this disease burden. However, the results from Scotland show that disease attributable to the NVT serotype 1 was increasing prior to PCV-7 use and that the NVT serotype 3

CFR was the highest in Scotland. This leads to concerns about the effectiveness of the vaccine in reducing disease and fatalities attributable to pneumococcal disease. However, it is acknowledged that no co-morbidities were recorded for use in the mortality association analysis and that some serotypes are likely to be opportunistic in behaviour when causing disease, commonly targeting those weaker individuals within the population. Thus, further studies are required in the UK in which co-morbidities are recorded in addition to the mortality information.

Furthermore, data on cases of disease and the serotypes and MLSTs associated with IPD following the introduction of PCV-7 should be analysed in order to properly gauge the effect of vaccine use. This information could not only affect the results of the cost-effectiveness analysis of PCV-7 but could have an impact on the development of future pneumococcal vaccines.

The analysis of the vaccine uptake in Scotland could perhaps be improved by adopting a spatial approach in the analysis since problems were encountered using the multi-level approach with multivariate response variables. In addition, it would be useful to obtain information relating to the specific household of each child in order to clearly examine associations between the educational level and employment of the parents or care-givers and deprivation, as well as information regarding to numbers of children within the household which was identified to be important in previous studies relating to vaccine uptake and timing.

9.4 Conclusions

In conclusion, in this thesis preliminary models of the association between MLSTs and serotypes have been examined. These are the first mathematical models involving both MLSTs and serotypes to have been considered. These models have shown that the vaccine does not result in elimination of a MLST in carriage unless the MLST is solely associated with a VT serotype and that coexistence of MLSTs or serotypes within an individual is a necessary prerequisite for coexistence of MLSTs within the population. The models considered in this thesis could be extended to consider more biologically appropriate models which explore the possibility of coexistence of bacterial types within an individual.

In addition, in this thesis, various analyses were carried out to examine factors important in determining the both the health and cost benefits of the use of PCV-7 and the elements which could prevent the vaccine from having long-term efficacy. The analyses carried out relating to disease could be extended by consideration of more recent post-vaccination data to determine the effect of PCV-7.

Appendix A

Global stability analysis

A.1 Model of two MLSTs with transmission due to MLST

The following global stability analysis is for $R_e > 1$ when $(T_1 + V_{T_1})(0) > 0$ and $(T_2 + V_{T_2})(0) > 0$.

Let $N = X + V + T_1 + V_{T_1} + T_2 + V_{T_2}$. Given $\epsilon > 0$, $\exists\ t_1$ such that for $t \geqslant t_1$, $\frac{L}{u} - \frac{\epsilon}{\beta_2} \leqslant N \leqslant \frac{L}{u} + \frac{\epsilon}{\beta_1}$ and $X + V \leqslant N$. Consider

$$\begin{aligned}
\frac{d}{dt}(T_1 + V_{T_1} + T_2 + V_{T_2}) &= \beta_1(X+V)(T_1 + V_{T_1}) + \beta_2(X+V)(T_2 + V_{T_2}) \\
&\quad - (\gamma + u)(T_1 + V_{T_1} + T_2 + V_{T_2}), \\
&= (\beta_1(T_1 + V_{T_1}) + \beta_2(T_2 + V_{T_2}))\,(X+V) \\
&\quad - (\gamma + u)(T_1 + V_{T_1} + T_2 + V_{T_2}), \\
&= (\beta_1(T_1 + V_{T_1}) + \beta_2(T_2 + V_{T_2})) \times \\
&\quad (N - T_1 - V_{T_1} - T_2 - V_{T_2}) - \\
&\quad (\gamma + u)(T_1 + V_{T_1} + T_2 + V_{T_2}),
\end{aligned}$$

$$\leqslant \beta_1(T_1 + V_{T_1} + T_2 + V_{T_2})(N - T_1 - V_{T_1} - T_2 - V_{T_2})$$
$$- (\gamma + u)(T_1 + V_{T_1} + T_2 + V_{T_2}), \quad \text{since } \beta_1 \geqslant \beta_2,$$

so

$$\frac{d}{dt}(T_1 + V_{T_1} + T_2 + V_{T_2}) \leqslant \left(\beta_1 \left(\frac{L}{u} + \frac{\epsilon}{\beta_1} - T_1 - V_{T_1} - T_2 - V_{T_2}\right) - (\gamma + u)\right)(T_1 + V_{T_1} + T_2 + V_{T_2}).$$

Let $T = T_1 + V_{T_1} + T_2 + V_{T_2}$. By the usual argument (see, for example, the argument on page 64), $\exists\ t_2 \geqslant t_1$ such that for $t \geqslant t_2$

$$T \leqslant \frac{(\gamma + u)(R_e - 1)}{\beta_1} + \epsilon\left(1 + \frac{1}{\beta_1}\right).$$

Therefore,

$$N - X - V \leqslant \frac{(\gamma + u)(R_e - 1)}{\beta_1} + \epsilon\left(1 + \frac{1}{\beta_1}\right),$$

so, since $N \geqslant \frac{L}{u} - \frac{\epsilon}{\beta_2}$,

$$X + V \geqslant \frac{L}{u} - \frac{(\gamma + u)(R_e - 1)}{\beta_1} - \left(1 + \frac{1}{\beta_1} + \frac{1}{\beta_2}\right)\epsilon,$$

$$= \frac{\gamma + u}{\beta_1} - \left(1 + \frac{1}{\beta_1} + \frac{1}{\beta_2}\right)\epsilon.$$

Similarly,

$$\frac{1}{T}\frac{dT}{dt} \geqslant \beta_2\left(\frac{L}{u} - \frac{\epsilon}{\beta_2} - (T_1 + V_{T_1} + T_2 + V_{T_2})\right) - (\gamma + u).$$

By the usual argument $\exists\ t_3 \geqslant t_2$ such that for $t \geqslant t_3$,

$$T \geqslant \frac{(\gamma+u)(R_{e2}-1)}{\beta_2} - \epsilon\left(1+\frac{1}{\beta_2}\right).$$

Hence arguing as above,

$$X+V \leqslant \frac{L}{u} - \frac{(\gamma+u)(R_{e2}-1)}{\beta_2} + \epsilon\left(1+\frac{1}{\beta_2}+\frac{1}{\beta_1}\right),$$

$$= \frac{\gamma+u}{\beta_2} + \epsilon\left(1+\frac{1}{\beta_1}+\frac{1}{\beta_2}\right).$$

When $\beta_1 = \beta_2$, it has been shown that $X+V \to \frac{\gamma+u}{\beta_1}$ and $T_1+V_{T_1}+T_2+V_{T_2} \to \frac{L}{u} - \frac{\gamma+u}{\beta_2}$ as $t \to \infty$.

When $\beta_1 > \beta_2$, $\exists\ t_4$ such that for $t \geqslant t_4$, $X+V \geqslant \frac{1}{2}\frac{\gamma+u}{\beta_1} > 0$.

Consider

$$\begin{aligned}
\frac{d}{dt}\frac{(T_1+V_{T_1})}{(T_2+V_{T_2})} &= \frac{(\dot{T_1}+\dot{V_{T_1}})(T_2+V_{T_2}) - (T_1+V_{T_1})(\dot{T_2}+\dot{V_{T_2}})}{(T_2+V_{T_2})^2}, \\
&= \frac{\beta_1(X+V)(T_1+V_{T_1})(T_2+V_{T_2}) - (\gamma+u)(T_1+V_{T_1})(T_2+V_{T_2})}{(T_2+V_{T_2})^2} \\
&- \frac{\beta_2(X+V)(T_1+V_{T_1})(T_2+V_{T_2}) - (\gamma+u)(T_1+V_{T_1})(T_2+V_{T_2})}{(T_2+V_{T_2})^2}, \\
&= \frac{(\beta_1-\beta_2)(X+V)(T_1+V_{T_1})}{T_2+V_{T_2}}, \\
&\geqslant (\beta_1-\beta_2)\frac{1}{2}\frac{(\gamma+u)}{\beta_1}\left(\frac{T_1+V_{T_1}}{T_2+V_{T_2}}\right), \quad \text{for } t \geqslant t_4.
\end{aligned}$$

Let $\xi = \frac{T_1+V_{T_1}}{T_2+V_{T_2}}$. ξ is able to be defined as such since $T_2(t)+V_{T_2}(t) > 0$. Therefore,

$$\frac{1}{\xi}\frac{d\xi}{dt} \geqslant (\beta_1-\beta_2)\frac{1}{2}\frac{(\gamma+u)}{\beta_1} > 0.$$

So, $\xi \to \infty$ as $t \to \infty$. Hence, $\frac{1}{\xi} \to 0$ as $t \to \infty$. This means that

$$\frac{T_2 + V_{T_2}}{T_1 + V_{T_1}} \to 0 \quad \text{as } t \to \infty.$$

It can be deduced that, given $\epsilon > 0$, $\exists\ t_5$ such that

$$\left| \frac{T_2(t) + V_{T_2}(t)}{T_1(t) + V_{T_1}(t)} - 0 \right| < \epsilon \quad \forall t \geqslant t_5.$$

So, $0 \leqslant T_2 + V_{T_2} \leqslant \epsilon(T_1 + V_{T_1})$ and $T_1 + V_{T_1} \leqslant 2\frac{L}{u}$ for $t \geqslant t_6 > t_5$. Therefore, for $t \geqslant t_6$, $0 \leqslant T_2 + V_{T_2} \leqslant 2\epsilon\frac{L}{u}$. So $T_2 + V_{T_2} \to 0$ as $t \to \infty$ since ϵ can be made as small as required.

Given $\epsilon > 0$, $\exists\ t_7 \geqslant t_1$ such that for $t \geqslant t_7$, $T_2 + V_{T_2} \leqslant \epsilon$. For $t \geqslant t_7$,

$$\begin{aligned} \frac{1}{T_1 + V_{T_1}} \frac{d}{dt}(T_1 + V_{T_1}) &= \beta_1(X + V) - (\gamma + u), \\ &\leqslant \beta_1 \left(\frac{L}{u} + \frac{\epsilon}{\beta_1} - (T_1 + V_{T_1}) - \frac{(\gamma + u)}{\beta_1} \right). \end{aligned} \tag{A.1}$$

This is true since $X + V = N - T_1 - V_{T_1} - T_2 - V_{T_2}$ and

$$\begin{aligned} \beta_1(X + V) &= \beta_1(N - T_1 - V_{T_1} - T_2 - V_{T_2}), \\ &\leqslant \beta_1(N - T_1 - V_{T_1}), \quad \text{since } T_2 + V_{T_2} > 0. \end{aligned}$$

Hence, from inequality (A.1) by the usual argument it can be deduced that $\exists$ $t_8 \geqslant t_7$ such that for $t \geqslant t_8$

$$T_1 + V_{T_1} \leqslant \frac{L}{u} - \frac{\gamma + u}{\beta_1} + \left(1 + \frac{1}{\beta_1}\right)\epsilon. \tag{A.2}$$

By a similar argument using $N \geqslant \frac{L}{u} - \frac{\epsilon}{\beta_2}$ for $t \geqslant t_7$,

$$\frac{1}{T_1 + V_{T_1}} \frac{d}{dt}(T_1 + V_{T_1}) \geqslant \beta_1 \left(\frac{L}{u} - \frac{\epsilon}{\beta_2} - T_1 - V_{T_1} - \epsilon - \frac{\gamma + u}{\beta_1} \right),$$

so $\exists\ t_9 \geqslant t_8$ such that

$$T_1 + V_{T_1} \geqslant \frac{L}{u} - \frac{\gamma + u}{\beta_1} - \left(2 + \frac{1}{\beta_2}\right)\epsilon \text{ for } t \geqslant t_9, \qquad \text{(A.3)}$$

again by the usual argument. Hence, combining (A.2) and (A.3) and letting $\epsilon \to 0$, $T_1 + V_{T_1} \to \frac{L}{u} - \frac{\gamma+u}{\beta_1}$ as $t \to \infty$ and $X + V = N - T_1 - V_{T_1} - T_2 - V_{T_2} \to \frac{\gamma+u}{\beta_1}$ as $t \to \infty$.

To determine the limiting values of X, V, T_1, T_2, V_{T_1} and V_{T_2}, first consider

$$\frac{d}{dt}(X + T_1 + T_2) = L(1 - f) - u(X + T_1 + T_2).$$

Therefore,

$$X + T_1 + T_2 \to \frac{L(1-f)}{u} \text{ as } t \to \infty.$$

It is known that $T_2 + V_{T_2} \to 0$ as $t \to \infty$. It can be deduced that, $T_2 \to 0$ and $V_{T_2} \to 0$ as $t \to \infty$ since $T_2 \geqslant 0$ and $V_{T_2} \geqslant 0$. Hence, $X + T_1 \to \frac{L(1-f)}{u}$ as $t \to \infty$.

Next, consider

$$\frac{d}{dt}(V + V_{T_1} + V_{T_2}) = Lf - u(V + V_{T_1} + V_{T_2}).$$

As a consequence,

$$V + V_{T_1} + V_{T_2} \to \frac{Lf}{u} \text{ as } t \to \infty,$$

so

$$V + V_{T_1} \to \frac{Lf}{u} \text{ since } V_{T_2} \to 0 \text{ as } t \to \infty.$$

Let $X = (1-f)\bar{X}$, $T_1 = (1-f)\bar{T_1}$, $T_2 = (1-f)\bar{T_2}$, $V = f\bar{V}$ and $V_{T_1} = f\bar{V}_{T_1}$ and $V_{T_2} = f\bar{V}_{T_2}$. Therefore,

$$\frac{d\bar{X}}{dt} = L - u\bar{X} - \beta_1\bar{X}(T_1 + V_{T_1}) + \gamma(\bar{T_1} + \bar{T_2}) - \beta_2\bar{X}(T_2 + V_{T_2}),$$

$$\frac{d\bar{T_1}}{dt} = \beta_1\bar{X}(T_1 + V_{T_1}) - (\gamma + u)\bar{T_1},$$

and

$$\frac{d\bar{T_2}}{dt} = \beta_2\bar{X}(T_2 + V_{T_2}) - (\gamma + u)\bar{T_2}.$$

Since $X + T_1 \to \frac{L(1-f)}{u}$ as $t \to \infty$,

$$(1-f)\bar{X} + (1-f)\bar{T_1} \to \frac{L(1-f)}{u}.$$

Therefore,

$$\bar{X} + \bar{T_1} \to \frac{L}{u}.$$

It can be deduced that

$$\frac{d\bar{T_1}}{dt} \to \beta_1\left(\frac{L}{u} - \bar{T_1}\right)\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right) - (\gamma+u)\bar{T_1} = \beta_1\frac{L}{u}\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right) - \beta_1\frac{L}{u}\bar{T_1}.$$

Consider some arbitrary $\epsilon > 0$. $\exists\ t_{10}$ such that for $t \geqslant t_{10}$,

$$\frac{d\bar{T_1}}{dt} \leqslant \beta_1\frac{L}{u}\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right) - \frac{\beta_1 L}{u}\bar{T_1} + \epsilon.$$

To begin, consider $\bar{T_1}$ decreasing. $\bar{T_1}$ is decreasing when $\frac{d\bar{T_1}}{dt} < 0$. $\frac{d\bar{T_1}}{dt} < 0$ when

$$\beta_1\frac{L}{u}\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right) - \beta_1\frac{L}{u}\bar{T_1} + \epsilon < 0.$$

This is the case when

$$\bar{T_1} > \frac{L}{u} - \frac{\gamma + u}{\beta_1} + \frac{\epsilon u}{\beta_1 L}. \tag{A.4}$$

By the usual argument, $\exists\ t_{11} > t_{10}$ such that

$$\bar{T_1} \leqslant \frac{L}{u} - \frac{\gamma + u}{\beta_1} + \frac{2\epsilon u}{\beta_1 L} \quad \forall t \geqslant t_{11}.$$

Once $\bar{T_1}$ goes beneath the level $\frac{L}{u} - \frac{\gamma+u}{\beta_1} + \frac{2\epsilon u}{\beta_1 L}$ it can never rise above this level.

Similarly, $\exists\ t_{12} > t_{11}$ such that

$$\bar{T_1} \geqslant \frac{L}{u} - \frac{\gamma + u}{\beta_1} - \frac{2\epsilon u}{\beta_1 L} \qquad \forall\ t \geqslant t_{12}.$$

So, once $\bar{T_1}$ is above $\frac{L}{u} - \frac{\gamma+u}{\beta_1} - \frac{2\epsilon u}{\beta_1 L}$ it can never drop beneath it. Therefore, for $t \geqslant \max(t_{11}, t_{12})$,

$$\left| \bar{T_1} - \left(\frac{L}{u} - \frac{\gamma + u}{\beta_1} \right) \right| \leqslant \frac{2\epsilon u}{\beta_1 L}.$$

But $\epsilon > 0$ is arbitrary. It can thus be observed that $\bar{T_1} \to \frac{L}{u} - \frac{\gamma+u}{\beta_1}$ as $t \to \infty$ so

$$T_1 \to (1 - f) \left(\frac{L}{u} - \frac{\gamma + u}{\beta_1} \right) \quad \text{as } t \to \infty.$$

Since $X + T_1 \to \frac{L(1-f)}{u}$,

$$X \to \frac{L}{u}(1 - f) - (1 - f) \left(\frac{L}{u} - \frac{\gamma + u}{\beta_1} \right) \quad \text{as } t \to \infty.$$

Therefore,

$$X \to (1 - f)\frac{\gamma + u}{\beta_1}.$$

It has been established that $X + V \to \frac{\gamma+u}{\beta_1}$. So,

$$V \to \frac{\gamma+u}{\beta_1} - (1-f)\frac{(\gamma+u)}{\beta_1} = f\frac{(\gamma+u)}{\beta_1} \text{ as } t \to \infty.$$

Since $V + V_{T_1} \to \frac{Lf}{u}$ as $t \to \infty$,

$$V_{T_1} \to f\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right).$$

In conclusion, when $R_e \geqslant 1$, $T_1(t) + V_{T_1}(t) > 0$ and $T_2(t) + V_{T_2}(t) > 0$,

$$X \to (1-f)\frac{\gamma+u}{\beta_1},\ V \to f\frac{\gamma+u}{\beta_1},\ T_1 \to (1-f)\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right),$$
$$V_{T_1} \to f\left(\frac{L}{u} - \frac{\gamma+u}{\beta_1}\right),\ T_2 \to 0 \text{ and } V_{T_2} \to 0 \text{ as } t \to \infty.$$

A.2 Model of two MLSTs with transmission due to serotype

A.2.1 Global stability analysis for $R_e > 1$ when $T_1(0) > 0$, or $V_{T_1}(0) > 0$ and $T_2(0) = V_{T_2}(0) = 0$.

Let $T = (P\beta_1 + (1-P)\beta_2)T_1 + \beta_2 V_{T_1}$. Then

$$\frac{dT}{dt} = \left[(P\beta_1 + (1-P)\beta_2)X + \beta_2 V - (\gamma+u)\right]T.$$

Since $X + T_1 \to \frac{L(1-f)}{u}$ and $V + V_{T_1} \to \frac{Lf}{u}$ as $t \to \infty$,

$$\frac{1}{T}\frac{dT}{dt} \to \left[(P\beta_1 + (1-P)\beta_2)\left(\frac{L(1-f)}{u} - T_1\right) + \beta_2\left(\frac{Lf}{u} - V_{T_1}\right) - (\gamma + u)\right], \tag{A.5}$$

i.e.

$$\frac{1}{T}\frac{dT}{dt} \to (P\beta_1 + (1-P)\beta_2)\frac{L(1-f)}{u} + \beta_2\frac{Lf}{u} - (\gamma + u) - T \text{ as } t \to \infty.$$

Hence, if $R_{e1} \leqslant 1$, $X \to \frac{(1-f)L}{u}$, $T_1 \to 0$, $V \to \frac{Lf}{u}$, and $V_{T_1} \to 0$ as $t \to \infty$.

When $R_{e1} > 1$ it can be shown that

$$T \;\to\; (P\beta_1 + (1-P)\beta_2)\frac{L(1-f)}{u} + \beta_2\frac{Lf}{u} - (\gamma + u) \text{ as } t \to \infty.$$

$$\begin{aligned}\frac{dT_1}{dt} &\to \left(\frac{L(1-f)}{u} - T_1\right)\left((P\beta_1 + (1-P)\beta_2)\frac{L(1-f)}{u} + \beta_2\frac{Lf}{u} - (\gamma + u)\right) \\ &- (\gamma + u)T_1, \\ &= \frac{L(1-f)}{u}\left((P\beta_1 + (1-P)\beta_2)\frac{L(1-f)}{u} + \beta_2\frac{Lf}{u} - (\gamma + u)\right) \\ &- \left((P\beta_1 + (1-P)\beta_2)\frac{L(1-f)}{u} + \beta_2\frac{Lf}{u}\right)T_1.\end{aligned}$$

Therefore,

$$T_1 \to \frac{\frac{L(1-f)}{u}\left((P\beta_1 + (1-P)\beta_2)\frac{L(1-f)}{u} + \beta_2\frac{Lf}{u} - (\gamma + u)\right)}{(P\beta_1 + (1-P)\beta_2)\frac{L(1-f)}{u} + \beta_2\frac{Lf}{u}}$$

$$= (1-f)\left(\frac{L}{u} - \frac{\gamma + u}{(P\beta_1 + (1-P)\beta_2)(1-f) + \beta_2 f}\right), \text{ as } t \to \infty.$$

Similarly,

$$V_{T_1} = f\left(\frac{L}{u} - \frac{\gamma + u}{(P\beta_1 + (1-P)\beta_2)(1-f) + \beta_2 f}\right), \text{ as } t \to \infty.$$

Hence,

$$X \to^{\lim}_{t\to\infty} \left(\frac{L(1-f)}{u} - T_1\right) = \frac{(1-f)(\gamma+u)}{(P\beta_1 + (1-P)\beta_2)(1-f) + \beta_2 f}$$

and

$$V \to^{\lim}_{t\to\infty} \left(\frac{Lf}{u} - V_{T_1}\right) = \frac{f(\gamma+u)}{(P\beta_1 + (1-P)\beta_2)(1-f) + \beta_2 f} \text{ as } t \to \infty.$$

A.2.2 Global stability analysis for $R_e > 1$ when $T_1(0) + V_{T_1}(0) > 0$ and $T_2(0) + V_{T_2}(0) > 0$

Similarly to previous arguments, it can be shown that when $T_1(0) + V_{T_1}(0) > 0$ and $T_2(0) + V_{T_2}(0) > 0$, $T_1(t) + V_{T_1}(t) > 0$ and $T_2(t) + V_{T_2}(t) > 0 \ \forall \ t$. Let $a = P\beta_1 + (1-P)\beta_2$, $b = \beta_2$ and $c = Q\beta_1 + (1-Q)\beta_2$.

Consider

$$\begin{aligned}
\frac{d}{dt}\left(\frac{aT_1 + bV_{T_1}}{cT_2 + bV_{T_2}}\right) &= \frac{(a\dot{T_1} + b\dot{V_{T_1}})(cT_2 + bV_{T_2}) - (aT_1 + bV_{T_1})(c\dot{T_2} + b\dot{V_{T_2}})}{(cT_2 + bV_{T_2})^2}, \\
&= (a-c)X\frac{(aT_1 + bV_{T_1})}{(cT_2 + bV_{T_2})}, \\
&= ((P\beta_1 + (1-P)\beta_2) - (Q\beta_1 + (1-Q)\beta_2))\, X\frac{(aT_1 + bV_{T_1})}{(cT_2 + bV_{T_2})}.
\end{aligned}$$

That is, if it can be shown that $X \geqslant \epsilon > 0$ for $t \geqslant t_0$ for some ϵ, t_0 then for $P\beta_1 + (1-P)\beta_2 > Q\beta_1 + (1-Q)\beta_2$

$$\frac{aT_1 + bV_{T_1}}{cT_2 + bV_{T_2}} \to \infty, \text{ as } t \to \infty,$$

whereas for $Q\beta_1 + (1-Q)\beta_2 > P\beta_1 + (1-P)\beta_2$,

$$\frac{aT_1 + bV_{T_1}}{cT_2 + bV_{T_2}} \to 0, \text{ as } t \to \infty.$$

If $Q\beta_1 + (1-Q)\beta_2 = P\beta_1 + (1-P)\beta_2$,

$$\frac{d}{dt}\left(\frac{aT_1 + bV_{T_1}}{cT_2 + bV_{T_2}}\right) = 0.$$

So, in this case, $\frac{aT_1+bV_{T_1}}{cT_2+bV_{T_2}}$ is constant.

However, $X + V + T_1 + V_{T_1} + T_2 + V_{T_2} \to \frac{L}{u}$ as $t \to \infty$. Therefore, $\exists\ t_1$ such that for $t \geqslant t_1$, $X + V + T_1 + V_{T_1} + T_2 + V_{T_2} \leqslant 2\frac{L}{u}$.

For $t \geqslant t_1$,

$$\begin{aligned}\frac{dX}{dt} &\geqslant L(1-f) - (u + aT_1 + bV_{T_1} + cT_2 + bV_{T_2})X, \\ &\geqslant L(1-f) - \left(u + \max(a,b,c)2\frac{L}{u}\right)X.\end{aligned}$$

For

$$X \leqslant \frac{L(1-f)}{2\left(u + \max(a,b,c)2\frac{L}{u}\right)},\ \ \frac{dX}{dt} \geqslant \frac{1}{2}L(1-f) > 0.$$

So, $\exists\ t_2 > t_1$ such that

$$X \geqslant \frac{\frac{1}{2}L(1-f)}{u + \max(a,b,c)2\frac{L}{u}} \text{ for } t \geqslant t_2.$$

That is, $X \geqslant \epsilon > 0$ for $t \geqslant t_2$ as required.

Write $\hat{\xi} = \frac{aT_1+bV_{T_1}}{cT_2+bV_{T_2}}$. $\hat{\xi}$ may be defined this way since $cT_2(t) + bV_{T_2}(t) > 0$ as $T_2(t) + V_{T_2}(t) > 0$ and $b, c > 0$.

$$\frac{1}{\hat{\xi}}\frac{d\hat{\xi}}{dt} = ((P\beta_1 + (1-P)\beta_2) - (Q\beta_1 + (1-Q)\beta_2))X.$$

Suppose first that $P\beta_1 + (1-P)\beta_2 > Q\beta_1 + (1-Q)\beta_2$. Then, since $X \geqslant \epsilon > 0$, $((P\beta_1 + (1-P)\beta_2) - (Q\beta_1 + (1-Q)\beta_2))X > 0$ so $\hat{\xi} \to \infty$ as $t \to \infty$. Therefore, $\frac{1}{\hat{\xi}} \to 0$ as $t \to \infty$. That is,

$$\frac{cT_2 + bV_{T_2}}{aT_1 + bV_{T_1}} \to 0 \text{ as } t \to \infty.$$

It can be deduced that, given $\epsilon > 0$, $\exists\ t_3$ such that

$$\left|\frac{cT_2(t) + bV_{T_2}(t)}{aT_1(t) + bV_{T_1}(t)} - 0\right| < \epsilon \ \forall t \geqslant t_3.$$

Hence, $0 \leqslant cT_2 + bV_{T_2} \leqslant \epsilon(aT_1 + bV_{T_1})$ and $aT_1 + bV_{T_1} \leqslant 2\max(a,b)\frac{L}{u}$. So, $0 \leqslant cT_2 + bV_{T_2} \leqslant 2\epsilon\frac{L}{u}$. As a consequence, T_2 and $V_{T_2} \to 0$ as $t \to \infty$ since b and c are strictly positive constants.

Consider

$$\frac{d}{dt}(X + T_1 + T_2) = L(1-f) - u(X + T_1 + T_2).$$

So,

$$X + T_1 + T_2 \to \frac{L(1-f)}{u} \text{ as } t \to \infty.$$

Similarly,

$$\frac{d}{dt}(V + V_{T_1} + V_{T_2}) = Lf - u(V + V_{T_1} + V_{T_2}).$$

So,

$$V + V_{T_1} + V_{T_2} \to \frac{Lf}{u} \text{ as } t \to \infty.$$

However, $T_2 \to 0$ and $V_{T_2} \to 0$ as $t \to \infty$. Therefore, $X + T_1 \to \frac{L(1-f)}{u}$ and $V + V_{T_1} \to \frac{Lf}{u}$ as $t \to \infty$. Then the same argument as that shown previously in Chapter 3 for the GSA of $R_e > 1$ and $T_2(0) = V_{T_2}(0) = 0$ (given in Appendix A.2.1) can be applied to show that as $t \to \infty$

$$X \to \frac{(1-f)(\gamma+u)}{(P\beta_1 + (1-P)\beta_2)(1-f) + \beta_2 f},$$
$$T_1 \to (1-f)\left(\frac{L}{u} - \frac{\gamma+u}{(P\beta_1 + (1-P)\beta_2)(1-f) + \beta_2 f}\right),$$

$$V \to \frac{f(\gamma+u)}{(P\beta_1 + (1-P)\beta_2)(1-f) + \beta_2 f},$$

and

$$V_{T_1} \to f\left(\frac{L}{u} - \frac{\gamma+u}{(P\beta_1 + (1-P)\beta_2)(1-f) + \beta_2 f}\right).$$

A similar argument shows that if $Q\beta_1 + (1-Q)\beta_2 > P\beta_1 + (1-P)\beta_2$, as $t \to \infty$

$$X \to \frac{(1-f)(\gamma+u)}{(Q\beta_1 + (1-Q)\beta_2)(1-f) + \beta_2 f},$$

$$T_2 \to (1-f)\left(\frac{L}{u} - \frac{\gamma+u}{(Q\beta_1 + (1-Q)\beta_2)(1-f) + \beta_2 f}\right),$$

$$V \to \frac{f(\gamma+u)}{(Q\beta_1+(1-Q)\beta_2)(1-f)+\beta_2 f},$$

and

$$V_{T_2} \to f\left(\frac{L}{u} - \frac{\gamma+u}{(Q\beta_1+(1-Q)\beta_2)(1-f)+\beta_2 f}\right).$$

When $Q\beta_1+(1-Q)\beta_2 = P\beta_1+(1-P)\beta_2$ then the equations become

$$\begin{aligned}\frac{dX}{dt} &= L(1-f) - uX - (P\beta_1+(1-P)\beta_2)X(T_1+T_2)\\ &\quad - \beta_2 X(V_{T_1}+V_{T_2}) + \gamma(T_1+T_2),\end{aligned}$$

$$\begin{aligned}\frac{d(T_1+T_2)}{dt} &= (P\beta_1+(1-P)\beta_2)X(T_1+T_2) + \beta_2 X(V_{T_1}+V_{T_2})\\ &\quad - (\gamma+u)(T_1+T_2),\end{aligned}$$

$$\begin{aligned}\frac{dV}{dt} &= Lf - uV - \beta_2 V(V_{T_1}+V_{T_2}) - (P\beta_1+(1-P)\beta_2)V(T_1+T_2)\\ &\quad + \gamma(V_{T_1}+V_{T_2}),\end{aligned}$$

and

$$\begin{aligned}\frac{d(V_{T_1}+V_{T_2})}{dt} &= \beta_2 V(V_{T_1}+V_{T_2}) + (P\beta_1+(1-P)\beta_2)V(T_1+T_2)\\ &\quad - (\gamma+u)(V_{T_1}+V_{T_2}).\end{aligned}$$

The next result follows from that of the global stability argument described in the previous section of this appendix (A.2), with T_1 replaced by T_1+T_2 and V_{T_1} replaced by V_{T_1} replaced by $V_{T_1}+V_{T_2}$. That is,

$$T_1+T_2 \to (1-f)\left(\frac{L}{u} - \frac{\gamma+u}{(P\beta_1+(1-P)\beta_2)(1-f)+\beta_2 f}\right),$$

$$X \to \frac{(1-f)(\gamma+u)}{(P\beta_1+(1-P)\beta_2)(1-f)+\beta_2 f},$$

$$V_{T_1}+V_{T_2} \to f\left(\frac{L}{u}-\frac{(\gamma+u)}{(P\beta_1+(1-P)\beta_2)(1-f)+\beta_2 f}\right),$$

and

$$V \to \frac{f(\gamma+u)}{(P\beta_1+(1-P)\beta_2)(1-f)+\beta_2 f} \text{ as } t\to\infty.$$

X, T_1, T_2, V, V_{T_1}, V_{T_2} tend to the surface given by $X=(1-f)(\frac{L}{u}-\xi)$, $V=f(\frac{L}{u}-\xi)$, $V_{T_1}+V_{T_2}=f\xi$ and $T_1+T_2=(1-f)\xi$, where

$$\xi=\frac{L}{u}-\frac{\gamma+u}{(P\beta_1+(1-P)\beta_2)(1-f)+\beta_2 f}.$$

Furthermore, $aT_2+bV_{T_2}=k(aT_1+bV_{T_1})$ where k is a constant. On the above surface, if $V_{T_1}=\alpha_1 f\xi$ then $0\leqslant\alpha_1\leqslant 1$ and $V_{T_2}=(1-\alpha_1)f\xi$. Similarly, if $T_1=\alpha_2(1-f)\xi$ then $0\leqslant\alpha_2\leqslant 1$ and $T_2=(1-\alpha_2)(1-f)\xi$. Now,

$$\begin{aligned}\frac{d}{dt}\left(\frac{T_1}{T_2}\right)&=\frac{\dot{T_1}T_2-T_1\dot{T_2}}{{T_2}^2},\\ &=\frac{1}{{T_2}^2}\left(T_2(aXT_1+bXV_{T_1}-(\gamma+u)T_1)-T_1(aXT_2+bXV_{T_2}\right.\\ &\left.-(\gamma+u)T_2)\right),\\ &=\frac{X}{{T_2}^2}(T_2T-T_1kT),\end{aligned} \tag{A.6}$$

where, as before, $T=aT_1+bV_{T_1}$

$$= \frac{XT}{T_2}\left(1 - \frac{T_1 k}{T_2}\right).$$

It can be asserted that $\exists\ t_5 > 0$ and $\epsilon_1 > 0$ such that for $t \geqslant t_5$, $T \geqslant \epsilon_1 > 0$. Note that

$$X + V \to \frac{\gamma + u}{(P\beta_1 + (1-P)\beta_2)(1-f) + \beta_2 f} < \frac{L}{u}.$$

Hence, $\exists\ \epsilon_2 > 0$ and $t_6 > 0$ such that for $t \geqslant t_6$

$$\left|X + V - \frac{L}{u}\right| \geqslant \epsilon_2 > 0.$$

$\exists\ t_5 \geqslant t_6$ such that for $t \geqslant t_5$, $\left|N - \frac{L}{u}\right| \leqslant \frac{\epsilon_2}{2}$. For $t \geqslant t_5$, $(k+1)T = |aT_1 + bV_{T_1} + aT_2 + bV_{T_2}| \geqslant \min(a,b)\ |T_1 + V_{T_1} + T_2 + V_{T_2}| = \min(a,b)|N - (X+V)|$. But $|N - (X+V)| + |N - \frac{L}{u}| \geqslant |\frac{L}{u} - (X+V)|$. Hence, for $t \geqslant t_5$,

$$|N - (X+V)| \geqslant \left|\frac{L}{u} - (X+V)\right| - \left|N - \frac{L}{u}\right| \geqslant \frac{\epsilon_2}{2},$$

so

$$T \geqslant \epsilon_1 = \frac{\min(a,b)}{2(k+1)}\epsilon_2 > 0.$$

Hence $\exists\ t_7 > 0$ and $\epsilon, \epsilon_1 > 0$ such that for $t \geqslant t_7$, $X \geqslant \epsilon > 0$, $T \geqslant \epsilon_1 > 0$ and $T_2 \leqslant 2\frac{L}{u}$.

Thus, if $k\frac{T_1}{T_2} > 1$ then $\frac{d}{dt}\frac{T_1}{T_2} < 0$ and if $k\frac{T_1}{T_2} < 1$ then $\frac{d}{dt}\frac{T_1}{T_2} > 0$. Using a similar argument to one used previously it is straightforward to show that $\frac{T_1}{T_2} \to \frac{1}{k}$ as $t \to \infty$.

Now,

$$b(kV_{T_1} - V_{T_2}) = k(aT_1 + bV_{T_1}) - (aT_2 + bV_{T_2}) - a(kT_1 - T_2),$$

$$= -a(kT_1 - T_2) \to 0 \text{ as } t \to \infty.$$

So

$$\frac{V_{T_1}}{V_{T_2}} \to \frac{1}{k} \text{ as } t \to \infty.$$

Hence

$$\frac{\alpha_1}{1-\alpha_1} \to \frac{1}{k} \text{ and } \frac{\alpha_2}{1-\alpha_2} \to \frac{1}{k} \text{ as } t \to \infty.$$

So $\alpha_1 \to \alpha_2$ as $t \to \infty$. Moreover, for all times t,

$$\begin{aligned} aT_2 + bV_{T_2} &= k(aT_1 + bV_{T_1}), \\ a(1-f)(1-\alpha_2) + b(1-\alpha_1)f &= k(a\alpha_2(1-f) + b\alpha_1 f), \\ a(1-f) + bf &= \alpha_1 bf + \alpha_2 a(1-f) + ka\alpha_2(1-f) + kb\alpha_1 f, \\ &= \alpha_1(bf + a(1-f))(k+1) + \\ &\quad (\alpha_2 - \alpha_1)a(1-f)(k+1). \end{aligned}$$

So,

$$\alpha_1 = \frac{a(1-f) + bf - (\alpha_2 - \alpha_1)a(1-f)(k+1)}{(k+1)(a(1-f) + bf)}.$$

Thus,

$$\alpha_1 \to \frac{1}{k+1} \quad \text{as } t \to \infty.$$

So, $\alpha_1, \alpha_2 \to \frac{1}{k+1}$ as $t \to \infty$. So X, V, T_1, T_2, V_{T_1} and V_{T_2} approach the equilib-

rium point where

$$X = (1-f)\left(\frac{L}{u} - \xi\right), \;\; T_1 = (1-f)\alpha\xi, \;\; T_2 = (1-f)(1-\alpha)\xi,$$
$$V = f\left(\frac{L}{u} - \xi\right), \;\; V_{T_1} = \alpha f\xi, \;\; \text{and} \;\; V_{T_2} = f\xi(1-\alpha), \;\; \text{for} \;\; \alpha = \frac{1}{1+k}$$

and $0 < \alpha < 1$. k is given in terms of the initial conditions by

$$\frac{aT_2(0) + bV_{T_2}(0)}{aT_1(0) + bV_{T_1}(0)}.$$

References

Abuelreish, M., A. Subedar, T. Chiu, P. Wludyka, and M. Rathore (2007). Increase In Invasive Pneumococcal Disease In Children Associated With Shortage Of Heptavalent Pneumococcal Conjugate Vaccine. *Clinical Pediatrics 46*(1), 45–52.

Adrian, P. V., D. Bogaert, M. Oprins, S. Rapola, M. Lahdenkari, T. Kilpi, R. de Groot, H. Käyhty, and P. W. M. Hermans (2004). Development Of Antibodies Against Pneumococcal Proteins α-Enolase, Immunoglobulin A1 Protease, Streptococcal Lipoprotein Rotamase A, And Putative Proteinase Maturation Protein A In Relation To Pneumococcal Carriage And Otitis Media. *Vaccine 22*, 2737–2742.

Agresti, A. (2007). *An Introduction To Categorical Data Analysis* (Second ed.). John Wiley and Sons, Inc., Hoboken, New Jersey.

Aguiar, S. I., I. Serrano, F. R. Pinto, J. Melo-Cristino, and M. Ramirez (2008). Changes In *Streptococcus Pneumoniae* Serotypes Causing Invasive Disease With Non-Universal Vaccination Coverage Of The Seven-Valent Conjugate Vaccine. *Clinical Microbiology and Infection 14*(9), 835–843.

Akmatov, M. K., M. Kretzschmer, A. Krämer, and R. T. Mikolajczyk (2008). Timeliness Of Vaccination And Its Effects On Fraction Of Vaccinated Population. *Vaccine 26*(31), 3805–3811.

Alanee, S. R. J., L. McGee, D. Jackson, C. C. Chiou, C. Feldman, A. J. Morris, A. Ortqvist, J. Rello, C. M. Luna, L. M. Baddour, M. Ip, V. L. Yu, and K. P. Klugman (2007). Association Of Serotypes Of *Streptococcus Pneumoniae* With Disease Severity And Outcome In Adults: An International Study. *Clinical Infectious Diseases 45*, 46–51.

Albrich, W. C., W. Baughman, B. Schmotzer, and M. M. Farley (2007). Changing Characteristics Of Invasive Pneumococcal Disease In Metropolitan Atlanta, Georgia, After Introduction Of A 7-Valent Pneumococcal Conjugate Vaccine. *Clinical Infectious Diseases 44*, 1569–1576.

Ali, M., V. D. Thiem, J. Park, R. L. Ochiai, D. G. Canh, M. C. Danovaro-Holliday, L. M. Kaljee, J. D. Clemens, and C. J. Acosta (2007). Geographic Analysis Of Vaccine Uptake In A Cluster-Randomized Controlled Trial In Hue, Vietnam. *Health and Place 13*, 577–587.

Amrine-Madsen, H., J. Van Eldere, R. M. Mera, L. A. Miller, J. A. Poupard, E. S. Thomas, W. S. Halsey, J. A. Becker, and F. P. O'Hara (2008). Temporal And Spatial Distribution Of Clonal Complexes Of *Streptococcus Pneumoniae* Isolates Resistant To Multiple Classes Of Antibiotics In Belgium, 1997 To 2004. *Antimicrobial Agents and Chemotherapy 52*(9), 3216–3220.

Andersson, M., K. Ekdahl, S. Mölstad, K. Persson, H. B. Hansson, and J. Giesecke (2005). Modelling The Spread Of Penicillin-Resistant *Streptococcus Pneumoniae* In Day-Care And Evaluation Of Intervention. *Statistics in Medicine 24*, 3593–3607.

Ardanuy, C., F. Tubau, R. Pallares, L. Calatayud, M. A. Domínguez, D. Rolo, I. Grau, R. Martín, and J. Liñares (2009). Epidemiology Of Invasive Pneumococcal Disease Among Adult Patients In Barcelona Before And After Pediatric 7-Valent Pneumococcal Conjugate Vaccine Introduction, 1997-2007. *Clinical Infectious Diseases 48*, 57–64.

Bailey, N. T. J. (1957). *The Mathematical Theory Of Epidemics*. Griffin, London, England.

Balakrishnan, I., P. Crook, R. Morris, and S. H. Gillespie (2000). Early Predictors Of Mortality In Pneumococcal Bacteraemia. *Journal of Infection 40*(3), 256–261.

Bardenheier, B. H., A. Shefer, L. Barker, C. A. Winston, and C. K. Sionean (2005). Public Health Application Comparing Multilevel Analysis With Logistic Regression: Immunization Coverage Among Long-Term Care Facility Residents. *Annals of Epidemiology 15*(10), 749–755.

Bardenheier, B. H., H. R. Yusuf, J. Rosenthal, J. M. Santoli, A. M. Shefer, D. L. Rickert, and S. Y. Chu (2004). Factors Associated With Underimmunization At 3 Months Of Age In Four Medically Underserved Areas. *Public Health Reports 119*(5), 479–485.

Bartsch, G., W. Horninger, H. Klocker, A. Pelzer, J. Bektic, W. Oberaigner, H. Schennach, G. Schäfer, F. Frauscher, M. Boniol, G. Severi, C. Robertson, and P. Boyle (2008). Tyrol Prostate Cancer Demonstration Project: Early Detection, Treatment, Outcome, Incidence And Mortality. *British Journal of Urology International 101*(7), 809–816.

Beall, B., M. C. McEllistrem, R. E. Gertz Jr., S. Wedel, D. J. Boxrud, A. L. Gonzalez, M. Medina, R. Pai, T. A. Thompson, L. H. Harrison, L. McGee, C. G. Whitney, and the Active Bacterial Core Surveillance Team (2006). Pre- And Postvaccination Clonal Compositions Of Invasive Pneumococcal Serotypes For Isolates Collected In The United States In 1999, 2001, And 2002. *Journal of Clinical Microbiology 44*(3), 999–1017.

Bergman, A., J. Hjelmgren, Å. Örtqvist, T. Wisløff, I. S. Kristiansen, L. D. Högberg, K. M. Persson, and U. Persson (2008). Cost-effectiveness Analysis Of A Universal Vaccination Programmeme With The 7-Valent Pneumococcal Conjugate Vaccine (PCV-7) In Sweden. *Scandinavian Journal of Infectious Diseases 40*(9), 721–729.

Bernoulli, D. (1760). Essai D'Une Nouvelle Analyse De La Mortalité Causée Par La Petite Vérole Et Les Avantages De L'Inoculation Pour La Prévenir. *Mémoires de Mathématique et de Physique, Académie Royale des Sciences, Paris*, 1–45.

Beutels, P., N. Thiry, and P. V. Damme (2007). Convincing Or Confusing? Economic Evaluations Of Childhood Pneumococcal Conjugate Vaccination-A Review (2002-2006). *Vaccine 25*(8), 1355–1367.

Bobo, J. K., J. L. Gale, P. B. Thapa, and S. G. F. Wassilak (1993). Risk

Factors For Delayed Immunization In A Random Sample Of 1163 Children From Oregon And Washington. *Pediatrics 91*(2), 308–314.

Bogaert, D., M. Sluijter, N. L. Toom, T. J. Mitchell, W. H. F. Goessens, S. C. Clarke, R. de Groot, and P. W. M. Hermans (2006). Dynamics Of Pneumococcal Colonization In Healthy Dutch Children. *Microbiology 152*, 377–385.

Bos, J. M., H. Rümke, R. Welte, and M. J. Postma (2003). Epidemiologic Impact And Cost-effectiveness Of Universal Infant Vaccination With A 7-Valent Conjugated Pneumococcal Vaccine In The Netherlands. *Clinical Therapeutics 25*(10), 2614–2630.

Bowman, A. W. and A. Azzalini (1997). *Applied Smoothing Techniques For Data Analysis*. Oxford University Press, Oxford, England.

Briggs, A., K. Claxton, and M. Sculpher (2006). *Decision Modelling For Health Economic Evaluation*. Oxford University Press, Oxford, England.

Briggs, A. and M. Sculpher (1998). An Introduction To Markov Modelling For Economic Evaluation. *Pharmacoeconomics 13*(4), 397–409.

Brisson, M. and W. J. Edmunds (2002). The Cost-effectiveness Of Varicella Vaccination In Canada. *Vaccine 20*, 1113–1125.

Brueggemann, A. B. and G. V. Doern (2000). Resistance Among *Streptococcus Pneumoniae*: Patterns, Mechanisms, Interpreting The Breakpoints. *The American Journal of Managed Care 6*(23), S1189–S1196.

Brueggemann, A. B., D. T. Griffiths, E. Meats, T. Peto, D. W. Crook, and B. G. Spratt (2003). Clonal Relationships Between Invasive And Carriage *Streptococcus Pneumoniae* And Serotype- And Clone-Specific Differences In Invasive Disease Potential. *The Journal of Infectious Diseases 187*(9), 1424–1432.

Brueggemann, A. B., R. Pai, D. W. Crook, and B. Beall (2007). Vaccine Escape Recombinants Emerge After Pneumococcal Vaccination In The United States. *Public Library of Science Pathogens 3*(11), 1628–1636.

Brueggemann, A. B., T. E. A. Peto, D. W. Crook, J. C. Butler, K. G. Kristinsson, and B. G. Spratt (2004). Temporal And Geographic Stability Of The

Serogroup-Specific Invasive Disease Potential Of *Streptococcus Pneumoniae* In Children. *The Journal of Infectious Diseases 190*, 1203–1211.

Brueggemann, A. B. and B. G. Spratt (2003). Geographic Distribution And Clonal Diversity Of *Streptococcus Pneumoniae* Serotype 1 Isolates. *Journal of Clinical Microbiology 41*(11), 4966–4970.

Brugger, S. D., L. J. Hathaway, and K. Mühlemann (2009). Detection Of *Streptococcus Pneumoniae* Strain Cocolonization In The Nasopharynx. *Journal of Clinical Microbiology 47*(6), 1750–1756.

Butler, J. C., R. F. Breiman, H. B. Lipman, J. Hofmann, and R. R. Facklam (1995). Serotype Distribution Of *Streptococcus Pneumoniae* Infections Among Preschool Children In The United States, 1978-1994: Implications For Development Of A Conjugate Vaccine. *The Journal of Infectious Diseases 171*(4), 885–889.

Butler, J. R. G., P. McIntyre, C. R. MacIntyre, R. Gilmour, A. L. Howarth, and B. Sander (2004). The Cost-effectiveness Of Pneumococcal Conjugate Vaccination In Australia. *Vaccine 22*, 1138–1149.

Cairns, J. (2003). Providing Guidance To The NHS: The Scottish Medicines Consortium And The National Institute For Clinical Excellence Compared. *Health Policy 76*(2), 134–143.

Cameron, J. C., V. Friederichs, and C. Robertson (2007). Vaccine Uptake: New Tools For Investigating Changes In Age Distribution And Predicting Final Values. *Vaccine 25*(32), 6078–6085.

Census Dissemination Unit (2001). General Register Office For Scotland, 2001 Census: Standard Area Statistics (Scotland). `http://casweb.mimas.ac.uk`. ESRC/JISC Census Programme, Census Dissemination Unit, Mimas (University of Manchester).

Chen, C., C. Lin, Y. Chen, C. Wang, C. Chiu, T. Lin, and Y. Huang (2009). Host And Microbiologic Factors Associated With Mortality In Taiwanese Children With Invasive Pneumococcal Diseases, 2001 To 2006. *Diagnostic Microbiology and Infectious Disease 63*(2), 194–200.

Claes, C. and J. M. Graf von der Schulenburg (2009). Cost Effectiveness Analysis Of Heptavalent Pneumococcal Conjugate Vaccine In Germany Consid-

ering Herd Immunity Effects. *European Journal of Health Economics 10*(1), 25–38.

Claes, C., R. R. Reinert, and J. M. von der Schulenburg (2003). Cost Effectiveness Of Pneumococcal Vaccination For Infants And Children With The Conjugate Vaccine PnC-7 In Germany. *Pharmacoeconomics 21*(8), 587–600.

Clark, A. and C. Sanderson (2009). Timing Of Childrens' Vaccinations In 45 Low-Income And Middle-Income Countries: An Analysis Of Survey Data. *The Lancet 373*, 1543–1549.

Clarke, S. C., K. J. Scott, and S. M. McChlery (2004). Serotypes And Sequence Types Of Pneumococci Causing Invasive Disease In Scotland Prior To The Introduction Of Pneumococcal Conjugate Polysaccharide Vaccines. *Journal of Clinical Microbiology 42*(10), 4449–4452.

Coffey, T. J., M. C. Enright, M. Daniels, J. K. Morona, R. Morona, W. Hryniewicz, J. C. Paton, and B. G. Spratt (1998). Recombinational Exchanges At The Capsular Polysaccharide Biosynthetic Locus Lead To Frequent Serotype Changes Among Natural Isolates Of *Streptococcus Pneumoniae*. *Molecular Microbiology 27*(1), 73–83.

Cohen, J. (1960). A Coefficient Of Agreement For Nominal Scales. *Educational and Psychological Measurement 20*, 37–46.

Cohen, R., C. Levy, F. de La Rocque, N. Gelbert, A. Wollner, B. Fritzell, E. Bonnet, R. Tetelboum, and E. Varon (2006). Impact Of Pneumococcal Conjugate Vaccine And Of Reduction Of Antibiotic Use On Nasopharyngeal Carriage Of Nonsusceptible Pneumococci In Children With Acute Otitis Media. *The Pediatric Infectious Disease Journal 25*(11), 1001–1007.

Colantonio, L., J. A. Gómez, N. Demarteau, B. Standaert, A. Pichón-Rivière, and F. Augustovski (2009). Cost-effectiveness Analysis Of A Cervical Cancer Vaccine In Five Latin American Countries. *Vaccine 27*(40), 5519–5529.

Crook, D. W., A. B. Brueggemann, K. Sleeman, and T. E. A. Peto (2004). *Pneumococcal Carriage*. ASM Press, Washington DC, USA. In: The Pneumococcus, Edited by E.I. Tuomanen, T.J. Mitchell, D.A. Morrison and B.G. Spratt.

Cullen, M. and E. Gillooly (2006). Pathways: The Clinical Awareness Newsletter From South Manchester Pathology. `http://elearn-uhsm.co.uk/pathhandbook/Pathways_Winter06.pdf`. Accessed on 8th May 2009.

Dagan, R., N. Givon-Lavi, O. Zamir, and D. Fraser (2003). Effect Of A Nonavalent Conjugate Vaccine On Carriage Of Antibiotic-Resistant *Streptococcus Pneumoniae* In Day-Care Centers. *The Pediatric Infectious Disease Journal 22*(6), 532–540.

Dagan, R., N. Givon-Lavi, O. Zamir, M. Sikuler-Cohen, L. Guy, J. Janco, P. Yagupsky, and D. Fraser (2002). Reduction Of Nasopharyngeal Carriage Of *Streptococcus Pneumoniae* After Administration Of A 9-Valent Pneumococcal Conjugate Vaccine To Toddlers Attending Day Care Centers. *The Journal of Infectious Diseases 185*(7), 927–936.

Dagan, R., R. Melamed, M. Muallem, L. Piglansky, D. Greenberg, O. Abramson, P. M. Mendelman, N. Bohidar, and P. Yagupsky (1996). Reduction Of Nasopharyngeal Carriage Of Pneumococci During The Second Year Of Life By A Heptavalent Conjugate Pneumococcal Vaccine. *The Journal of Infectious Diseases 174*(6), 1271–1278.

Dasbach, E. J., R. P. Insinga, and E. H. Elbasha (2008). The Epidemiological And Economic Impact Of A Quadrivalent Human Papillomavirus Vaccine (6/11/16/18) In The UK. *BJOG: An International Journal of Obstetrics and Gynaecology 115*(8), 947–956.

Datar, A., A. Mukherji, and N. Sood (2007). Health Infrastructure And Immunisation Coverage In Rural India. *Indian Journal of Medical Research 125*(1), 31–42.

Davis, M. M., S. M. Ndiaye, G. L. Freed, and S. J. Clark (2003). One-Year Uptake Of Pneumococcal Conjugate Vaccine: A National Survey Of Family Physicians And Pediatricians. *The Journal of the American Board of Family Practice 16*(5), 363–371.

Dayan, G. H., K. M. Shaw, A. L. Baughman, L. C. Orellana, R. Forlenza, A. Ellis, J. Chaui, S. Kaplan, and P. Strebel (2006). Assessment Of Delay In Age-Appropriate Vaccination Using Survival Analysis. *American Journal of Epidemiology 163*(6), 561–570.

De Carvalho Gomes, H., M. Muscat, D. L. Monnet, J. Giesecke, and P. L. Lopalco (2009). Use Of Seven-Valent Pneumococcal Conjugate Vaccine (PCV7) In Europe, 2001-2007. *Eurosurveillance 14*(12), 1–6.

De Wals, P., G. Petit, L. J. Erickson, M. Guay, T. Tam, B. Law, and A. Framarin (2003). Benefits And Costs Of Immunization Of Children With Pneumococcal Conjugate Vaccine In Canada. *Vaccine 21*, 3757–3764.

Department of Health (2007). Changes To The Green Book: Changes To The Discount Rate. `http://www.dh.gov.uk/en/Procurementandproposals/Publicprivatepartnership/Privatefinanceinitiative/Changestotreasurygreenbook/DH_4016196`. Accessed 28th July 2008.

Diekmann, O. and J. A. P. Heesterbeek (2000). *Mathematical Epidemiology Of Infectious Diseases: Model Building, Analysis, And Interpretation*. Wiley, West Sussex, England.

Dietz, K. and J. A. P. Heesterbeek (2000). Bernoulli Was Ahead Of Modern Epidemiology. *Nature 408*, 513–514.

Dombkowski, K. J., P. M. Lantz, and G. L. Freed (2002). The Need For Surveillance Of Delay In Age-Appropriate Immunization. *American Journal of Preventive Medicine 23*(1), 36–42.

Dombkowski, K. J., P. M. Lantz, and G. L. Freed (2004). Risk Factors For Delay In Age-Appropriate Vaccination. *Public Health Reports 119*(2), 144–155.

Drummond, M. F., M. J. Sculpher, G. W. Torrance, B. J. O'Brien, and G. L. Stoddart (2005). *Methods For The Economic Evaluation Of Health Care Programmes* (Third ed.). Oxford University Press, Oxford, England.

Durrant, G. B. and F. Steele (2009). Multilevel Modelling Of Refusal And Non-Contact In Household Surveys: Evidence From Six UK Government Surveys. *Journal of the Royal Statistical Society, Series A 172*(2), 361–381.

Enright, M. C. and B. G. Spratt (1998). A Multilocus Sequence Typing Scheme For *Streptococcus Pneumoniae*: Identification Of Clones Associated With Serious Invasive Disease. *Microbiology 144*, 3049–3060.

Ess, S. M., U. B. Schaad, A. Gervaix, S. Pinösch, and T. D. Szucs (2003).

Cost-effectiveness Of A Pneumococcal Conjugate Immunisation Program For Infants In Switzerland. *Vaccine 21*(23), 3273–3281.

Evans, D. (2003). Hierarchy Of Evidence: A Framework For Ranking Evidence Evaluating Health Care Interventions. *Journal of Clinical Nursing 12*(1), 77–84.

Ewig, S., M. Ruiz, A. Torres, F. Marco, J. A. Martinez, M. Sanchez, and J. Mensa (1999). Pneumonia Acquired In The Community Through Drug-Resistant *Streptococcus Pneumoniae*. *American Journal of Respiratory and Critical Care Medicine 159*(6), 1835–1842.

Farrell, D. J., D. Felmingham, J. Shackcloth, L. Williams, K. Maher, R. Hope, D. M. Livermore, R. C. George, G. Brick, S. Martin, and R. Reynolds (2008). Non-Susceptibility Trends And Serotype Distributions Among *Streptococcus Pneumoniae* From Community-Acquired Respiratory Tract Infections And From Bacteraemias In The UK And Ireland, 1999 To 2007. *Journal of Antimicrobial Chemotherapy 62*(S2), ii87–ii95.

Fedson, D. S. (1999). The Clinical Effectiveness Of Pneumococcal Vaccination: A Brief Review. *Vaccine 17*(S1), S85–S90.

Feikin, D. R. and K. P. Klugman (2002). Historical Changes In Pneumococcal Serogroup Distribution: Implications For The Era Of Pneumococcal Conjugate Vaccines. *Clinical Infectious Diseases 35*(5), 547–555.

Fiks, A. G., E. A. Alessandrini, A. A. Luberti, S. Ostapenko, X. Zhang, and J. H. Silber (2006). Identifying Factors Predicting Immunization Delay For Children Followed In An Urban Primary Care Network Using An Electronic Health Record. *Pediatrics 118*(6), e1680–e1686.

Finkelstein, J. A., S. S. Huang, J. Daniel, S. L. Rifas-Shiman, K. Kleinman, D. Goldmann, S. I. Pelton, A. DeMaria, and R. Platt (2003). Antibiotic-Resistant *Streptococcus Pneumoniae* In The Heptavalent Pneumococcal Conjugate Vaccine Era: Predictors Of Carriage In A Multicommunity Sample. *Pediatrics 112*(4), 862–869.

Foster, D., K. Knox, A. S. Walker, D. T. Griffiths, H. Moore, E. Haworth, T. Peto, A. B. Brueggemann, and D. W. Crook (2008). Invasive Pneumococcal Disease: Epidemiology In Children And Adults Prior To Imple-

mentation Of The Conjugate Vaccine In The Oxfordshire Region, England. *Journal of Medical Microbiology 57*, 480–487.

Friederichs, V., J. C. Cameron, and C. Robertson (2006). Impact Of Adverse Publicity On MMR Vaccine Uptake: A Population Based Analysis Of Vaccine Uptake Records For One Million Children, Born 1987-2004. *Archives of Disease in Childhood 91*(6), 465–468.

George, A. C. and A. Melegaro (2001). Invasive Pneumococcal Infection In England And Wales 1999. *Communicable Disease Report Weekly 11*, 4–17.

Ghaffar, F., I. R. Friedland, and G. H. McCracken Jr. (1999). Dynamics Of Nasopharyngeal Colonization By *Streptococcus Pneumoniae*. *The Pediatric Infectious Disease Journal 18*(7), 638–646.

Ginsberg, G. M., M. Fisher, I. Ben-Shahar, and J. Bornstein (2007). Cost-utility Analysis Of Vaccination Against HPV In Israel. *Vaccine 25*, 6677–6691.

Giorgi-Rossi, P., M. Merito, and P. Borgia (2009). Cost-effectiveness Of Introducing The Conjugated Pneumococcal Vaccine To Routine Free Immunizations For Infants In Lazio, Italy. *Health Policy 89*(2), 225–238.

GlaxoSmithKline (2009). Synflorix^{TM}, GlaxoSmithKline's Pneumococcal Vaccine, Receives European Authorisation. Protection Now Possible Against Three Pneumococcal Strains Not Covered By Currently Available Vaccine. `http://www.gsk.com/media/pressreleases/2009/2009_pressrelease_10039.htm#_ednref3`. Press release. Accessed 31st July 2009.

Global Forum for Health Research (2008). Interventions: Chapter 4- Modelling Cost-effectiveness Techniques And Methods. `http://www.globalforumhealth.org/filesupld/interventions/Interventions%20Chpt%204.pdf`. Accessed 1st February 2008.

Gold, M. R., J. E. Siegel, L. B. Russel, and M. C. Weinstein (Eds.) (1996). *Cost-effectiveness In Health And Medicine*. Oxford University Press, Oxford, England.

Gonzalez, B. E., K. G. Hulten, L. Lamberth, S. L. Kaplan, E. O. Mason Jr., and the U.S. Pediatric Multicenter Pneumococcal Surveillance Group

(2006). *Streptococcus Pneumoniae* Serogroups 15 And 33. An Increasing Cause Of Pneumococcal Infections In Children In The United States After The Introduction Of The Pneumococcal 7-Valent Conjugate Vaccine. *The Pediatric Infectious Disease Journal 25*(4), 301–305.

Goossens, L. M. A., B. Standaert, and N. Hartwig (2008). The Cost-utility Of Rotavirus Vaccination With Rotarix^{TM}(RIX4414) In The Netherlands. *Vaccine 26*(8), 1118–1127.

Gray, B. M., G. M. Converse 3rd, and H. C. Dillon Jr. (1980). Epidemiologic Studies Of *Streptococcus Pneumoniae* In Infants: Acquisition, Carriage, And Infection During The First 24 Months Of Life. *The Journal of Infectious Diseases 142*(6), 923–933.

Griffin, J. P. and J. O'Grady (Eds.) (2006). *The Textbook Of Pharmaceutical Medicine* (Fifth ed.). Blackwell Publishing Ltd., Oxford, UK. Part IV: Pharmacoeconomic and other issues.

Grundmann, H., S. Hori, and G. Tanner (2001). Determining Confidence Intervals When Measuring Genetic Diversity And The Discriminatory Abilities Of Typing Methods For Microorganisms. *Journal of Clinical Microbiology 39*(11), 4190–4192.

Guttmann, A., D. Manuel, P. T. Dick, T. To, K. Lam, and T. A. Stukel (2006). Volume Matters: Physician Practice Characteristics And Immunization Coverage Among Young Children Insured Through A Universal Health Plan. *Pediatrics 117*(3), 595–602.

Hamer, W. H. (1906). Epidemic Disease In England. *The Lancet 167*, 733–739.

Harboe, Z. B., R. W. Thomsen, A. Riis, P. Valentiner-Branth, J. J. Christensen, L. Lambertsen, K. A. Krogfelt, H. B. Konradsen, and T. L. Benfield (2009). Pneumococcal Serotypes And Mortality Following Invasive Pneumococcal Disease: A Population-Based Cohort Study. *Public Library of Science Medicine 6*(5).

Heffernan, H. M., D. R. Martin, R. E. Woodhouse, J. Morgan, and T. K. Blackmore (2008). Invasive Pneumococcal Disease In New Zealand 1998-2005: Capsular Serotypes And Antimicrobial Resistance. *Epidemiology and Infection 136*(3), 352–359.

Henrichsen, J. (1995). Six Newly Recognized Types Of *Streptococcus Pneumoniae*. *Journal of Clinical Microbiology 33*(10), 2759–2762.

Henriques, B., M. Kalin, Å. Örtqvist, B. O. Liljequist, M. Almela, T. J. Marrie, M. A. Mufson, A. Torres, M. A. Woodhead, S. B. Svenson, and G. Källenius (2000). Molecular Epidemiology Of *Streptococcus Pneumoniae* Causing Invasive Disease In 5 Countries. *The Journal of Infectious Diseases 182*(3), 833–839.

Hethcote, H. W. (2000). The Mathematics Of Infectious Diseases. *Society for Industrial and Applied Mathematics Review 42*(4), 599–653.

Hollingshead, S. K. and D. E. Briles (2001). *Streptococcus Pneumoniae*: New Tools For An Old Pathogen. *Current Opinions in Microbiology 4*(1), 71–77.

Hospital Episode Statistics (2007). HES Tabulation For Wyeth Europa Limited. `http://www.hesonline.nhs.uk`. Data request for hospital episodes of pneumococcal and unspecified pneumonia, septicaemia and meningitis between 1993/94 and 2005/06.

Hospital Episode Statistics (2008). HES Online. `http://www.hesonline.nhs.uk`.

Hsu, H., R. S. Lin, T. H. Tung, and T. H. Chen (2003). Cost-benefit Analysis Of Routine Childhood Vaccination Against Chickenpox In Taiwan: Decision From Different Perspectives. *Vaccine 21*, 3982–3987.

Huang, S. S., J. A. Finkelstein, and M. Lipsitch (2005a). Modeling Community- And Individual- Level Effects Of Child-Care Center Attendance On Pneumococcal Carriage. *Clinical Infectious Diseases 40*(9), 1215–1222.

Huang, S. S., R. Platt, S. L. Rifas-Shiman, S. I. Pelton, D. Goldmann, and J. A. Finkelstein (2005b). Post-PCV7 Changes In Colonizing Pneumococcal Serotypes In 16 Massachusetts Communities, 2001 And 2004. *Pediatrics 116*(3), e408–413.

Hubben, G. A., J. M. Bos, D. M. Glynn, A. van der Ende, L. van Alphen, and M. J. Postma (2007). Enhanced Decision Support For Policy Makers Using A Web Interface To Health-economic Models: Illustrated With A Cost-effectiveness Analysis Of Nationwide Infant Vaccination With The 7-Valent

Pneumococcal Conjugate Vaccine In The Netherlands. *Vaccine 25*(18), 3669–3678.

Huss, A., P. Scott, A. E. Stuck, C. Trotter, and M. Egger (2009). Efficacy Of Pneumococcal Vaccination In Adults: A Meta-Analysis. *Canadian Medical Association Journal 180*(1), 48–58.

Ihekweazu, C. A., D. A. Dance, R. Pebody, R. C. George, M. D. Smith, P. Waight, H. Christensen, K. A. Cartwright, and J. M. Stuart (2008). Trends In Incidence Of Pneumococcal Disease Before Introduction Of Conjugate Vaccine: South West England, 1996-2005. *Epidemiology and Infection 136*(8), 1096–1102.

Immunization Practices Advisory Committee (1984). Update: Pneumococcal Polysaccharide Vaccine Usage United States. *Morbidity and Mortality Weekly Report 33*(20), 273–276.

Immunization Practices Advisory Committee (2000). Preventing Pneumococcal Disease Among Infants And Young Children. *Morbidity and Mortality Weekly Report 49*(RR-9), 1–35.

Imöhl, M., R. R. Reinert, and M. van der Linden (2009). Regional Differences In Serotype Distribution, Pneumococcal Vaccine Coverage, And Antimicrobial Resistance Of Invasive Pneumococcal Disease Among German Federal States. *International Journal of Medical Microbiology*. Article in press.

Jackson, L. A., K. M. Neuzil, O. Yu, P. Benson, W. E. Barlow, A. L. Adams, C. A. Hanson, L. D. Mahoney, D. K. Shay, and W. W. Thompson (2003). Effectiveness Of Pneumococcal Polysaccharide Vaccine In Older Adults. *New England Journal of Medicine 348*(18), 1747–1755.

Jansen, A. G., G. D. Rodenburg, A. van der Ende, L. van Alphen, R. H. Veenhoven, L. Spanjaard, E. A. Sanders, and E. Hak (2009). Invasive Pneumococcal Disease Among Adults: Associations Among Serotypes, Disease Characteristics, And Outcome. *Clinical Infectious Diseases 49*(2), e23–e29.

Jefferies, J. M. C., A. Smith, S. C. Clarke, C. Dowson, and T. J. Mitchell (2004). Genetic Analysis Of Diverse Disease-Causing Pneumococci Indicates High Levels Of Diversity Within Serotypes And Capsule Switching. *Journal of Clinical Microbiology 42*(12), 5681–5688.

Jit, M., Y. H. Choi, and W. J. Edmunds (2008). Economic Evaluation Of Human Papillomavirus Vaccination In The United Kingdom. *British Medical Journal 337*(12), a769.

Johnson-Masotti, A. P., P. W. Laud, R. G. Hoffmann, M. J. Hayat, and S. D. Pinkerton (2001). Probabilistic Cost-effectiveness Analysis Of HIV Prevention: Comparing A Bayesian Approach With Traditional Deterministic Sensitivity Analysis. *Evaluation Review 25*(4), 474–502.

Kalin, M. (1998). Pneumococcal Serotypes And Their Clinical Relevance. *Thorax 53*, 159–162.

Kaplan, S. L., E. O. Mason Jr., E. Wald, T. Q. Tan, G. E. Schutze, J. S. Bradley, L. B. Givner, K. S. Kim, R. Yogev, and W. J. Barson (2002). Six Year Multicenter Surveillance Of Invasive Pneumococcal Infections In Children. *The Pediatric Infectious Disease Journal 21*(2), 141–147.

Kaplan, S. L., E. O. Mason Jr., E. R. Wald, G. E. Schutze, J. S. Bradley, T. Q. Tan, J. A. Hoffman, L. B. Givner, R. Yogev, and W. J. Barson (2004). Decrease Of Invasive Pneumococcal Infections In Children Among 8 Children's Hospitals In The United States After The Introduction Of The 7-Valent Pneumococcal Conjugate Vaccine. *Pediatrics 113*(3), 443–449.

Kermack, W. O. and A. G. McKendrick (1927). A Contribution To The Mathematical Theory Of Epidemics. *Proceedings of the Royal Society A London 115*, 700–721.

Kim, S. Y., S. J. Goldie, and J. A. Salomon (2009). Cost-effectiveness Of Rotavirus Vaccination In Vietnam. *BioMed Central Public Health 9*(29).

Kirkham, L. A., J. M. Jefferies, A. R. Kerr, Y. Jing, S. C. Clarke, A. Smith, and T. J. Mitchell (2006). Identification Of Invasive Serotype 1 Pneumococcal Isolates That Express Nonhemolytic Pneumolysin. *Journal of Clinical Microbiology 44*(1), 151–159.

Knuf, M., D. Grunert, J. Wysocki, J. Brzostek, J. C. Tejedor, J. Aristegui, P. Lommel, I. Dieussaert, and L. Schuerman (2008). Safety And Reactogenicity Of The New 10-Valent Pneumococcal Non-typeable *Haemophilus Influenzae* Protein D-Conjugate Vaccine (PHID-CV). In *ISPPD-6: 6th In-*

ternational Symposium on Pneumococci and Pneumococcal Diseases- Abstracts, Reykjavik, Iceland, June 2008, pp. 361.

Kobelt, G. (2002). *Health Economics: An Introduction To Economic Evaluation* (Second ed.). Office of Health Economics, London, England.

Konradsen, H. B. and M. S. Kaltoft (2002). Invasive Pneumococcal Infections In Denmark From 1995 To 1999: Epidemiology, Serotypes, And Resistance. *Clinical and Diagnostic Laboratory Immunology 9*(2), 358–365.

Kyaw, M. H., S. Clarke, G. F. S. Edwards, I. G. Jones, and H. Campbell (2000). Serotypes/Groups Distribution And Antimicrobial Resistance Of Invasive Pneumococcal Isolates : Implications For Vaccine Strategies. *Epidemiology and Infection 125*(3), 561–572.

Lebel, M. H., J. D. Kellner, E. L. Ford-Jones, K. Hvidsten, E. C. Wang, V. Ciuryla, S. Arikian, and R. Casciano (2003). A Pharmacoeconomic Evaluation Of 7-Valent Pneumococcal Conjugate Vaccine In Canada. *Clinical Infectious Diseases 36*(3), 259–268.

Leiberman, A., R. Dagan, E. Leibovitz, P. Yagupsky, and D. M. Fliss (1999). The Bacteriology Of The Nasopharynx In Childhood. *International Journal of Pediatric Otorhinolaryngology 49*, S151–S153.

Lexau, C. A., R. Lynfield, R. Danila, T. Pilishvili, R. Facklam, M. M. Farley, L. H. Harrison, W. Schaffner, A. Reingold, N. M. Bennett, J. Hadler, P. R. Cieslak, and C. G. Whitney (2005). Changing Epidemiology Of Invasive Pneumococcal Disease Among Older Adults In The Era Of Pediatric Pneumococcal Conjugate Vaccine. *Journal of the American Medical Association 294*(16), 2043–2051.

Lieu, T. A., S. B. Black, P. Ray, M. Chellino, H. R. Shinefield, and N. E. Adler (1994). Risk Factors For Delayed Immunization Among Children In An HMO. *American Journal of Public Health 84*(10), 1621–1625.

Lieu, T. A., G. T. Ray, S. B. Black, J. C. Butler, J. O. Klein, R. F. Breiman, M. A. Miller, and H. R. Shinefield (2000). Projected Cost-effectiveness Of Pneumococcal Conjugate Vaccination Of Healthy Infants And Young Children. *Journal of the American Medical Association 283*(11), 1460–1468.

Lindberg, A. A. (1999). Glycoprotein Conjugate Vaccines. *Vaccine 17*, S28–S36.

Lipsitch, M. (1997). Vaccination Against Colonizing Bacteria With Multiple Serotypes. *Proceedings of the National Academy of Sciences of the United States of America 94*(12), 6571–6576.

Lipsitch, M. (1999). Bacterial Vaccines And Serotype Replacement: Lessons From *Haemophilus Influenzae* And Prospects For *Streptococcus Pneumoniae*. *Emerging Infectious Diseases 5*(3), 336–345.

Lipsitch, M. (2001). Measuring And Interpreting Associations Between Antibiotic Use And Penicillin Resistance In *Streptococcus Pneumoniae*. *Clinical Infectious Diseases 32*(7), 1044–1054.

Lipsitch, M., C. Colijn, T. Cohen, W. P. Hanage, and C. Fraser (2009). No Coexistence For Free: Neutral Null Models For Multistrain Pathogens. *Epidemics 1*(1), 2–13.

Littlejohns, P. (2001). The Establishment Of The National Institute For Clinical Excellence: Its Importance And Implications For The Pharmaceutical Industry.

Lloyd, A., N. Patel, D. A. Scott, C. Runge, C. Claes, and M. Rose (2008). Cost-effectiveness Of Heptavalent Conjugate Pneumococcal Vaccine (Prevenar) In Germany: Considering A High-Risk Population And Herd Immunity Effects. *European Journal of Health Economics 9*(1), 7–15.

Lloyd-Evans, N., T. J. O'Dempsey, I. Baldeh, O. Secka, J. E. Todd, T. F. McArdle, W. S. Banya, and B. M. Greenwood (1996). Nasopharyngeal Carriage Of Pneumococci In Gambian Children And In Their Families. *The Pediatric Infectious Disease Journal 15*(10), 866–871.

Marchetti, A., L. Wang, R. Magar, H. B. Grossman, D. L. Lamm, P. F. Schellhammer, and P. Erwin-Toth (2000). Management Of Patients With Bacilli Calmette-Guérin-Refractory Carcinoma In Situ Of The Urinary Bladder: Cost Implications Of A Clinical Trial For Valrubicin. *Clinical Therapeutics 22*(4), 422–438.

Marchetti, M. and G. L. Colombo (2005). Cost-effectiveness Of Universal Pneumococcal Vaccination For Infants In Italy. *Vaccine 23*(37), 4565–4576.

Martens, P., S. W. Worm, B. Lundgren, H. B. Konradsen, and T. Benfield (2004). Serotype-Specific Mortality From Invasive *Streptococcus Pneumoniae* Disease Revisited. *BioMed Central Infectious Diseases 4*(21).

Mbelle, N., R. Huebner, A. Wasas, A. Kimura, I. Chang, and K. Klugman (1999). Immunogenicity And Impact On Nasopharyngeal Carriage Of A Nonavalent Pneumococcal Conjugate Vaccine. *The Journal of Infectious Diseases 180*(4), 1171–1176.

McChlery, S. M., K. J. Scott, and S. C. Clarke (2005). Clonal Analysis Of Invasive Pneumococcal Isolates In Scotland And Coverage Of Serotypes By The Licensed Conjugate Polysaccharide Pneumococcal Vaccine: Possible Implications For UK Vaccine Policy. *European Journal of Clinical Microbiology and Infectious Diseases 24*(4), 262–267.

McCormick, A. W., C. G. Whitney, M. M. Farley, R. Lynfield, L. H. Harrison, N. M. Bennett, W. Schaffner, A. Reingold, J. Hadler, P. Cieslak, M. H. Samore, and M. Lipsitch (2003). Geographic Diversity And Temporal Trends Of Antimicrobial Resistance In *Streptococcus Pneumoniae* In The United States. *Nature Medicine 9*(4), 424–430.

McIntosh, E. D. G. (2004). Cost-effectiveness Studies Of Pneumococcal Conjugate Vaccines. *Expert Review Vaccines 3*(4), 433–442.

McIntosh, E. D. G., P. Conway, J. Willingham, R. Hollingsworth, and A. Lloyd (2005). Pneumococcal Pneumonia In The UK- How Herd Immunity Affects The Cost-effectiveness Of 7-Valent Pneumococcal Conjugate Vaccine (PCV). *Vaccine 23*(14), 1739–1745.

McIntosh, E. D. G., P. Conway, J. Willingham, and A. Lloyd (2003). The Cost-burden Of Paediatric Pneumococcal Disease In The UK And The Potential Cost-effectiveness Of Prevention Using 7-Valent Pneumococcal Conjugate Vaccine. *Vaccine 21*, 2564–2572.

McKenzie, H., N. Reid, and R. S. Dijkhuizen (2000). Clinical And Microbiological Epidemiology Of *Streptococcus Pneumoniae* Bacteraemia. *Journal of Medical Microbiology 49*(4), 361–366.

Melegaro, A. and W. J. Edmunds (2004a). Cost-effectiveness Analysis Of Pneumococcal Conjugate Vaccination In England And Wales. *Vaccine 22*, 4203–

4214.

Melegaro, A. and W. J. Edmunds (2004b). The 23-Valent Pneumococcal Polysaccharide Vaccine. Part I. Efficacy Of PPV In The Elderly: A Comparison Of Meta-Analyses. *European Journal of Epidemiology 19*(4), 353–363.

Melegaro, A., N. J. Gay, and G. F. Medley (2004). Estimating The Transmission Parameters Of Pneumococcal Carriage In Households. *Epidemiology and Infection 132*(3), 433–441.

Mills, A. (1997). Leopard Or Chameleon? The Changing Character Of International Health Economics. *Tropical Medicine and International Health 2*(10), 963–977.

Mollison, D. (Ed.) (1995). *Epidemic Models: Their Structure And Relation To Data*. Cambridge University Press, Cambridge, England.

Mooney, J. D., A. Weir, J. McMenamin, L. D. Ritchie, T. V. Macfarlane, C. R. Simpson, S. Ahmed, C. Robertson, and S. C. Clarke (2008). The Impact And Effectiveness Of Pneumococcal Vaccination In Scotland For Those Aged 65 And Over During Winter 2003/2004. *BMC Infectious Diseases 8*(53).

Munõz-Almagro, C., I. Jordan, A. Gene, C. Latorre, J. J. Garcia-Garcia, and R. Pallares (2008). Emergence Of Invasive Pneumococcal Disease Caused By Nonvaccine Serotypes In The Era Of 7-Valent Conjugate Vaccine. *Clinical Infectious Diseases 46*(2), 174–182.

Musher, D. M. (2006). Pneumococcal Vaccine- Direct And Indirect ("Herd") Effects. *The New England Journal of Medicine 354*(14), 1522–1524.

Navas, E., L. Salleras, R. Gisbert, A. Domínguez, E. Timoner, D. Ibáñez, and A. Prat (2005). Cost-benefit And Cost-effectiveness Of The Incorporation Of The Pneumococcal 7-Valent Conjugated Vaccine In The Routine Vaccination Schedule Of Catalonia (Spain). *Vaccine 23*, 2342–2348.

NHS Immunisation (2008). Routine Immunisation Programme In 2008. `http://www.immunisation.nhs.uk/Immunisation_Schedule`. Accessed 22nd January 2009.

NHS NICE (2009). Measuring Effectiveness and Cost Effectiveness: the QALY. `http://www.nice.org.uk/newsroom/features/`

measuringeffectivenessandcosteffectivenesstheqaly.jsp. Accessed 14th August 2009.

NHS Scotland (2003). Pneumococcal Vaccine For People Aged 65 And Over. http://www.scotland.gov.uk/Publications/2003/09/18142/26154. Accessed 29th July 2009.

NHS Scotland (2009). NHS Scotland. http://en.wikipedia.org/wiki/NHS_Scotland. Accessed 15th May 2009.

NICE (2005). A Guide To NICE. http://www.nice.org.uk/aboutnice/guidetonice/a_guide_to_nice.jsp. Accessed 19th December 2007.

Noakes, K., R. G. Pebody, U. Gungabissoon, J. Stowe, and E. Miller (2006). Pneumococcal Polysaccharide Vaccine Uptake In England, 1989-2003, Prior To The Introduction Of A Vaccination Programme For Older Adults. *Journal of Public Health 28*(3), 242–247.

Nold, A. (1980). Heterogeneity In Disease-Transmission Modeling. *Mathematical Biosciences 52*, 227–240.

Normark, B. H., M. Kalin, Å. Örtqvist, T. Åkerlund, B. O. Liljequist, J. Hedlund, S. B. Svenson, J. Zhou, B. G. Spratt, S. Normark, and G. Källenius (2001). Dynamics Of Penicillin-Susceptible Clones In Invasive Pneumococcal Disease. *The Journal of Infectious Diseases 184*(7), 861–869.

Nuorti, J. P., S. W. Martin, P. J. Smith, J. S. Moran, and B. Schwartz (2008). Uptake Of Pneumococcal Conjugate Vaccine Among Children In The 1998-2002 United States Birth Cohorts. *American Journal of Preventive Medicine 34*(1), 46–53.

Obaro, S. and R. Adegbola (2002). The Pneumococcus: Carriage, Disease And Conjugate Vaccines. *Journal of Medical Microbiology 51*(2), 98–104.

Office for National Statistics (2008). Population Estimates. http://www.statistics.gov.uk/CCI/nugget.asp?ID=6. Accessed on 20th February 2008.

O'Hagan, A. and B. R. Luce (2003). A Primer On Bayesian Statistics In Health Economics And Outcomes Research. http://www.shef.ac.uk/content/1/c6/02/55/92/primer.pdf. Accessed 18th February 2008.

Öktem, H. (2005). A Survey On Piecewise-Linear Models Of Regulatory Dynamical Systems. *Nonlinear Analysis 63*, 336–349.

Østergaard, C., C. Brandt, H. B. Konradsen, and S. Samuelsson (2004). Differences In Survival, Brain Damage, And Cerebrospinal Fluid Cytokine Kinetics Due To Meningitis Caused By 3 Different *Streptococcus Pneumoniae* Serotypes: Evaluation In Humans And In 2 Experimental Models. *The Journal of Infectious Diseases 190*(7), 1212–1220.

Ozcirpici, B., S. Sahinoz, S. Ozgur, A. I. Bozkurt, T. Sahinoz, A. Ceylan, E. Ilcin, G. Saka, H. Acemoglu, Y. Palanci, M. Ak, and F. Akkafa (2006). Vaccination Coverage In The South-East Anatolian Project (SEAP) Region And Factors Influencing Low Coverage. *Public Health 120*(2), 145–154.

Palmer, A. J. (2005). Health Economics- What The Nephrologist Should Know. *Nephrology Dialysis Transplantation 20*(6), 1038–1041.

Park, I. H., D. G. Pritchard, R. Cartee, A. Brandao, M. C. C. Brandileone, and M. H. Nahm (2007). Discovery Of A New Capsular Serotype (6C) Within Serogroup 6 Of *Streptococcus Pneumoniae*. *Journal of Clinical Microbiology 45*(4), 1225–1233.

Pasteur, L. (1881). Note Sur La Maladie Nouvelle Provoquée Par La Salive D'Un Enfant Mort De La Rage. *Comptes Rendus 92*, 159–165.

Pelton, S. I., H. Huot, J. A. Finkelstein, C. J. Bishop, K. K. Hsu, J. Kellenberg, S. S. Huang, R. Goldstein, and W. P. Hanage (2007). Emergence Of 19A As Virulent And Multidrug Resistant Pneumococcus In Massachusetts Following Universal Immunization Of Infants With Pneumococcal Conjugate Vaccine. *The Pediatric Infectious Disease Journal 26*(6), 468–472.

Prosser, L. A., G. T. Ray, M. O'Brien, K. Kleinman, J. Santoli, and T. A. Lieu (2004). Preferences And Willingness To Pay For Health States Prevented By Pneumococcal Conjugate Vaccine. *Pediatrics 113*(2), 283–290.

Raftery, J. (2000). Economics Notes: Costing In Economic Evaluation. *British Medical Journal 320*, 1597.

Rasbash, J., F. Steele, W. Browne, and H. Goldstein (2008). *A User's Guide To MLwiN Version 2.10*. Centre for Multilevel Modelling, University of Bristol, Bristol, England.

Ray, G. T., C. G. Whitney, B. H. Fireman, V. Ciuryla, and S. B. Black (2006). Cost-effectiveness Of Pneumococcal Conjugate Vaccine: Evidence From The First 5 Years Of Use In The United States Incorporating Herd Effects. *The Pediatric Infectious Disease Journal 25*(6), 494–501.

Reading, R., H. Surridge, and R. Adamson (2004). Infant Immunization And Family Size. *Journal of Public Health 26*(4), 369–371.

Riley, I. D., F. A. Everingham, D. E. Smith, and R. M. Douglas (1981). Immunisation With A Polyvalent Pneumococcal Vaccine. Effect On Respiratory Mortality In Children Living In The New Guinea Highlands. *Archives of Disease in Childhood 56*(5), 354–357.

Ripley, B. D. and W. N. Venables (2002). *Modern Applied Statistics With S*. Springer, New York, USA.

Robinson, W. S. (1950). Ecological Correlations And The Behaviour Of Individuals. *American Sociological Review 15*, 351–357.

Rodewald, L., E. Maes, J. Stevenson, B. Lyons, S. Stokley, and P. Szilagyi (1999). Immunization Performance Measurement In A Changing Immunization Environment. *Pediatrics 103*(4), 889–897.

Ross, R. (1916). An Application Of The Theories Of Probabilities To The Study Of A Priori Pathometry- Part 1. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 92*, 204–230.

Rückinger, S., R. von Kries, A. Siedler, and M. van der Linden (2009). Association Of Serotype Of *Streptococcus Pneumoniae* With Risk Of Severe And Fatal Outcome. *The Pediatric Infectious Disease Journal 28*(2), 118–122.

Ruedin, H. J., S. Essb, H. P. Zimmermanna, and T. Szucs (2003). Invasive Meningococcal And Pneumococcal Disease In Switzerland: Cost-utility Analysis Of Different Vaccine Strategies. *Vaccine 21*, 4145–4152.

Saha, S. K., A. Naheed, S. E. Arifeen, M. Islam, H. Al-Emran, R. Amin, K. Fatima, W. A. Brooks, R. F. Breiman, D. A. Sack, S. P. Luby, and the Pneumococcal Study Group (2009). Surveillance For Invasive *Streptococcus Pneumoniae* Disease Among Hospitalized Children In Bangladesh:

Antimicrobial Susceptibility And Serotype Distribution. *Clinical Infectious Diseases 48*, S75–S81.

Salisbury, D., M. Ramsay, and K. Noakes (Eds.) (2006). *Immunisation Against Infectious Disease - "The Green Book"*. Department of Health, The Stationery Office, Norwich, England.

Salo, H., H. Sintonen, J. P. Nuorti, M. Linna, H. Nohynek, J. Verho, and T. Kilpi (2005). Economic Evaluation Of Pneumococcal Conjugate Vaccination In Finland. *Scandinavian Journal of Infectious Diseases 37*, 821–832.

Schutze, G. E., N. C. Tucker, and E. O. Mason Jr. (2004). Impact Of The Conjugate Pneumococcal Vaccine In Arkansas. *The Pediatric Infectious Disease Journal 23*(12), 1125–1129.

Sculpher, M. (2004). The Use Of A Probabilistic Sensitivity Analysis For Decision Making: The Example Of Drug-Eluting Stents. `http://www.york.ac.uk/inst/che/pdf/teehtacosteff04.pdf`. Slides from a seminar presented at the Harvard Clinical Research Institute. Accessed 28th July 2008.

Serrano, I., J. Melo-Cristino, J. A. Carriço, and M. Ramirez (2005). Characterization Of The Genetic Lineages Responsible For Pneumococcal Invasive Disease In Portugal. *Journal of Clinical Microbiology 43*(4), 1706–1715.

Silfverdal, S. A., S. Berg, C. Hemlin, and I. Jokinen (2009). The Cost-burden Of Paediatric Pneumococcal Disease In Sweden And The Potential Cost-effectiveness Of Prevention Using 7-Valent Pneumococcal Vaccine. *Vaccine 27*(10), 1601–1608.

Simon, S. (2000). STATS: Steve's Attempt To Teach Statistics- Fisher's Exact Test. `http://www.childrensmercy.org/stats/ask/fishers.asp`. Accessed 24th September 2008.

Simpson, E. H. (1949). Measurement Of Diversity. *Nature 163*, 688.

Singleton, R. J., T. W. Hennessy, L. R. Bulkow, L. L. Hammitt, T. Zulz, D. A. Hurlburt, J. C. Butler, K. Rudolph, and A. Parkinson (2007). Invasive Pneumococcal Disease Caused By Nonvaccine Serotypes Among Alaska Native Children With High Levels of 7-Valent Pneumococcal Conjugate Vaccine Coverage. *Journal of the American Medical Association 297*(16), 1784–1792.

Sinha, A., O. Levine, M. D. Knoll, F. Muhib, and T. A. Lieu (2007). Cost-effectiveness Of Pneumococcal Conjugate Vaccination In The Prevention Of Child Mortality: An International Economic Analysis. *The Lancet 369*(9559), 389–396.

Sjöström, K., C. Spindler, A. Ortqvist, M. Kalin, A. Sandgren, S. Kühlmann-Berenzon, and B. Henriques-Normark (2006). Clonal And Capsular Types Decide Whether Pneumococci Will Act As A Primary Or Opportunistic Pathogen. *Clinical Infectious Diseases 42*(4), 451–459.

Sloan, F. A. (1996). *Valuing Health Care: Costs, Benefits And Effectiveness Of Pharmaceuticals And Other Medical Technologies*. Cambridge University Press, Cambridge, England.

Smith, P. J., J. P. Nuorti, J. A. Singleton, Z. Zhao, and K. M. Walter (2007). Effect Of Vaccine Shortages On Timeliness Of Pneumococcal Conjugate Vaccination: Results From The 2001–2005 National Immunization Survey. *Pediatrics 120*(5), e1165–e1173.

Snijders, T. A. B. and R. J. Bosker (1999). *Multilevel Analysis: An Introduction To Basic And Advanced Multilevel Modeling*. SAGE Publications, London, England.

Spiegelhalter, D. J., K. R. Abrams, and J. P. Myles (2004). *Bayesian Approaches To Clinical Trials And Health-Care Evaluation*. Wiley, West Sussex, England.

Spratt, B. G. and B. M. Greenwood (2000). Prevention Of Pneumococcal Disease By Vaccination: Does Serotype Replacement Matter? *The Lancet 356*(9237), 1210–1211.

Spratt, B. G., W. P. Hanage, and A. B. Brueggemann (2004). *Evolutionary And Population Biology Of Streptococcus Pneumoniae*. ASM Press, Washington DC, USA. In: The Pneumococcus, Edited by E.I. Tuomanen, T.J. Mitchell, D.A. Morrison and B.G. Spratt.

Sternberg, G. M. (1881). A Fatal Form Of Septicaemia In The Rabbit, Produced By Subcutaneous Injection Of Human Saliva. An Experimental Research. *National Board of Health Bulletin 2*, 781–783.

Steyer, T. E., A. G. Mainous III, and M. E. Geesey (2005). The Effect Of Race And Residence On The Receipt Of Childhood Immunizations: 1993-2001. *Vaccine 23*(12), 1464–1470.

Stokley, S., K. M. Shaw, L. Barker, J. M. Santoli, and A. Shefer (2006). Impact Of State Vaccine Financing Policy On Uptake Of Heptavalent Pneumococcal Conjugate Vaccine. *American Journal of Public Health 96*(7), 1308–1313.

Sutton, K. L., H. T. Banks, and C. Castillo-Chávez (2008). Estimation Of Invasive Pneumococcal Disease Dynamics Parameters And The Impact Of Conjugate Vaccination In Australia. *Mathematical Biosciences and Engineering 5*(1), 175–204.

Szucs, T. (2000). Cost-effectiveness Of Hepatitis A And B Vaccination Programme In Germany. *Vaccine 18*, S86–S89.

Temime, L., P. Boëlle, P. Courvalin, and D. Guillemot (2003). Bacterial Resistance To Penicillin G By Decreased Affinity Of Penicillin-Binding Proteins: A Mathematical Model. *Emerging Infectious Diseases 9*(4), 411–417.

Temime, L., P. Y. Boëlle, L. Opatowski, and D. Guillemot (2008). Impact Of Capsular Switch On Invasive Pneumococcal Disease Incidence In A Vaccinated Population. *Public Library of Science ONE 3*(9), e3244.

Temime, L., P. Y. Boëlle, A. J. Valleron, and D. Guillemot (2005). Penicillin-Resistant Pneumococcal Meningitis: High Antibiotic Exposure Impedes New Vaccine Protection. *Epidemiology and Infection 133*(3), 493–501.

Temime, L., D. Guillemot, and P. Y. Boëlle (2004). Short- And Long-Term Effects Of Pneumococcal Conjugate Vaccination Of Children On Penicillin Resistance. *Antimicrobial Agents and Chemotherapy 48*(6), 2206–2213.

The Scottish Government (2005). Health Services In Argyll And Clyde. `http://www.scotland.gov.uk/News/Releases/2005/12/08084136`. Accessed 15$^{\text{th}}$ May 2009.

The Scottish Government (2009). Scottish Index Of Multiple Deprivation. `http://www.scotland.gov.uk/Topics/Statistics/SIMD`. Accessed 12$^{\text{th}}$ June 2009.

Tilson, L., C. Usher, K. Butler, J. Fitzsimons, F. O'Hare, S. Cotter, D. O'Flanagan, H. Johnson, and M. Barry (2008). Economic Evaluation Of A Universal Childhood Pneumococcal Conjugate Vaccination Strategy In Ireland. *Value in Health 11*(5), 898–903.

Torun, S. D. and N. Bakırcı (2006). Vaccination Coverage And Reasons For Non-Vaccination In A District Of Istanbul. *BioMedCentral Public Health 6*(125).

Towse, A., C. Pritchard, and N. Devlin (Eds.) (2002). *Cost-effectiveness Thresholds: Economic And Ethical Issues*. King's Fund: Office of Health Economics, London, England.

TreeAge Software Inc. (2008). Markov Models - Markov Analysis - State Transition Models. `http://www.treeage.com/learnMore/MarkovModels.html`. Accessed 28th July 2008.

Trotter, C. L., J. M. Stuart, R. George, and E. Miller (2008). Increasing Hospital Admissions For Pneumonia, England. *Emerging Infectious Diseases 14*(5), 727–733.

U.S. Department of Health & Human Services Food & Drug Administration (2008). Guidance For The Use Of Bayesian Statistics In Medical Device Clinical Trials. `http://www.fda.gov/cdrh/osb/guidance/1601.html#4`. Accessed 18th February 2008.

Vogt, A. and J. G. Bared (1998). Accident Models For Two Lane Rural Roads: Segments And Intersections; Chapter 5: Modeling. `http://www.tfhrc.gov/safety/98133/ch05/ch05_01.html`. Accessed 22nd May 2008.

Wang, Y. C. and M. Lipsitch (2006). Upgrading Antibiotic Use Within A Class: Tradeoff Between Resistance And Treatment Success. *Proceedings of the National Academy of Sciences of the United States of America 103*(25), 9655–9660.

Weisstein, E. W. (1999). Fisher's Exact Test From MathWorld- A Wolfram Web Resource. `http://mathworld.wolfram.com/FishersExactTest.html`. Accessed 24th September 2008.

Whitney, C. G., M. M. Farley, J. Hadler, L. H. Harrison, N. M. Bennett, R. Lynfield, A. Reingold, P. R. Cieslak, T. Pilishvili, D. Jackson, R. R.

Facklam, J. H. Jorgensen, and A. Schuchat (2003). Decline In Invasive Pneumococcal Disease After The Introduction Of Protein-Polysaccharide Conjugate Vaccine. *The New England Journal of Medicine 348*(18), 1737–1746.

Whitney, C. G., T. Pilishvili, M. M. Farley, W. Schaffner, A. S. Craig, R. Lynfield, A. Nyquist, K. A. Gershman, M. Vazquez, N. M. Bennett, A. Reingold, A. Thomas, M. P. Glode, E. R. Zell, J. H. Jorgensen, B. Beall, and A. Schuchat (2006). Effectiveness Of Seven-Valent Pneumococcal Conjugate Vaccine Against Invasive Pneumococcal Disease: A Matched Case-Control Study. *The Lancet 368*(9546), 1495–1502.

Williams, E. J., S. Thorson, M. Maskey, S. Mahat, M. Hamaluba, S. Dongol, A. M. Werno, B. K. Yadav, A. S. Shah, D. F. Kelly, N. Adhikari, A. J. Pollard, and D. R. Murdoch (2009). Hospital-Based Surveillance Of Invasive Pneumococcal Disease Among Young Children In Urban Nepal. *Clinical Infectious Diseases 48*, S114–S122.

Winstead, E. R. (2001). Genome News Network: *Streptococcus Pneumoniae* Strain R6 Is Sequenced. `http://www.genomenewsnetwork.org/articles/10_01/strep_R6_strain.shtml`. Accessed 26th May 2009.

Wisløff, T., T. G. Abrahamsen, M. A. R. Bergsaker, Ø. Løvoll, P. Møller, M. K. Pedersen, and I. S. Kristiansen (2006). Cost Effectiveness Of Adding 7-Valent Pneumococcal Conjugate (PCV-7) Vaccine To The Norwegian Childhood Vaccination Program. *Vaccine 24*, 5690–5699.

World Health Organization (2007). Pneumococcal Conjugate Vaccine For Childhood Immunization WHO Position Paper. *Weekly Epidemiological Record 82*(12), 93–104.

Wyeth Pharmaceuticals (2009a). Israel Adds Wyeth's 7-Valent Pneumococcal Conjugate Vaccine (PCV7) To Their National Immunisation Programme. `http://www.presseportal.de/pm/56060/1439744/wyeth_pharmaceuticals`. Press report. Accessed 31st July 2009.

Wyeth Pharmaceuticals (2009b). Wyeth's Prevenar 13™ Receives First Approval. `http://www.presseportal.de/pm/56060/1440178/wyeth_pharmaceuticals`. Press report. Accessed 31st July 2009.

Wyss, G. D. and K. H. Jorgensen (1998). A User's Guide To LHS: Sandia's Latin Hypercube Sampling Software. http://www.prod.sandia.gov/cgi-bin/techlib/access-control.pl/1998/980210.pdf. Accessed 8th September 2008.

Zimmerman, R. K., T. A. Mieczkowski, and M. Michel (1999). Are Vaccination Rates Higher If Providers Receive Free Vaccines And Follow Contraindication Guidelines? *Family Medicine 31*(5), 317–323.