Respiratory Syncytial Virus in Young Children in England: Burden and Risk Factors for Severe Disease

A thesis presented for the degree of

Doctor of Philosophy

University College London

Rachel Melanie Reeves

2017

I, Rachel Melanie Reeves, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Respiratory syncytial virus (RSV) is a major cause of respiratory infection in infants and young children. With potential vaccines on the horizon it is essential that RSV burden is accurately calculated by age and risk group to identify optimal target populations. My PhD investigates the secondary care burden of RSV in children younger than five years in England using routinely collected national laboratory surveillance and hospital admissions data.

First, I explore the use of the individual datasets to describe RSV epidemiology. Second, I use ecological time series modelling to estimate the number of weekly hospital admissions attributable to RSV. Third, I use probabilistically linked laboratory and hospital data to describe laboratory-confirmed RSV-associated hospital admissions in England for the first time, and to determine risk factors for severe RSV-associated disease. Finally, from the linked data I generate a predictive model for RSV-associated admissions in infants, and use this to estimate the national burden of RSV-associated admissions by patient and clinical characteristics.

I estimate an annual average of 33,500 (95% CI: 30,400-38,500) RSV-associated admissions in children younger than five years in England. 82% (95% CI: 79-87%) of admissions for bronchiolitis in children younger than six months could be attributed to RSV. My results highlight the importance of young age (<3 months) and birth around the beginning of RSV season in the risk of RSV-associated admission, and the importance of young age, prematurity and comorbidities in the increasing severity of disease.

This is the first study to use linked laboratory surveillance and hospital admissions data for RSV in England. I have produced detailed, recent estimates of RSV-associated admissions in infants and young children using multiple methods, and highlight the strengths and limitation of using routinely collected data for RSV research. My results provide essential baseline epidemiological data required for vaccine impact studies.

Acknowledgements

I would firstly like to thank my supervisors – Dr Richard Pebody, Dr Pia Hardelid and Professor Ruth Gilbert – for their invaluable support, guidance and encouragement during my PhD. I feel privileged to have worked with each of my supervisors and to have learned from their expertise – I have no doubt that I am a better researcher for it.

I would also like to thank colleagues at Public Health England for their support. I owe thanks to Fiona Warburton, Nikolaos Panagiotopoulos, and Mehdi Minaji for their vital contributions and valuable advice at various stages of my project. Joanna Ellis, Hongxin Zhao and Matthew Donati have also provided much-appreciated input as co-authors on published papers of my analysis, and I extend my sincere thanks to the Farr Institute of Health Informatics for providing funding for this PhD.

Completing this PhD would have been much more difficult and much less enjoyable were it not for my fellow PhD students. These wonderful friends have provided advice, encouragement and laughter, and deserve thanks for helping me keep things in perspective. I must also thank Professor Mark Pearce for being the first to encourage me to undertake a PhD.

Finally, without three people this thesis would not have been possible. Dan, you have been there through all the ups and downs of this PhD experience and I am continually amazed by your patience and support. Mum and Dad, your incredible and unwavering support and encouragement has allowed me to achieve beyond my dreams and ambitions – thank you for everything.

This project has been supported by awards to establish the Farr Institute of Health Informatics Research at UCLPartners, from the Medical Research Council, Arthritis Research UK, British Heart Foundation, Cancer Research UK, Chief Scientist Office, Economic and Social Research Council, Engineering and Physical Sciences Research Council, National Institute for Health Research, National Institute for Social Care and Health Research, and Wellcome Trust (grant MR/K006584/1).

Contents

Figures

Tables

Appendices

Abbreviations

A&E	Accident and emergency
CHD	Congenital heart disease
CI	Confidence interval
CIDSC	Centre for Infectious Disease Surveillance
CLD	Chronic lung disease
COPD	Chronic obstructive pulmonary disease
CPRD	Clinical Practice Research Datalink
EM	Expectation-maximization
FCE	Finished consultant episodes
GP	General practitioner
GSK	GlaxoSmithKline
HES	Hospital Episode Statistics
hMPV	Human metapneumovirus
ICD-10	International Statistical Classification of Diseases and Related Health Problems 10th Revision
IQR	Interquartile range
JCVI	Joint Committee on Vaccination and Immunisation
LRTI	Lower respiratory tract infection
LTV	Long-term ventilated
NHS	National Health Service

NPA	Nasopharyngeal aspirate
NPV	Negative predictive value
ONS	Office for National Statistics
OPIE	Organism-Patient-Illness-Episode
OR	Odds ratio
PbR	Payment by Results
PCR	Polymerase chain reaction
PHE	Public Health England
PICANet	Paediatric Intensive Care Audit Network
PICU	Paediatric intensive care unit
PII	Patient identifiable information
PPV	Positive predictive value
QALY	Quality-adjusted life-years
RCGP	Royal College of General Practitioners
RDS	Respiratory DataMart System
RESCEU	Respiratory Syncytial Virus Consortium in Europe
ReSViNET	Respiratory Syncytial Virus Network
ROC	Receiver operator curve
RR	Relative risk
RSV	Respiratory syncytial virus
RTI	Respiratory tract infection
RT-PCR	Reverse transcriptase polymerase chain reaction

SCID	Severe combined immunodeficiency syndrome
SE	Standard error
SEIR	Susceptible-Exposed-Infectious-Removed
SGSS	Second Generation Surveillance System
SUS	Secondary Uses Service
UCLH	University College London Hospitals
URTI	Upper respiratory tract infection
WHO	World Health Organization

Chapter 1

Introduction and background

Chapter 1 Introduction and background

1.1 Introduction

In this thesis, I use routinely collected data to investigate the secondary care burden of respiratory syncytial virus (RSV) in children younger than five years of age in England. RSV is the most important cause of viral lower respiratory tract illness in young children worldwide and considered one of the world's greatest unmet vaccine needs. Advances in molecular virology, immunology and vaccinology, as well as increased knowledge of pathology and pathogenesis, suggest that an RSV vaccine will be commercially available within the next 5-10 years. As of August 2017, there are around 60 vaccine candidates in development, with 16 currently advancing through Phase 1 to Phase 3 clinical trials. However, there are crucial gaps in knowledge of RSV burden that need to be addressed before optimal target populations can be identified and vaccine programmes introduced once these vaccines are licensed. The work presented in this thesis aims to fill some of these crucial gaps, and is therefore timely and relevant.

1.2 Thesis structure and content

This chapter sets out the rationale for the research undertaken for this PhD, describing the context of this thesis within the field of RSV research. The aims of this work are outlined in Chapter 2, a description of the data sources in Chapter 3, results presented in Chapters 4 to 8, and the discussion and conclusions in Chapter 9.

1.3 An introduction to RSV

1.3.1 The virus

RSV was first isolated in 1956 from a laboratory chimpanzee (1). Soon after its isolation it was discovered to be of human origin, and shown to be the leading viral cause of serious respiratory illness in children worldwide (2). Within the past 30 years, advances in molecular biology (such as molecular cloning, reverse genetics, and studies of protein function and structure) have significantly increased our understanding of RSV (3).

RSV is a *Pneumovirus* belonging to the family Paramyxoviridae (4). This family also contains *Paramyxoviruses* (including parainfluenza and mumps virus) and *Morbilliviruses* (including the measles virus) (5). Two distinct RSV antigenic subgroups have been identified: RSV A and RSV B. There is considerable genetic diversity between and within RSV subgroups, and the distribution of genotypes can be highly distinct in different communities (6).

RSV is an enveloped virus with a genome consisting of single-stranded negative-sense RNA (4). The genome encodes 10 proteins (7). Eight of these ten proteins are structural proteins, also present in infected cells (6). RSV has three surface envelope glycoproteins: G (the attachment protein), F (the fusion protein) and SH (the small hydrophobic protein (7) (Figure 1-1). These surface proteins induce protective antibodies (6). The G and F proteins regulate the initial phases of RSV infection: the G protein targets the ciliated cells of the airways and the F protein directs viral penetration to the host plasma membrane (7). Variations in the G protein structure characterise the A and B subtypes (8). Despite the considerable genetic diversity between RSV isolates, the F protein is highly conserved (including in both A and B subtypes) and is therefore a key target for neutralising antibody and vaccine development (9).

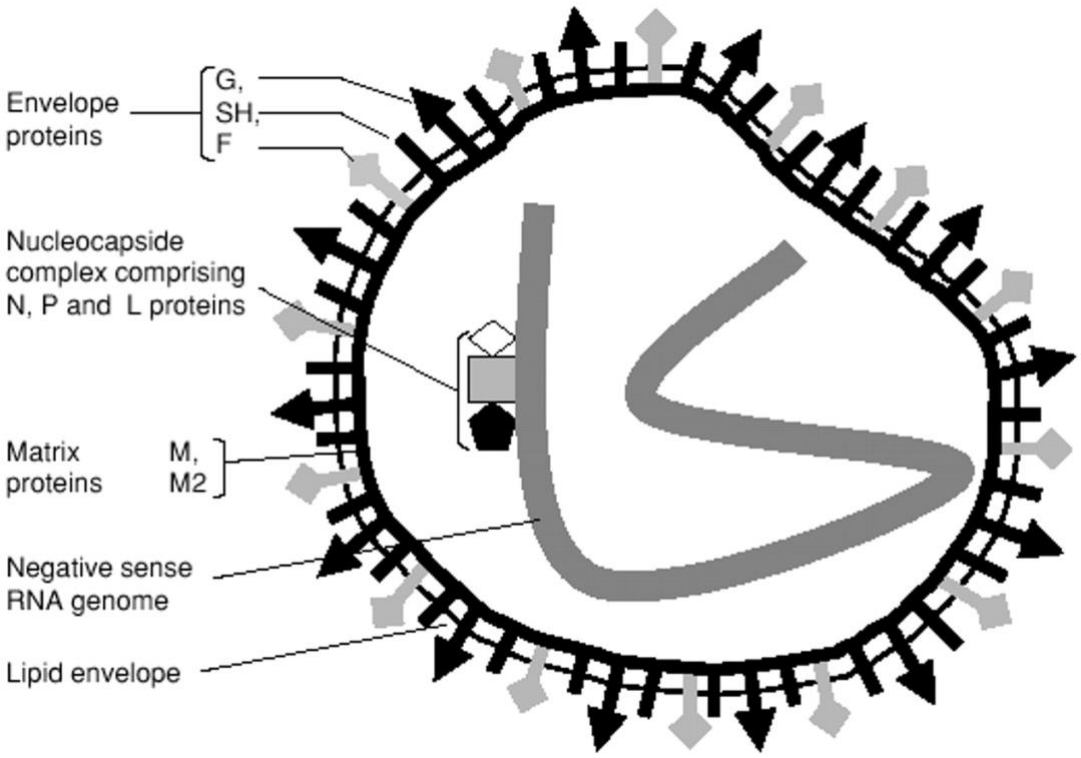

Figure 1-1. A schematic diagram of the RSV virion. From McNamara and Smyth (2002) (7). Permission to reuse granted from Oxford University Press (July 2017).

1.3.2 Epidemiology

RSV is the most important cause of viral lower respiratory tract illness in young children worldwide (10,11). Approximately 60% to 80% of children are infected by RSV during their first year of life, with almost all infected by two years of age (12,13). It has been estimated that over 33 million episodes of RSV-associated lower respiratory tract infection (LRTI) occur in children younger than 5 years worldwide each year, representing 22% of all acute LRTI episodes; at least 3 million of these represent episodes severe enough to require hospital admission (10,14). Furthermore, RSV contributes to the global burden of child mortality with approximately 66,000-199,000 RSV-associated deaths in children younger than 5 years each year, 99% of which occur in low-income countries (10). In the UK, RSV-associated mortality rates are very low in previously healthy infants (0.5-1.5%), but

significantly higher in children with comorbidities and chronic conditions (2-23%) (15,16).

RSV has the ability to infect very young infants; children younger than 6 months of age are at particularly high risk of severe RSV infection, with a peak in RSV-associated hospital admissions around 1 to 2 months of age (2,15,17). Other respiratory viruses – such as influenza, rhinovirus and human metapneumovirus – can also infect children younger than 6 months. However, RSV causes more frequent and severe disease in these young infants (2). There are several reasons for the increased severity with young age upon infection with RSV, including: narrow airways, immaturity of the immune system, and a bias towards a TH2-type immune response which is associated with higher severity of illness (T helper (TH) 2 cells of the immune system control immunity to extracellular parasites and all allergic inflammatory responses (18)) (17,19).

The majority of infants and young children hospitalised with RSV are born at term and are previously healthy (20–22). However, children born prematurely (<35 weeks gestational age) or those with pre-existing chronic conditions are at increased risk of severe RSV-associated disease (more likely to be admitted to paediatric intensive care or require respiratory support) (23). Pre-existing conditions that are associated with a significantly higher severity and risk of death from RSV infection include those with chronic lung disease (CLD), congenital heart disease (CHD), metabolism abnormalities, neuromuscular disease, liver disease, immunodeficiency, and chromosomal abnormalities (8,24–27). In one study at the Royal Liverpool Children's Hospital in England, all 35 RSV-associated deaths over eight consecutive RSV seasons were in children with pre-existing medical conditions (24).

1.3.2.1 *Seasonality of epidemics*

RSV circulation is characterised by marked seasonality in temperate climates (28). Each year, large numbers of infants with moderate-to-severe RSV-associated LRTI are admitted to paediatric wards during the annual RSV epidemics. The timing of RSV epidemics vary with climate, latitude and altitude (29,30). In temperate regions, RSV epidemics occur during late autumn, winter and early spring (28). In tropical climates north of the equator, the seasonality of RSV is associated with decreased temperature and increased rainfall (29). Southern tropical climates have RSV seasons associated with both decreased temperature and rainfall (29). In equatorial countries, RSV is detected throughout the year, with some increase during the dry months (30). Furthermore, in tropical regions multiple epidemics can occur within the same year (30). While most countries report annual cycles of RSV circulation, several studies report biennial cycles of RSV activity (e.g. in Sweden and Finland, with biennial cycles of minor and major epidemic peaks) (30).

The factors driving the occurrence of repeated RSV epidemics are not fully understood. Meteorological factors are strongly associated with epidemics, but it is not clear whether the driving factors behind these associations are related to the spread of the virus, climate-dependent behavioural factors (such as indoor crowding, which increase the intensity of exposure), or cyclical changes in immunity within populations (29). Furthermore, RSV epidemics do not appear to result from the spread in neighbouring areas; similar RSV strains occur simultaneously in different communities (29). There is no clear explanation for this simultaneous appearance of similar RSV strains, however, it has been shown that RSV can be isolated from patients with chronic obstructive pulmonary disease (COPD) (31) or those with HIV (32,33) almost year-round. Therefore, it has been hypothesised that these patients act as reservoirs for RSV infection, allowing the beginning of RSV epidemics when climatic and immunologic factors interact (29). However, conflicting evidence on this demonstrates the need for further investigation into the factors driving RSV epidemics (34).

In the UK, RSV season occurs during the winter months (during the period from October to March) (8). Whilst the size of the epidemic can vary by year, the timing of the peak consistently occurs between late December and early January (28). RSV A and B subtypes circulate simultaneously during outbreaks and, though group A is slightly more transmissible, the predominating subgroup can vary each season (28,35) . The seasonality of RSV in England and Wales (demonstrated using national laboratory surveillance data) is illustrated in Figure 1-2.

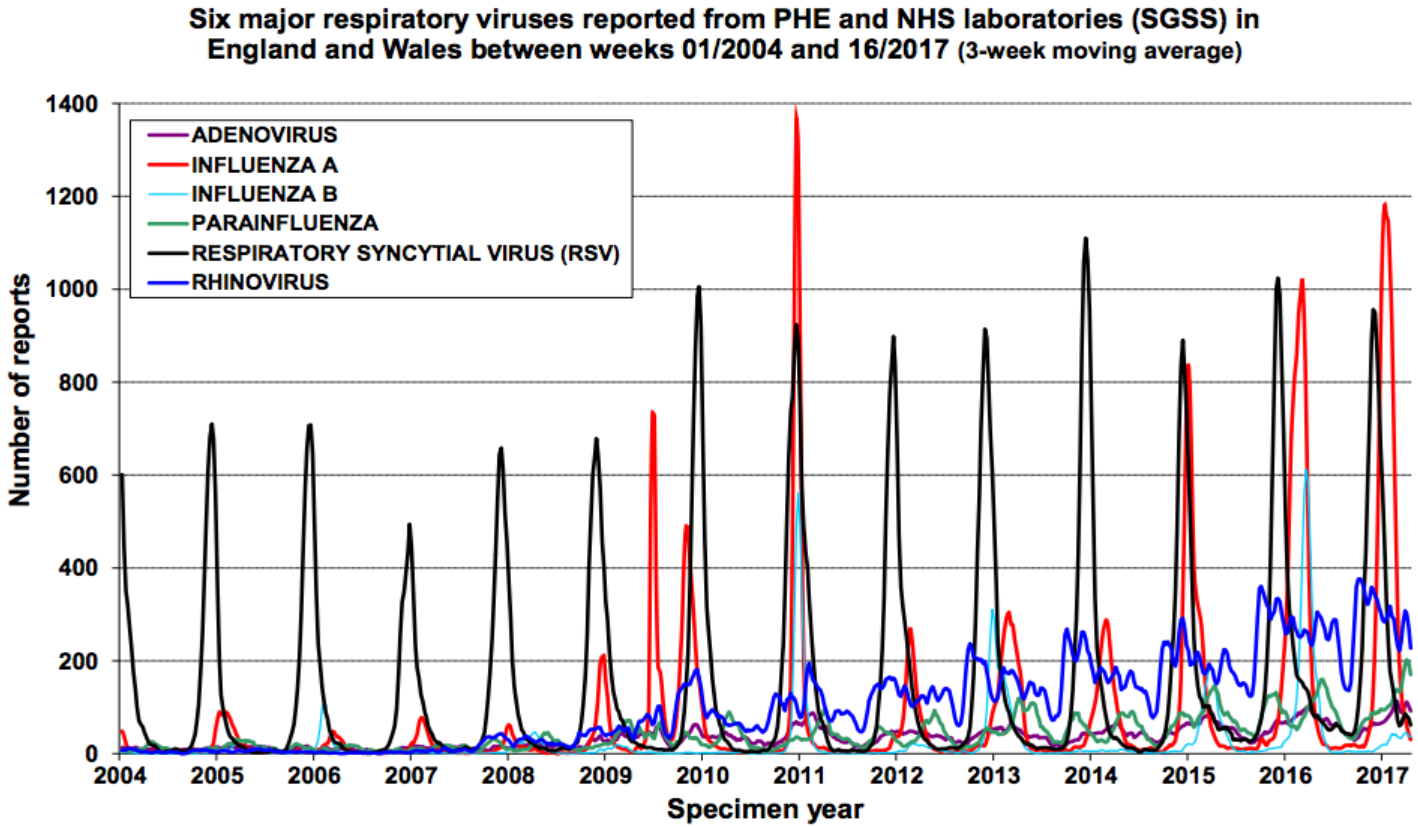

Figure 1-2. Laboratory reports of major respiratory viruses (adenovirus, influenza A and B, parainfluenza, RSV and rhinovirus) from the national laboratory surveillance system (Second Generation Surveillance System, SGSS) held by Public Health England (PHE). From Public Health England RSV guidance (2017) (36).

1.3.3 Transmission

RSV is highly contagious (2). The virus is usually transmitted by direct or close contact with contaminated respiratory secretions. The virus can remain viable for up to an hour on hands and for up to approximately seven hours on surfaces such as countertops, beds and toys (37). RSV commonly infects the ciliated epithelial cells of the respiratory mucosa, disseminating locally and replicating in the nasopharynx (3). RSV can also infect the eye, where the infection can spread to the lungs (likely via the tear ducts, into the nose, trachea and then the lung) (38). The mechanism for the development of LRTI is thought to be through aspiration of nasopharyngeal secretions, direct spread along the respiratory epithelium, and/or the infection of macrophages which migrate to the lower airways (5,39). Individuals can remain contagious for up to 8 days following symptom onset, though infants and young children may remain contagious for up to one month (40).

On average, one RSV-infected person can transmit the virus to between 5 and 25 people (41). Nosocomial infection is therefore a major risk on paediatric wards. Early identification of RSV-associated infections (through laboratory testing, see section 1.3.7) allows for cohorting or isolation of infected infants and children to minimise nosocomial transmission. Strict infection control practices can also reduce nosocomial RSV infection rates (41,42).

RSV also easily spreads in households, day-care facilities and other close-contact settings (2). In one study, the (natural) introduction of RSV into a day-care facility resulted in more than 90% of infants and children becoming infected (2). Studies conducted in rural Africa have demonstrated the role of parents and siblings in transmitting RSV to infants within the household (43,44). However, there are limited studies in alternative geographical and social contexts of the patterns of RSV spread during outbreaks (45). Therefore, many aspects of RSV transmission remain unknown.

1.3.4 Reinfection

Despite the annual occurrence of RSV outbreaks suggesting that continuous transmission is prevented by acquired immunity, reinfection with RSV is common throughout life (13). The antigenic differences between RSV A and B are thought to facilitate the persistent reinfection with RSV (13). However, even reinfection with the same strain group of RSV within the same season has been shown to be possible (though is rare) (46). Lower levels of neutralising F and G antibodies have been shown to be associated with increased risk of reinfection (47). In addition, reinfection may be facilitated by the restriction of the virus life cycle to the epithelium, where the systemic immune system has little access (3).

The severity of RSV infection decreases with increasing age (6). This decreased severity was conventionally thought to be due to a combination of an accumulation of immunity following previous exposure, increased maturity of the immune system and physiological changes such as widening of the airways (13). However, as there is only a temporary 60-70% reduction in the rate of reinfection during the six-months following RSV infection, there is now increasing evidence that physiological changes due to increasing age is the most important driving factor for the decreased severity of RSV with increasing age (13).

1.3.5 Clinical presentation

The primary RSV infection (typically before 2 years of age) is usually the most severe (48). RSV causes a wide spectrum of disease, from mild self-limiting upper respiratory tract infection (URTI) to otitis media or severe LRTI. RSV first infects the upper respiratory tract, then clinical symptoms occur after a typical incubation period of 2-8 days (peaking around day 5 of infection) (37,41,49). If the infection spreads to the lower respiratory tract, this will usually occur within a further 1 to 3 days (41). It has been estimated that 25-40% of infants and young children develop LRTI during their first RSV infection (50).

The most common type of LRTI caused by RSV is bronchiolitis. In the UK, 5% of all infants are admitted to hospital with bronchiolitis each year (51). Bronchiolitis is a clinical syndrome characterised by inflammation of the bronchioles (the smallest air passages of the lungs), though diagnostic criteria varies by country (6). In more severe bronchiolitis cases the bronchioles can become blocked, causing breathing problems. Although RSV is the most common cause of bronchiolitis, other causal agents include rhinovirus, parainfluenza, adenovirus and influenza A (6). Estimates of the percentage of bronchiolitis episodes caused by RSV vary widely, from approximately 45 to 90% (15,49,52).

RSV is also the most common single viral cause of pneumonia (53). Pneumonia is an inflammatory condition of the lung affecting the alveoli (microscopic air sacs of the lungs). Other viral causal agents include influenza, rhinovirus, human metapnuemovirus (hMPV) and parainfluenza (54,55). Pneumonia can also be caused by bacterial agents, such as *Steptococcus pneumoniae* and *Haemophilus influenzae* (56). Between 5 and 55% of pneumonia cases are estimated to be caused by RSV (6,15). Bronchiolitis and pneumonia have overlapping clinical presentations in young children, and are difficult to differentiate without a chest radiograph (40). As a result, there are discrepancies in diagnosis of bronchiolitis and pneumonia between countries, regions, and even hospitals (57).

The presentation of RSV-associated illness varies with age (6). An estimated one-fifth of RSV-positive hospital admissions in pre-term or very young infants are associated with apnoea (58,59). Preterm infants with RSV infection are also likely to present with poor feeding or lethargy (49). Cough and wheeze occur in approximately half of RSV-positive infants, and rhinorrhoea, fever, pharyngitis or respiratory distress are also common (49). Older children and adults are likely to present with typical cold symptoms, for example nasal congestion, ear infections, cough and fever (49).

The duration of RSV-associated illness also varies with age (6). In most patients, even those admitted to hospital, symptoms resolve within a few days (49). The

typical duration of an RSV-associated hospital admission in children ranges from approximately 2 to 7 days, though there is significant international variation in the duration RSV-associated hospital admissions and in England there has been a large increase in short-term admissions (of less than 1 day) for bronchiolitis over recent years (60,61). Infants younger than 2 months and those with underlying comorbidities usually require the longest hospital admissions, with the sickest children hospitalised for 7 days or more (6). RSV-associated respiratory tract infections (RTIs) are also significantly associated with higher disease severity compared to non-RSV-associated RTIs, with longer length of hospital stay, increased risk of intensive care admission and increased use of supplementary oxygen or mechanical ventilation (22).

1.3.6 RSV-associated morbidity

A number of follow-up studies have demonstrated that patients hospitalised with RSV-associated LRTI are more likely to experience subsequent long-term abnormalities such as recurring wheezing or pulmonary function deficits (48,62). It has been suggested that over half of infants hospitalised with RSV LRTI have subsequent wheezing (63). These wheezing episodes can persist up to adolescence (48). RSV infection in early life is also associated with the development of asthma (63). It is unclear whether these links are causal, or whether the children who experience more severe RSV infection have underlying functional or anatomical abnormalities that predispose them to both severe RSV infection and long term respiratory abnormalities (62). There is growing evidence of a causal role of RSV in the development of subsequent asthma (48,62,63). However, not all infants who experience severe RSV infections in early life will develop subsequent asthma or other respiratory abnormalities (48). Therefore, it is likely that some infants are predisposed to both due to genetic susceptibility or unfavourable birth characteristics (e.g. prematurity).

1.3.7 Diagnosis and virus detection

Specific viral aetiologies of respiratory tract infections cannot be differentiated using clinical diagnosis alone, as the symptoms of respiratory virus infections are non-specific (as outlined in section 1.3.5) (64,65). Laboratory testing is therefore required to distinguish RSV-associated respiratory tract infections (RTIs) from non-RSV-associated RTIs.

Accurate diagnosis of RSV infection relies on virus detection in respiratory secretions (66). Nasal washes and nasopharyngeal aspirates the most sensitive specimens for RSV detection, however nasopharyngeal swabs are also used as they are usually less unpleasant for patients and easier to carry out (67). RSV can be detected via multiple methods, including: direct or indirect immunofluorescent staining, enzyme-linked immunosorbent assays, or amplification assays (most commonly reverse transcription polymerase chain reaction (RT-PCR) (67).

RT-PCR is now the most commonly used method of detecting RSV in microbiology laboratories in the UK and considered the gold-standard test due to its high sensitivity and specificity (67). Multiplex RT-PCR allows for the simultaneous testing of multiple viruses, with results available within a few hours (66). Many commercial RT-PCR assays are available, with similar sensitivity and specificity (approaching 100%) (68). RT-PCR also allows quantification of the viral load, which may correspond with more severe disease (67).

For most infants and children infected with RSV who experience mild symptoms, diagnostic testing is of very limited clinical value (41). Even in secondary care, while diagnostic testing can help in infection control practices (isolating or cohorting patients with RSV) and in the monitoring of epidemics, it rarely alters the course of treatment (49). Laboratory testing is therefore only carried out in a minority of infants with respiratory illness, including among those hospitalised (69).

Though RSV is a non-notifiable infection in England, laboratory-confirmed RSV infections are routinely reported to Public Health England (PHE), the national

agency responsible for infectious disease surveillance, prevention and control. This reporting builds up a national epidemiological picture of seasonal epidemics (e.g. Figure 1-2) that can be used in the preparation and management of healthcare resources.

1.3.8 Treatment

Currently there is no specific, effective treatment for RSV infection (50). Furthermore, there are no currently available treatments that reduce the duration of infection or improve symptoms of RSV infection (41,70). Bronchodilators, corticosteroids and the antiviral medication Ribavirin (effective against select members of the *Paramyxoviridae* family) have failed to show effectiveness in randomised control trials, and in most countries are not currently recommended as treatments for RSV infection (71). There is high variability in the management of RSV-associated hospital admissions (72). Supportive care – including hydration, saline nose drops to clear nasal obstruction, nasal bulb suction, feeding support, and supplementary oxygen – is the mainstay of treatment for RSV-infection (49). Approximately 5% of healthy infants and 20% of infants with underlying chronic conditions with RSV-associated hospital admissions require mechanical ventilation (49). In the absence of a specific treatment for RSV, preventative strategies are vital.

1.3.9 Palivizumab prophylaxis: prevention of severe RSV infection

The monoclonal antibody Palivizumab (Synagis®, MedImmune) provides short-term passive protection against RSV (8,73,74). Palivizumab binds to the F glycoprotein of RSV (described in section 1.3.1) which prevents invasion of the host cells in the airway, reducing viral activity and cell-to-cell transmission (37). This mechanism can therefore prevent the development of LRTI, confining the virus to the upper airway (37). Due to its high cost (over £1,000 per single dose (70)) and short half-life

(requiring multiple intramuscular injections to be given in five monthly doses from the beginning of the RSV season), Palivizumab is only considered to be cost-effective in certain high-risk populations (75,76).

The Joint Committee on Vaccination and Immunisation (JCVI), an independent expert advisory committee of the UK Department of Health, currently recommends Palivizumab be given to the following high-risk groups (74):

a) Preterm infants with CLD at the chronological ages at the start of RSV season and gestational ages at birth within the shaded area in Table 1-1.
b) Preterm infants with haemodynamically significant CHD at the chronological ages at the start of RSV season and gestational ages at birth within the shaded area in Table 1-2.
c) Children under the age of 24 months with severe combined immunodeficiency syndrome (SCID).
d) Long term ventilated (LTV) children aged less than 12 months at the start of RSV season, or aged less than 24 months with an additional comorbidity.

Table 1-1. Cost effective use of Palivizumab (shaded area) for preterm infants with CLD by chronological age (months) at the start of the RSV season (beginning of October) and gestational age at birth (weeks). The definition of CLD is oxygen dependency for at least 28 days from birth. Table from the Joint Committee on Vaccination and Immunisation Statement on immunisation for Respiratory Syncytial Virus (74)**.**

	Gestational age at birth (weeks)						
Age (months)	≤24	>24 to ≤26	>26 to ≤28	>28 to ≤30	>30 to ≤32	>32 to ≤34	≥35
1 to <1.5							
1.5-3							
3-6							
6-9							
>9							

Table 1-2. Cost effective use of Palivizumab (shaded area) for preterm infants with haemodynamically significant CHD by chronological age (months) at the start of the RSV season (beginning of October) and gestational age at birth (weeks). Table from the Joint Committee on Vaccination and Immunisation Statement on immunisation for Respiratory Syncytial Virus (74)**.**

	Gestational age at birth (weeks)						
Age (months)	≤24	>24 to ≤26	>26 to ≤28	>28 to ≤30	>30 to ≤32	>32 to ≤34	≥35
<1.5							
1.5-3							
3-6							
6-9							
>9							

In a randomised, double-blind, placebo-controlled trial, it was estimated that Palivizumab prevents 39-78% of RSV-associated admissions in high-risk infants and young children, with premature children without CLD benefiting the most from this prophylaxis (77). Palivizumab has also been shown to reduce severity of RSV-associated disease (i.e. length of hospital stay, respiratory support, and incidence of intensive care unit admissions) in high-risk infants and young children (77). However, there has been significant controversy around its cost-effectiveness, with highly inconsistent economic evaluations ranging from not cost-effective to highly cost-effective (and many of the cost-effectiveness studies funded by MedImmune) (37,76). In England, there is no routinely collected national data from hospital pharmacies on prescriptions, therefore it is not possible to ascertain the proportion of recommended children who receive palivizumab prophylaxis. Despite the recommendations for palivizumab use, the burden of RSV in high-risk children (particularly preterm infants) remains high (37,78).

1.3.10 RSV vaccine development

Due to the high hospital burden globally, and the substantial mortality burden in low-income countries, RSV has been a priority for vaccine development for over half a century (4). In the 1960s, a formalin-inactivated RSV vaccine was tested in infants in four studies in the United States with severe consequences (79–82). Vaccine-enhanced respiratory disease was observed; upon natural RSV infection following vaccination, the hospitalisation rate of vaccinated infants was almost 80% compared to only 5% in controls in one study, and two children died (79). These failed trials halted the development of RSV vaccines for a significant period of time.

Despite extensive research into the cause of the vaccine-enhanced illness, the mechanism is not completely understood (4,83). However, major advances in the understanding of the biology of RSV and innovations in immunogen design have resulted in large numbers of potential vaccine candidates under development within recent years, with 16 candidates now in phase 1 to phase 3 clinical trials as of August 2017 (84). A World Health Organization (WHO) consultation in early 2015 on the development of RSV vaccines suggested that it is likely that an RSV vaccine will be available commercially within 5-10 years (85). The primary focus of an RSV vaccine is the prevention of severe RSV-associated disease at the extremes of the age spectrum. However, there is no universal, standardised definition of severe RSV-associated disease (86). As regulatory agencies expect primary endpoints for RSV vaccine trials to reflect clinically relevant disease prevention, primary endpoints include RSV-associated LRTI, RSV-associated hospital admission, or severe RSV-associated LRTI requiring intensive care admission (85).

1.3.10.1 Overview of current vaccine environment

Around 60 diverse RSV vaccine candidates are currently in the development pipeline (71). These vaccines target different populations (further described in section 1.3.10.2 below): infants and young children, pregnant women, and the elderly. Though most candidates are in the preclinical stage, 16 are in phase 1 to phase 3 clinical trials (Figure 1-3) (71). Preclinical vaccine developers include academic institutions, pharmaceutical companies, government agencies, and biotechnology organisations (71). Developers with maternal vaccine candidates in clinical trials include Novovax (Phase 3) and GlaxoSmithKline (GSK) (Phase 2). Developers with candidate vaccines in clinical trials targeted at the paediatric population include GSK (Phase 2), Novovax (Phase 1), Sanofi (Phase 1), Janssen Pharmaceutical (Phase 1) and Mucosis (Phase 1) (Figure 1-3). These vaccines are diverse, using a number of different platforms.

Live-attenuated vaccines (currently in Phase 1 clinical trials, Figure 1-3) have been in development for decades. The immune response to these vaccines closely resembles the response to natural infection, and there has been no evidence of these vaccines causing enhanced disease in RSV-naïve infants (4,71). Live-attenuated vaccines should protect against both URTI and LRTI, as they induce systemic as well as local immunity (4). Furthermore, these live-attenuated vaccines have been shown to be infective in very young infants (despite the presence of maternally acquired antibody) for other respiratory viruses: influenza A and parainfluenza (4). Therefore, live-attenuated RSV vaccines currently in development target infants and young children.

During recent years there has been a significant surge in the development of RSV vaccine candidates using other platforms, particularly particle-based vaccines, subunit vaccines and gene-based vectors (71). These vaccine candidates are in various stages of clinical development (Figure 1-3). A number of these candidates target the F glycoprotein of RSV – already shown to be an effective immunogen for Palivizumab. Particle and protein-based vaccines are currently being developed with the aim of targeting pregnant women. These types of vaccines are often

formulated with an adjuvant, which enhance the immune response to give longer-lasting protection. It is likely that the only acceptable adjuvant for a maternal RSV vaccines will be aluminium, due to its demonstrated history of safe use during pregnancy (71). Vector- and subunit-based vaccines aim to simulate the safe immune responses that occur during live RSV virus infection, reducing the risk of enhanced disease in RSV-naïve infants and young children (87).

Of all vaccine candidates currently in development, the Novavax RSV F protein nanoparticle vaccine candidate (RSV F Vaccine) has advanced furthest in clinical trials (Figure 1-3). This vaccine candidate has been shown to be well tolerated and immunogenic in women of child bearing age (88). However, it did not show efficacy in a Phase 3 trial in older adults (89). A low RSV attack rate (the percentage of people who get infected) during the study period has been hypothesised for this lack of efficacy, which is a reasonable likely cause (89). Phase 2 trials of other vaccine candidates have also shown some success (71). In June 2017, Bavarian Nordic announced positive results for their RSV vaccine candidate (Modified Vaccinia Ankara, MVA-BN RSV), which elicited a broad immune response against RSV in a Phase 2 clinical trial (90). Therefore, considering the number and variety of RSV vaccine candidates currently in clinical development, and the promising Phase 2 clinical trials, it is highly likely that an effective RSV vaccine will be available within the next decade (85).

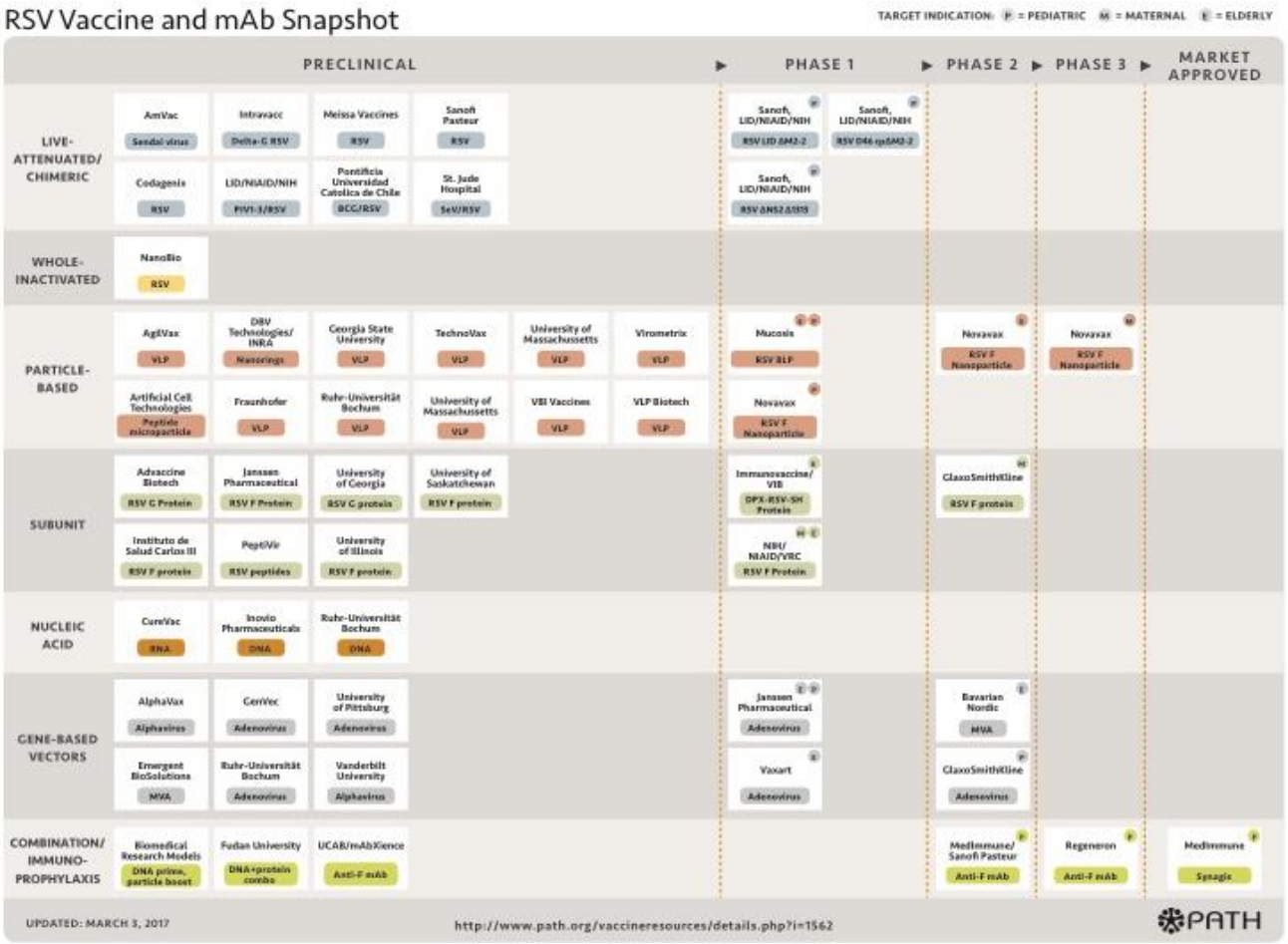

Figure 1-3. Snapshot of RSV vaccines currently in development. Published by PATH (2017), available at http://www.path.org/vaccineresources/details.php?i=1562

1.3.10.2 Potential target populations

As the peak in RSV-associated hospital admissions occurs at just a few months of age (section 1.3.2) and natural immunity is so short (section 1.3.4), RSV is a difficult virus to design a vaccine for. There are a number of key target populations for an RSV vaccine that aims to prevent severe infections in infants and young children: young infants (<6 months old), older infants and children 6-24 months of age, and pregnant women (Table 1-3) (87,91). Other potential vaccination strategies could include: (a) targeting older siblings (or other household members) to block transmission to infants or young children in the household, and (b) targeting healthcare workers to prevent transmission in healthcare settings (87). Though a further key target population for a potential RSV vaccine is the elderly, this population will not be discussed in this thesis as my work focusses on RSV in children younger than 5 years of age only. The target population determines the type of vaccine used (92).

Table 1-3. Key target populations for an RSV vaccine. Adapted from Anderson et al. (2013) (87)**.**

Target population	**Aim**	**Primary vaccine approaches**
0-6 month old infants	To prevent severe infections and hospital admissions	a) Live-attenuated RSV b) Live chimeric virus vectors c) Gene-based vectors d) Live or gene-based vector followed by subunit or particle-based vaccine
6-24 month old children	To prevent severe infections and reduce transmission to at-risk household contacts	a) Gene-based vectors b) Live-attenuated RSV c) Live chimeric virus vectors d) Live or gene-based vector followed by subunit or particle-based vaccine
Pregnant women	To increase passive protection to newborn infants	a) Subunit protein with standard adjuvants b) Particle (including virus-like particle) with standard adjuvants

Vaccinating pregnant women is likely to be the most feasible approach to prevent severe RSV disease during the high-risk period in early infancy, as very young infants (≤3 months old) may not respond adequately to vaccination and are at risk of vaccine-enhanced disease (85,92). Higher concentrations of RSV-specific maternal antibodies are associated with reduced incidence of RSV disease during the first months of life (92). Therefore, increasing maternal RSV antibodies may effectively reduce the number or severity of infections in early infancy. Vaccinating pregnant women also may also reduce infant infections by reducing virus transmission (4). There are existing platforms for maternal immunisation (e.g. for influenza), with the vast majority of pregnant women having some antenatal health service contact (even in the least-developed countries) (92). These existing platforms could enable the introduction of a future RSV vaccine for pregnant women without amending existing antenatal care delivery. However, there are significant challenges of maternal vaccination. First, there is the difficulty of inducing an immune response that is more robust and longer lasting than that induced by natural infection, able to protect the infant for the whole duration of the high-risk young infancy period (93). Second, although vaccination late in pregnancy would ensure that antibodies last as long as possible in the infant, this could mean that premature babies vulnerable to severe RSV infection would not be protected. A further consideration of a maternal vaccine strategy is there is a low uptake of maternal vaccines in the UK; for example, uptake rates for the influenza and pertussis vaccines are 45% and 60% respectively (94).

An alternative potential target population could be children aged 6 to 24 months old. These children have a more mature immune system and less maternal antibodies, and are therefore likely to respond better to vaccines than very young infants (87). Though these children are not at the highest risk of severe RSV-associated disease, a substantial burden of disease still occurs in these children. Delaying the primary RSV infection in these children until their airways are larger, through vaccination, is likely to reduce the severity of that infection (95). In addition, vaccinating these children could potentially have important herd effects and reduce transmission to the more vulnerable younger infants. Similarly,

vaccinating siblings, parents or other contacts of infants could also reduce transmission of RSV and consequently reduce the burden or severity of RSV infection in young infants; however, as these people may receive minimal benefit themselves from the vaccine, this strategy raises ethical considerations (45).

1.3.10.3 Determining optimal target populations

To determine optimal target populations for a potential future RSV vaccine programme, economic evaluations must be carried out. In the economic evaluation of a vaccine programme the costs of different vaccination strategies are compared with a reference strategy (i.e. no vaccination) and the health benefits of vaccination are studied (96). Health benefits can be expressed in monetary units, natural units (cases avoided or life-years saved) or in quality-adjusted life-years (QALYs) gained from vaccination (96).

Economic evaluations are based on models, which can be *static* or *dynamic* (97). Static models have been conventionally used in health economics to explore the potential direct effect of vaccination on the recipient. However, these models do not account for the indirect or herd effects of vaccination (98). For transmissible infectious diseases, dynamic models are more appropriate. Dynamic models take into account that vaccination not only reduces the probability of the vaccinated individual developing the infection, but also reduces the exposure of others to the infection (97).

In all cost-effectiveness models, a crucial component is data on the incidence of healthcare utilisation (i.e. hospital admissions and primary care attendances), usually stratified by key demographic factors such as age. This data is typically estimated from burden of disease studies, with the secondary care burden particularly important for RSV (as the primary focus of an RSV vaccine is the prevention of severe RSV-associated disease including RSV-associated hospital admissions, section 1.3.10). Data also needs to be inputted on factors such as

vaccine efficacy, age at vaccination, duration of antibody protection, QALYs gained from vaccination (or other measure of health benefit), and all associated costs (i.e. health-care associated costs, and the costs of vaccination) (98). The most well-known dynamic model for the spread of infection is the Susceptible-Exposed-Infectious-Removed (SEIR) model, which also requires information on the contact rates among and within health groups, the length of the infectious period, and the probability of transmission of the virus during a contact (in addition to the data mentioned above) (97).

The aforementioned data can be used within a mathematical framework to investigate alternative vaccine strategies, modelling their projected impact and cost-effectiveness. PHE have a long history of carrying out this type of work to inform vaccine policy (for example, for influenza (99,100)), and will carry out similar modelling studies to inform the introduction of a potential future RSV vaccine.

1.4 PhD Rationale

In order to design optimal interventions, such as a future vaccine programme, the burden of RSV on the health service first needs to be accurately determined. Information on the pre-vaccine era burden of disease can be used as a baseline to measure the impact of a new vaccine once it is introduced, and act as a platform for cost-effectiveness studies in key target groups (section 1.3.10.3). However, advances in RSV vaccine candidate development are currently occurring at a faster rate than action to gather epidemiological data on RSV-associated disease (101). A number of key gaps in knowledge of the overall health-care burden, potential long-term health effects, associated costs, and transmission characteristics of RSV remain unanswered (45,85,87,101,102). These key gaps in knowledge are:

- Precise estimates of RSV-associated morbidity by age and risk group, including:
    - RSV-associated hospital admissions
    - Paediatric intensive care unit (PICU) admissions
    - Accident and emergency (A&E) admissions
    - General practitioner (GP) consultations
- Information on chains of RSV transmission to populations at risk of severe RSV-associated disease,
- Information on the long-term health impact of RSV infection, such as recurrent wheezing and asthma, and the subsequent associated health care utilisation,
- Precise estimates of RSV-associated mortality (less relevant for the UK which has low mortality due to RSV, but important in developing countries).

These gaps must be filled to provide a robust evidence base to guide national, regional and global policy decisions regarding future interventions for RSV (101). This information will enable calculation of the costs and benefits of potential future vaccines, and determine whether a direct vaccination programme (e.g. targeted to

young infants, or older infants) or an indirect vaccination programme (e.g. maternal immunisation, or school-age siblings) will be the most effective approach.

It is important to have national-level estimates of RSV morbidity and mortality in order guide national vaccine policy. As there is significant international variation in the management of infants hospitalised with RSV, cost-effectiveness of vaccine strategies may differ between countries (60). In addition, transmission may vary due to different mixing patterns among young infants and other age groups, which could impact the cost-effectiveness of different potential vaccine strategies (as well as informing the implementation of other preventative strategies, e.g. improving hygiene practices such as handwashing). Furthermore, as infants born in close proximity to the beginning of RSV season are thought to be at highest risk of severe RSV infection requiring hospital admission (21,103–106), and the timing of RSV season differs by geographical location (section 1.3.2.1), country-specific burden estimates are essential in tailoring interventions to optimal target groups in each country.

RSV research networks have recently been established in order to better understand RSV epidemiology and the burden of RSV-associated disease (101,102). For example, the Respiratory Syncytial Virus Network (ReSViNET) was established as an independent research network in 2014 to combine expertise on RSV across the globe (107). In addition, the Respiratory Syncytial Virus Consortium in Europe (RESCEU) – a Europe-wide consortium from academia, public health, regulatory agencies and industry - was launched in January 2017 to develop evidence on RSV disease burden and economic impact (108). These networks will provide a platform for measuring the potential health and economic benefits of a future RSV vaccine and monitor the impact of vaccination once introduced. However, new surveillance systems will take time to implement, due to the unavoidable financial and technical challenges of setting up such large systems (101). Existing surveillance systems – particularly those focussed on influenza – may provide an alternate opportunity to implement RSV surveillance. However, this presents its own challenges, due to the differing age patterns of the viruses and different clinical presentation of cases with

influenza and RSV (101). Furthermore, depending on this design of these surveillance systems and the subsequent representativeness and generalisability of their data, they may not necessarily capture the national-level burden of disease (for example, due to biases in testing). Either way, the time required to implement such surveillance systems further delays the gathering of baseline epidemiological data on RSV, required to guide mathematical and economic models to define vaccine strategy.

1.4.1 Using administrative data to inform vaccine strategy

Retrospective studies utilising routinely collected (or administrative) datasets (e.g. electronic health records) provide the opportunity to answer key epidemiological questions traditionally answered using large-scale, longitudinal studies (109). The increasing availability of such information-rich routinely collected datasets can, relatively inexpensively and quickly, supply large samples and long-term follow-up (109). There are limitations to these datasets, such as the data not being collected for research purposes, which may mean certain information of interest is lacking or incomplete. In addition, there remain important issues surrounding data governance, privacy protection and access to data. However, the limitations of these datasets do not outweigh their potential benefits for answering important epidemiological and public health questions (109). For example, analysis of routinely collected laboratory surveillance datasets can give important information on virus circulation and laboratory-confirmed episodes of infection. In addition, routinely collected hospital data is an information-rich source of data on hospital admissions. Routinely collected datasets are therefore a valuable source of information which can be utilised to fill key gaps in knowledge of RSV burden.

1.4.2 Estimating the secondary care burden of RSV using routinely collected data

In this thesis, I use routinely collected datasets to estimate the secondary care burden of RSV in infants and young children in England. As outlined above, national-level data on RSV burden is required to highlight key target populations for interventions, such as a potential vaccine programme, and for economic evaluations to drive vaccine policy decisions. I focus on the secondary care burden of RSV due to the significant health and economic impact of hospital admissions, the subsequent importance of secondary care burden estimates in RSV vaccine cost-effectiveness studies, and the use of RSV-associated hospital admission as a primary endpoint for RSV vaccine trials.

Calculating the national secondary care burden of RSV is not straightforward. As explained in section 1.3.7, only a minority of patients presenting with respiratory infections undergo laboratory testing to determine the causal pathogen. The burden of RSV in secondary care therefore has to be estimated, and has been done so (using routinely collected datasets) in a number of different ways:

a) Using hospital admissions for bronchiolitis as a proxy for all RSV-associated hospital admissions (110),
b) Using time-series modelling of routinely collected laboratory surveillance and hospital admissions data (111–113),
c) Using population-based linkage of laboratory surveillance and hospital admissions data (114–117).

Method (a) has been used previously in England to identify risk factors for RSV-associated hospital admission using a population-based cohort (110). While this method is easy to implement, it is not specific for RSV as not all bronchiolitis admissions are caused by RSV (although, in the height of RSV season, bronchiolitis admissions have a high positive predictive value (PPV) for RSV). Furthermore, not all RSV-associated hospital admissions are coded as bronchiolitis. Therefore, this method is likely to underestimate the true burden of RSV-associated hospital admissions.

Method (b) has been used to estimate the national secondary care burden of RSV in in multiple previous studies in England (111–113,118). This method has been used to determine the number of RSV-associated hospital admissions by age group (e.g. <6 months, 6-23 months) (113). As these estimates are derived from models based on aggregate data and are ecological in design, they are limited in the detail they can provide. These estimates are therefore not broken down by risk factors such as the presence of underlying chronic conditions, which is important as children with underlying chronic illness are more likely to develop more severe disease and have longer hospital stays (section 1.3.2). Furthermore, none of the studies estimating the secondary care burden of RSV in England include more recent data than 2009. Since the 2009 influenza A (H1N1) pandemic, surveillance of respiratory viruses including RSV has increased through the implementation of a new respiratory virus surveillance system, the Respiratory DataMart System (RDS) (119). This dataset collects positive and negative laboratory test results for major respiratory viruses, providing an opportunity to explore RSV epidemiology while accounting for patterns and biases in laboratory testing for RSV. In addition, as a recent study in England demonstrates that rates of hospital admissions coded as due to bronchiolitis are increasing over time, this emphasises the need for more recent estimates of RSV burden including the post-pandemic era, as previous estimates may not reflect the current burden of disease (51).

Worldwide, few studies have used method (c) and utilised population-based data linkage to describe the aetiology or pathogen-specific burden of acute lower respiratory infections (115). To date, the only such studies focusing on RSV have been based in Australia, Canada and Denmark (114–117). Linkage of routinely collected laboratory surveillance and hospital admissions data on RSV has not previously been carried out in England. This linkage facilitates the analysis of more complete information without the time and cost burden of primary data collection (for example, through a prospective surveillance study) (120). Therefore, linkage of routinely collected laboratory surveillance and hospital admissions data is a more timely and cost-effective method of describing laboratory-confirmed RSV-associated hospital admissions in more detail than population-based time-series modelling

estimates (i.e. by narrower age groupings such as age in months, and by individual-level clinical risk factors such as comorbidities) (115).

In this thesis, I produce more recent estimates of the secondary care burden of RSV in infants and young children in England using method (b). In addition, I expand on method (c) by using linked laboratory surveillance and hospital admissions data for RSV in England for the first time, and develop a novel methodology of using this linked data to develop a predictive model of RSV-positivity – I then use this predictive model to determine the probability of a hospital admission without a laboratory test being caused by RSV, enabling me to estimate the secondary care burden of RSV in infants and young children in England in detail by demographic and clinical characteristics. Using both method (b) and an expansion of method (c) allows comparison of both results, as well as validation of the novel methodology that I have used. My PhD demonstrates the usefulness of utilising routinely collected datasets in generating epidemiological data on the secondary-care burden of RSV and risk factors for severe RSV-associated disease, filling crucial gaps in the knowledge of RSV morbidity that are required to inform decisions regarding future vaccine strategy.

Chapter 2

Aims and objectives

Chapter 2 Aims and objectives

2.1 Aim

The overall aim of my PhD is to estimate the burden of RSV in infants and young children in secondary care in England using routinely collected laboratory surveillance and hospital admissions data.

2.1.1 Objectives

This overall aim can be broken down into five objectives:

1. Describe the epidemiology of RSV in children younger than five years in England using laboratory surveillance and hospital administrative data.

2. Estimate the number of RSV-associated hospital admissions in children younger than five years in England (stratified by age, primary diagnosis and calendar week) using time-series modelling of national laboratory surveillance and hospital administrative data.

3. Describe, using linked laboratory surveillance and hospital admissions data, laboratory-confirmed RSV-associated hospital admissions in children younger than five years in England by key patient and clinical characteristics.

4. Determine risk factors for severe disease (indicated by prolonged hospital stay or use of invasive ventilation) among laboratory-confirmed RSV-associated hospital admissions in children younger than 5 years in England.

5. Use the linked dataset to estimate the total national burden of RSV-associated hospital admissions in children younger than five years of age in England (stratified by age, primary diagnosis, calendar week and risk group).

Chapter 3

Routinely collected data on RSV in England

Chapter 3 Routinely collected data on RSV in England

3.1 Introduction

The overall aim of this thesis is to estimate the burden of RSV in children younger than 5 years in secondary care in England using routinely collected data. There are two types of routinely collected data on RSV that can be used for this purpose in England, and I have used both for this work:

1) Laboratory surveillance data,
2) Secondary care data, in the form of hospital administrative data on admissions.

This Chapter gives an overview of these datasets, and outlines the study populations used in the analysis in this thesis.

3.2 Laboratory surveillance data on RSV in England

The Centre for Infectious Disease Surveillance and Control (CIDSC) at PHE is the national agency responsible for infectious disease surveillance, prevention and control. There are two infection surveillance datasets held by the CIDSC which record information on laboratory-confirmed RSV infections in England: the Second Generation Surveillance System (SGSS) (formerly known as LabBase2) and the Respiratory DataMart System (RDS).

3.2.1 Overview of laboratory surveillance datasets

The main characteristics of the SGSS and RDS laboratory surveillance databases are outlined in Table 3-1. The major difference between the datasets – that has influenced their use in my analysis – is that SGSS only contains positive RSV test results, whereas RDS contains both positive and negative results. There are also distinct differences between the datasets in terms of geographical coverage, the period of available data, and the characterisation of infection episodes (i.e. deduplication of multiple records for the same patient within a certain time period).

Table 3-1. Main characteristics of the Second Generation Surveillance System (SGSS) and the Respiratory DataMart System (RDS).

	Second Generation Surveillance System (SGSS)	**Respiratory DataMart System (RDS)**
Coverage	All PHE, NHS and private microbiology laboratories in England and Wales	14 laboratories only (all major PHE regional laboratories and four local NHS laboratories)
Inclusion of RSV test results	Positive results only	Positive and negative results
Timescale	Since 1990	Since 2010
Testing	All diagnostic tests	Multiplex PCR (respiratory viruses) only
Deduplication	Records within 2 weeks in the same patient for the same organism grouped under a unique ID (all records retained)	Tests for all respiratory viruses within 6 weeks in the same patient are merged into a single record

3.2.2 Second Generation Surveillance System (SGSS)

SGSS (formerly known as LabBase2) holds records of all RSV-positive laboratory tests – as well as positive test results for other viruses and bacteria – from PHE, NHS and private microbiology laboratories in England and Wales (64). All microbiology reports defined as "clinically significant" (by laboratory microbiologists) are requested although no guidelines for the judgement of clinical significance are defined (121). Laboratory reporting to SGSS of RSV positive tests is voluntary. PHE does not collect data on the ascertainment of RSV tests in SGSS, but studies considering the ascertainment of other clinically significant infections (blood-stream infections, rotavirus, methicillin-resistant and methicillin-susceptive *Staphylococcus aureus*) estimate ascertainment to be high (approximately 70-90%) and consistent all year round (122–124). Most of the laboratories reporting to SGSS are based within NHS hospitals, and all provide clinical diagnostic services to both primary and secondary healthcare providers (122). There is no reliable variable in SGSS that indicates whether the sample was collected from primary or secondary care. Records have mandatory data fields for patient identifiable information (e.g. NHS number, date of birth), laboratory information, specimen type (e.g. serum, urine, throat) and sample date. Test type (e.g. antigen detection, electron microscopy) is specified in 65-80% of RSV records each year.

3.2.2.1 SGSS data extraction

Data was initially extracted on 03/12/2014 for all records in children aged younger than 5 years in England during the period 01/01/1999 to 26/11/2014 on all major respiratory viruses (RSV, influenza A and B, human metapneumovirus (hMPV), adenovirus, parainfluenza and rhinovirus). Table 3-2 shows the variables extracted.

Table 3-2. Variables extracted from SGSS on all records in children <5 years of age.

Variable	Description
OPIE ID[1]	Identification number (OPIE = Organism-Patient-Illness-Episode)
Organism name	Name of virus tested positive for: RSV Influenza A Influenza B hMPV Adenovirus Parainfluenza Rhinovirus
Specimen date	Date specimen taken (01/01/1999 – 26/11/2014)
Specimen type	Description of specimen type
Method	Description of test type
Sex	Male, Female or Unknown
Date of birth	In format: dd/mm/yyyy
Surname	Surname of the patient
First name	First name of the patient
NHS number	NHS number of the patient
Hospital number	Hospital number of the patient (can also be clinic or laboratory number)
Postcode prefix	First part of patient's postcode
Postcode suffix	Second part of patient's post code
Laboratory name	Name of original source laboratory
Current region name	Current region name of laboratory
Specimen source	Description of specimen source (Hospital outpatient, General practitioner etc.)

[1] See section 3.2.2.2 for full definition

3.2.2.2 Episodes of infection

An individual may have multiple tests for a single virus during a single episode of infection. A single episode of infection is termed an Organism-Patient-Infection-Episode (OPIE) in SGSS. Records are grouped under one OPIE if:

1. The patient is the same,
2. The organism is the same,
3. The record is within 14 days of the previous record (in otherwise healthy individuals, including those hospitalised, recovery is usually within 10 to 14 days (68)),
4. The source laboratory is the same.

In order to carry out this deduplication process and identify tests carried out during a single episode of infection, probabilistic record linkage is used to identify whether a newly submitted record matches any existing record. This method calculates match weights (or match probabilities) that measure the similarity of patient identifiable information between records, while also taking into account possible errors or missing values in these variables (125) (Chapter 6). A unique identifier (OPIE ID) is assigned to records with a new OPIE. For OPIE duplicates, the same OPIE ID is assigned to all of the matching records.

An example of the process of merging OPIE duplicates into records with the same OPIE ID is shown in Figure 3-1. In this figure, the first table shows the data before the deduplication/linkage process. Records 1-3 are from the same patient and the records fulfil the criteria for OPIE duplicates outlined above (i.e. same organism, same source laboratory and records within 14 days of each other). During the linkage process, a new variable is created called “Earliest specimen date”. This variable takes the value of the earliest “Specimen date” within an OPIE. The only fields allowed to differ within an OPIE are specimen date, specimen type and the reporting laboratory (which may differ from the source laboratory).

Record no.	Specimen date	Specimen type	Source laboratory	Organism	NHS no.	Date of birth	Sex
1	10/08/2013	NPA	Northern General Hosp (Sheffield)	RSV	9434765919	31/07/2011	M
2	12/08/2013	NPA	Northern General Hosp (Sheffield)	RSV	9434765919	31/07/2011	M
3	17/08/2013	Upper respiratory tract	Northern General Hosp (Sheffield)	RSV	9434765919	31/07/2011	M

Deduplication/linkage process

OPIE ID	Earliest specimen date	Specimen date	Specimen type	Source laboratory	Organism	NHS no.	Date of birth	Sex
1234567	10/08/2013	10/08/2013	NPA	Northern General Hosp (Sheffield)	RSV	9434765919	31/07/2011	M
1234567	10/08/2013	12/08/2013	NPA	Northern General Hosp (Sheffield)	RSV	9434765919	31/07/2011	M
1234567	10/08/2013	17/08/2013	Upper respiratory tract	Northern General Hosp (Sheffield)	RSV	9434765919	31/07/2011	M

Figure 3-1. Example of SGSS data before and after the deduplication/linkage process. Records 1-3 are from the same patient. Circle highlights created variables. NPA = Nasopharyngeal aspirate.

Only multiple-record OPIEs are affected by this merging process, single-record OPIEs are not. The number of records and unique OPIE IDs per virus is outlined in Table 3-3. The average number of records per OPIE ID is similar for each virus, with RSV having the lowest average number of records at 1.07 per OPIE ID.

Table 3-3. Number of records and unique OPIE IDs per respiratory virus in the SGSS data extract.

Virus	No. of records	No. of unique OPIE IDs	Average no. of records per OPIE ID
RSV	133,065	124,207	1.07
Influenza A	36,095	32,707	1.10
Influenza B	9,496	8,685	1.09
Rhinovirus	44,874	39,305	1.14
Parainfluenza	21,556	18,836	1.14
hMPV	5,330	4,622	1.15
Adenovirus	70,080	65,156	1.08

The majority of RSV records (94.5%) in children younger than 5 years were single-record OPIEs (Figure 3-2). For multiple-record OPIEs, only the first record was included for analysis in this thesis.

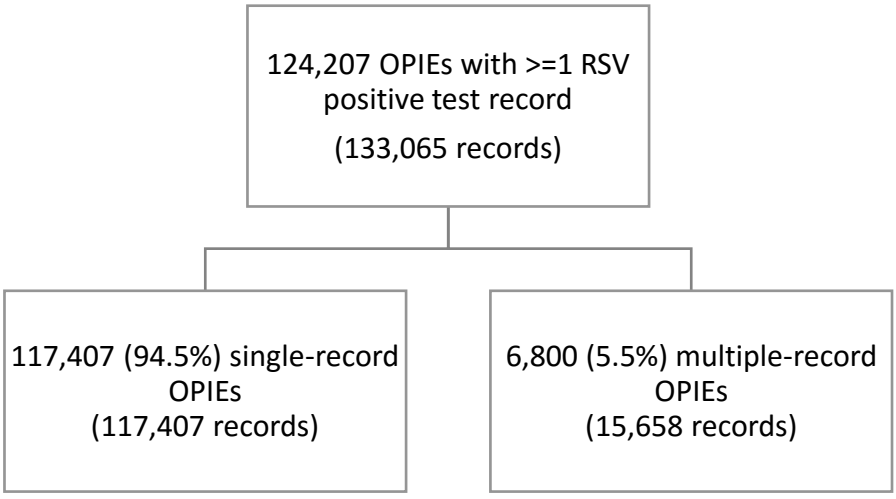

Figure 3-2. Make-up of RSV Organism-Patient-Infection-Episodes (OPIEs) in SGSS data extract (01/01/1999 – 26/11/2014).

3.2.2.3 Inclusion and exclusion criteria for SGSS analyses

SGSS collects data from England and Wales; therefore – as the population of interest for this thesis is England – records were restricted to those from laboratories in England only. In all analyses using SGSS in this thesis, each unique OPIE ID was considered to be one episode of infection and therefore only the first record of each OPIE ID was included. This ensured that multiple tests for the same virus within 14 days in the same individual were not counted as separate infections. Furthermore, only detections from respiratory samples were included in the analysis in this thesis, as detections from other sources are unlikely to be clinically relevant. These respiratory samples include nasopharyngeal aspirates (NPAs), throat, nasal, upper respiratory tract, or pharyngeal samples.

There has been an increase in RSV testing over time, as well as significant changes in the types of laboratory testing carried out. Testing has changed from mainly light microscopy and antigen detection (from 1998-2006) to mainly genomic detection (multiplex PCR) and antigen detection (Figure 3-3). Over time, there has been an increase in the proportion of RSV tests in SGSS labelled as 'unknown' test type – it is likely that the majority of these are multiplex respiratory PCR assays as over the last decade this has become the most frequently used diagnostic test in virology laboratories in the English NHS for diagnosing respiratory viral infections such as RSV (67). Multiplex PCR assays are the gold standard detection method for RSV as they have almost 100% specificity and are highly sensitive (section 1.3.7). Antibody detection and electron microscopy are not considered appropriate tests for RSV in young children (67), therefore these tests were excluded from my analyses.

Due to the changes in testing over time, only data from mid-2007 onwards (30/06/2007 to 01/07/2014) was considered for analysis in this thesis, covering six consecutive RSV seasons. Using this time-period ensured overlap with previous studies (to allow comparison of results), while maximising consistency in testing (allowing clearer interpretation of results). Figure 3-4 shows the final SGSS study population taken from the data extract, from which the study populations for analysis in this thesis are derived.

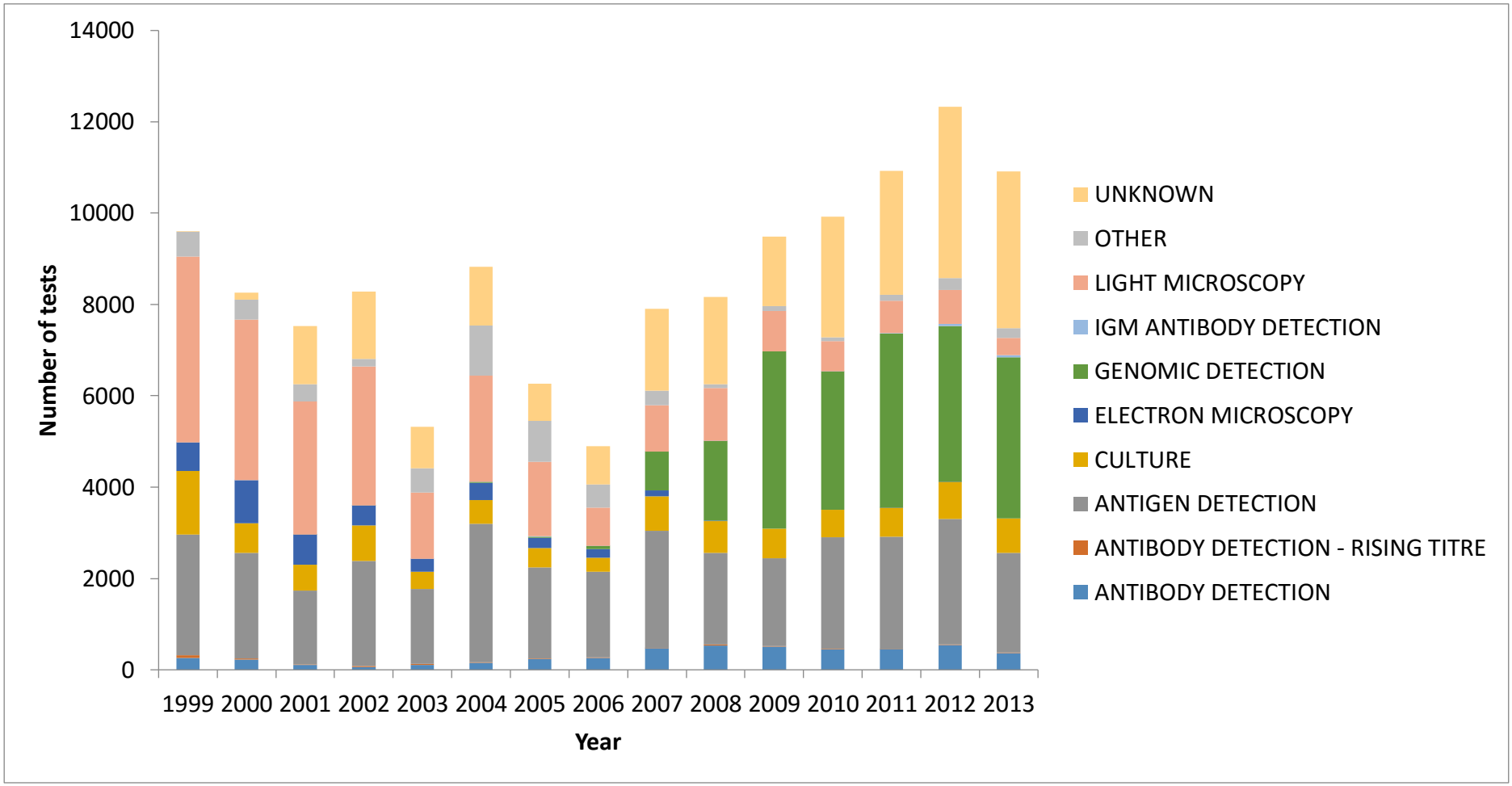

Figure 3-3. Number of laboratory tests for RSV in children <5 years old recorded in SGSS each year, by type of test.

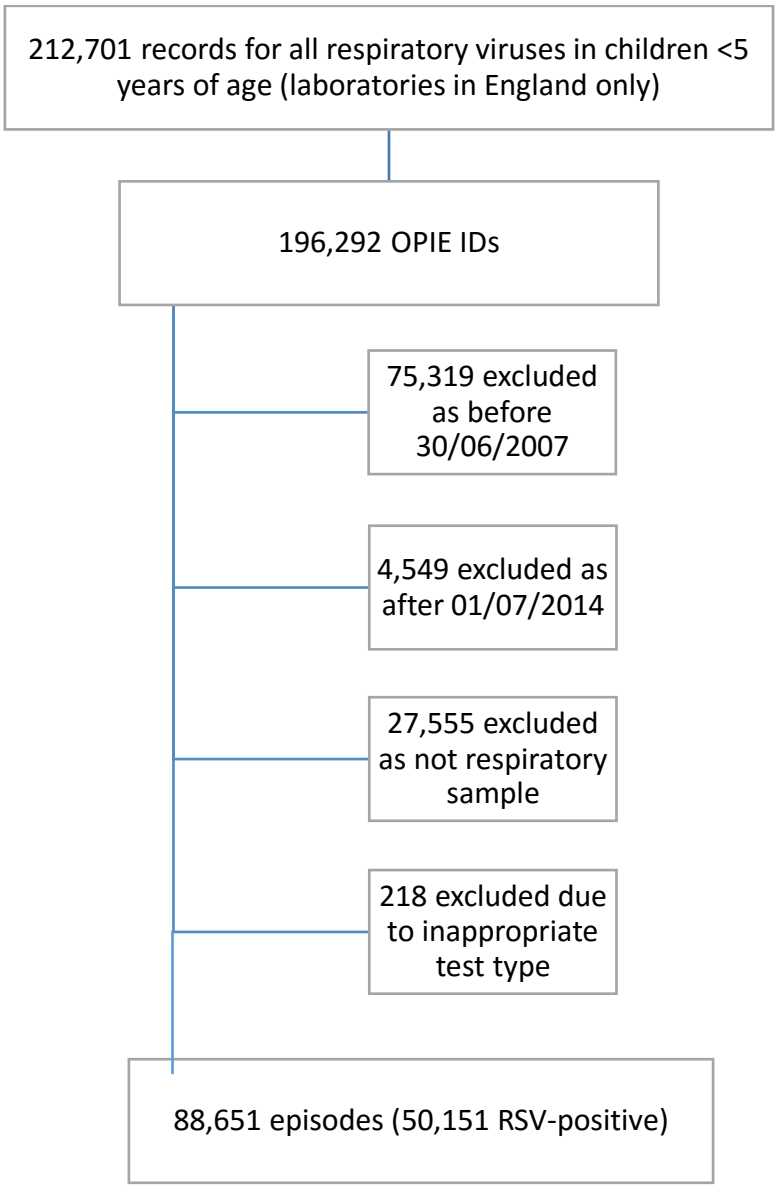

Figure 3-4. Flowchart of final SGSS study population based on inclusion/exclusion criteria.

3.2.2.4 Strengths and limitations of SGSS

SGSS is the largest dataset on laboratory-confirmed RSV infections in England as it covers all clinically significant RSV-positive test results from PHE, NHS and private laboratories. It is important for surveillance purposes and monitoring trends in RSV circulation over time, identifying epidemics in real-time and monitoring abnormal increases in circulation. The CIDSC at PHE produces weekly, monthly and annual reports on laboratory-confirmed respiratory infections in England and Wales, stratified by age group (<1 year, 1-4 years, 5-14 years, 15-44 years, 25-64 years, >65 years) and calendar week (126). Ascertainment of laboratory reports is generally considered to be high in SGSS; however, no studies have been carried out that estimate the ascertainment of RSV laboratory reports. SGSS does not contain data on RSV subtypes or strains.

SGSS has some important limitations for research use as there are no negative tests recorded. With only positive test results, it is impossible to know the denominator of children tested (i.e. we don't know the total number of tests for RSV that have been carried out). This means that testing practices cannot be examined, and any increase in virus circulation cannot be disentangled from an increase in testing. Furthermore, there is no reliable information on which samples were collected in primary or secondary care. Therefore, any analysis carried out on SGSS data must be interpreted with caution as any apparent differences in the number of positive samples may be due to differences or changes in testing practice. It is for this reason that – although patient identifiable information is included in the dataset – linkage between SGSS data and hospital admission data has not been carried out for this thesis.

3.2.2.5 Use of SGSS data in this thesis

In this thesis I have used SGSS data in two ways:

1. I have used SGSS data to address Objective 1 of this thesis and analyse the epidemiology of RSV-positive laboratory-confirmed infections in children younger than 5 years in England, summarising the number of RSV positive episodes in SGSS by age, sex, birth month, calendar week and year. For this analysis, I used data from calendar week 27 in 2010 to week 26 in 2014 in order to have a comparable study population to the RDS population (described in section 3.2.3) which only has data available from 2010 onwards. This work is presented in Chapter 4 of this thesis.

2. To address Objective 2 of this thesis, I have used SGSS data aggregated by calendar week in the time-series modelling of hospital admissions, which aims to estimate the total number of RSV-associated respiratory hospital admissions in children younger than 5 years in England by age and primary diagnosis. For this analysis, SGSS data from calendar week 27 in 2007 to calendar week 28 in 2012 was used (the period of analysis was restricted by the availability of the hospital admission data). This work is presented in Chapter 5.

3.2.3 The Respiratory DataMart System (RDS)

The RDS was established by PHE during the 2009 influenza A (H1N1) pandemic to collect positive and negative results for the detection and confirmation of the new influenza A (H1N1)pdm09 virus infection (119). RDS was extended in 2010 to include these other common respiratory viruses – including RSV – tested for using multiplex PCR, from both primary and secondary healthcare settings, in order to strengthen respiratory virus surveillance in England. In this thesis, I include all the test data available on RSV in RDS at the time of data extraction, from mid-2010 to mid-2014.

The RDS currently uses automatic weekly electronic outputs from all major PHE regional laboratories and four local NHS laboratories: Barts and The London, Birmingham, Bristol, Cambridge, Kings, Leeds, Leicester, Manchester (including Preston), Newcastle, Nottingham, Royal Free Hospital, Southampton, Truro, University College London Hospitals (UCLH), and the national reference laboratory (PHE Respiratory Virus Unit, Colindale, London). Figure 3-5 illustrates the geographical spread of these laboratories.

The RDS laboratories detect from a selection of influenza, RSV, parainfluenza, rhinovirus, human metapneumovirus (hMPV) and adenovirus using reverse transcription real time PCR (119). Not all laboratories test for all viruses, but all test for RSV and influenza A and B. Reports of both positive and negative results are normally submitted weekly to RDS. Laboratories can be requested to submit daily during pandemics, and a subset of seven major participating laboratories undertook daily submissions from April to September 2012 for the respiratory virus surveillance for the London 2012 Olympic and Paralympic Games (119). As all participating laboratories employ different information management systems and coding systems, data items such as sample date, date of birth and test results for each virus are standardised before combination in the central database.

Figure 3-5. Laboratories that submit to the Respiratory DataMart System (RDS).

3.2.3.1 RDS data extraction

Data was initially extracted on 13/12/2014 for all records in children aged younger than 5 years during the period 01/07/2010 to 30/06/2014, to cover four consecutive RSV seasons and including the earliest available data on RSV. Data on the test results were extracted for all viruses (i.e. whether the specimen tested positive for RSV, and whether it tested positive or negative for any other respiratory virus included in the multiplex PCR assay). Table 3-4 shows the variables extracted from RDS.

Table 3-4. Variables extracted from RDS on all records in children <5 years of age

Variable	Description
ID number	Identification number (unique to RDS)
Specimen date	Date specimen tested (01/07/2010 – 30/06/2014)
Sex	Male, Female or Unknown
Date of birth	In format: dd/mm/yyyy
Surname	Surname of the patient
First name	First name of the patient
NHS number	NHS number of the patient
Hospital number	Hospital number of the patient (can also be clinic or laboratory number)
Postcode prefix	First part of patient's postcode
Postcode suffix	Second part of patient's post code
Laboratory name	Name of original source laboratory
RSV test result	RSV test result (Positive, Negative, Unknown)
Other test result	Test result for other viruses (Positive, Negative, Unknown)

3.2.3.2 Episodes of infection

As in SGSS, deduplication of records is carried out during the data importation process as an individual may have multiple tests during a single episode of infection. Samples taken from the same individual within a six-week period are grouped as one record to capture a single episode of infection in an individual. This single record contains the date of the first test only, and contains one test result for each virus (positive test results override a negative result during the deduplication process). Figure 3-6 gives an example of this deduplication process.

Record no.	Specimen date	Test result						
		RSV	Influenza A	Influenza B	Parainfluenza	Rhinovirus	hMPV	Adenovirus
1	10/08/2013	P	N	N	N	N	N	N
2	12/08/2013	P	N	N	P	N	N	N

Deduplication process

ID no.	Specimen date	Test result						
		RSV	Influenza A	Influenza B	Parainfluenza	Rhinovirus	hMPV	Adenovirus
123456	10/08/2013	P	N	N	P	N	N	N

Figure 3-6. Example of RDS data before and after the deduplication process. Records 1-2 are from the same patient. Circles highlight created/merged variables.

In all analysis in this thesis, each RDS record is considered to be a single episode of infection.

3.2.3.3 Inclusion and exclusion criteria for RDS analyses

RSV test results were included in the RDS from 2010 onwards. All RDS RSV records (representing both positive and negative RSV results) from mid-2010 (calendar week 27) to mid-2014 (calendar week 26) were included in the analysis in this thesis, to cover four consecutive RSV seasons.

All records in RDS have one of three possible results for each of the respiratory viruses (RSV, influenza, parainfluenza, rhinovirus, hMPV, adenovirus):

1. Positive
2. Negative, or
3. Unknown.

All records with unknown test result for RSV (*n*=6,877/75,227, 9%) were excluded from analysis, as it was not possible to determine whether they tested positive for RSV nor whether they should be included in the denominator population containing all children tested for RSV.

A summary of the total number of RSV tests, the total number of RSV-positive tests and the positivity rate (number of RSV positives as a percentage of the total number of RSV tests) for each laboratory in the extracted RDS dataset is shown in Table 3-5. I identified a low number of RSV-positive records at the King's College Hospital laboratory. This was investigated by colleagues at PHE, who subsequently identified a problem with the reporting of positive test results from this laboratory. As this problem was unable to be rectified for my data extract, all records from the King's laboratory were excluded from analysis in this thesis.

Table 3-5. Total number of RSV tests, total number of RSV-positive tests in children <5 years of age in the RDS data extract, by laboratory.

Laboratory name	Total	Positive	Positivity rate
Birmingham	9,606	1,841	19%
Bristol	9,547	2,017	21%
Barts and the London	5,027	925	18%
Cambridge	2,733	542	20%
PHE Colindale	1,063	117	11%
King's	4,407	5	0%
Leeds	8,227	1,893	23%
Leicester	4,378	601	14%
Manchester	9,573	1,996	21%
Newcastle	4,790	829	17%
Nottingham	4,849	1,276	26%
Preston	346	93	27%
Royal Free Hospital	444	67	15%
Southampton	1,878	472	25%
Truro	579	237	41%
UCLH	903	130	14%

Figure 3-7 shows the final RDS study population for analysis. The data linkage work presented in Chapter 6 and Chapter 7 of this thesis uses a subset of this data (records from mid-2010 to mid-2012 only) which is described in the methodology sections of those chapters.

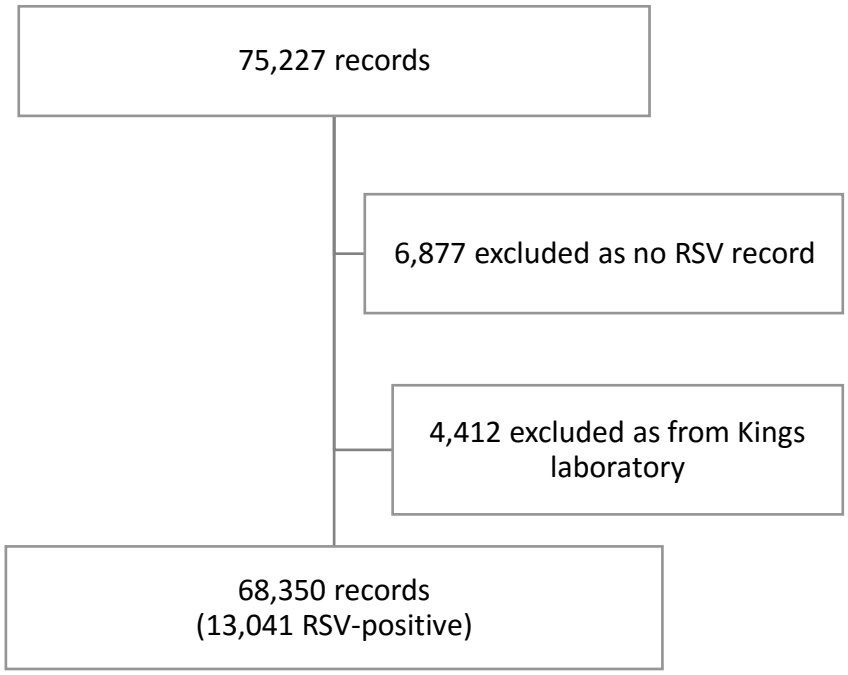

Figure 3-7. Flowchart of final RDS study population based on inclusion/exclusion criteria outlined in section 3.2.3.3.

3.2.3.4 Strengths and limitations of the RDS dataset

RDS – like SGSS – is used for surveillance purposes and monitoring trends in respiratory virus circulation. RDS does not contain data on RSV subtypes or strains. The CIDSC at PHE produces weekly reports based on the RDS data for influenza, tracking the weekly number of positive results and weekly positivity (proportion of positive results) by calendar week and age group and virus (127).

RDS is a powerful dataset for RSV research as it is the only routinely collected dataset in the UK to contain information on positive and negative RSV laboratory tests and therefore has a clear denominator population, unlike SGSS. This allows testing practices to be analysed, and variation in RSV positivity to be disentangled from variation in testing. In addition, as all RDS laboratories test for all major respiratory viruses – not just RSV – selection bias for RSV testing may be reduced.

A limitation of the dataset is that it does not cover all laboratories in England. It is not known where the tested population is drawn from, or where the samples are collected, as CIDSC does not record which hospitals (or primary healthcare providers) submit samples to RDS laboratories. However, patient identifiable information is included in the dataset which allows linkage with other data sources in order to enhance the information available for analysis. In this thesis RDS data has been linked to hospital admissions data, which has allowed me to estimate a population denominator for RDS (by identifying hospitals that submit data to RDS laboratories) and to explore laboratory-confirmed RSV-associated hospital admissions in detail.

3.2.3.5 Use of RDS data in this thesis

In this thesis I have used RDS data in two ways:

1. I have used RDS data to address Objective 1 of this thesis and analyse the epidemiology of RSV-positive laboratory-confirmed infections in children younger than 5 years in England, summarising the number of RSV positive episodes in SGSS by age, sex, birth month, calendar week and year. Comparing the distribution of key demographic characteristics of the RDS study population to the SGSS study population demonstrates the representativeness of RDS. I have also used the RDS to compare RSV testing practices, exploring differences by age, sex, month of birth and year. This work is presented in Chapter 4 of this thesis.

2. To address Objectives 3, 4 and 5 of this thesis, I used RDS data probabilistically linked to administrative hospital data by NHS number, date of birth, sex and post code. Objective 3 is addressed in Chapter 6 of this thesis, where I use the linked data to describe the epidemiology of laboratory-confirmed RSV-associated hospital admissions in children younger than 5 years in England in terms of age, primary diagnosis, calendar week and known clinical risk factors. In Chapter 7 I address Objective 4, determining risk factors for severe-RSV associated disease among laboratory-confirmed RSV-positive admissions. In Chapter 8I address Objective 5, estimating the total number of RSV-associated hospital admissions in young children in England (stratified by age, primary diagnosis, calendar week and known clinical risk factors).

3.3 Hospital admission data (Hospital Episode Statistics, HES)

Hospitals collect clinical and administrative information on all of their patients. This information is not only used locally to support patient care, but submitted to the Secondary Uses Service (SUS) data warehouse held by NHS Digital (128) where copies of the data can be extracted for one of two purposes: (1) for the Payment by Results (PbR) system (which allows hospitals be paid for the care that they deliver, taking into account the complexity of the patient's needs (129)), and (2) for research, in the form of Hospital Episode Statistics (HES) data.

The HES admitted patient care database contains routinely collected data on all admissions to all NHS hospitals in England, and all patients treated in independent hospital providers where their care was paid for by the NHS. At the end of each financial year, a final extract of the yearly data is produced (130). A wide range of information on each patient is available, including administrative information (e.g. admission and discharge dates), clinical information (i.e. diagnoses and operations), patient identifiable information (e.g. NHS number, date of birth, gender, ethnicity), and geographical information (e.g. where patients are treated and the area they live).

3.3.1 Available HES data extract

The HES extract available for analysis in this thesis is held at the PHE Respiratory Diseases Department, Colindale, London. The extract covers all respiratory admissions from HES years 2007/08 to 2011/12 (financial year, April to March, as explained above), containing all finished consultant episodes (FCEs)[2] with any mention of ICD-10 codes J00-J99 (Chapter X: Diseases of the Respiratory System)[3] in the HES diagnosis fields. More recent data was not available for this project.

[2] Defined in section 3.3.2.
[3] See section 3.3.3.1 for full details.

3.3.2 Identifying patients and admissions in HES data

HES records are created for each period of care in hospital under a single consultant, known as finished consultant episodes (FCEs). A patient's first episode is their admission episode. If the patient then moves to the care of a different consultant then a second FCE will be created, and this continues until the discharge episode, when the patient is discharged from hospital. A spell includes all FCEs between the admission and discharge episodes (Figure 3-8).

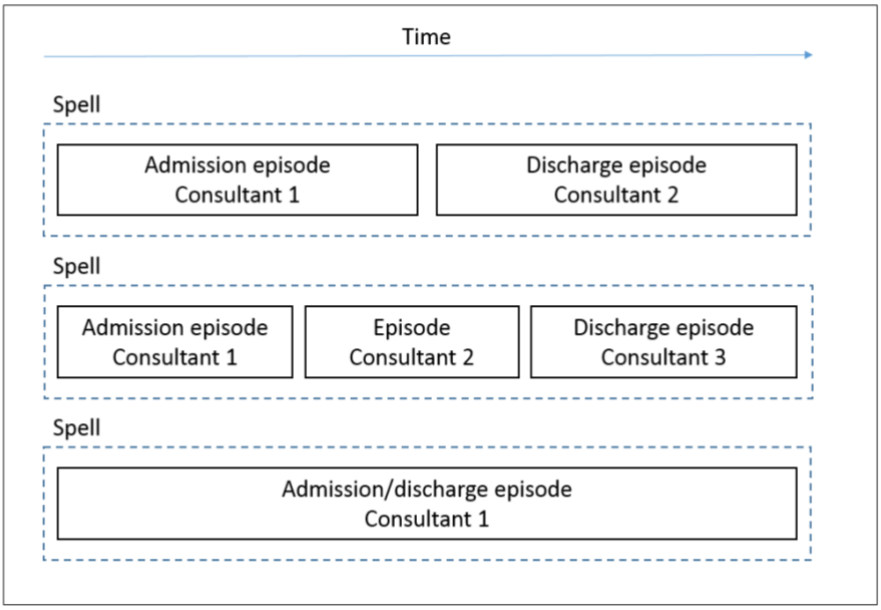

Figure 3-8. Diagram illustrating the concept of HES spells and episodes.

A longitudinal record for each patient can be created by linking individual records using the HESID pseudonymisation algorithm (131). This algorithm identifies records referring to the same patient (using patient identifiers such as date of birth, sex, postcode, NHS number) and labels them with a unique HESID number, allowing them to be tracked over time in the HES database (132). In all analysis in this thesis, individuals are identified using the HESID variable, and admission refers to a single spell (from admission to discharge).

The unit of analysis of HES data in this thesis is admissions. Therefore, data cleaning was carried out in order to link FCEs into admissions (using the HES_ID variable, admission and discharge dates, and episode start and end dates) so only one record per admission remained in the final dataset. Transfers from one hospital to another were counted as one admission. Similarly, if one admission in an individual began on the same day that another admission in that same individual finished, these were counted as a single admission.

3.3.3 HES data extraction

Data was extracted from 01/08/2007 to 31/07/2012 on all respiratory admissions in children aged younger than 5 years at the time of admission, in order to cover five consecutive RSV seasons. The HES extract contained 111 variables (the full HES data dictionary is available online [4]); Table 3-6 outlines some of the main variables utilised in this thesis.

Table 3-6. Variables extracted from HES on all records younger than 5 years of age at admission.

Variable	Description
HES ID	Unique identification number
Date of birth	In format: dd/mm/yyyy
NHS number	NHS number of the patient
Hospital number	Hospital number of the patient
Postcode prefix	First part of patient's postcode
Postcode suffix	Second part of patient's post code
Admission date	Date of admission
Episode start date	Date of start of FCE
Episode end date	Date of end of FCE
Discharge date	Date of discharge
Diag_01	Primary ICD-10 diagnosis code
Diag_02 – Diag_20	Secondary ICD-10 diagnosis codes
Oper_01 – Oper_24	Operative procedure codes (oper_01 is the main, most resource intensive procedure)
Provider code	Unique provider code that identifies the health care provider (e.g. NHS trust)

[4] http://content.digital.nhs.uk/media/1358/HES-Hospital-Episode-Statistics-HES-Admitted-Patient-Data-Dictionary/pdf/HES_Inpatients_DD_Sept10.pdf

3.3.3.1 ICD-10 diagnosis codes

Diagnoses are recorded in HES using the International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) coding system, devised by the WHO to classify diseases, symptoms, abnormal findings, complaints, social circumstances and external causes of injury or disease (133). Up to 20 diagnosis codes are allowed per HES episode. The first diagnosis field indicates the primary diagnosis. All admissions in the extract used in this thesis had one or more diagnosis codes from ICD-10 Chapter X: Diseases of the respiratory system in any of the 20 diagnosis fields available in HES (ICD-10 codes J00-J99). This chapter contains the diagnoses outlined in Table 3-7.

Table 3-7. ICD-10 codes in Chapter X: Diseases of the Respiratory System.

ICD-10 codes	Diagnosis
J00-J06	Acute upper respiratory infections
J09-J18	Influenza and Pneumonia
J20-J22	Other acute lower respiratory infections
J30-J39	Other diseases of the upper respiratory tract
J40-J47	Chronic lower respiratory diseases
J60-J70	Lung diseases due to external agents
J80-J84	Other respiratory diseases principally affecting the interstitium
J85-J86	Suppurative and necrotic conditions of lower respiratory tract
J90-J94	Other diseases of pleura
J95-J99	Other diseases of the respiratory system

3.3.4 Inclusion and exclusion criteria for HES analyses

The diagnoses considered relevant for identifying admissions potentially caused by RSV (given the extract I had access to) are outlined in Table 3-8. I also identified admissions with RSV-specific codes, in order to carry out sub-analyses on admissions identified as being caused by RSV. ICD-10 diagnosis codes used to identify RSV as the specific cause of each type of infection were: J21.0 (Acute bronchiolitis due to respiratory syncytial virus), J12.1 (Respiratory syncytial virus pneumonia) and J20.5 (Acute bronchitis due to respiratory syncytial virus). An additional diagnosis code used to identify RSV-associated admissions was B97.4 (Respiratory syncytial virus as the cause of diseases classified to other chapters). However, my HES extract does not include records which have diagnosis code B97.4 unless they also have an ICD-10 code from Chapter X in any of the other diagnosis fields.

Table 3-8. ICD-10 codes used to identify hospital admissions with a diagnosis of URTI, pneumonia, bronchitis, bronchiolitis or unspecified LRTI, and RSV-specific ICD-10 codes used to identify RSV-associated hospital admissions.

Diagnosis	ICD-10 code	RSV-specific ICD-10 code(s)
Acute upper respiratory infections (URTI)	J00-06	n/a
Pneumonia	J12-18	J12.1
Bronchitis	J20	J20.5
Bronchiolitis	J21	J21.0
Unspecified acute lower respiratory tract infection (unspecified LRTI)	J22	n/a

Some analyses in this thesis only considers the primary diagnosis of admission whereas some also considers secondary diagnoses (i.e. any diagnosis); this will be specified in the relevant methodology section for the analysis.

The analyses in this thesis uses different subsets of this HES extract. The work presented in Chapter 4 and Chapter 5 of this thesis uses all available HES data from mid-2007 to mid-2012. The data linkage work presented in Chapter 6 and Chapter 7 of this thesis uses a subset of this data (records from mid-2010 to mid-2012 only). The exact study population used in each analysis is described in the methodology sections of those chapters.

3.3.5 Strengths and limitations of the HES dataset

HES is a rich dataset that is very useful for research as it covers all hospital admissions in England. As the information in HES was collected for administrative purposes, and submitting high quality data is beneficial for the healthcare provider as it results in more accurate payment for the care they provide, the information is generally well completed. However, there are important limitations that need to be taken into account during analysis and interpretation of results. The major limitation of HES is the diagnostic coding – this is important for identifying admissions caused by RSV as RSV-specific codes can be used based on clinical diagnosis alone (which is not sufficient to distinguish respiratory viruses from one another, section 1.3.5) and, more often, non-specific codes are used (e.g. Bronchiolitis due to unspecified cause). Using a specific code mentioning RSV as the cause of the respiratory infection does not result in higher payment (the main purpose of the hospital data held by NHS is for the PbR system, explained at the beginning of this section), therefore there is also no financial incentive to determine the causal pathogen. Coding practices also vary between hospitals according to factors such as testing practices and clinician preferences. A specific limitation of the HES dataset that was available for my PhD work is that only respiratory admissions were available. Therefore, any admissions due to RSV but not coded as a respiratory admission (e.g. viral wheeze coded using ICD-10 codes B34.9 and R06.2) will not be included and the subsequent analysis may underestimate the burden of RSV in secondary care.

3.3.6 Use of HES data in this thesis

In this thesis, I have used HES data in three ways:

1. I have used HES data to address Objective 1 of this thesis and analyse the epidemiology of respiratory hospital admissions (potentially caused by RSV) in children younger than 5 years in England, summarising the number of respiratory hospital admissions from mid-2007 to mid-2012 by age, birth month, calendar week, year, primary diagnosis, length of stay and comorbidities. This work is presented in Chapter 4 of this thesis.

2. To address Objective 2 of this thesis, I have used HES data aggregated by calendar week in time-series modelling analysis which aims to estimate the total number of RSV-associated respiratory hospital admissions in children younger than 5 years in England by age and primary diagnosis, using SGSS data on laboratory-confirmed respiratory virus infections. For this analysis, HES data from calendar week 27 in 2007 to calendar week 28 in 2012 was used. This work is presented in Chapter 5.

3. To address Objectives 3, 4 and 5 of this thesis, I used HES data probabilistically linked to RDS data by NHS number, date of birth, sex and post code. Objective 3 is addressed in Chapter 6 of this thesis, where I use the linked data to describe the epidemiology of laboratory-confirmed RSV-associated hospital admissions in children younger than 5 years in England in terms of age, primary diagnosis, calendar week and known clinical risk factors. In Chapter 7 I address Objective 4, determining risk factors for severe-RSV associated disease among laboratory-confirmed RSV-positive admissions. In Chapter 8 I address Objective 5, estimating the total number of RSV-associated hospital admissions in young children in England (stratified by age, primary diagnosis, calendar week and known clinical risk factors).

Chapter 4

Describing RSV epidemiology using routinely collected data

Chapter 4 Describing RSV epidemiology using routinely collected data

4.1 Introduction

This Chapter addresses Objective 1 of this thesis: describe the epidemiology of RSV in children younger than five years in England using laboratory surveillance data and hospital administrative data. Section 4.2 of this Chapter contains the analysis of laboratory-confirmed RSV infections using the RDS and SGSS laboratory surveillance datasets, and section 4.3 contains the analysis of hospital administrative data on respiratory hospital admissions using the HES dataset.

I have presented the work in this Chapter at The Farr Institute International Conference 2015 (St Andrews, Scotland) and the 1st International Meeting on Respiratory Pathogens 2015 (Singapore). The analysis of RSV laboratory surveillance data (section 4.2) is published in the peer-reviewed journal *Epidemiology and Infection*:

> *RM Reeves, P Hardelid, R Gilbert, J Ellis, H Zhao, M Donati, R Pebody (2016) Epidemiology of laboratory-confirmed respiratory syncytial virus infection in young children in England, 2010-2014: the importance of birth month. Epidemiology and Infection, 144: 2049-56*

A copy of this paper is included in Appendix 1. I carried out all of the analysis in this paper (and that presented in this Chapter), and drafted the manuscript. All co-authors commented on the final manuscript draft.

4.2 Epidemiology of laboratory-confirmed RSV infections

4.2.1 Background

As explained in Chapter 1 (section 1.4), a robust evidence base of epidemiological data on RSV-associated disease is required for RSV vaccine impact studies to identify target populations for a potential future vaccine and to design optimal interventions. Detailed information on the burden of RSV on the health service, and particularly the distribution of RSV-associated illness by age, is required to determine the optimal target population(s) for a potential future vaccine.

Since the risk of RSV hospital admission peaks in the first few months of life and then declines, data by month of age is required in infants and young children. However, a further risk factor for RSV-associated hospitalisation in infancy has been shown to be birth near the beginning of the RSV season (21,103–106). Therefore, vaccine impact studies may consider the cost-benefit of targeting infants born in specific months of the year. However, the previous studies that demonstrate this association rely solely on clinical diagnoses, are based outside of the UK or are of small study populations (64). Relying on clinical diagnoses without laboratory confirmation of RSV to explore RSV epidemiology may lead to misclassification of outcome (i.e. non-RSV illness being classified as RSV illness, and vice versa), particularly at times that RSV is less common, and therefore bias any observed associations (section 1.3.7).

Describing the epidemiology of RSV using only laboratory-confirmed infections has the advantage of being highly specific for RSV infection. Testing for RSV is not routinely carried out in UK primary care nor in hospital (though testing in secondary care is much more common) and therefore only a minority of cases undergo laboratory testing (69). However, understanding the epidemiology of laboratory-confirmed RSV infection is important in generating baseline data on the demographic distribution of children who have tested positive for RSV, and it is the starting point from which estimates of the burden of RSV in England can be

calculated. The epidemiology of laboratory-confirmed RSV infection in infants and young children in England has not previously been described in detail.

This analysis uses the SGSS and RDS laboratory surveillance datasets (described in Chapter 3) to describe the epidemiology of laboratory-confirmed RSV infection in children younger than five years in England from 2010-2014, and examines the relationship between laboratory-confirmed RSV infection, age and birth month. The results are important in helping to determine target populations for a potential future vaccine programme, particularly if it is not feasible or cost-effective for a future RSV vaccine to be delivered throughout the year. In addition, I carried out this analysis to gain an understanding of the epidemiology of RSV in infants and young children as the starting point from which I estimated the secondary care burden of RSV in England using statistical modelling (Chapter 5 of this thesis) and data linkage methods (Chapter 8 of this thesis).

4.2.2 Methodology

4.2.2.1 Data extraction

This analysis uses the RDS and SGSS datasets described in Chapter 3. Data was extracted from both datasets from calendar week 27 in 2010 (mid-2010) to week 26 in 2014 (mid-2014) in all children younger than five years of age. Using this time period ensured that I could include all of the available RSV data in RDS, and allowed a full comparison (over the same time period) with SGSS data to explore the representativeness of RDS data (by comparing characteristics of the laboratory-confirmed infections). The extracted data included information on patient's date of birth, sex, date of sample, and from RDS also included information on whether the RSV laboratory test was positive for RSV, negative for RSV but positive for another respiratory virus, or negative for all viruses including RSV. Only one record per episode of infection was included in this analysis (as described in section 3.2).

4.2.2.2 Descriptive analysis

I summarised the total number of RSV tests (positive and negative) in the RDS extract, the number of positive RSV tests in the RDS extract and the number of positive episodes in the SGSS extract by age, month of birth and sex. I grouped admissions into one-year periods from 01/08/2007 (i.e. 2007/8 season refers to the period 01/08/2007-31/07/2008, etc.). The RSV positivity rate was calculated as the number of RSV positive tests divided by the total number of RSV tests in the RDS extract (stratified by age, month of birth, sex and year).

RSV season was defined as the first of two consecutive weeks in which the mean percentage of samples testing positive for RSV in the RDS was ≥10%, and the end of RSV season defined as the last of two consecutive weeks in which the mean percentage of samples testing positive for RSV was ≥10%. This method of defining the RSV season has been used in previous studies (134,135).

4.2.2.3 Statistical analysis

Multivariable logistic regression models were used to estimate the odds of a positive result by birth month, using the RDS extract. Infants born in January were used as the baseline group. Age group (0, 1, 2, 3 and 4 years), sex (male, female and unknown) and year (2010-2011, 2011-2012, 2012-2013, 2013-2014) were investigated as potential confounders and were added to the model in a forward stepwise manner. An interaction term between age group and birth month was included to examine whether the relationship between RSV-positivity and timing of birth in relation to RSV season varied with age. Likelihood ratio tests were used to determine whether the inclusion of a variable significantly improved the fit of the model; a likelihood ratio test *p*-value of <0.05 was considered significant. Robust standard errors were used to allow for clustering by laboratory – this method produces less biased standard errors than classical standard error measurements even if the observations from each laboratory are correlated [5] [6].

[5] http://www.stata.com/manuals13/rregress.pdf (Page 10)
[6] http://www.stata.com/manuals13/u20.pdf#u20.21Obtainingrobustvarianceestimates

4.2.3 Results

4.2.3.1 Seasonality of laboratory-confirmed RSV infections

In the RDS there was an average of 15,986 tests and 3,259 RSV positives per year in children younger than five years of age during the study period. Laboratory-confirmed RSV positivity showed a clear and consistent seasonal pattern (Figure 4-1, Figure 4-2); RSV season onset was in October each year (ranging from calendar week 41 to 43) during the study period and the end of RSV season ranged from January to March (week 4 to week 10). Overall testing peaked during December in each RSV season of the four years studied (Figure 4-2). The week with the highest proportion of positive RSV laboratory tests in children less than five years of age each season was week 48 in 2010-2011 (42% positive), week 1 in 2011-2012 (55% positive), and week 49 in 2012-2013 and 2013-2014 (52% and 55% positive, respectively). Of the four years of the study, the highest number of RSV laboratory tests was carried out during the 2010-2011 RSV season.

Of the RSV positive tests in the RDS during the study period (*n*=13,034), 16% were also positive for at least one other respiratory virus. Of the RSV negative tests in the RDS during the study period (*n*=50,793), 41% were positive for at least one other respiratory virus.

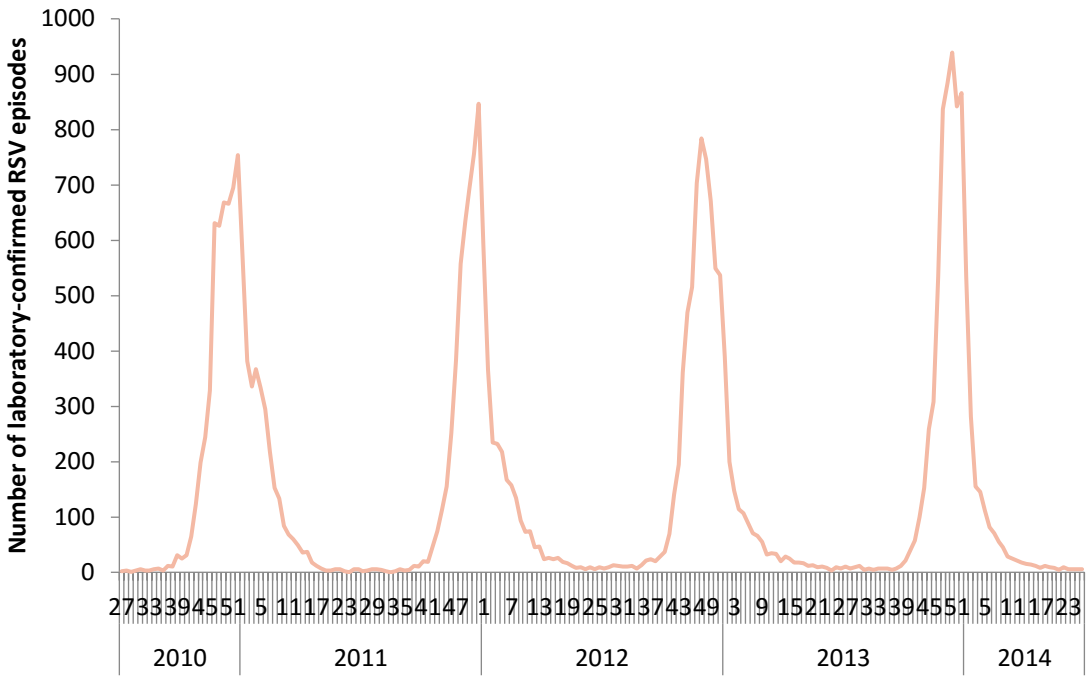

Figure 4-1. Total number of RSV-positive episodes in children aged <5 years in England recorded in SGSS from week 27 (2010) to week 26 (2014), per calendar week.

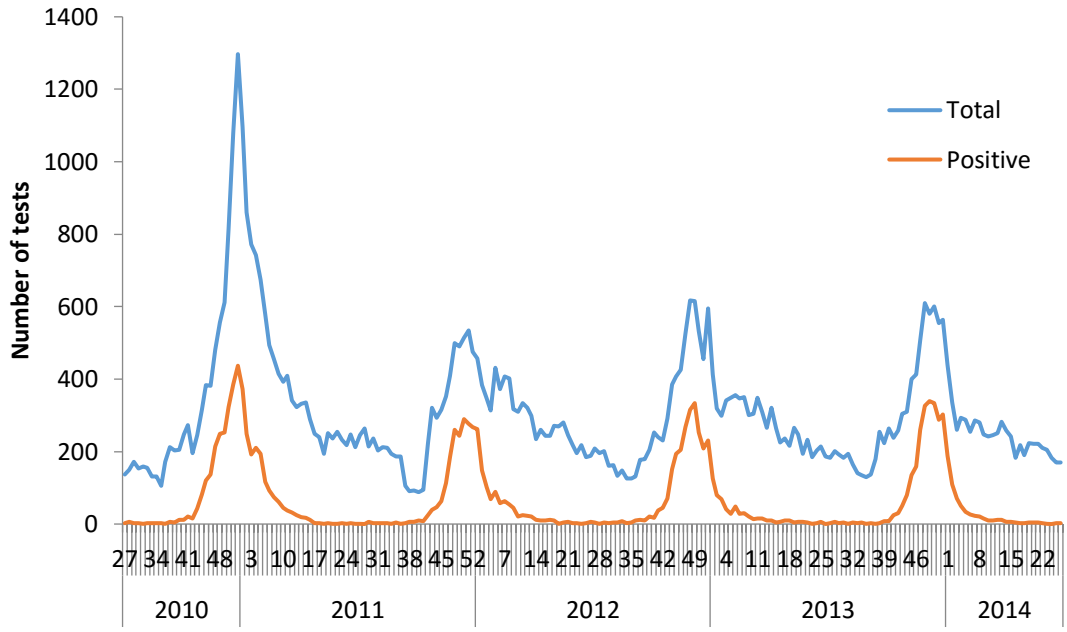

Figure 4-2. Total number (blue) and number of positive (red) RSV tests in children aged <5 years recorded in RDS from week 27 (2010) to week 26 (2014), per calendar week.

4.2.3.2 Similarity of laboratory-confirmed RSV infections in SGSS and RDS

The characteristics of the study population are shown in Table 4-1. The distribution of laboratory-confirmed RSV infections by age, birth month and sex were very similar in the RDS and SGSS extracts. In SGSS, 82% (25,283/30,669) of positive results in children younger than five years of age over the study period were in infants younger than one year of age; only 1% of tests (251/30,669) were in children aged four years**.** In RDS, 76% (9,933/13,034) of positive results were in infants younger than one year of age, compared to 2% (214/13,034) in children aged four years. The number of RSV positive tests in SGSS and RDS peaked at age one month (*n*=5,326 and *n*=2,198, respectively). 13% (3,297/25,283) of infants with a positive RSV test in their first year of life recorded in SGSS were born in September, 16% (4,088/25,283) were born in October and 15% (3,743/25,283) were born in November. This is comparable to the RDS dataset, where 12% (1,210/9,933) of infants with a positive RSV test in their first year of life were born in September, 16% (1,593/9,933) were born in October and 16% (1,565/9,933) were born in November (Figure 4-3). In both datasets the sex ratio (M:F) was 1.3:1.

Table 4-1. Total number of tests (RDS), number of RSV-positive tests (RDS), RSV positivity rate (RDS) and number of RSV-positive tests (SGSS) in children <5 years of age from week 27 in 2010 to week 26 in 2014 by sex, age, birth month if tested for RSV in the first year of life, and season.

	Respiratory DataMart System			SGSS
	Total tests N (%)	RSV positive N (%)	RSV positivity rate	RSV positive
Total	63,827	13,034	20%	30,669
Sex				
Male	31,278 (49%)	6,165 (47%)	20%	17,050 (56%)
Female	23,577 (37%)	4,848 (37%)	21%	13,332 (43%)
Unknown	8,970 (14%)	2,021 (16%)	23%	287 (1%)
Sex ratio (M:F)	1.3:1	1.3:1		1.3:1
Age				
<3 months	20,467 (32%)	4,982 (38%)	24%	12,641 (41%)
3-5 months	9,384 (15%)	2,423 (19%)	26%	6,526 (21%)
6-11 months	11,712 (18%)	2,528 (19%)	22%	6,116 (20%)
1 year	10,439 (16%)	1,815 (14%)	17%	3,703 (12%)
2 years	4,905 (8%)	670 (5%)	14%	922 (3%)
3 years	3,929 (6%)	402 (3%)	10%	510 (2%)
4 years	2,991 (5%)	214 (2%)	7%	251 (1%)
Birth month[7]				
January	3,271 (8%)	586 (6%)	18%	1,384 (5%)
February	2,774 (7%)	404 (4%)	15%	994 (4%)
March	2,970 (7%)	436 (4%)	15%	1,066 (4%)
April	2,818 (7%)	448 (5%)	16%	1,132 (4%)
May	3,038 (7%)	525 (5%)	17%	1,344 (5%)
June	3,056 (7%)	646 (7%)	21%	1,519 (6%)
July	3,173 (8%)	700 (7%)	22%	1,906 (8%)
August	3,484 (8%)	906 (9%)	26%	2,484 (10%)
September	3,833 (9%)	1,210 (12%)	32%	3,297 (13%)
October	4,626 (11%)	1,593 (16%)	34%	4,088 (16%)
November	4,575 (11%)	1,565 (16%)	34%	3,743 (15%)
December	3,945 (9%)	914 (9%)	23%	2,326 (9%)
Season				
2010-2011	19,751 (31%)	4,103 (31%)	21%	8,327 (27%)
2011-2012	14,804 (23%)	2,919 (22%)	20%	7,228 (24%)
2012-2013	15,021 (24%)	3,013 (23%)	20%	7,495 (24%)
2013-2014	14,251 (22%	2,999 (23%)	21%	7,619 (25%)

[7] Birth month if <1 year old only. Percentage denominator is the total number in infants aged <1 year [i.e. total tests (RDS)=41 563, RSV positive (RDS) = 9933 and RSV positive (SGSS)=25 283].

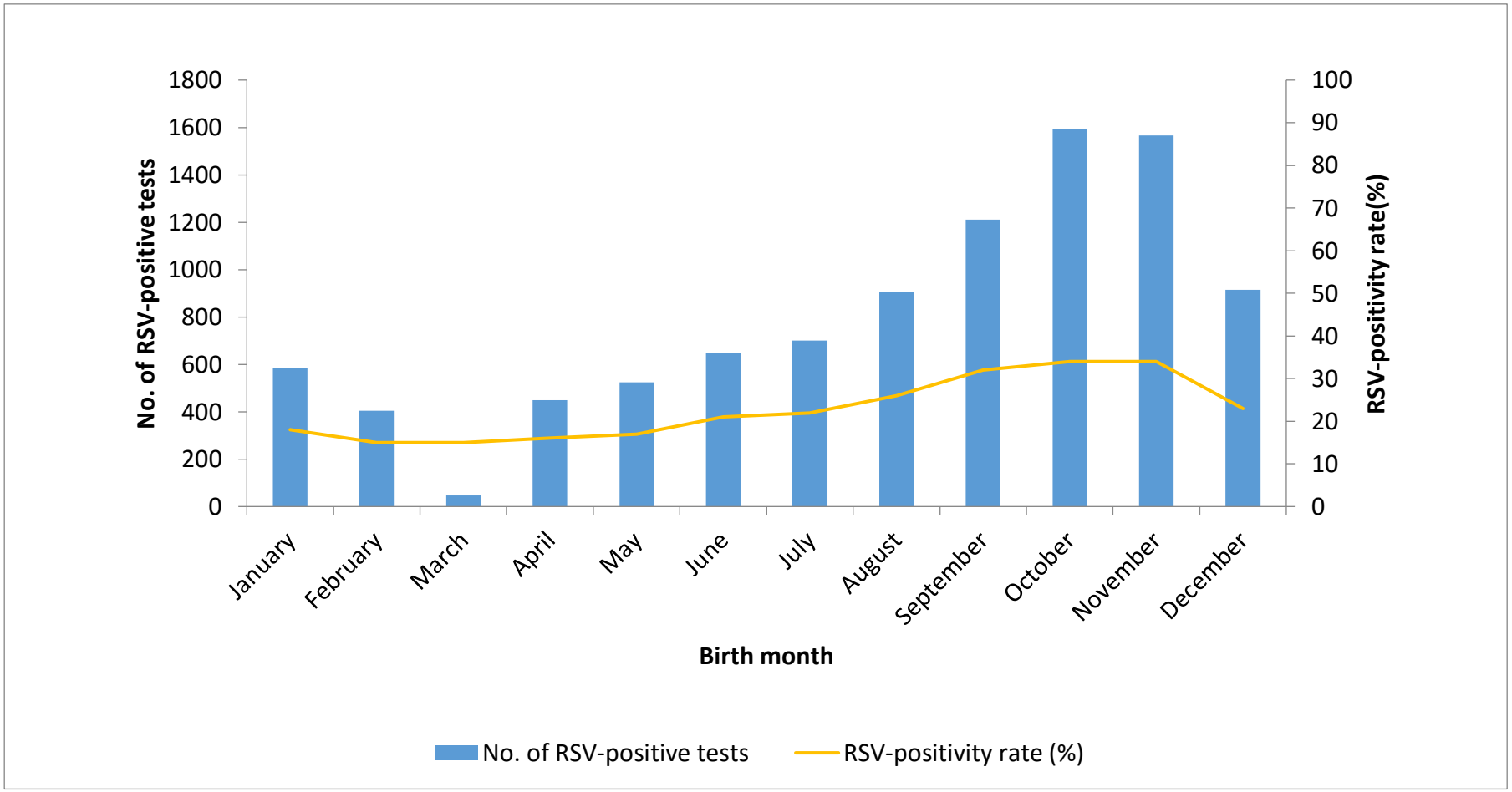

Figure 4-3. No. of RSV-positive tests (blue) and RSV-positivity rate (%) (orange) in children <1 year of age in the RDS extract, by month of birth.

4.2.3.3 Age distribution of laboratory-confirmed RSV infections

The RDS results demonstrate that both the total number of tests and the number of positive tests decreased with increasing age (Figure 4-4). 47% of tests (29,851/63,827) and 57% (7,405/13,034) of positives in children less than five years of age over the study period were in infants less than six months of age. The number of RSV positives peaked at age one month (*n*=2,198). Infants aged 1, 2 and 3 months had the highest rate of RSV positivity: 29% (2,198/7,551) and 29% (1,507/5,194) and 27% (996/3,697) tested positive for RSV, respectively. The highest number of tests was in infants aged less than one month (*n*=7,722).

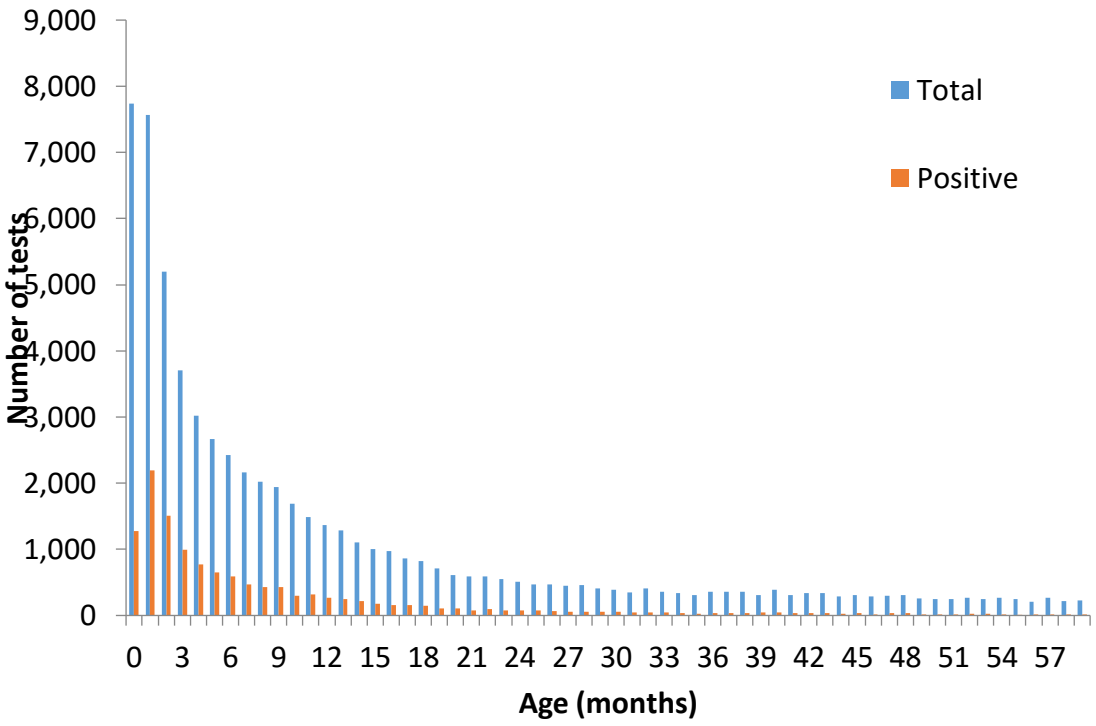

Figure 4-4. Total number (blue) and number of positive (red) RSV tests in children <5 years of age recorded in RDS from week 27 in 2010 to week 26 in 2014, by age in months.

4.2.3.4 The effect of birth month on RSV-positivity

Infants younger than one year of age who were born in September, October and November had the highest number and proportion of positive test results (Table 4-1, Figure 4-3). The best fitting multivariable logistic regression model to estimate the odds of a positive result (if tested for RSV) included sex, calendar year, age group and birth month as well as an age group:birth month interaction term. Infants younger than one year of age born in September (odds ratio (*OR*)=2.1, 95% CI 1.7, 2.7), October (*OR*=2.4, 95% CI 2.1, 2.8) or November (*OR*=2.4, 95% CI 2.1, 2.7) had the highest odds of a positive result if tested for RSV in the first year of life compared to infants born in January (Figure 4-5). The effect of birth month on the odds of a RSV positive test result decreased with increasing age (Figure 4-6). For example, children aged four years born in September and January had the same odds of a positive result (both *OR*s=0.4, 95% CI 0.3, 0.6).

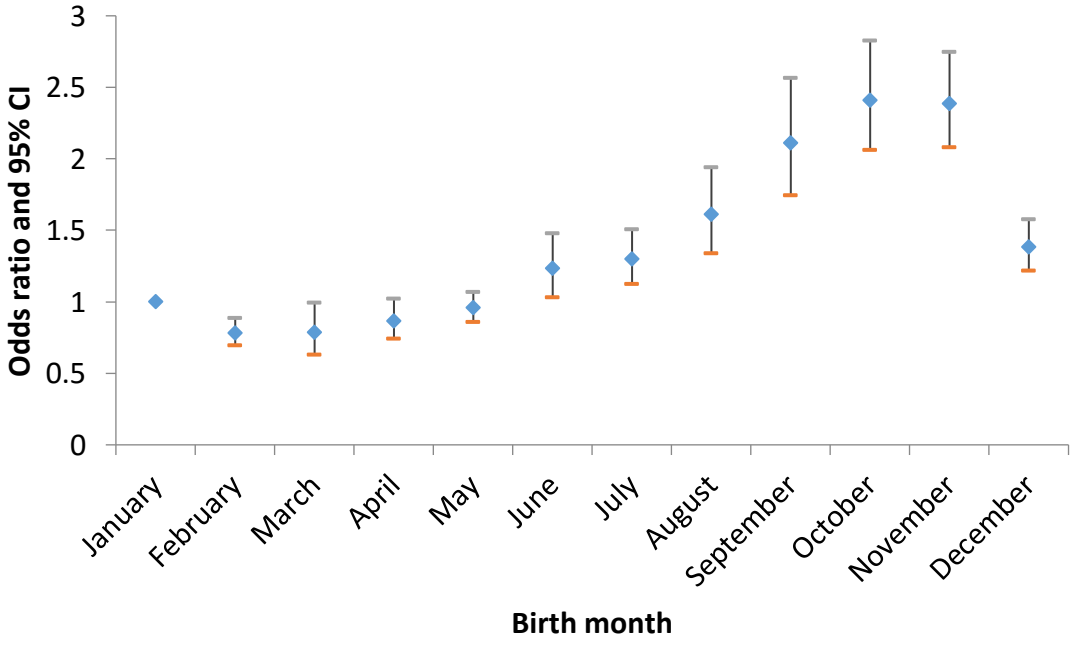

Figure 4-5. Odds ratios and 95% confidence intervals (CI) from final multiple logistic regression model using RDS data to compare odds of a positive result if tested for respiratory syncytial virus by birth month (showing results for infants aged <1 year only). Infants born in January are the baseline group.

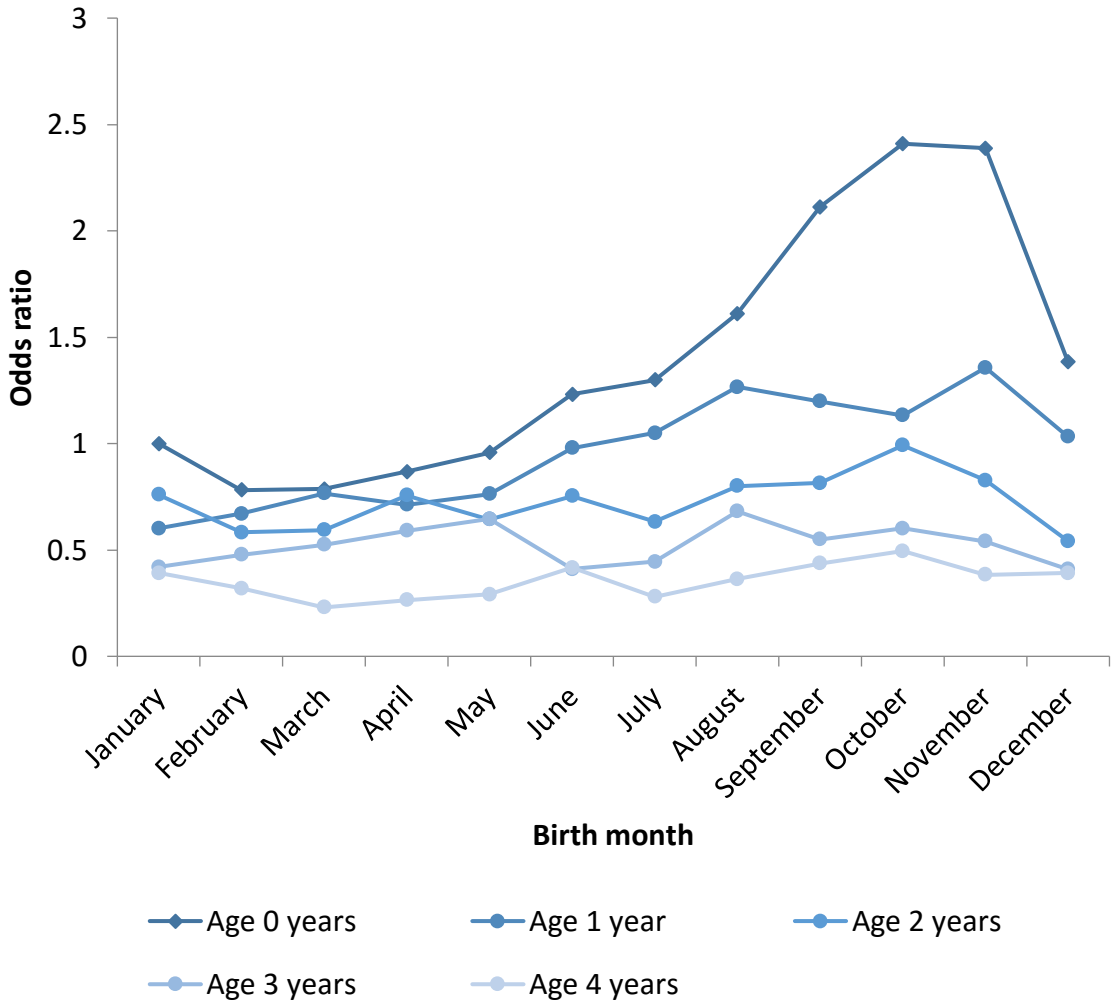

Figure 4-6. Odds ratios from final multiple logistic regression model using RDS data to compare odds of a positive result if tested for respiratory syncytial virus by birth month, stratified by age in years. Infants born in January are the baseline group.

4.2.4 Discussion

This analysis shows that 62% of laboratory-confirmed RSV infections in children younger than 5 years in England recorded in SGSS and 57% in RDS from 2010 to 2014 were in infants aged younger than 6 months. 41% of SGSS records and 38% of RSV-positive RDS records were in children younger than 3 months. In both datasets, there was a peak in RSV-positive tests in infants aged 1 month. RSV circulation was very consistent in timing each year, with a 3-week range in season onset over the study period. Moreover, infants born from September to November (near the beginning of the onset of RSV circulation) had significantly increased odds of a positive result if tested for RSV during the first year of life. The consistency between results from the RDS and SGSS data suggest generalisability of the RDS results to the long established national surveillance system, SGSS.

My analysis found RSV circulation to be highly consistent in timing each year, with only a 3-week range in season onset over the study period. The large peak in infants being tested for RSV during 2010–2011 can be attributed to the intense influenza A(H1N1)pdm09 season during this first post-pandemic winter period, as the RDS holds records of the results of samples tested simultaneously for multiple respiratory viruses including influenza and RSV (119,136). The large number of tests performed outside of the RSV season can also be attributed to this simultaneous testing for other viruses.

The considerable number of tests and RSV-positive results in infants aged <6 months and the peak in number and percentage of positive RSV tests in infants aged 1 month is consistent with existing literature that reports age <6 months as a significant risk factor for severe RSV infection (17) and a peak in RSV bronchiolitis at age 1 month (110,137). Young infants (<6 months) have been a high priority for future vaccines, particularly due to the serious complications and subsequent morbidity that can occur following RSV infection in early life (91). (138). My results highlight the importance of developing optimal strategies to prevent disease in young infants, particularly those aged <3 months, and indicate that a maternal vaccine would need to induce protection in newborn infants for at least three

months in order to have a substantial impact on the burden of RSV-associated illness in young infants.

That birth during months September to December was associated with increased odds of a positive result if tested for RSV in the first year of life in England supports the results of previous studies that show birth around the beginning of the RSV season is a risk factor for RSV-associated hospitalisation (21,103–105). These previous studies investigating month of birth as a risk factor for RSV-associated hospitalisation were limited by the use of clinical diagnosis or sample size, and all were based outside of the UK. The largest study was carried out in the United States and suggests children born during December and January had a two- and threefold higher risk, respectively, of RSV-confirmed hospitalisation during infancy than those born in July, though these cases were identified by International Classification of Diseases (Ninth Revision) codes only (103). The increased risk of severe RSV infection in infants born close to the beginning of the RSV season is likely due to these infants having a longer exposure to RSV at a young age in combination with lower levels of maternal antibodies during the beginning of the RSV seasons (17). The birth months with the highest risk of severe RSV infection in infancy varies between countries due to differences in the timing of RSV season, which highlights the importance of country-specific epidemiology studies including birth month as a potential risk factor when analysing severe RSV infection. If a potential future vaccine could only be delivered during certain times of the year in the UK, my results suggest that it would need to protect babies born from September to December.

4.2.4.1 Strengths and limitations of this analysis

This analysis is the first to describe the epidemiology of laboratory-confirmed RSV infection in children younger than five years of age in England in detail and on a national scale. A strength of this analysis is the use of laboratory-confirmed RSV infection rather than clinically diagnosed RSV. This eliminates potential bias in associations between patient characteristics and RSV infection due to

misclassification of outcome that may occur when using clinical diagnosis alone (as specific respiratory viruses cannot be differentiated clinically). However, a limitation is that no clinical information on the individuals tested was available in either dataset. The majority of RSV records in the RDS and SGSS are from hospitalised patients (119), and it is likely that only severe or complex cases requiring hospital admission will require laboratory confirmation of RSV infection as it would usually be unnecessary to investigate mild infection for the presence of RSV. However, without clinical information it is not possible to confirm whether this assumption is correct. In addition, most tests in both RDS and LabBase2 were carried out on young children (<1 year). It is therefore possible that differences in testing according to age means that RSV infection is less likely to be confirmed in older children. The linkage of hospital admission data and laboratory surveillance data (presented in Chapter 6 of this thesis) allows analysis of the potential association between clinical presentation, patient characteristics, the probability of being tested, and RSV positivity.

4.2.4.2 Implications for future analyses in this thesis

This analysis is the first to explore the representativeness of RDS data on RSV to the national laboratory surveillance database, SGSS. The distribution of key demographic characteristics of the RDS and SGSS study population are similar, which suggests generalisability of the RDS results to the long established national surveillance system. As the RDS dataset contains records of both positive and negative RSV tests it is the preferential dataset for analysis of RSV-associated hospital admissions (see section 3.2.3.4). Therefore, the representativeness of RDS is an important finding that permits the use of RDS instead of SGSS in this thesis when required.

This analysis highlights the importance of young age (<6 months) and birth near the beginning of the RSV season as a risk for laboratory-confirmed RSV infection in infants. In this thesis, I will further explore the relationship between RSV-positivity, age and birth month through linked laboratory surveillance and hospital admission

data (Chapter 6 of this thesis). In addition, this analysis is also important in highlighting potential biases in RSV testing. As I have identified a potential testing bias towards infants (i.e. young infants may be more likely to be tested), I will further explore this potential bias during the analysis of the linked laboratory surveillance and hospital admission data.

4.3 Epidemiology of RSV-associated hospital admissions

4.3.1 Background

RSV is a major cause of hospital admission for acute respiratory tract infection (RTI) in infants and young children (86). It has been estimated that between 25 and 40% of infants and young children will develop lower respiratory tract infection (LRTI) during their first RSV infection (50). One of the most common LRTIs caused by RSV is bronchiolitis. Some studies suggest that up to 80% of bronchiolitis cases in infants are caused by RSV (15,52), as well as around 20% of childhood pneumonia cases (52).

The HES dataset contains information on all admissions to NHS hospitals in England (see section 3.3). However, specific RSV diagnoses are significantly under-recorded in HES as laboratory testing is not routine for all children presenting to hospital with symptoms of respiratory infection and hospital coding for RSV-associated illness commonly uses non-specific diagnosis codes (139). Furthermore – and despite laboratory testing being the only method of accurately diagnosing RSV infection – RSV-specific diagnosis codes may be used based on clinical diagnosis alone. Subsequently, RSV-specific diagnosis codes have low specificity for RSV infection, and there is potential for biases in RSV-specific diagnostic coding. In order to estimate the secondary care burden of RSV in children younger than five years of age in England (the overall aim of this thesis), it is important to first gain an understanding of: (a) the total number of hospital admissions potentially caused by RSV, and (b) the potential under-recording of RSV in diagnostic coding.

In this chapter I use the HES database to describe the epidemiology of hospital admissions for RTI in children younger than five years of age in England (in terms of calendar week, age, diagnosis, and risk group), and calculate incidence rates for RTI admissions, stratified by diagnosis.

4.3.2 Methodology

4.3.2.1 Data extraction

This analysis used the HES dataset described in Chapter 3 of this thesis. Data was extracted on all admissions with any URTI or LRTI (bronchiolitis, bronchitis, pneumonia, unspecified LRTI) diagnosis (section 3.3.4, Table 3-8) beginning between 01/08/2007 and 31/07/2012 in children younger than 5 years of age (at admission).

4.3.2.2 Descriptive analysis

I summarised the average annual number of hospital admissions for respiratory tract infections (RTIs), and calculated the percentage of admissions with an RSV-specific diagnosis code. I stratified admissions by RTI diagnosis – first by primary diagnosis, then by any mention of the RTI diagnosis in any of the 20 HES diagnosis fields (Chapter 3). I defined RSV-associated hospital admissions using the RSV-specific ICD-10 diagnosis codes defined previously in section 3.3.4 (Table 3-8). I grouped admissions into one-year periods from 01/08/2007 (i.e. 2007/8 season refers to the period 01/08/2007-31/07/2008, etc.). I calculated incidence rates for RTI admissions (stratified by diagnosis) using Office for National Statistics (ONS) mid-year population estimates for England by age group (<1 year, 1-4 years) (140,141) – the ONS mid-year population estimate for 2007 was used for the 2007/8 season and the ONS mid-year population estimate for 2008 was used for the 2008/9 season, etc.

I described RTI admissions over time by age, risk group and length of stay (from admission to discharge), for each diagnosis and for RSV-associated hospital admissions. High-risk children were those with any code in their longitudinal HES record of respiratory admissions indicating chronic lung disease (including bronchopulmonary dysplasia), congenital heart disease, prematurity, neurological disorders and immunodeficiency – these risk groups were chosen after a thorough search of the literature, and the list of ICD-10 codes used to identify high-risk

children (shown in Table 4-2) based on code lists developed previously to identify groups at risk of severe respiratory illness.

Table 4-2. ICD-10 codes used to identify high-risk children in HES.

Risk Factor	ICD-10 code
Chronic lung disease including bronchopulmonary dysplasia	J41-J44, J47 – chronic lower respiratory disease Q30-Q34 – congenital malformations of the respiratory system P27 – chronic respiratory disease originating in the perinatal period P28 – other respiratory conditions originating in the perinatal period E84 – cystic fibrosis
Congenital heart disease	Q20-Q26, Q89.3 – congenital malformations of the circulatory system I00-I28, I31-I39, I41, I42.0-I42.5, I42.7-I42.9, I43.0, I43.1, I43.2-I43.8, I44.1-I44.7, I45.1-I45.9, I46-I51, I52.8, I70-I71, I72.1-I72.4, I72.8, I72.9, I73-I77, I79.0, I79.1, I79.8, I81-I82, I98-I99, M03.6, N08.8, Q27, Q28, S26, T82.0-T82.3, T82.5-T82.9, T86.2, Y60.5, Y61.5, Y62.5, Y84.0, Z45.0, Z94.1, Z95 – other chronic heart conditions[8]
Prematurity	P07 – disorders related to short gestation and low birth weight
Neurological disorders	G80-G83 – cerebral palsy P10, P21.0, P52, P57, P90, P91.1, P91.2, P91.6 – perinatal conditions Q00-Q07, Q10.4, Q10.7, Q11-Q12, Q13.0-Q13.4, Q13.8, Q13.9, Q14-Q16, Q75.0, Q75.1, Q85, Q86.0, Q86.1, Q86.8, Q90-Q93, Q95.2, Q95.3, Q97, Q99 – congenital anomalies of neurological or sensory systems
Immunodeficiency[9]	C00-C97, D37-D39, D40 – malignant neoplasms [affecting the immune system] B20-B24 – human immunodeficiency virus [HIV] disease Z94 – transplanted organ and tissue status Z85 – personal history of malignant neoplasm D56.1, D57, D61, D70-D73, D76, D80-D84, K90.0 – conditions affecting the immune system

[8] ICD-10 codes for chronic heart conditions from Hardelid et al. (2014) http://bmjopen.bmj.com/content/4/8/e005331.long

[9] ICD-10 codes from Cromer et al. (2014) http://www.sciencedirect.com/science/article/pii/S0163445313003733

4.3.3 Results

4.3.3.1 Summary of RTI admissions

In total, there was an average of 143,151 hospital admissions for URTI, bronchiolitis, unspecified LRTI, pneumonia and bronchitis in children younger than five years of age in England per year, from 2007-2012 (Table 4-3). There was an average of 30,083 hospital admissions with any mention of bronchiolitis in children younger than 5 years of age in England per year; 25% (7,478/30,083) of these admissions had a code indicating RSV as the cause of disease. There was an average of 81,709 hospital admissions with any mention of URTI; less than 1% (143/81,709) of these hospital admissions had a code indicating RSV as the reason for admission.

Table 4-3. Number of hospital admissions for LRTI (bronchiolitis, pneumonia, bronchitis, unspecified LRTI) and URTI in children younger than five years of age in England, from 2007-2012.

Diagnosis (any)	Number of hospital admissions		Number of hospital admissions with RSV-specific code		
	Total	Average per year	Total	Average per year	% of total admissions with RSV code
URTI	408,546	81,709	713	143	0.2%
Bronchiolitis	150,417	30,083	37,388	7,478	25%
Unspecified LRTI	97,122	19,424	1,187	237	1%
Pneumonia	56,516	11,303	2,233	447	4%
Bronchitis	3,161	632	337	67	11%
Total	715,762	143,152	41,858	8,372	6%

URTI was the most common diagnosis in children aged 1-4 years, and bronchiolitis the most common diagnosis in children aged <1 year (Table 4-4). In the majority of cases (72-94%), hospital admissions with any mention of RTI had the RTI diagnosis as the primary diagnosis (Table 4-5).

Table 4-4. Average annual hospital admissions in children <5 years of age in England with any mention of the RTI diagnosis of interest or primary diagnosis of the diagnosis of interest, stratified by age group, from 2007-2012.

Diagnosis	Average annual number of hospital admissions			
	Any mention of the diagnosis		Primary diagnosis[10]	
	<1 year	1-4 years	<1 year	1-4 years
URTI	25,278 (41%)	56,431 (69%)	21,631 (40%)	46,934 (68%)
Bronchiolitis	27,743 (46%)	2,340 (3%)	25,967 (48%)	2,079 (3%)
Unspecified LRTI	4,756 (8%)	14,668 (18%)	3,794 (7%)	11,912 (17%)
Pneumonia	2,832 (5%)	8,472 (10%)	2,033 (4%)	7,552 (11%)
Bronchitis	312 (<1%)	320 (<1%)	273 (<1%)	276 (<1%)
Total	60,921	82,231	53,698	68,753

[10] The admissions with a primary diagnosis of URTI, bronchiolitis, unspecified LRTI, pneumonia and bronchitis are a subset of the admissions with any mention of the aforementioned diagnoses.

Table 4-5. Percentage of average annual respiratory hospital admissions in children <5 years of age in England with the RTI diagnosis as the primary diagnosis, stratified by age, from 2007-2012.

Diagnosis	Percentage of respiratory hospital admissions with the RTI diagnosis as the primary diagnosis	
	<1 year	1-4 years
URTI	86%	83%
Bronchiolitis	94%	89%
Unspecified LRTI	80%	81%
Pneumonia	72%	89%
Bronchitis	88%	86%

4.3.3.2 Admission rates

From 2007/8 to 2011/12, the annual rate of hospital admissions with any mention of bronchiolitis increased from 36.2/1,000 to 44.7/1,000 children <1 year of age (Table 4-6). Bronchiolitis admissions in children aged 1-4 years remained relatively constant with a rate of 1.0 to 1.3 per 1,000 children. The annual rate of hospital admissions with any mention of unspecified LRTI increased from 5.8 to 9.2/1,000 children 1-4 years of age, and increased from 6.2 to 7.8/1,000 children <1 year of age from 2007/8 to 2011/12. Admission rates for other diagnoses remained relatively constant over the study period. Average annual admission rates for each RTI diagnosis is shown in Figure 4-7.

Table 4-6. Annual hospital admission rates for URTI, bronchiolitis, unspecified LRTI, pneumonia and bronchitis per 1,000 children in England.

Hospital admission rate (per 1,000 children)						
Diagnosis (any)	**Age group**	**2007/8**	**2008/9**	**2009/10**	**2010/11**	**2011/12**
URTI	**<1y**	37.2	38.5	38.1	37.7	38.3
	1-4y	28.3	29.9	29.4	29.3	30.2
Bronchiolitis	**<1y**	36.2	38.8	42.2	46.1	44.7
	1-4y	1.0	1.1	1.3	1.4	1.3
Unspecified LRTI	**<1y**	6.2	6.9	7.2	7.5	7.8
	1-4y	5.8	7.1	7.6	8.3	9.2
Pneumonia	**<1y**	4.0	4.1	4.5	4.4	4.3
	1-4y	4.0	4.1	4.6	4.8	4.6
Bronchitis	**<1y**	0.5	0.5	0.4	0.5	0.5
	1-4y	0.2	0.1	0.1	0.2	0.2

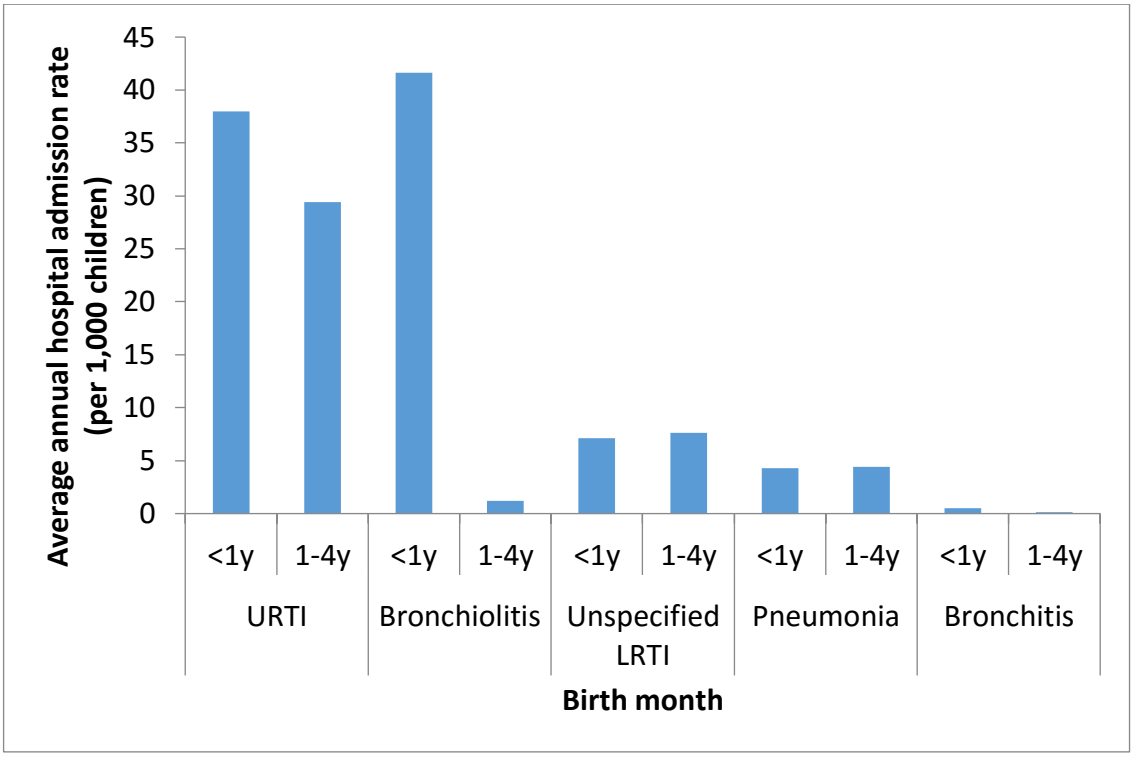

Figure 4-7. Average annual hospital admission rate (per 1,000 children), by diagnosis and age group.

4.3.3.3 Seasonal variations in diagnoses

Figure 4-8 shows the weekly number of hospital admissions with any mention of URTI, bronchiolitis, pneumonia, bronchitis or unspecified LRTI. Hospital admissions for bronchiolitis are markedly seasonal, mirroring the pattern of laboratory-confirmed RSV episodes (section 4.2.3.1) (Figure 4-8). Hospital admissions for pneumonia had a very similar seasonal pattern to hospital admissions for unspecified LRTI, with peaks also occurring at the same time as the peaks in laboratory-confirmed RSV episodes (in SGSS).

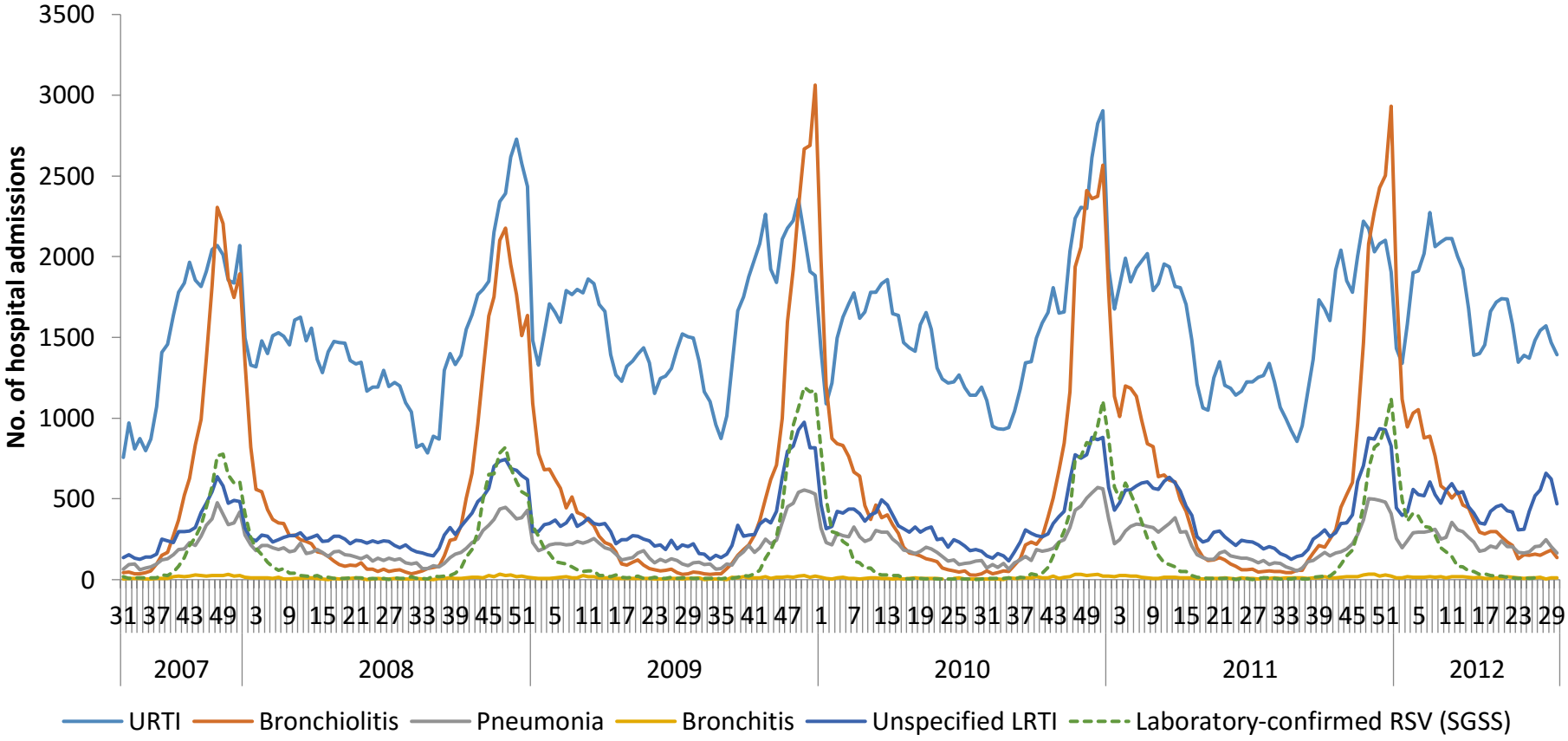

Figure 4-8. Weekly number of hospital admissions in children <5 years of age in England with any diagnosis of URTI, bronchiolitis, pneumonia, bronchitis or unspecified LRTI. Weekly number of laboratory-confirmed RSV infections in children <5 years of age in England from SGSS.

4.3.3.4 Age at admission

There were substantial differences in diagnosis by age (Figure 4-9). Hospital admissions with any mention of bronchiolitis or URTI peaked at age 1 month. 65% (98,239/150,417) of hospital admissions in children aged younger than 5 years with any mention of bronchiolitis were in children younger than 6 months old.

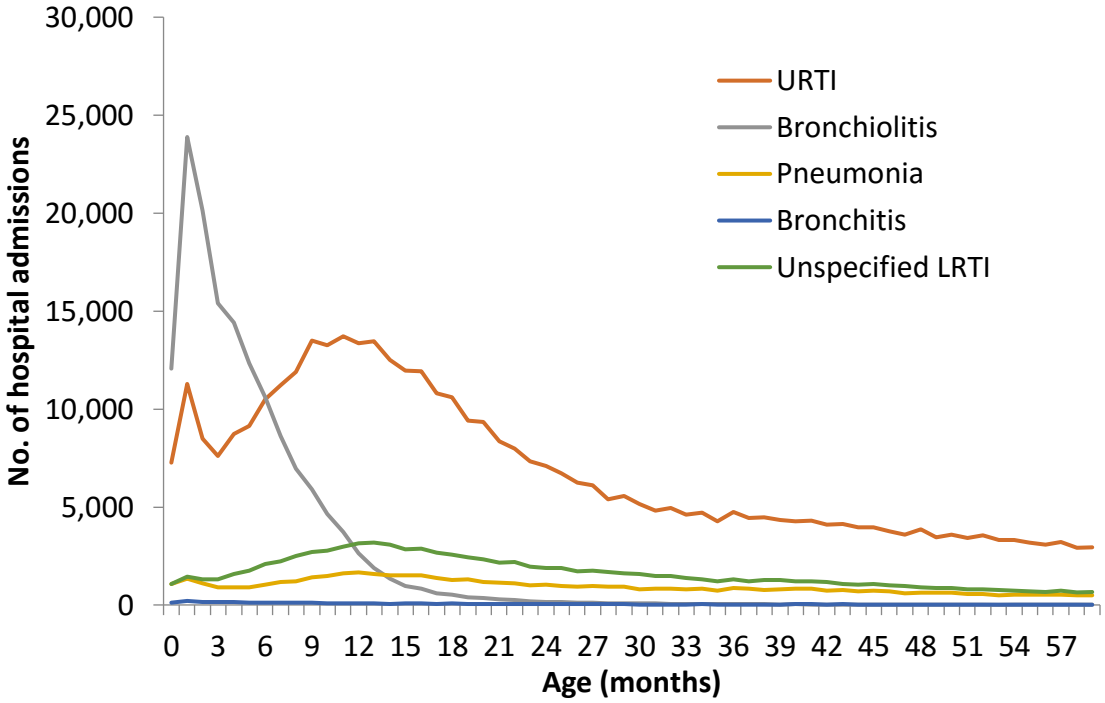

Figure 4-9. Total number of hospital admissions with any mention of URTI, bronchiolitis, pneumonia, bronchitis or unspecified LRTI, from 2007-2012 in England, by age in months.

The number of RTI admissions, as well as the number of RTI admissions with an RSV-specific ICD-10 diagnosis code, peaked at age 1 month (Figure 4-10). The percentage of RTI admissions with an RSV-specific code decreased with increasing age; 33% of RTI admissions in children aged younger than 1 month had an RSV-specific diagnosis code, compared to <1% of admissions in children aged 20 months or older (Figure 4-11).

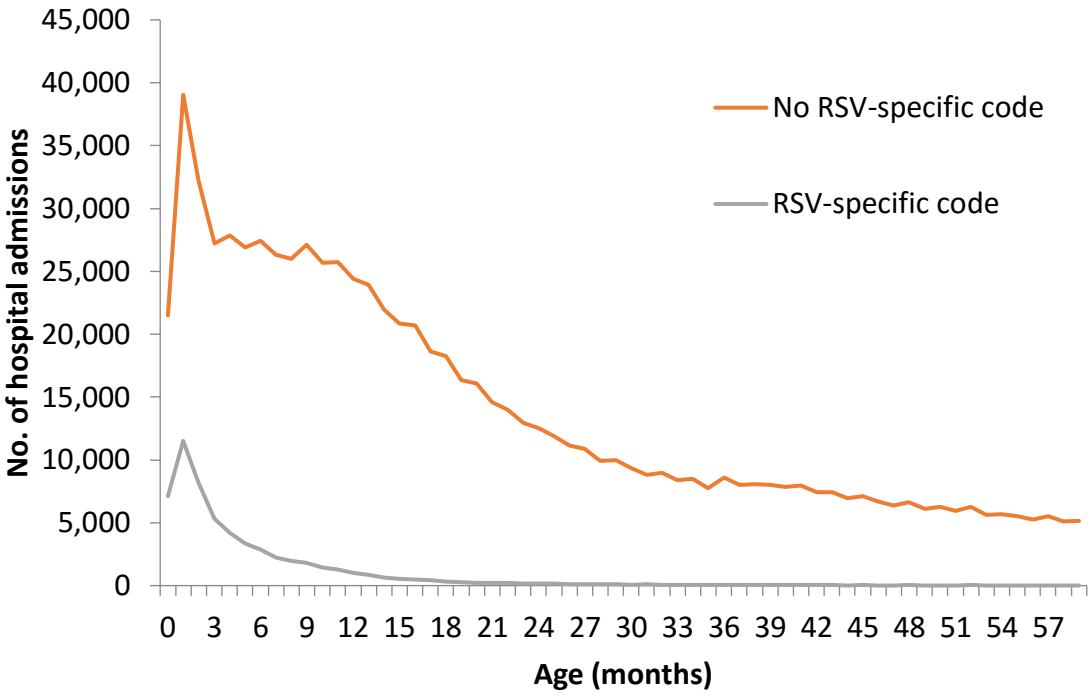

Figure 4-10. Total number of RTI hospital admissions with and without an RSV-specific ICD-10 diagnosis code, from 2007-2012 in England, by age in months.

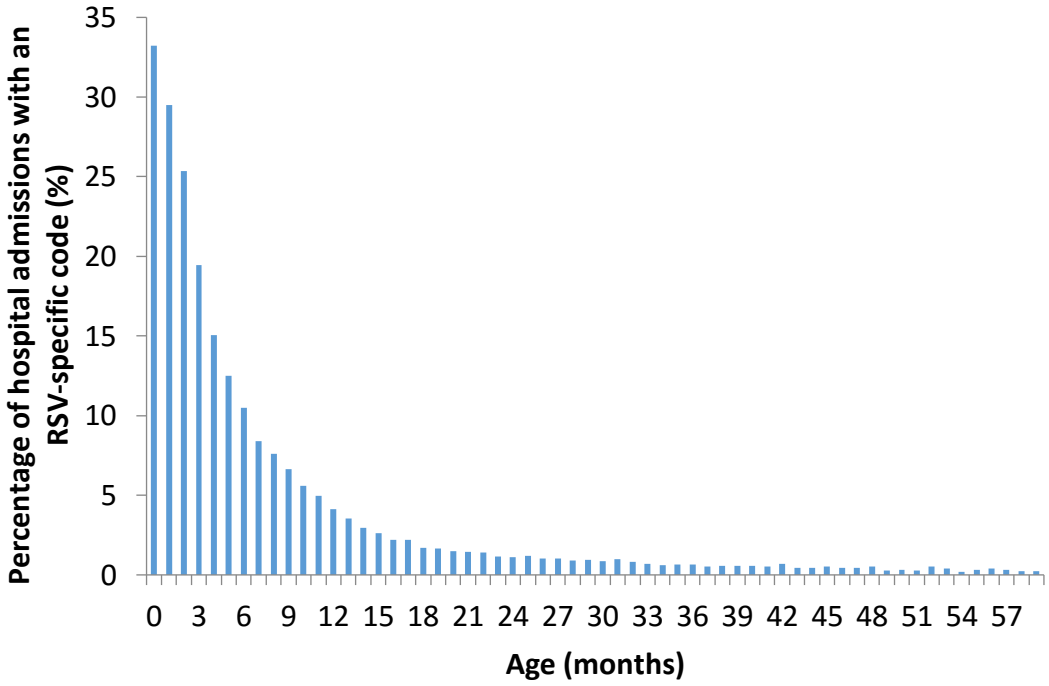

Figure 4-11. Percentage of RTI hospital admissions in children younger than 5 years of age in England with an RSV-specific ICD-10 diagnosis code, from 2007-2012, by age in months.

4.3.3.5 Risk group

An average of 6% (3,917/60,921) of RTI admissions in children aged <1 year and 8% (6,332/82,231) of RTI admissions in children aged 1-4 years had one or more ICD-10 code indicating high-risk status per year. Table 4-7 shows the percentage of admissions in high-risk children, stratified by primary diagnosis and age group. For all diagnosis except bronchiolitis, there was a higher percentage of admissions with high-risk ICD-10 codes in children aged <1 year compared to those aged 1-4 years.

Table 4-7. Percentage of hospital admissions for URTI, bronchiolitis, pneumonia, bronchitis or unspecified LRTI with ICD-10 codes indicating high-risk status, in infants aged <1 year and children aged 1-4 years in England, from 2007-2012.

Diagnosis	Percentage of admissions with ICD-10 codes indicating high-risk status	
	<1 year	1-4 years
URTI	6%	5%
Bronchiolitis	5%	14%
Pneumonia	14%	11%
Bronchitis	8%	7%
Unspecified LRTI	14%	14%

The most common comorbidities were CHD (3.7% children <5 years had an ICD-10 code for CHD) followed by neurological disorder (3.1%) and CLD (2.9%).

4.3.3.6 Length of hospital stay

Table 4-8 shows that, apart from URTI, children with ICD-10 codes indicating high risk status had higher median length of stay than those without. Admissions with any mention of pneumonia had a higher median length of stay in children aged <1 year, although in high risk children aged 1-4 years those with any mention of bronchiolitis had a longer median length of stay.

Table 4-8. Median length of stay (days) for hospital admissions with any mention of URTI, bronchiolitis, unspecified LRTI, pneumonia or bronchitis in children <5 years of age in England, stratified by age and risk group. Interquartile range (IQR) shown in brackets.

Primary diagnosis	Median length of stay, in days (IQR)			
	<1 year		1-4 years	
	No risk	High risk	No risk	High risk
URTI	<1 (<1-1)	<1 (<1-1)	<1 (<1-1)	<1 (<1-1)
Bronchiolitis	1 (<1-3)	3 (1-7)	1 (<1-2)	4 (1-7)
Unspecified LRTI	1 (<1-2)	3 (1-6)	1 (<1-2)	2 (<1-4)
Pneumonia	3 (1-4)	8 (2-11)	2 (1-3)	3 (1-7)
Bronchitis	1 (<1-3)	5 (1-8)	<1 (<1-1)	1 (<1-3)

Table 4-9 shows the median length of stay in days for only admissions with RSV-specific ICD-10 codes. The median length of stay was higher for RSV-coded admissions for all groups and diagnoses, compared to the overall median length of stay (Table 4-8).

Table 4-9. Median length of stay (days) for hospital admissions with an ICD-10 code indicating RSV as the cause of disease in children <5 years of age in England, stratified by age and risk group. Interquartile range (IQR) shown in brackets.

Primary diagnosis	Median length of stay, in days (IQR)			
	<1 year		1-4 years	
	No risk	High risk	No risk	High risk
URTI	1 (<1-2)	2 (<1-3)	1 (<1-2)	2 (1-7)
Bronchiolitis	3 (1-5)	6 (3-11)	2 (1-4)	5 (3-9)
Unspecified LRTI	2 (1-4)	8 (3-22)	2 (1-4)	5 (3-9)
Pneumonia	2 (4-6)	10 (5-15)	3 (2-5)	7 (4-12)
Bronchitis	2 (1-4)	4 (2-9)	2 (1-4)	4 (3-9)

4.3.4 Discussion

Hospital admissions for bronchiolitis had markedly similar epidemiology to laboratory-confirmed RSV episodes, mirroring the seasonality and age distribution with a peak in infants aged one month. One-quarter of bronchiolitis admissions had an ICD-10 code indicating RSV as the cause of disease, but very few URTI, pneumonia and unspecified LRTI admissions had an ICD-10 code indicating RSV as the cause of disease. Admissions with RSV-specific ICD-10 codes were in younger children that stayed in hospital for longer on average and who were more likely to have a known clinical risk factor (i.e. prematurity or comorbidity).

Bronchiolitis admissions in children younger than one year of age increased throughout the study period, peaking in 2010/11. This is consistent with previous studies that show an increase in bronchiolitis admissions in England, as well as those showing a general increase in admissions in children for all causes (51,61). Potential causes of this increase could include the lowering of thresholds for hospital admission (an inadvertent effect of changes to accident and emergency waiting time targets and clinical protocol), increases in the number of parents taking their child straight to hospital without seeking NHS advice elsewhere, parents being advised by NHS Direct to take their child straight to hospital, failure to manage these acute illnesses in the community care setting, and clinicians being increasingly risk-adverse due to an increased perceived risk of legal claims (8,51).

As expected, only a minority of RTI admissions had RSV-specific diagnosis codes. Only 25% of bronchiolitis admissions had an RSV-specific code, though it has been estimated that between 45 and 90% of bronchiolitis admissions are due to RSV (15,49,52). RSV is known to be under-recorded in hospital admissions data, with a large number of admissions recorded at the syndromic level (i.e. unspecified bronchiolitis, or unspecified LRTI) (118). This under-recording is particularly important when using clinical data alone to explore RSV burden, as I found that admissions with specific RSV-diagnostic coding are longer than those without (an indicator of increased severity of illness). The longer length of stay for admissions with an RSV-specific code found in this analysis may reflect bias in RSV-coding

towards children with more severe illness (indicated by the increased length of stay), or children in hospital for longer may be more likely to undergo laboratory testing which then informs the diagnostic coding of these admissions. The linkage of hospital admission data and laboratory surveillance data (presented in Chapter 5 and 6 of this thesis) will allow the potential association between RSV positivity and illness severity to be further investigated.

4.3.4.1 Strengths and limitations of this analysis

This is the most recent analysis of RTI admissions (and those coded as RSV-associated) in infants and young children in England. HES is the national hospital admissions database, and my results are therefore representative of all RTI admissions in children younger than 5 years in England. However, this analysis only covers admission with diagnosis code from Chapter X of the ICD-10 (that is, ICD-10 codes J00-J99); therefore, admissions with diagnostic codes from other ICD-10 chapters which may also be used for RSV–related illness such as R06.0-R06.8 (Abnormalities of breathing) and B34.9 (or R06.2) (Wheeze)– are not included in the dataset. A limitation of using hospital admissions data in isolation to investigate RSV-associated hospital admissions is the lack of laboratory testing data, compounded by the underreporting of RSV infection. However, this analysis is important in highlighting potential biases in diagnostic coding of RSV-associated illness, which I will investigate further in this thesis.

4.3.4.2 Implications for further analysis in this thesis

This analysis describes the epidemiology of RTI admissions in children younger than 5 years of age in England, which form the baseline population used in the statistical modelling estimates of RSV burden presented in Chapter 5 of this thesis. This analysis demonstrated that the majority of RTI admissions have RTI as the primary diagnosis, and therefore the analysis considering only the primary diagnosis (required to ensure that admissions with multiple RTI diagnoses are not counted as separate admissions, potentially overestimating the burden of RSV) will include the majority of RTI admissions.

This analysis highlights the difference in clinical coding of RTI admissions in children younger than 5 years in England, particularly the high number of bronchiolitis admissions in children younger than 6 months. In this thesis, I will use linked laboratory surveillance and hospital admission data to explore the relationship between patient characteristics, diagnostic coding, RSV testing and RSV positivity (Chapter 6). Furthermore, as this analysis highlights potential bias in RSV-specific coding towards younger children staying in hospital for longer, I will explore this potential bias during analysis of the linked data; it may have important implications when interpreting the results of the linked data analysis.

4.4 Conclusions

The analysis presented in this Chapter describes the epidemiology of laboratory-confirmed RSV infection in children younger than 5 years in England from 2010-2014, highlighting the impact of young age (<3 months) and birth near the beginning of the RSV season (September, October and November) on RSV-positivity, but also highlighting potential biases in RSV testing towards young infants. The analysis presented in this Chapter also describes the epidemiology of hospital admissions for RTI in children younger than five years of age in England, and highlights the under-recording of RSV-specific diagnoses in HES. In this thesis, I will further explore the relationship between RSV-positivity, age and birth month through linked laboratory surveillance and hospital admission data (Chapter 6 of this thesis), and use two methods to more robustly estimate the hospital burden of RSV in infants and young children in secondary care in England: time-series modelling (Chapter 5) and data linkage (Chapter 8).

Chapter 5

Estimating the secondary care burden of RSV using time-series modelling

Chapter 5 Estimating the secondary care burden of RSV using time-series modelling

5.1 Introduction

This chapter addresses Objective 2 of this thesis: estimate the number of RSV-associated hospital admissions in children younger than five years in England (stratified by age, primary diagnosis and calendar week) using time-series modelling of national laboratory surveillance and hospital administrative data.

I have presented the work in this chapter at the 10th International Respiratory Syncytial Virus Symposium (Patagonia, Argentina). In addition, this work is published in the peer-reviewed journal *Influenza and Other Respiratory Viruses*:

> *RM Reeves, P Hardelid, R Gilbert, F Warburton, J Ellis, R Pebody (2017) Estimating the burden of respiratory syncytial virus (RSV) on respiratory hospital admissions in children less than five years of age in England, 2007-2012.* Influenza and Other Respiratory Viruses *00, 1–8. doi: 10.1111/irv.12443*

A copy of this paper is included in Appendix 2. I carried out all of the analysis presented in this paper and chapter, and wrote the manuscript. Co-authors provided statistical advice and/or commented on the manuscript drafts.

5.2 Background

Evidence presented so far in this thesis demonstrates that calculating the national secondary care burden of disease due to RSV in young children is not straightforward. A reliable diagnosis of RSV infection can only be achieved through laboratory testing of respiratory samples but, as explained previously, only a minority of children hospitalised with an acute respiratory infection will be tested to identify the causal pathogen, as explained in section 1.3.7. Furthermore, most respiratory infections are recorded in hospital admission data using non-specific diagnosis codes such as unspecified pneumonia or bronchiolitis (Chapter 4.3). Hospital admission data or laboratory surveillance data alone cannot therefore be used to accurately describe the burden of RSV in secondary care.

Respiratory pathogens have varying temporal patterns which can be observed in the laboratory surveillance data. Statistical models which utilise the week on week variation in laboratory reports by pathogen can be constructed to attribute hospital admissions to different viruses (111,112). This method of estimating the hospital burden of RSV has previously been used in the UK; however, the most recent study only considers data up to 2009 (111–113). Since the 2009 influenza A (H1N1) pandemic, surveillance of respiratory viruses (including RSV) has been strengthened in England; there has been more widespread use of laboratory confirmation for respiratory viruses, with multiplex PCR methods allowing a whole panel of viruses to be tested for using a single sample, and a greater degree of reporting to national surveillance schemes (119). In addition, a recent study in the UK demonstrates that hospital admission due to bronchiolitis is increasing over time (51). These developments emphasise the need for more up to date estimates of RSV-associated hospital admissions as previous estimates may not reflect the current burden of disease.

In this chapter I aim to estimate the number of hospital admissions attributable to RSV in children <5 years of age in England in the period from mid-2007 to mid-2012 using ecological time series modelling of national laboratory surveillance and

hospital administrative data. This work provides updated estimates of RSV burden, including the post-pandemic period, for the first time in England. The results can be used as baseline epidemiological data in vaccine impact studies. In addition, the results will be compared to the estimated secondary care burden of RSV produced using novel data linkage methodology (Objective 5 of this thesis) which I present in Chapter 8 of this thesis, enabling validation of those results.

5.3 Methodology

5.3.1 Data extraction

This analysis used the SGSS and HES datasets described in Chapter 3. Data was extracted from both datasets from calendar week 27 in 2007 (beginning of July 2007) to calendar week 26 in 2012 (end of June 2012). Using this time-period allowed for an overlap with the most recent previous study (112) (to allow comparison) and used the most recently available hospital admission data. Calendar weeks were defined as blocks of 7 days beginning on 1st January each year, with week 52 allowed to have more than 7 days.

Weekly aggregated counts of laboratory-confirmed episodes of the following pathogens extracted from SGSS were included in this analysis: RSV, influenza A, influenza B, rhinovirus, parainfluenza, hMPV, adenovirus, *Streptococcus pneumoniae*, *Mycoplasma pneumoniae* and *Haemophilus influenza*. Only HES hospital admissions with a primary diagnosis of ICD-10 codes for bronchiolitis (ICD-10 J21), pneumonia (J12-18), unspecified LRTI (J22), bronchitis (J20) and URTI (J00-06) in children <5 years of age (at admission) in England from calendar week 27 in 2007 to calendar week 26 in 2012 were included. Only the primary diagnosis of each HES admission was considered to avoid double counting of admissions that had two or more of the RTI diagnoses. The weekly number of hospital admissions were stratified into three age groups: <6 months, 6-11 months and 1-4 years; smaller age groupings resulted in too few weekly admissions to build the statistical model, therefore these were the optimal age groups. Laboratory data in children aged <5 years was not stratified by age group.

5.3.2 Statistical analysis

This analysis used the observed temporal variation in weekly laboratory reports of potential causal pathogens to estimate the number of RTI hospital admissions that could be attributed to RSV, building on methods applied in previous modelling

studies (111,118,142). Separate models were developed for each primary diagnosis, using the weekly number of hospital admissions for each respective diagnosis as the dependent variable. For each diagnosis, separate models were constructed by age group (<6 months, 6-11 months, 1-4 years).

Multiple linear regression models were used to estimate the number of hospital admissions due to RSV from HES data coded as acute bronchiolitis, pneumonia, unspecified LRTI, bronchitis and URTI. These types of models have been used in similar studies (111,118,143,144), however no previous study has stratified admissions by diagnosis to build a more detailed picture of RSV burden by clinical presentation. All models used the weekly number of laboratory-confirmed episodes in children <5 years of age in England for the following pathogens as the independent variables: RSV, influenza A, influenza B, rhinovirus, parainfluenza, hMPV and adenovirus. The weekly number of laboratory-confirmed episodes of *S. pneumoniae*, *M. pneumoniae* and *H. influenza* in children <5 years of age in England was also included as independent variables in the models for pneumonia and unspecified LRTI. All variables were first included in the models, then those with negative coefficients removed (in order of decreasing significance) due to biological implausibility (as pathogens cannot cause a negative number of hospital admissions, as that would be to prevent them), followed by those with positive coefficients that did not contribute significantly to the model (F-test $p>0.05$). Interactions between all pathogens in the final model were investigated ($p\leq0.01$ was considered significant) due to the potential for co-circulation of pathogens. Interaction terms between pathogens in the final model (i.e. the remaining significant pathogens) and an indicator variable taking the value 0 for the pre-pandemic period (before week 20 2009) and 1 for the pandemic and post-pandemic period were investigated to account for potential changes in testing practice following the 2009 influenza pandemic ($p\leq0.01$ was considered significant). I also explored whether including a linear term to account for the underlying increase in admissions over time, or including a lag term between test and hospital admission (-2, -1, +1 and +2 week

lags were investigated) – however, including these terms did not improve the fit of the model.

Final estimates of the number of hospital admissions attributable to each pathogen were calculated by multiplying the coefficient from the final model by the total number of weekly laboratory-confirmed episodes for each relevant pathogen. 95% confidence intervals (CIs) were calculated as ±1.96*SE. The total number of hospital admissions for all children <5 years old was calculated as the sum of the number of hospital admissions calculated in each age group for each pathogen.

Admission rates were calculated using ONS mid-year population estimates for England by age group (<1 year, 1-4 years) (140,141). The average of ONS mid-year population estimates for 2007 and 2008 was used as the denominator for the 2007/2008 epidemiological year and the average of ONS mid-year population estimates for 2008 and 2009 used for the 2008/2009 epidemiological year, etc.

5.4 Results

5.4.1 Seasonality of laboratory reports and hospital admissions

The temporal trend in hospital admissions in children <5 years old in England with a primary diagnosis of bronchiolitis, pneumonia, bronchitis, unspecified LRTI or URTI are shown in Figure 5-1. Hospital admissions with a primary diagnosis of bronchiolitis (coloured red in Figure 5-1) were markedly seasonal and directly mirror the pattern of laboratory-confirmed RSV infections. Hospital admissions with a primary diagnosis of pneumonia had a very similar seasonal pattern to hospital admissions with a primary diagnosis of unspecified LRTI, with peaks also occurring at the same time as the peaks in laboratory- confirmed RSV infections in SGSS each year (Figure 5-1). The temporal variation in laboratory reports by pathogen is shown in Figure 5-2.

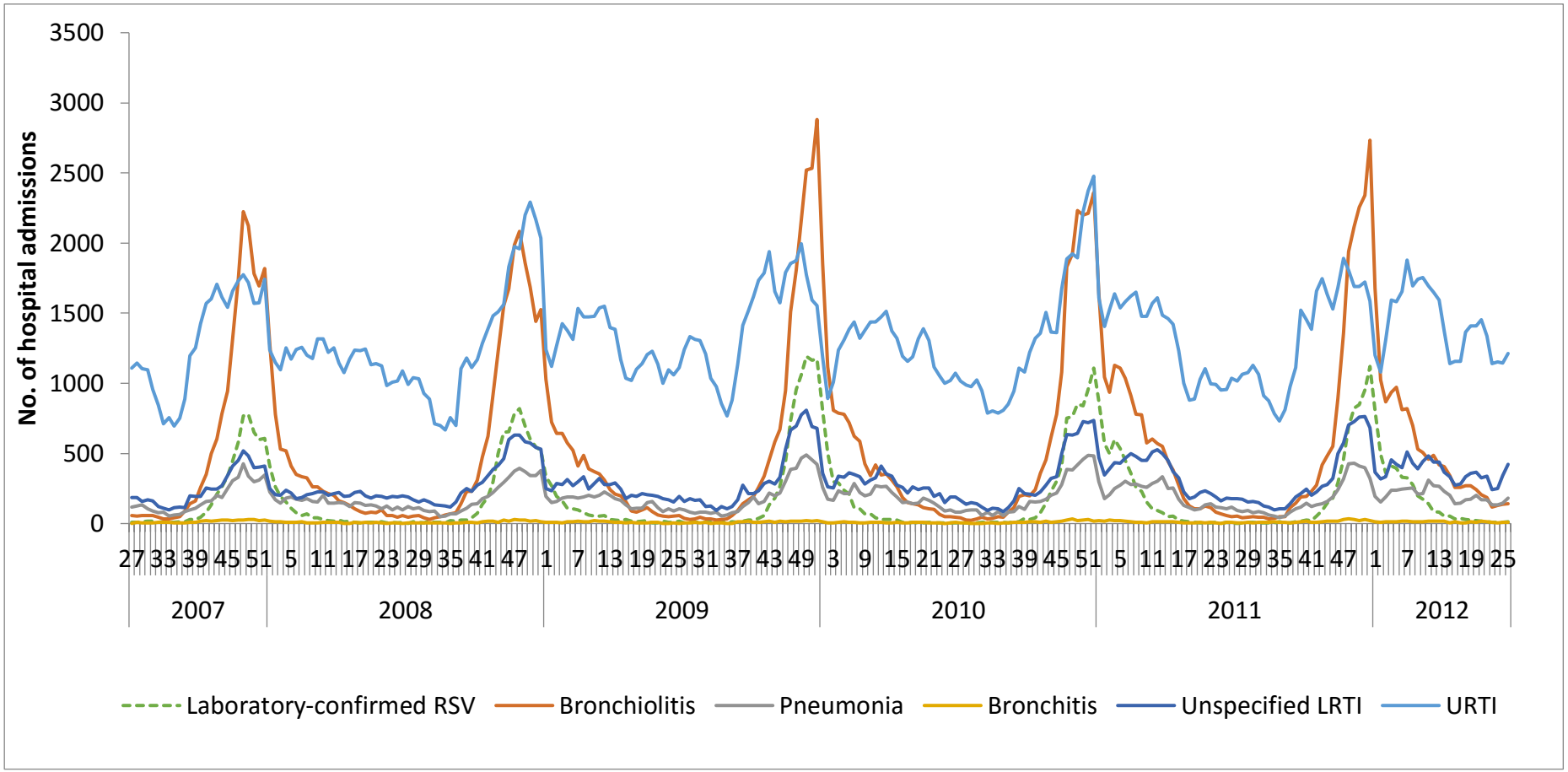

Figure 5-1. Weekly number of hospital admissions in children <5 years of age in England with any diagnosis of URTI, bronchiolitis, pneumonia, bronchitis or unspecified LRTI. Weekly number of laboratory-confirmed RSV infections in children <5 years of age in England (SGSS – section 4.2.3.1) shown in orange.

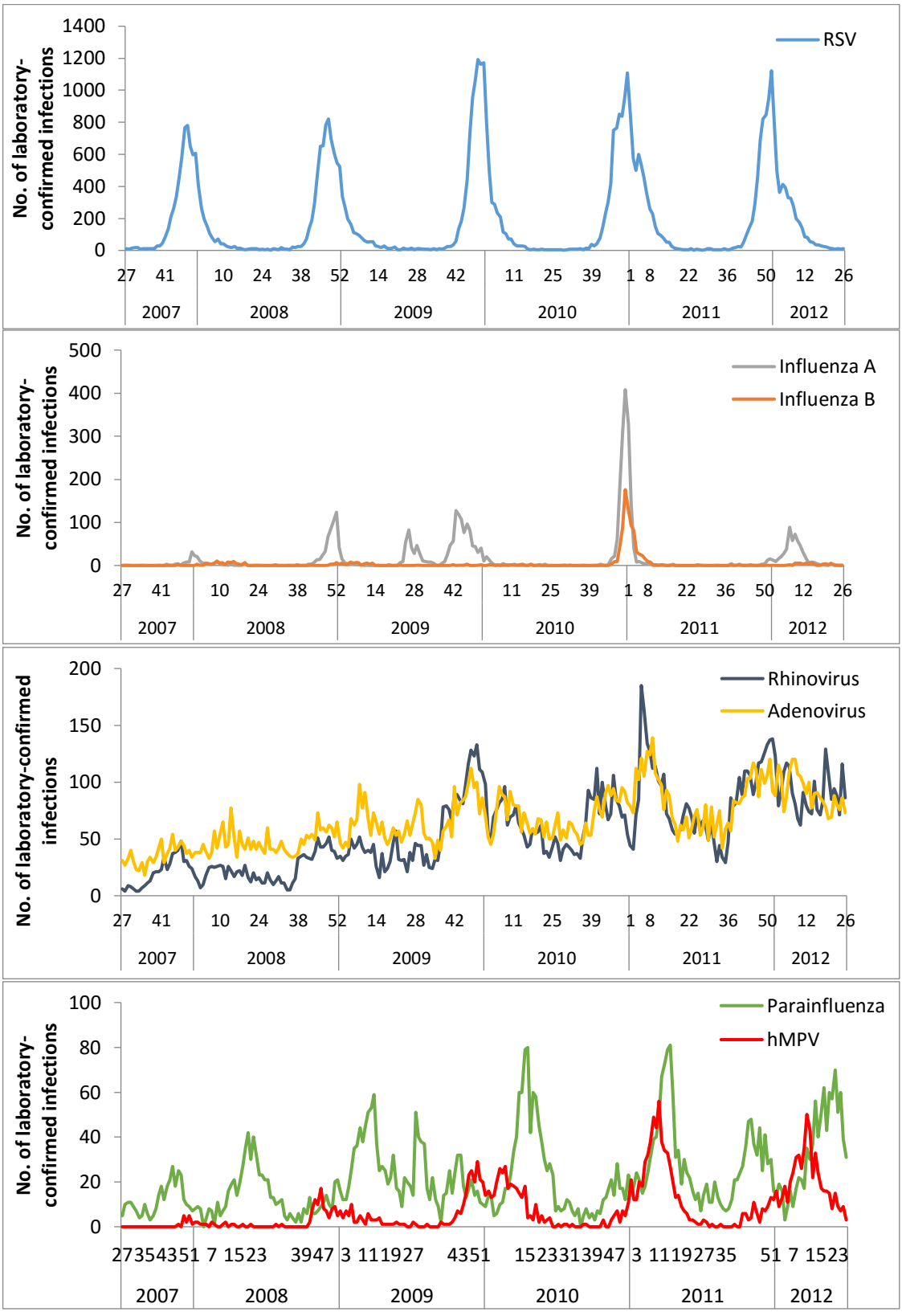

Figure 5-2. Weekly number of laboratory-confirmed cases of major respiratory viruses recorded in SGSS for children <5 years of age, over time.

5.4.2 Estimated RSV-associated hospital admissions

5.4.2.1 All RTI admissions

Of the 121,968 hospital admissions with a primary diagnosis of bronchiolitis, pneumonia, bronchitis, unspecified LRTI or URTI that occurred annually in children <5 years of age from mid-2007 to mid-2012, I estimate that 28% (33,561/121,968, 95% CI: 25-32%) were due to RSV. Of the RSV-associated RTI hospital admissions, 48% (16,202/33,561, 95% CI: 46-52%) were in children <6 months of age, 21% (7,108/33,561, 95% CI: 19-25%) were in children 6-11 months of age, and 31% (10,251/33,561, 95% CI: 26-38%) were in children aged 1-4 years (Table 5-1). The majority, 84% (28,111/33,561, 95% CI: 81-91%), of RSV-associated RTI hospital admissions were for LRTI. About 65% (21,418/33,561, 95% CI: 62-70%) of RSV-associated hospital admissions had a primary diagnosis of bronchiolitis (Table 5-1, Table 5-2). The fit of the final models for bronchiolitis, pneumonia and unspecified LRTI are demonstrated in Figure 5-3, Figure 5-4 and Figure 5-5, respectively.

On average, the estimated admission rate of any RSV-associated RTI hospital admission was 35.1 (95% CI: 32.9- 38.9) per 1,000 children <1 year of age and 5.31 (95% CI: 4.5-6.6) per 1,000 children 1- 4 years of age, per year (Table 5-3). Estimated rates of RSV-associated hospital admissions were, on average, higher in <1- year-olds for all diagnoses except unspecified LRTI, where admission rates in 1- to 4-year-olds were higher (Table 5-3). All RSV-associated LRTI admission rates increased in both age groups over time, peaking in 2010/2011 (the RSV season following the 2009 influenza A (H1N1) pandemic, which was an intense influenza season) and with a slight decrease in 2011/2012 for all diagnoses except pneumonia. Admissions for RSV-associated URTI decreased over the study period, particularly in the 1-4 year age group which saw a 70% decrease from 2.9 (95% CI: 2.5-3.4) admissions per 1,000 children in 2007/2008 to 0.9 (95% CI: 0.3-2.1) per 1,000 children in 2011/2012 (Table 5-3).

The percentage of weekly RTI hospital admissions attributed to RSV varied by calendar week for all primary diagnoses and age groups (Figure 5-6).

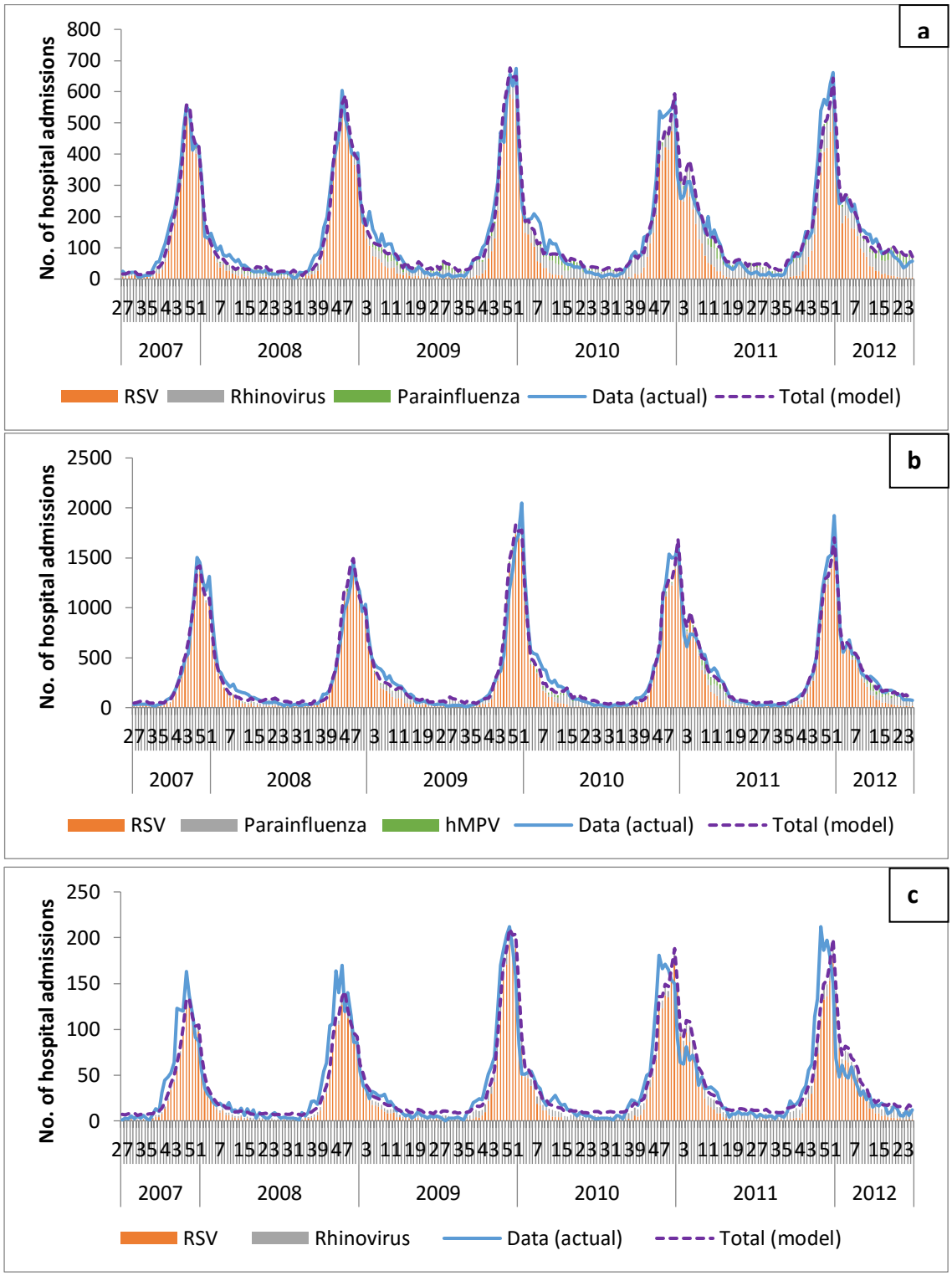

Figure 5-3. Weekly <u>bronchiolitis</u> hospital admissions due to respiratory syncytial virus (RSV), parainfluenza and human metapneumovirus (hMPV) in England from 2007-2012, as estimated by the final models, in: (a) children aged <6 months; (b) children aged 6-11 months; (c) children aged 1-4 years. Only pathogens significantly contributing to the age-group specific model (and therefore included in the final model) are included.

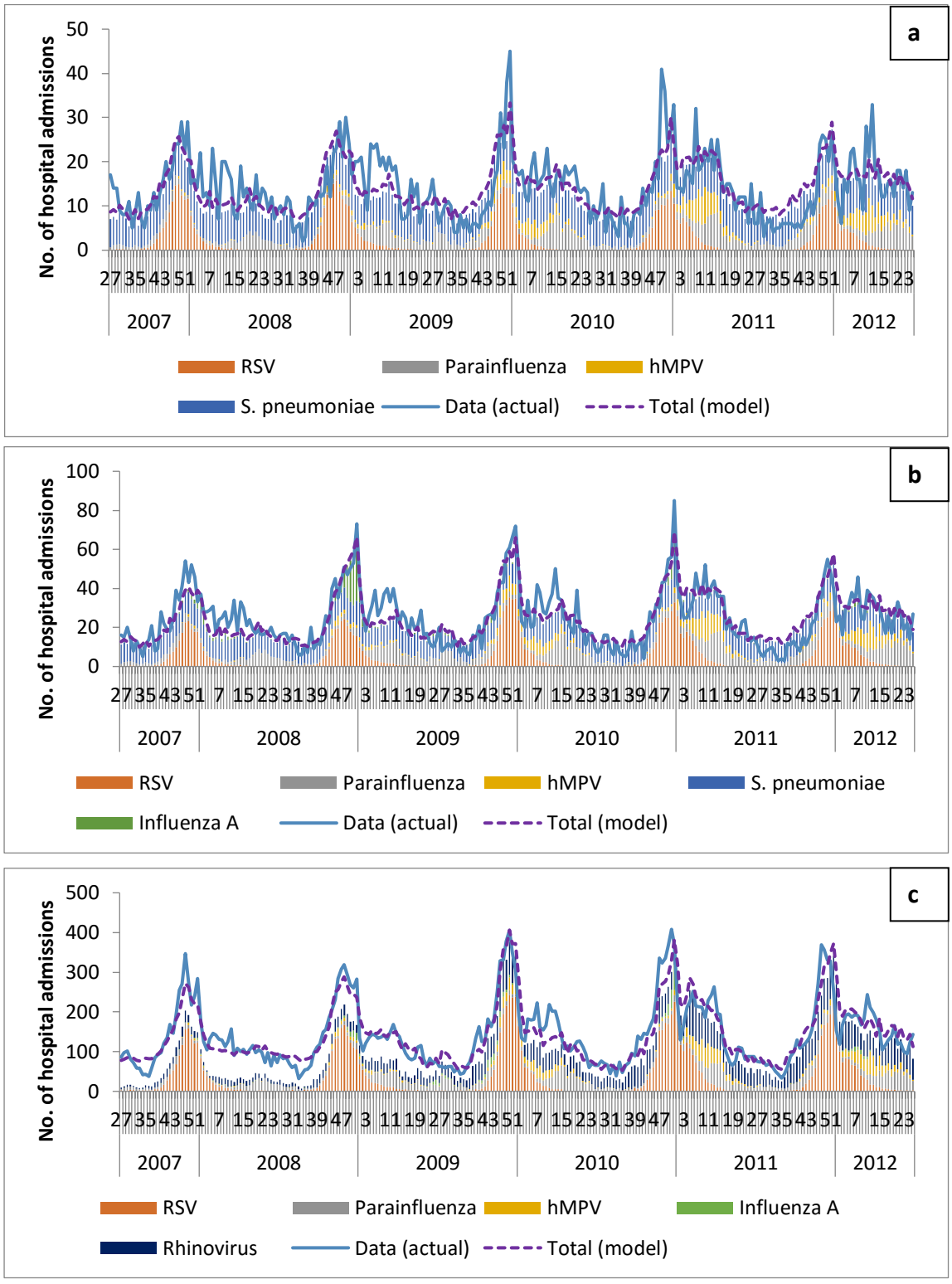

Figure 5-4. Weekly pneumonia hospital admissions due to respiratory syncytial virus (RSV), parainfluenza, human metapneumovirus (hMPV), *S. pneumonia*, influenza A and rhinovirus in England from 2007-2012, as estimated by the final models, in: (a) children aged <6 months; (b) children aged 6-11 months; (c) children aged 1-4 years. Only pathogens significantly contributing to the age-group specific model (and therefore included in the final model) are included.

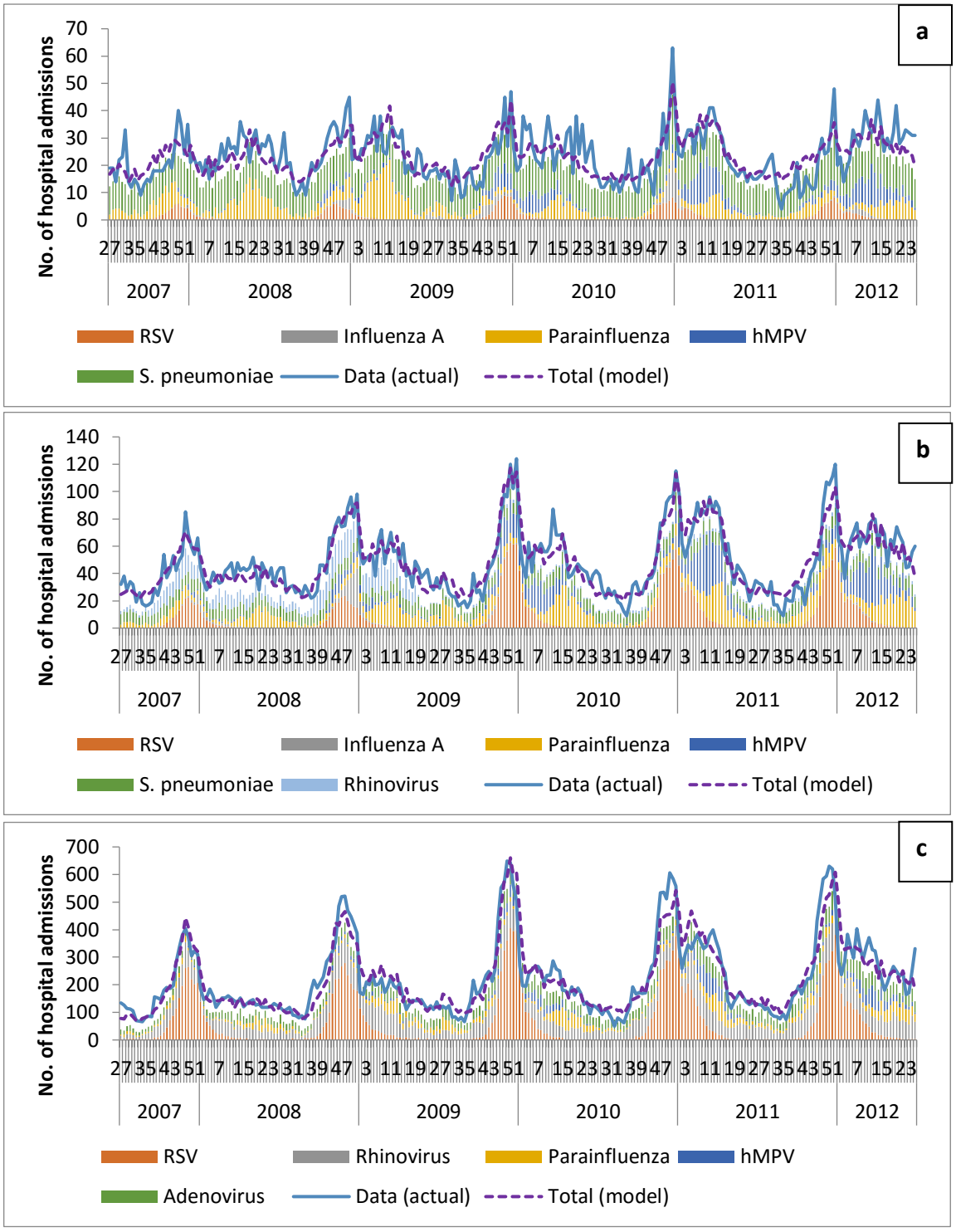

Figure 5-5. Weekly <u>unspecified LRTI</u> hospital admissions due to respiratory syncytial virus (RSV), influenza A, parainfluenza, human metapneumovirus (hMPV), *S. pneumonia*, rhinovirus and adenovirus in England from 2007-2012, as estimated by the final models, in: (a) children aged <6months; (b) children aged 6-11 months; (c) children aged 1-4 years. Only pathogens significantly contributing to the age-group specific model (and therefore included in the final model) are included.

5.4.2.2 LRTI admissions: bronchiolitis, pneumonia, bronchitis and unspecified LRTI

Overall, 53% (28,111/53,514, 95% CI: 51-57%) of LRTI admissions in children <5 years of age could be attributed to RSV, and 54% (15,260/28,111, 95% CI: 52-58%) of RSV-associated LRTIs were in infants <6 months old (Table 5-1). Of RSV-associated LRTI admissions, 78% (21,830/28,111, 95% CI: 74-83%) were coded as bronchiolitis, 13% (3,766/28,111, 95% CI: 12-15%) were coded as unspecified LRTI, 8% (2,346/28,111, 95% CI: 7-9%) were coded as pneumonia, and <1% (169/28,111) were coded as bronchitis.

There was an annual average of 27,969 hospital admissions with a primary diagnosis of bronchiolitis in children <5 years of age in England from 2007 to 2012. We estimate that approximately 78% (21,830/27,969, 95% CI: 75-83%) of these were due to RSV (Table 5-1, Table 5-2). There were differences by age, with approximately 82% 14,962/18,246, 95% CI: 79-87%) of all bronchiolitis admissions in children aged <6 months attributable to RSV, 70% (5,319/7,652, 95% CI: 66- %) of bronchiolitis admissions in children aged 6-11 months attributable to RSV and 75% (1,549/2,071, 95% CI: 71-79%) of bronchiolitis admissions in children aged 1-4 years attributable to RSV (Table 5-2). During the period from calendar week 46 to week 2, over 90% of bronchiolitis admissions in children aged <6 months and over 80% in children aged 6- 11 months and 1-4 years were attributable to RSV each week (Figure 5-3). The other explanatory pathogens for bronchiolitis admissions were parainfluenza, hMPV and rhinovirus, with differences in causal pathogens by age (Table 5-4).

There was an annual average of 9,537 hospital admissions with a primary diagnosis of pneumonia in children <5 years of age in England from 2007 to 2012. We estimate that approximately 25% (2,346/9,537, 95% CI: 22-27%) of these were due to RSV (Table 5-1). There were differences by age, with approximately 26% (1,923/7,503, 95% CI: 23-28%) of pneumonia admissions in children aged 1-4 years attributable to RSV compared to 19% in children aged <6 months (138/739, 95% CI: 14-28%). The other main explanatory pathogens for hospital admissions with a

primary diagnosis of pneumonia in children <5 years of age were *S. pneumoniae*, hMPV, parainfluenza, influenza A and rhinovirus, with differences in causal pathogens by age (Table 5-4).

There was an annual average of 547 hospital admissions with a primary diagnosis of acute bronchitis in children <5 years of age in England from 2007 to 2012. We estimate that approximately 31% (169/547, 95% CI: 23-42%) of these were due to RSV (Table 5-1). There were differences by age, with approximately 56% (89/159, 95% CI: 48-71%) of bronchitis hospital admissions in children <6 months of age attributed to RSV, compared to only 14% (39/276, 95% CI: 8-20%) in children aged 1-4 years (Table 5-2). The other explanatory pathogens for hospital admissions with a primary diagnosis of bronchitis were parainfluenza and adenovirus in children aged 1-4 years only (Table 5-4).

There was an annual average of 15,461 hospital admissions with a primary diagnosis of unspecified LRTI in children <5 years of age in England from 2007 to 2012. We estimate that approximately 24% (3,766/15,461, 95% CI: 22-27%) of these were due to RSV. There were differences by age, with approximately 28% (3,266/11,690, 95% CI: 26-30%) of unspecified LRTI hospital admissions in children aged 1-4 years of age attributed to RSV, compared to only approximately 6% (71/1,224, 95% CI: 3-8%) in children <6 months of age (Table 5-2). The other main explanatory pathogens for hospital admissions with a primary diagnosis of unspecified LRTI in children <5 years of age were adenovirus, *S. pneumoniae*, hMPV, parainfluenza, rhinovirus and influenza A, with differences in causal pathogens by age (Table 5-4).

5.4.2.3 URTI admissions

From 2007 to 2012, there was an annual average of 68,455 hospital admissions with a primary diagnosis of URTI in children <5 years of age in England. We estimate that approximately 8% (5,450/68,455, 95% CI: 5- 12%) of these were due to RSV (Table 5-1). There were no significant differences in the percentage of URTI admissions attributable to RSV by age group, although a slightly higher percentage of

admissions in children <6 months of age were attributed to RSV (11%, 942/8,868, 95% CI: 9-12%). The other explanatory pathogens for hospital admissions with a primary diagnosis of URTI were influenza A, parainfluenza and adenovirus. The number of hospital admissions attributed to each pathogen in each final model is shown in Table 5-4.

Table 5-1. Average annual number of hospital admissions in children <5 years of age in England: total per primary diagnosis and number estimated to be attributed to RSV by the final models, stratified by age group (<6 months, 6-11 months, 1-4 years).

Primary diagnosis		Annual number of hospital admissions (95% CI)			
		<6 months	6-11 months	1-4 years	Total
Bronchiolitis	Total	18,246	7,652	2,071	27,969
	RSV-associated	14,962 (14,396 – 15,942)	5,319 (5,066 – 5,775)	1,549 (1,467 – 1,631)	21,830 (20,929 – 23,348)
Pneumonia	Total	739	1,294	7,503	9,537
	RSV-associated	138 (101 – 158)	285 (245 – 324)	1,923 (1,718 – 2,128)	2,346 (2,064 – 2,610)
Bronchitis	Total	159	113	276	547
	RSV-associated	89 (76 – 113)	41 (29 – 62)	39 (23 – 55)	169 (128 – 230)
Unspecified LRTI	Total	1,224	2,546	11,690	15,461
	RSV-associated	71 (37 – 104)	429 (322 – 619)	3,266 (3,007 – 3,525)	3,766 (3,366 – 4,248)
URTI	Total	8,868	12,737	46,850	68,455
	RSV-associated	942 (802 – 1,082)	1,034 (742 – 1,559)	3,474 (2,398 – 5,412)	5,450 (3,222– 8,053)
Total		29,236	24,342	68,390	121,968
Total RSV-associated		16,202 (15,412 – 17,399)	7,108 (6,404 – 8,339)	10,251 (8,613 – 12,751)	33,561 (30,429 – 38,489)

Table 5-2. Total RTI hospital admissions estimated to be due to RSV by the final models, stratified by age group (<6 months, 6-11 months, 1-4 years) and primary diagnosis, as a percentage of the total hospital admissions for the respective primary diagnosis.

Primary diagnosis	Percentage of annual hospital admissions (per primary diagnosis) attributed to RSV (95% CI)			
	<6 months	6-11 months	1-4 years	Total
Bronchiolitis	82% (79-87%)	70% (66-75%)	75% (71-79%)	78% (75-83%)
Pneumonia	19% (14-28%)	22% (19-25%)	26% (23-28%)	25% (22-27%)
Bronchitis	56% (48-71%)	37% (28-55%)	14% (8-20%)	31% (23-42%)
Unspecified LRTI	6% (3-8%)	17% (13-24%)	28% (26-30%)	24% (22-27%)
URTI	11% (9-12%)	8% (6-12%)	7% (5-12%)	8% (5-12%)

Table 5-3. Estimated admission rates of RSV-associated hospital admissions per 1,000 children <5 years of age in England.

Estimated admission rate of RSV-associated hospital admissions (per 1,000) (95% CI)							
	Age group	**2007/8**	**2008/9**	**2009/10**	**2010/11**	**2011/12**	**Average**
URTI	**<1y**	3.00 (2.50-3.50)	3.41 (2.84-3.41)	2.73 (2.03-4.00)	3.00 (2.23-4.39)	2.59 (1.92-3.79)	**2.95 (2.31-3.93)**
	1-4y	2.94 (2.49-3.38)	3.29 (2.79-3.29)	0.95 (0.35-2.27)	1.04 (0.38-2.48)	0.90 (0.33-2.14)	**1.82 (1.27-2.81)**
Bronchiolitis	**<1y**	27.1 (26.2-28.1)	30.8 (29.8-31.9)	30.5 (29.2-33.4)	33.6 (32.1-36.7)	29.0 (27.7-31.7)	**30.2 (29.0-32.3)**
	1-4y	0.64 (0.60-0.67)	0.72 (0.68-0.76)	0.87 (0.82-0.91)	0.94 (0.89-0.99)	0.82 (0.77-0.86)	**0.90 (0.75-0.84)**
Unspecified LRTI	**<1y**	0.40 (0.24-0.57)	0.46 (0.27-0.65)	0.93 (0.70-1.36)	1.03 (0.78-1.50)	0.89 (0.67-1.29)	**0.74 (0.53-1.07)**
	1-4y	1.34 (1.24-1.45)	1.52 (1.40-1.64)	1.82 (1.68-1.97)	1.99 (1.83-2.15)	1.72 (1.58-1.86)	**1.68 (1.55-1.81)**
Pneumonia	**<1y**	0.54 (0.45-0.63)	0.62 (0.52-0.72)	0.65 (0.53-0.84)	0.72 (0.58-0.93)	2.69 (2.28-3.16)	**1.04 (0.87-1.26)**
	1-4y	0.79 (0.71-0.88)	0.89 (0.81-0.99)	1.07 (0.96-1.19)	1.17 (1.05-1.30)	1.01 (0.91-1.12)	**0.99 (0.88-1.09)**
Bronchitis	**<1y**	0.21 (0.18-0.24)	0.24 (0.20-0.27)	0.17 (0.13-0.26)	0.19 (0.15-0.29)	0.17 (0.13-0.25)	**0.20 (0.16-0.26)**
	1-4y	0.02 (0.01-0.02)	0.02 (0.01-0.03)	0.02 (0.01-0.03)	0.02 (0.01-0.03)	0.02 (0.01-0.03)	**0.02 (0.01-0.03)**
Total	**<1y**	**31.3 (29.5-33.0)**	**35.6 (33.6-37.6)**	**35.0 (32.6-39.8)**	**38.5 (35.9-43.8)**	**35.5 (32.7-40.1)**	**35.1 (32.9-38.9)**
	1-4y	**5.73 (5.05-6.40)**	**6.43 (5.67-7.20)**	**4.74 (3.82-6.37)**	**5.17 (4.17-6.95)**	**4.47 (3.61-6.01)**	**5.31 (4.46-6.59)**

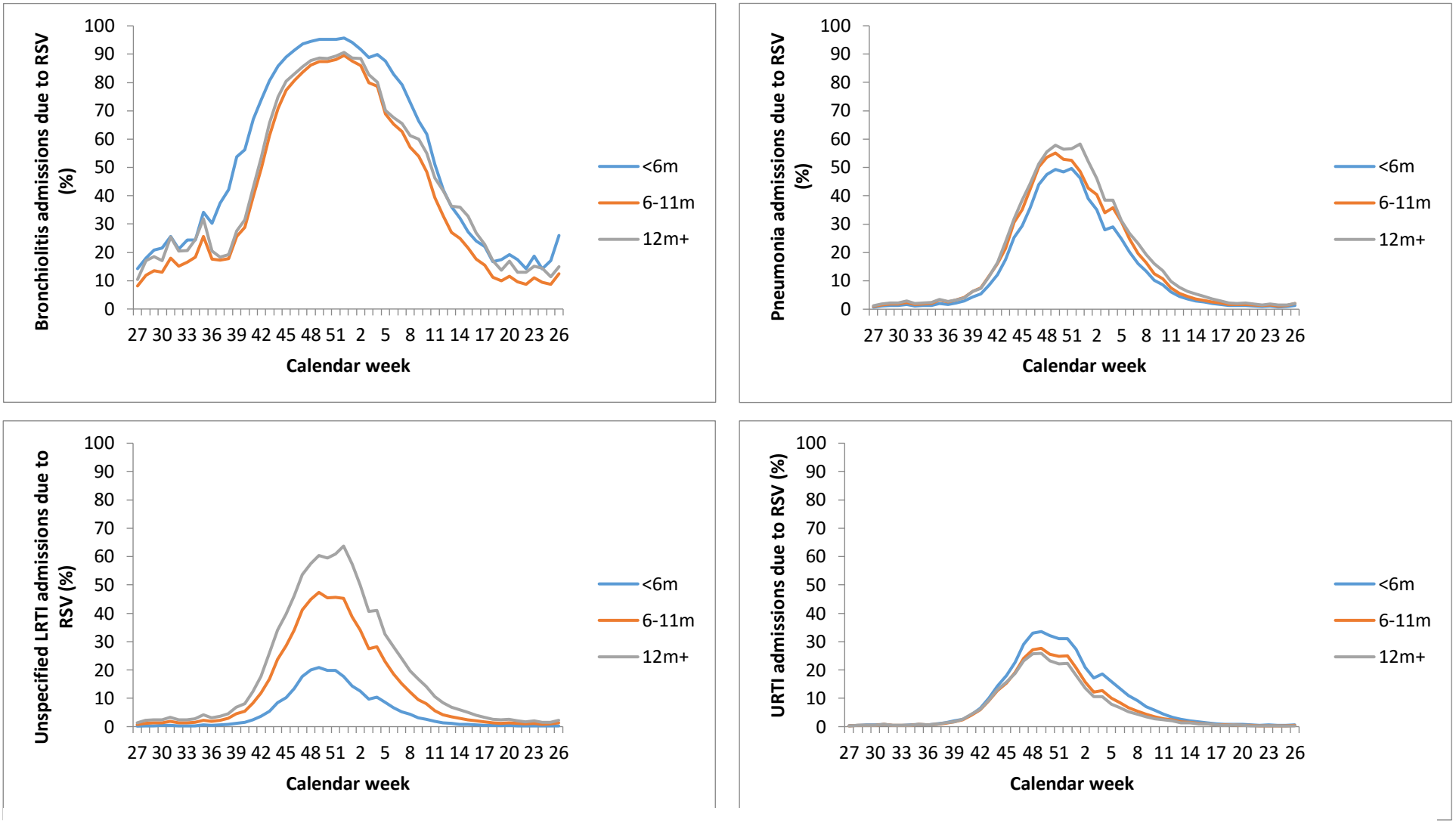

Figure 5-6. Percentage of total weekly RTI hospital admissions estimated to be due to RSV for each primary diagnosis, by age group.

Table 5-4. Average annual estimated hospital admissions in children <5 years of age in England from 2007 to 2012 attributed to each pathogen (RSV, influenza A, influenza B, rhinovirus, parainfluenza, hMPV, Adenovirus, S. pneumoniae, M. pneumoniae, H. influenza) by the final models, stratified by age group (<6 months, 6-11 months, 1-4 years). " - " denotes that the pathogen was not included in the final model.

	Average annual number of hospital admissions (95% CI)														
	Bronchiolitis			Pneumonia			Bronchitis			Unspecified LRTI			URTI		
	<6m	6-11m	1-4y	<6m	6-11m	1-4y	<6m	6-11m	1-4y	<6m	6-11m	1-4y	<6m	6-11m	1-4y
Total	18246	7652	2071	739	1294	7503	159	113	276	1225	2546	11690	8868	12737	26280
Explained by:															
RSV	14962 (14396 - 15942)	5319 (5066 - 5775)	1549 (1467 - 1631)	138 (101 - 204)	285 (245 - 324)	1923 (1718 - 2128)	89 (76 - 113)	41 (29 - 62)	39 (23 - 55)	71 (37 - 104)	429 (322 - 619)	3266 (3007 - 3525)	942 (802 - 1082)	1034 (742 - 1559)	3474 (2398 - 5412)
Influenza A	-	-	-	-	54 (9 - 168)	115 (23 - 207)	-	-	-	28 (12 - 45)	51 (43 - 127)	-	323 (74 - 774)	362 (272 - 453)	1259 (886 - 1633)
Influenza B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Rhinovirus	-	1640 (1185 - 2096)	279 (79 - 479)	-	-	1914 (529 - 4157)	-	-	-	-	251 (174 - 966)	2731 (708 - 5982)	-	-	-
Parainfluenza	1598 (902 - 2293)	505 (199 - 811)	-	112 (66 - 158)	236 (167 - 304)	736 (416 - 1056)	-	-	33 (1 - 64)	214 (121 - 369)	440 (350 - 530)	1040 (560 - 1521)	1046 (623 - 1742)	917 (594 - 1240)	2612 (1281 - 3944)
hMPV	593 (204 - 982)	-	-	56 (30 - 81)	105 (70 - 143)	469 (279 - 659)	-	-	-	114 (79 - 148)	323 (269 - 377)	612 (318 - 907)	-	-	-
Adenovirus	-	-	-	-	-	-	-	-	108 (38 - 179)	-	-	1911 (272 - 3551)	-	3408 (2159 - 5201)	15187 (11881 - 18494)
S. pneumoniae	-	-	-	349 (140 - 558)	550 (240 - 860)	-	-	-	-	566 (300 - 832)	432 (35 - 829)	-	-	-	-
M. pneumoniae	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
H. influenza	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

5.5 Discussion

Every year, 33,561 (95% CI: 30,429-38,489) hospital admissions for RTIs in children <5 years of age from 2007 to 2012 could be attributed to RSV. This represents annual RSV-associated RTI admission rates of 35.1 (95% CI: 32.9- 38.9) per 1,000 children <1 year of age and 5.31 (95% CI: 4.5-6.6) per 1,000 children 1-4 years of age. Most of these RSV-associated admissions (84%, 95% CI: 81-91%) were coded as LRTIs. Of the RSV-associated LRTIs, nearly half were in children aged <6 months (48%, 95% CI: 46-52%). Approximately 82% (95% CI: 79-87%) of hospital admissions for bronchiolitis in children aged <6 months could be attributed to RSV.

This study is the first to determine the burden of RSV-associated respiratory hospital admissions in children in England by primary diagnosis and age group. My analysis uses smaller age groupings than previous studies to more precisely reflect the differences in RSV-associated illness by age (previous studies have used age groupings of <1 and 1-4 years (118), or <5 years only (111)). In addition, my analysis is the first to estimate the burden of RSV in England in the post-2009 influenza A (H1N1) pandemic era, and the first to stratify based on diagnosis.

Previous estimates of RSV burden in secondary care in England range from 26,500 to 29,160 RSV-associated hospital admissions per year in children <5 years of age—all lower than my estimate of 33,561 (95% CI: 30,429-38,489) RSV-associated admissions (111–113,118). This difference is most likely due to those studies using earlier time periods, as my analysis found a steady, general increase in the admission rates of RSV-associated LRTI hospital admissions over time. From 1999 to 2010, there has been an increase of 28% in emergency hospital admissions in children <5 years of age, particularly admissions for acute illness, and the annual number of hospital admissions due to bronchiolitis in young children in the UK has also increased sevenfold between 1979 and 2011 (51,61). This evidence suggests that the increase in RSV-associated admissions that I have found is not due to an increase in the severity of infection, nor the virulence of RSV in the age group, because paediatric intensive care admission rates have changed little from 2004 to

2012 (51). Instead, this trend is likely due to a general increase in hospital admissions in children, as explained in section 4.3.4.

My analysis found a high burden of RSV-associated hospital admissions in children <6 months of age, a group well documented as being at high risk of RSV- associated hospital admission (145). However, I also found a high number of RSV-associated pneumonia and unspecified LRTI admissions in children aged 1-4 years. RSV was the pathogen associated with the highest number of admissions for all types of LRTI in all age groups compared to the other pathogens, except for pneumonia and unspecified LRTI in children aged <6 months and 6-11 months. Therefore, RSV-associated hospital admissions are more likely to be coded as bronchiolitis in young infants, but as pneumonia or unspecified LRTI in children aged 1-4 years. Studies of RSV-associated hospital admissions in young children therefore need to consider that RSV infection may manifest in other clinical syndrome leading to an ICD-10 code other than bronchiolitis being entered as the primary diagnosis, particularly in older children.

5.5.1 Strengths and limitations of this analysis

The methodology I have used here has several limitations. Firstly, the time-series models assume that the temporal variation in laboratory reports of the causative agents is an accurate representation of their relative incidence over time. The results will be biased if this is not the case (i.e. due to seasonal changes in laboratory testing or reporting). Secondly, there may be other reasons for temporal variations that are not accounted for in this analysis (e.g. meteorological or environmental variables), and it is possible that some hospital admissions could be attributed to other pathogens with similar temporal patterns to those that I have considered (e.g. other bacterial pathogens) (118). Thirdly, it is possible that age-related differences in testing practices may have impacted my estimates—particularly those for influenza-associated admissions. As there were only a very

small number of young children with laboratory-confirmed influenza infection, it is possible that some of the unattributed admissions in my models should have been attributed to influenza, and I have therefore underestimated the burden of influenza in my study population. However, that my results demonstrate a significantly higher burden of RSV-associated hospital admissions in young children compared to other respiratory pathogens, including influenza, is consistent with the results of previous studies (146–148). Finally, as my analysis is restricted to hospital admissions with a primary diagnosis of URTI or LRTI, the results are likely to underestimate the true burden of RSV-associated hospital admissions. Though my analysis in section 4.3 demonstrates that the majority of admissions with an RTI diagnosis have the URTI or LRTI diagnosis as the primary diagnosis (and therefore they would be included in this analysis), children with other non-respiratory ICD-10 codes (such as R06.0-R06.8 (Abnormalities of breathing) and B34.9 (or R06.2) (Wheeze)) would not be included.

5.5.2 Implications for further analysis in this thesis

This analysis estimates the number of RSV-associated hospital admissions in children younger than five years in England (stratified by age, primary diagnosis and calendar week) using time-series modelling of national laboratory surveillance data and hospital administrative data. As this analysis is based on weekly aggregate data, it is not possible to investigate individual-level risk factors for RSV-associated hospital admission (such as comorbidities) or additional outcomes indicating the severity of illness such as length of stay or admission to paediatric intensive care. The analysis of linked laboratory surveillance and hospital admissions data (presented in Chapter 6, 7 and 8 of this thesis) – will describe the number of RSV-associated hospital admissions in more detail than is achievable through this statistical modelling methodology. As there were previously no estimates of RSV burden that considered data more recent than the 2008/09 season – and data linkage has not been used to explore RSV burden in England before – the results of

this analysis are very important in providing estimates that can be compared to both previous estimates and my estimates calculated using the linked data.

5.5.3 Conclusions

RSV is a significant cause of hospital admissions for LRTI in children <5 years of age in England, and has varying presentation by age. This work provides a pre-introduction baseline for vaccine impact studies if such vaccines are introduced into the UK immunisation programme in the future. However, large-scale individual-level analysis of RSV-associated hospital admissions in young children in secondary care in England is required to describe the characteristics of admitted children, compare the severity of cases by age, and assess additional measure of secondary care burden such as length of hospital stay. The analysis of linked hospital admission data and laboratory surveillance data (presented in Chapter 6, 7 and 8 of this thesis) aims to fulfil this requirement.

Chapter 6

Linking RSV laboratory surveillance data and hospital admissions data

Chapter 6 Linking RSV laboratory surveillance and hospital admissions data

6.1 Introduction

This chapter addresses Objective 3 of this thesis: describe, using linked laboratory surveillance and hospital admissions data, laboratory-confirmed RSV-associated hospital admissions in children younger than five years in England by key patient and clinical characteristics. Nikolaos Panagiotopoulos and Mehdi Minaji (both from PHE) carried out the data linkage that I present in this chapter. I carried out all subsequent data cleaning and analysis.

The analysis presented in this chapter is an original contribution to knowledge because:

- I demonstrate that population-based data linkage between laboratory surveillance and hospital admissions data for RSV in England is possible, though assessment of linkage quality is difficult.
- I show that young age and having a known clinical risk factor (prematurity or comorbidity) were significantly associated with successful linkage, indicating that these children were more likely to undergo laboratory testing to identify the causal pathogen.
- I describe laboratory-confirmed RSV-associated hospital admissions in England, by key demographic and clinical characteristics, using linked data for the first time.

The linked dataset described in this chapter is also used for the analysis presented in Chapter 7 and Chapter 8 of this thesis.

6.2 Background

As shown in Chapter 4.3 of this thesis, the majority of respiratory infections are recorded in hospital admission data using non-specific diagnosis codes such as unspecified pneumonia or bronchiolitis. Whilst laboratory surveillance data contain data on laboratory-confirmed RSV infections, there is no clear denominator population and the datasets lack clinical information (i.e. on comorbidities and length of stay). Therefore, hospital admission data or laboratory surveillance data alone cannot be used to accurately describe the burden of RSV in secondary care. Previous studies of the national burden of RSV in secondary care in England have used time series regression modelling methods on aggregate data, such as those presented in Chapter 5 of this thesis. While these studies are informative as they can estimate the number of hospital admissions attributed to RSV at the population level, more detailed information on the burden of RSV-associated hospital admissions by age in months and risk group is required to determine optimal target populations for a potential future vaccine, as outlined in section 1.4 (101).

Linkage of routinely collected data facilitates the analysis of more complete information without the time and cost burden of primary data collection (for example, through a prospective surveillance study) (120). Consequently, linkage of routinely collected laboratory surveillance and hospital admissions data is a time and cost effective method of describing laboratory-confirmed RSV-associated hospital admissions in the detail required to inform vaccine policy (115). Worldwide, few studies have utilised population-based data linkage to describe the aetiology or pathogen-specific burden of acute lower respiratory infections (115). Previous data linkage studies focusing on RSV have been based in Australia, Canada and Denmark (114–117). Despite the availability of patient identifiable information (PII) in both data types in England, linkage of routinely collected laboratory surveillance and hospital admissions data on RSV has not previously been carried out.

Although both SGSS and RDS laboratory surveillance datasets contain PII, allowing linkage to HES, I have chosen to link the RDS dataset. This is because RDS contains

both RSV-positive and RSV-negative test results and therefore has a clear denominator population of tested children (unlike SGSS) (section 3.2). This has enabled me to disentangle variation in RSV-positivity from variation in RSV testing.

In this chapter, I aim to:

1. Describe the methodology used by NP and MM to link laboratory surveillance and hospital admission data for RSV in children <5 years old in England for the first time.
2. Describe all linked and linked RSV-positive admissions by patient characteristics (sex, age, birth month and comorbidities) and clinical characteristics (primary diagnosis, week of admission, recorded clinical risk factors and length of hospital stay).

6.3 Methodology

6.3.1 Data extraction

For this work, I used the RDS and HES datasets described in Chapter 3. As RDS data on RSV was only available from 2010 onwards and HES only up to 2012, data was extracted from HES from 01/08/2010 to 31/07/2012 and from RDS from 01/07/2010 to 31/08/2012. Using the time period from mid-2010 to mid-2012 allowed for analysis of 2 consecutive RSV seasons covering the post-2009 influenza A (H1N1) pandemic period. Including laboratory data from one month either side of the hospital data allowed for potential delays in testing, with restrictions on time between test and admission applied after linkage during data cleaning (see section 6.3.2).

6.3.2 Data linkage

6.3.2.1 Overview of data linkage techniques

There are two main data linkage techniques: deterministic and probabilistic. The technique chosen depends on multiple factors, such as the quality of linkage variables (completeness, as well as the degree to which they are individually and collectively able to uniquely identify an individual), time and resources, and the research question (149,150). For both methods, it is important that the datasets to be linked are first cleaned and standardised in the same way (151). This data preparation stage is important in ensuring that the linkage variables are stored in a consistent format which will facilitate successful linkage (for example, ensuring that dates are stored as DD/MM/YYYY or DD.MM.YYYY in both datasets, or ensuring that numeric variable contain no non-numeric characters).

6.3.2.1.1 Deterministic linkage

Deterministic linkage algorithms determine whether record pairs agree or disagree on one or more identifiers. Agreement is "all-or-nothing" – records are either classified as a match if the identifiers agree, or classified as a non-match if the identifiers do not agree (149). Linkage is either by a one-step strategy, where records are compared all at once on the full set of identifiers, or a multi-step strategy (also referred to as a stepwise or iterative strategy) in which records are matched in a series of progressively less restrictive steps (152). For example, the deterministic algorithm used by NHS Digital to link hospital admission records for the same individual within HES uses the following multi-step strategy (153):

1) Two records classified as a match if they match on NHS number, date of birth and sex.
2) Of the remaining records, records classified as a match if they match on local patient identifier, hospital provider, date of birth, sex and postcode.
3) Of the remaining records, two records classified as a match if they match on date of birth, sex and postcode.

However, deterministic linkage ignores that certain identifiers (or values within those identifiers) can have more discriminatory power than others do (e.g. NHS number has more discriminatory power than sex) (149). Furthermore, errors within identifiers (such as identifier "ABC1234" appearing as "ABC123A", which could be due to a typographical error during data input) cannot be accounted for. Therefore, while false matches (i.e. records from different individuals classified as a match) may be avoided during deterministic linkage, missed matches (i.e. records from the same individual classified as a non-match) are highly likely (131).

6.3.2.1.2 Probabilistic linkage

Unlike deterministic linkage, probabilistic linkage algorithms can (a) assess the discriminatory power of each identifier, and (b) determine the likelihood that two records are a match, based on the agreement and disagreement between the various identifiers (149). Probabilistic linkage therefore allows for missing values or possible errors in patient identifiers, and takes into account that certain identifiers can have more discriminatory power than others (121,125,154).

Probabilistic linkage follows the Felligi-Sunter model (155) (Equation 1). In this model, each identifier (or the component of each identifier – for example the components of date of birth are day, month and year) is assigned an agreement weight and a disagreement weight (149). These weights are log likelihood ratios based on the ability of the identifiers to discriminate between records and the probability that the values contain errors (156). The Felligi-Sunter model sums the component weights of each identifier in the j^{th} record pair to generate match likelihood scores for each record-pair (155).

Equation 1. Felligi-Sunter model of record linkage.

$$Score = \sum_{k=1}^{n} log\left[\frac{m_k}{u_k}\right]^{\gamma_k^j} \log\left[\frac{1-m_k}{1-u_k}\right]^{1-\gamma_k^j}$$

Where, for the k^{th} identifier in the j^{th} record pair:

n = number of identifiers per record

γ^j_k = observed agreement/disagreement value (1=agree, 0=disagree)

m_k = identifier agreement rate among true links (*m*-probability)

u_k = identifier agreement rate among false links (*u*-probability)

In the Felligi-Sunter model, the *m*-probability (m_k) is the probability that a field agrees given that the pair of records is a true match. For any given field, the same *m*-probability applies for all records (155). For example, an *m*-probability for the

identifier NHS number of 0.95 means that the probability of two records belonging to the same person agreeing on NHS number is 0.95. Therefore, data quality is quantified by the *m*-probabilities. Possible reasons for disagreement among records belonging to the same person include data entry errors, missing data, or instability of the value (i.e. changes over time). The *u*-probability (u_k) is the probability that a field agrees given that the pair of records is not a true match. The *u*-probability is value-specific and will often have multiple values for each field (155). It is typically estimated as the proportion of records with a specific value, based on the frequency that value occurs within the data. For example, if there are an equal number of boys and girls in the dataset, the *u*-probability for gender will be 0.5.

The *m*- and *u*-probabilities have to be estimated from the data (155). The *m*-probability is estimated using either a gold standard dataset (of known true links and non-links) or training dataset (a subset of the data, which aims to mimic a gold-standard dataset) (155). A widely used method of obtaining the *m*- and *u*-probabilities is the expectation-maximisation (EM) algorithm, which uses an iterative approach to maximise the match likelihood score for each comparison pair (157). In addition, the Felligi-Sunter model can be adapted to include an approximate comparator – an algorithm that determines how closely two values match - to allow for errors within identifiers (156). Approximate comparators can be used to calculate partial agreement weights for each identifier (for example, allowing for partial agreement between the names *Rachel* and *Rachael*, and *Angela* and *Angelina*) (149,156). These full and partial agreement weights are summed across each identifier (or, components of each identifier) to calculate the match likelihood score for each comparison pair (149).

The match likelihood score traditionally classifies record pairs into one of three groups: matches, possible matches, or non-matches (149). Possible matches are then classified as matches or non-matches using manual review (150). However, the need for manual review can be removed by the calculation of a single threshold, above which record pairs are classified as links (and below which they are classified as non-links) (155). A single cut-off threshold is the most appropriate method when

using partial agreement weights (156). This threshold can be estimated using a gold-standard or training dataset (defined above) (150). The EM algorithm has been shown to accurately calculate a single cut-off threshold in this way (155).

When linking large datasets, the number of comparison pairs can be very large (and often computationally impractical) (149). For example, if linking file A (1,000 records) with file B (10,000 records) there will be 1,000 * 10,000 = 10,000,000 comparison pairs. Blocking can be used to divide a large dataset into smaller datasets of individuals with at least one common characteristic. For example, blocking on geographical region restricts comparison pairs to those within the same defined geographical region. Though blocking can reduce the number of potential matches to a more manageable number, it can influence linkage success and is therefore not always appropriate to use (149).

In summary, the basic steps in probabilistic linkage are:

1) Estimate the *m*- and *u*- probabilities for each linking variable (e.g. using the EM algorithm).
2) Calculate agreement and disagreement weights for each identifier (or the components of each identifier) using the *m*- and *u*- probabilities.
3) Sum the individual linking weights for each variable for each record pair.
4) Compare the total linkage weight to a threshold, above which pairs are considered to be a match. This threshold is calculated using the information generated in step 1.

6.3.2.1.3 Our approach

For this analysis, we chose a probabilistic approach to allow for missing values and possible errors in patient identifiers, and to take into account that certain identifiers can have more discriminatory power than others. NP used the EM algorithm with approximate comparators to calculate the *m*- and *u*- probabilities for the components of each identifier, with a single threshold chosen for matches and non-matches (and not including a range of possible matches). NP linked patients in the RDS dataset to patients in the HES dataset (*not* laboratory records to admissions), without using any blocking criteria. I then used the date of test and date of admission (within the longitudinal RDS and HES records of each linked patient) to determine which laboratory records and admissions within these linked patients were related. The methodology and rationale for this is further described in the following sections.

6.3.3 Completeness of patient identifiable information in RDS and HES

The patient identifiable information (PII) available in both the RDS and HES extracts was NHS number, date of birth, postcode and sex. Completeness of PII is shown in Table 6-1. Of the 135,708 admissions in the HES extract, completeness of identifiers was very high, with only NHS number having <100% completeness. However, completeness of identifiers within RDS was much lower (of the 13,034 RSV-positive records and 23,052 RSV-negative records, NHS number was only 58-60% complete and postcode only 75-78% complete), with small differences between RSV-positive and RSV-negative records.

The completeness of PII also varied by laboratory (Figure 6-1). Three laboratories (Leeds, UCLH and PHE Colindale (CFI)) had no records with NHS number recorded. Completeness of NHS number amongst the other laboratories varied from 32% to 99%. Completeness of full post code varied from 34% to 100% by laboratory. NHS number was the most discriminatory identifier in the datasets, followed by postcode. Therefore, using deterministic linkage would be highly likely to result in a significant number of missed matches. Therefore, we used probabilistic linkage (using all available PII variables: NHS number, date of birth (split into components day, month and year), sex and postcode (split into components prefix and suffix)) in order to maximise the number of successful links between the datasets.

Table 6-1. Identifier completeness (%) of records in the two datasets to be linked: HES and RDS.

Identifier	Identifier completeness		
		The Respiratory DataMart System (RDS)	
	Hospital Episode Statistics (HES) *N* = 135,708 (%)	RSV-positive records *N* = 13,034 (%)	RSV-negative records *N* = 23,052 (%)
NHS number	98	60	58
Sex	100	85	86
Date of birth	100	100	100
Post code prefix	100	77	78
Post code suffix	100	75	75

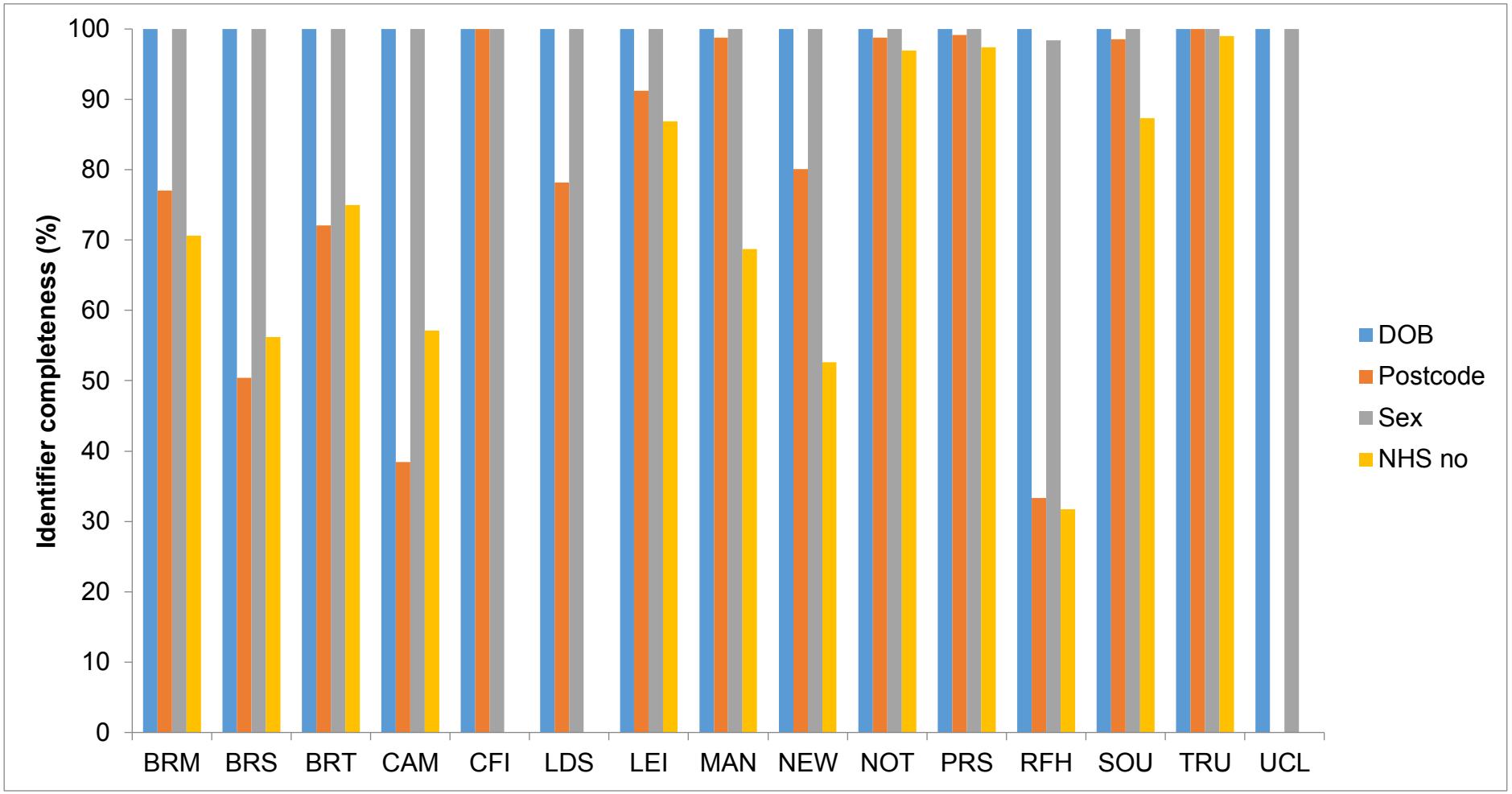

Figure 6-1. Identifier completeness (%) of RDS records, by laboratory.

6.3.4 Linkage methodology

6.3.4.1 Data pre-processing

For both datasets, the following cleaning and data preparation were undertaken: non-numeric characters within NHS numbers were removed; postcode prefix and suffix were merged into a single variable, and postcodes with invalid format set to missing (158); all dates were set to format DD/MM/YYYY; any records with unknown sex (sex is recorded in HES as M, F or U) were set to missing. All missing PII were treated as non-informative when calculating linkage weights (i.e. assigned a fixed weight of zero, rather than allowing a disagreement weight to be calculated).

The HESID variable was used to identify individual children within the HES dataset. As explained in section 3.3.2 and 6.3.2.1.1, the HESID is derived by NHS Digital using a deterministic algorithm. There are known errors within the HESID variable, with an estimated missed matched rate of 4% and an estimated false match rate of 0.2% (131). For computational reasons (to improve speed and efficiency of linkage), the HES data was divided into three datasets based on the number of episodes per person (i.e. per HESID) - one dataset for patients with only a single HES record ('Single-HES'), one for patients with multiple HES records who had no discrepancies in their identifiers in HES ('Multi-HES, no error'), and one for patients with multiple HES records who had one or more discrepancies in their identifiers in HES ('Multi-HES, with error'). For example. a patient in the 'multi-HES, with error' dataset may have 5 HES episodes within the same HESID, 3 which stated sex as Male and 2 which stated sex as Female. For these patients, any discrepancies in the within-HES PII were solved after linkage by choosing the most common value. Following linkage between each of these HES datasets and RDS, the three resulting linked datasets were merged back into one file.

6.3.4.2 Linking patients within HES and RDS

As there is no previous study linking laboratory and hospital data for RSV in England, and no published data exists on the standard timing between RSV laboratory test and hospital admission in the UK, we decided not to use blocking on date of admission and date of test. Instead, I applied restrictions on dates after linkage, following investigation of the pattern of timing between admission and test. This section therefore describes how **patients** in HES and RDS were linked probabilistically by NP.

As outlined in section 6.3.2.1.2, the EM algorithm is a widely used method of estimating the *m*- and *u*- probabilities in record linkage where no gold standard dataset is available (155). No gold-standard dataset exists to enable us to determine true links and false links within the RDS and HES data. Therefore, NP used the EM algorithm to calculate *m*- and *u*- probabilities using a training dataset.

The training dataset was generated by NP, with the aim of resembling a gold-standard dataset of true links and true non-links. Figure 6-2 outlines how this training dataset was generated: first, a dataset to resemble true links was generated using deterministic linkage (first using NHS number only, then using postcode and date of birth for the remaining records); then, of the remaining records (that did not link deterministically), a random combination of lab and HES records was used to resemble a dataset of non-links. These two datasets of links and non-links were then merged to create the training dataset.

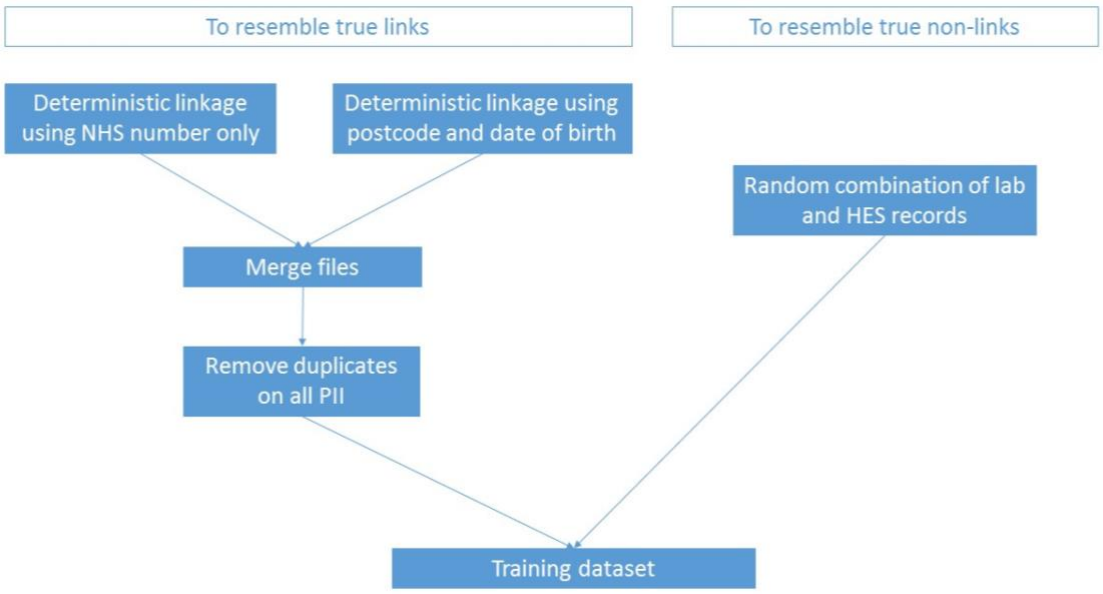

Figure 6-2. Methodology for creating a training dataset to resemble gold-standard data of true links and true non-links.

Using the training dataset, NP estimated *m*- and *u*-probabilities for the components of each identifier using the EM algorithm with approximate comparators (section 6.3.2.1.2) (156). The error rates for each identifier (for *u*-probabilities, calculated using the training data) are included in Appendix 3 (Table A3-1). NP used the *m*- and *u*-probabilities to derive the weights for each identifier component. The training dataset was then used to assign the weights derived from the EM-algorithm to each comparison record pair in the RDS-HES matched data, and the component weights for each comparison pair summed to obtain the total weight (match likelihood score) for each comparison pair.

The distribution of total weights for non-links and links in the RDS-HES matched dataset was distinct (Figure 6-3). The larger of the two peaks for links (red line, Figure 6-3) represents links that matched on NHS number as that was the most discriminatory identifier within the datasets, and the smaller peak represents the weight of record pairs that linked on other identifiers (and vice versa for non-links). A cut-off threshold for the total weight for links (calculated using the EM algorithm

(155)) was approximately the middle value between the two nearest peaks of links and non-links (weight = 15). Finally, NP applied the total weights calculated using the training dataset to all HES and RDS records; all record pairs with a total weight above the cut-off threshold were classified as links. NP extracted the longitudinal HES and RDS records for linked patients into a master dataset. I then used this master dataset to determine linked tests and admissions (the final linked dataset used for analysis) (section 6.3.4.3).

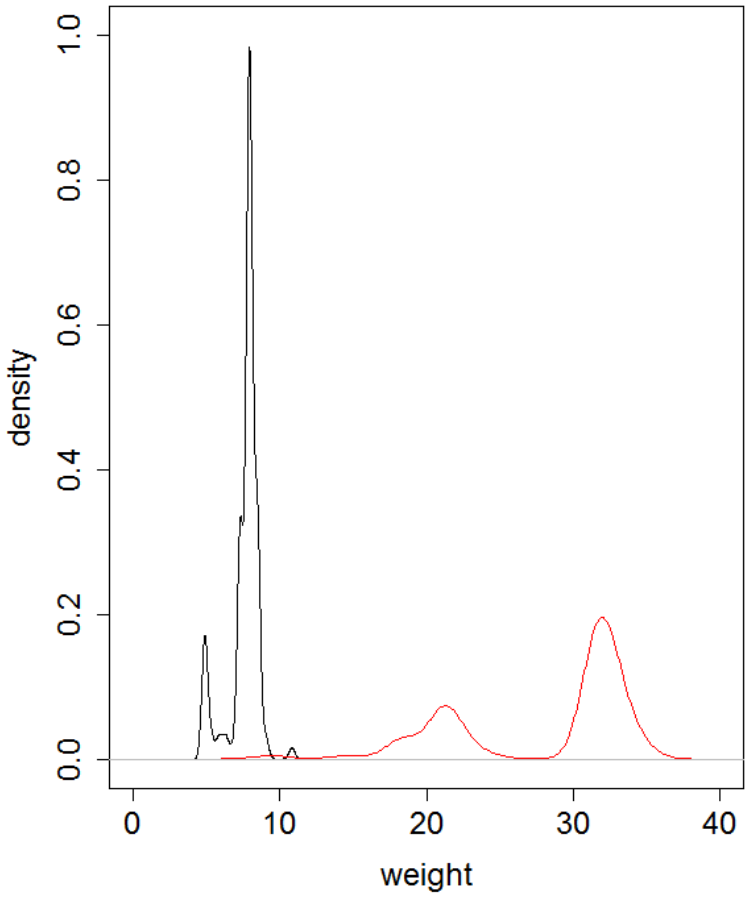

Figure 6-3. Distribution of final weights (match likelihood scores) of links (red) and non-links (black) in the training dataset.

6.3.4.3 Linking RSV tests and admissions

In the master dataset of all records from linked patients (described above), HES admissions could link to none, one or multiple RDS records, and vice versa. In addition, some admissions and tests were duplicated in this master dataset (as each possible test/admission combination for each linked patient was included in the master dataset) (Figure 6-4).

Not all of these admissions and tests in the master dataset were within a timeframe that meant the admission was plausibly due to RSV. The majority (82%) of linked RDS records were within 7 days of the start of a hospital admission (Appendix 3: Figure A3-1, Figure A3-2, Figure A3-3). Of these linked records, 70% had RSV tests within +/-1 day of admission (Figure 6-5). I restricted the final linked dataset for analysis to only those admissions beginning within +/-7 days of a linked RDS test (*So, for the example in Figure 6-4, links C, D and E in Patient 2 would be excluded*). The majority of links where the laboratory record was not within +/-7 days of admission were from the multi-HES datasets (Appendix 3: Figure A3-2, Figure A3-3).

Using a 7-day time restriction allow for short delays in testing, attempted to exclude obvious nosocomial infections (i.e. where RSV was not the cause of admission, for example in long-term hospitalised patients), and allowed for a child being tested in A&E and then later being admitted to hospital (e.g. following worsening of symptoms). I excluded all other linked records which were outside the time window, and removed duplicate admissions (i.e. admissions that appeared twice in the dataset, due to linking to >1 laboratory record within +/-7 days of the admission – the test with the date closest to the date of admission was retained).

The resulting dataset was the final linked dataset used for analysis of RSV-positive and RSV-negative hospital admissions.

(a)

HES records		
Patient number	**Admission number**	**Admission date**
Patient 1	Admission 1	01/02/2010
Patient 2	Admission 1	21/03/2010
Patient 2	Admission 2	14/10/2010

(b)

RDS records		
Patient number	**Test number**	**Test date**
Patient 1	Test 1	02/02/2010
Patient 2	Test 1	19/03/2010
Patient 2	Test 2	22/04/2010

(c)

	Linked record master dataset				
Link	**Patient number**	**Admission number**	**Admission date**	**Test number**	**Test date**
A	Patient 1	Admission 1	01/02/2010	Test 1	02/02/2010
B	Patient 2	Admission 1	21/03/2010	Test 1	19/03/2010
C	Patient 2	Admission 2	14/10/2010	Test 1	19/03/2010
D	Patient 2	Admission 1	21/03/2010	Test 2	22/04/2010
E	Patient 2	Admission 2	14/10/2010	Test 2	22/04/2010

Figure 6-4. Diagram demonstrating how duplication of records (from (a) HES and (b) RDS records) occurred in the master dataset of linked patients (c).

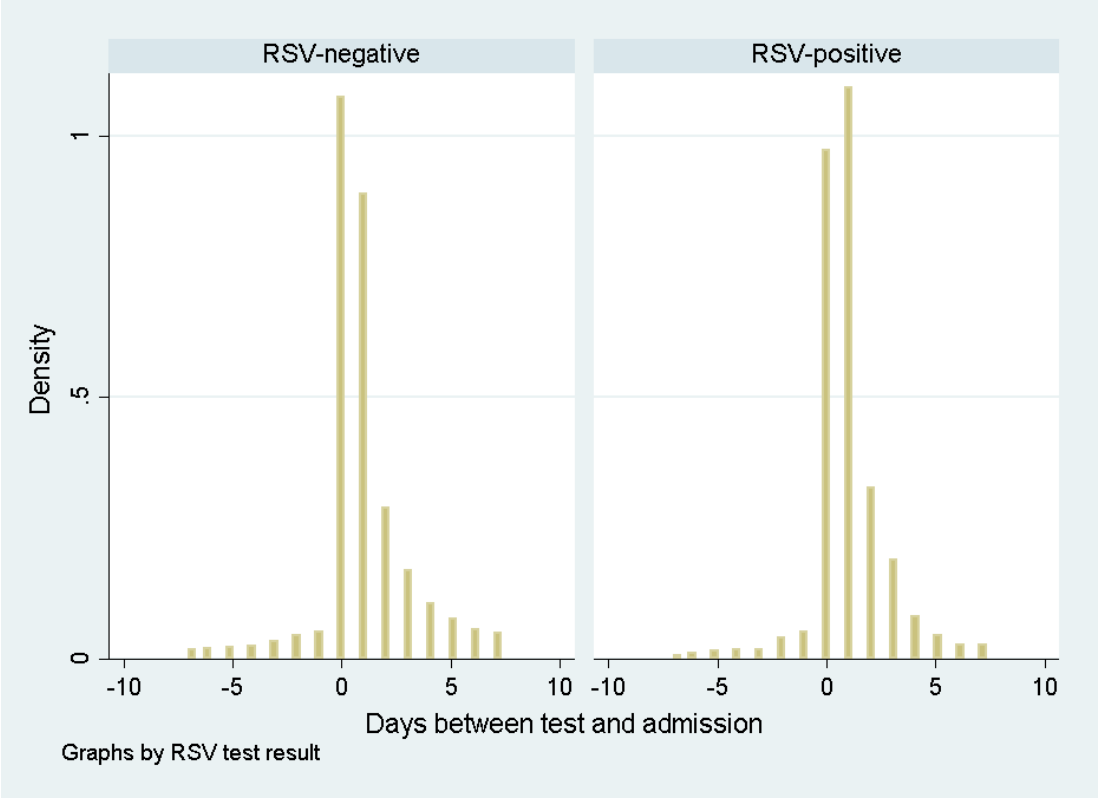

Figure 6-5. Time between test and admission date (days) for admissions included in the final linked data study population, by RSV test result (positive number of days means the test was *after* admission).

6.3.4.4 Validation of linkage

Assessment of linkage quality between RDS and HES is difficult. The most common method for validating data linkage is by comparison of the linked dataset with reference or 'gold-standard' datasets where the true match status is known (151). However, as explained previously in this chapter, no such dataset exists for RDS and HES. In addition, an unknown proportion of laboratory records in RDS will have been from primary care, A&E, and outpatient clinics; these records would not be expected to link to a HES admission.

When a gold-standard dataset is not available, potential sources of bias due to linkage error can usually be identified by comparing the characteristics of linked and unlinked data (151). However, there is no previously published literature on RSV testing practices in England, and the RDS dataset (which covers a subset of major laboratories in England, section 3.2.3) has an unknown denominator of hospitals. Therefore, it is not possible to estimate the proportion of HES admissions that should link to an RDS record.

In this chapter, I compare the characteristics of linked and unlinked data to highlight potential biases in linkage (or testing). In addition, to reduce the probability of having false matches in the final linked dataset, I use a +/-7 day cut-off (described above) to classify linked records for analysis.

6.3.5 Descriptive analysis

I summarised the admissions in the final linked dataset by laboratory, calculating the percentage of RDS records from each laboratory that linked to a HES record in the final, cleaned dataset of linked admissions. I then summarised the total number and proportion of linked admissions (total admissions (i.e. RSV-positive and negative) and RSV-positive admissions only) by the following characteristics:

- Calendar week (as defined in section 5.3),
- Sex (male or female),
- Age group at admission (<3 months, 3-5 months, 6-11 months, 1 year, 2 years, 3 years, 4 years),
- Length of stay (number of days from admission to discharge, admissions <1 day were allocated a length of stay of 0.5 days),
- Primary diagnosis (URTI, bronchiolitis, bronchitis, pneumonia, unspecified LRTI, other),
- Year (2010/11 or 2011/12, as defined in section 5.3),
- Risk group (ICD-10 diagnosis codes for prematurity, CLD, CHD, immunodeficiency and neurological disorder used to identify high-risk children, the same as used in Chapter 4 (Table 4-2) – any mention of these codes in any of the child's HES admissions (linked and unlinked) classified a child as high risk).

I calculated the RSV positivity rate as the number of RSV-positive linked admission divided by the total number of linked admissions (i.e. RSV-positive and RSV-negative).

I compared the characteristics of the linked admissions to the full dataset of HES respiratory admissions (from 01/08/2010 to 31/07/2012, cleaned as in section 3.3) to determine how the linked (i.e. laboratory confirmed RSV-positive and RSV-negative) population and linked RSV-positive population differed from the overall denominator population of respiratory admissions.

6.3.6 Statistical analysis

I used a multivariable logistic regression model to determine the patient and clinical factors associated with successful linkage between laboratory records and hospital admissions. The denominator was the original (unlinked) dataset of HES respiratory admissions (from 01/08/2010 to 31/07/2012, cleaned as in section 3.3), and the outcome was whether the admission linked or not (i.e. was included in the final linked study population, as the test was within +/- 7 days of the admission). Predictors included in the model were: sex, age group, primary diagnosis, and known clinical risk group (as a binary variable – known clinical risk group '1', no known clinical risk group '0'), and season. Non-significant (Wald test $p<0.05$) variables were removed in a backwards stepwise manner. I have reported odds ratios with 95% confidence intervals (calculated as ±1.96*SE) and p-values for each variable included in the final model.

6.4 Results

6.4.1 Linkage outcomes

Overall, 37% of RDS records in the extract linked to a HES admission with +/-7 days of the test. However, the percentage of linked records different by laboratory (Table 6-2). The laboratory with the highest percentage of linked admissions was Truro, with 88% (263/298) of records linked to (and within +/- 7 days of) a hospital admission. Manchester and Birmingham had the highest number of linked records (*n*=2,845 (42% linked) and *n*=2,503 (44% linked), respectively). The percentage of linked laboratory records was directly correlated with identifier completeness (particularly NHS number and postcode, which were the most discriminatory identifiers) (Figure 6-1), and both UCLH and PHE Colindale laboratories (which had poor PII completeness) had no linked records in the final study population.

Table 6-2. Total number of linked records per laboratory, compared to the total number of tests (RDS records) per laboratory (i.e. linked dataset compared to original unlinked dataset).

Laboratory	RDS records		
	Linked	Total	Percentage linked (%)
Birmingham	2,503	5,649	44
Bristol	1,555	5,641	28
Barts and the London	728	1,965	37
Cambridge	208	749	28
PHE Colindale	3	626	0
Leeds	1,091	4,565	24
Leicester	1,131	2,482	46
Manchester	2,845	6,713	42
Newcastle	1,051	2,859	37
Nottingham	1,259	2,466	51
Preston	252	346	73
Royal Free Hospital	23	126	18
Southampton	395	694	57
Truro	263	298	88
UCLH	0	907	0
Total	**13,307**	**36,086**	**37**

6.4.2 Comparing original dataset of (unlinked) admissions to linked admissions

The final study population consisted of 13,307 RSV positive and negative admissions in 10,626 children. Linked admissions were in younger children than admissions in the original (unlinked) dataset of all HES respiratory admissions (median age 18 months vs 7 months, respectively). Of the 13,307 linked admissions, 26% (3,469/13,307) were in children <3 months old, compared to only 11% (41,545/378,508) of admissions in the original HES dataset (Table 6-3). A higher percentage of linked admissions were in children with known clinical risk factors compared to the original dataset of HES admissions (30% vs 13%, respectively). 40% (5,368/13,307) of linked admissions were in children with a primary diagnosis of bronchiolitis, and 25% (3,312/13,307) had a primary diagnosis of 'other'.

The predictors of successful linkage included in the final model were age group, primary diagnosis, risk group and season. The odds of successful linkage were 5.04 (95% CI 4.81-5.29) times higher in children with one or more risk factors, and highest in admissions with a primary diagnosis of bronchiolitis or those in children aged <3 months (Table 6-3).

The median length of stay for all (original, unlinked) HES respiratory admissions was 1 day, compared to 2 days for linked admissions.

Table 6-3. Patient characteristics of admissions in the original HES dataset compared to the linked RDS-HES dataset. Odds Ratio's and 95% CIs for characteristics predicting successful linkage, from the final multivariate logistic regression model.

	All HES respiratory admissions	Linked admissions	Multivariable logistic regression Odds Ratio (95% CI)	P-value
Total	378,508	13,307	-	-
Sex				
Male	225,167 (59%)	7,739 (58%)	-	-
Female	153,310 (41%)	5,567 (42%)	-	-
Sex ratio (M:F)	1.4:1	1.3:1	-	-
Age group				
<3 months	41,545 (11%)	3,469 (26%)	Reference	-
3-5 months	33,291 (9%)	2,138 (16%)	0.99 (0.92-1.07)	0.784
6-11 months	60,872 (16%)	2,889 (22%)	0.93 (0.87-1.00)	0.043
1 year	89,227 (24%)	2,374 (18%)	0.65 (0.60-0.70)	<0.001
2 years	57,063 (15%)	1,039 (8%)	0.44 (0.40-0.49)	<0.001
3 years	52,109 (14%)	798 (6%)	0.38 (0.35-0.43)	<0.001
4 years	44,401 (12%)	600 (5%)	0.34 (0.30-0.38)	<0.001
Primary diagnosis				
Bronchiolitis	61,499 (16%)	5,368 (40%)	Reference	-
Bronchitis	1,260 (<1%)	46 (<1%)	0.54 (0.37-0.78)	0.001
Pneumonia	20,494 (5%)	978 (7%)	0.88 (0.80-0.96)	0.006
Unspecified LRTI	36,222 (10%)	1,318 (10%)	0.61 (0.56-0.67)	<0.001
URTI	141,344 (37%)	2,285 (17%)	0.25 (0.23-0.27)	<0.001
Other	117,689 (31%)	3,312 (25%)	0.43 (0.41-0.46)	<0.001
Risk group				
No risk	327,475 (87%)	9,367 (70%)	Reference	-
Risk factor	51,033 (13%)	3,940 (30%)	5.04 (4.81-5.29)	<0.001
Season				
2010/11	187,448 (50%)	7,074 (53%)	Reference	-
2011/12	191,060 (50%)	6,233 (47%)	0.81 (0.75-0.86)	<0.001

6.4.3 Characteristics of linked and RSV-positive admissions

There was a total of 4,476 RSV-positive admissions out of 13,307 admissions that linked to an RDS test result, giving a positivity rate of 34% (Table 6-4). Infants <3 months had the highest RSV-positivity rate (50%) and children aged 4 years the lowest (10%). RSV-positivity rate was higher in children with no known risk factors (39%) compared to children with known risk factors (20%).

Table 6-4. Linked admissions, RSV-positive linked admissions and RSV positivity rate, by patient and admission characteristics.

	Total RDS linked admissions N (%)	RSV-positive linked admissions N (%)	Positivity rate (%)
Total	13,307	4,476	34%
Sex			
Male	7,739 (58%)	2,514 (56%)	33%
Female	5,567 (42%)	1,961 (44%)	35%
Sex ratio (M:F)	1.4:1	1.3:1	
Age group			
<3 months	3,469 (26%)	1,750 (39%)	50%
3-5 months	2,138 (16%)	846 (19%)	40%
6-11 months	2,889 (22%)	950 (21%)	33%
1 year	2,374 (18%)	559 (12%)	24%
2 years	1,039 (8%)	200 (4%)	19%
3 years	798 (6%)	114 (3%)	14%
4 years	600 (5%)	57 (1%)	10%
Risk group			
No risk factor	9,367 (70%)	3,678 (82%)	39%
Risk factor	3,940 (30%)	798 (18%)	20%
Primary diagnosis			
Bronchiolitis	5,368 (40%)	3,066 (68%)	57%
Bronchitis	46 (<1%)	6 (<1%)	13%
Pneumonia	978 (%)	220 (5%)	22%
Unspec LRTI	1,317 (10%)	302 (7%)	23%
URTI	2,283 (17%)	327 (7%)	14%
Other	3,315 (25%)	555 (1%)	17%

Of the 13,307 linked admissions, 7,074 (53%) were from the 2010/11 season and 6,233 (47%) from the 2011/12 season. Of the 4,476 RSV-positive linked admissions, 2,361 (53%) were from the 2010/11 season and 2,115 (47%) from the 2011/12 season. There was a distinct seasonal pattern in RSV-positive admissions, but not in all linked admissions, following the overall pattern of RDS records (Figure 6-6, Figure 4-2). RSV-positive admissions peaked in week 52 (*n*=254) in the 2010/11 season and in week 50 (*n*=229) in the 2011/12 season.

The total number of linked admissions, the number of RSV-positive admissions, and the RSV positivity rate all decreased with increasing age (Figure 6-7). The median age of linked admissions was 7 months, compared to 4 months for RSV-positive admissions. Both linked admissions and RSV-positive admissions peaked in children aged 1 month (*n*=1,504 and *n*=780, respectively). 42% (5,607/13,307) of linked admissions and 58% (2,596/4,476) of RSV-positive admissions were in children <6 months old. 26% (3,469/13,307) of linked admission and 39% (1,750/4,476) of RSV-positive admissions were in children <3 months old. Only 5% (600/13,307) of linked admissions and 1% (57/4,476) of RSV-positive admissions were in children 4 years of age. Of the 3,546 linked RSV-positive admissions in infants aged <1 year, those in children born at the beginning of RSV season – in September, October and November – accounted for the highest number RSV-positive admissions (*n*=507, *n*=620, and *n*=707, respectively) (Figure 6-8).

The percentage of children with at least one recorded clinical risk factor generally increased with increasing age (Figure 6-9). 20% (354/1,750) of admissions in children <3 months were in children with recorded clinical risk factors. 11% (89/846) of admissions in children aged 3-5 months and 12% (117/950) of admission in children aged 6-11 months were in children with recorded clinical risk factors. However, 63% (36/57) of admissions in children aged 4 years were in children with recorded clinical risk factors. In total, 30% (3,940/13,307) of linked admissions and 18% (798/4,476) of RSV-positive admissions were in high-risk children. The most common comorbidities in the high-risk children with a linked RDS test were CHD (52% of the 3,940 linked admissions in high-risk children had CHD), CLD (44%) and

neurological disorders (31%). Similarly, the most common comorbidities recorded in the 798 RSV-positive admissions in high-risk children were CLD (48%), CHD (41%) and neurological disorders (21%).

The median length of stay for linked admissions was 2 days, compared to 3 days for RSV-positive admissions (Figure 6-10).

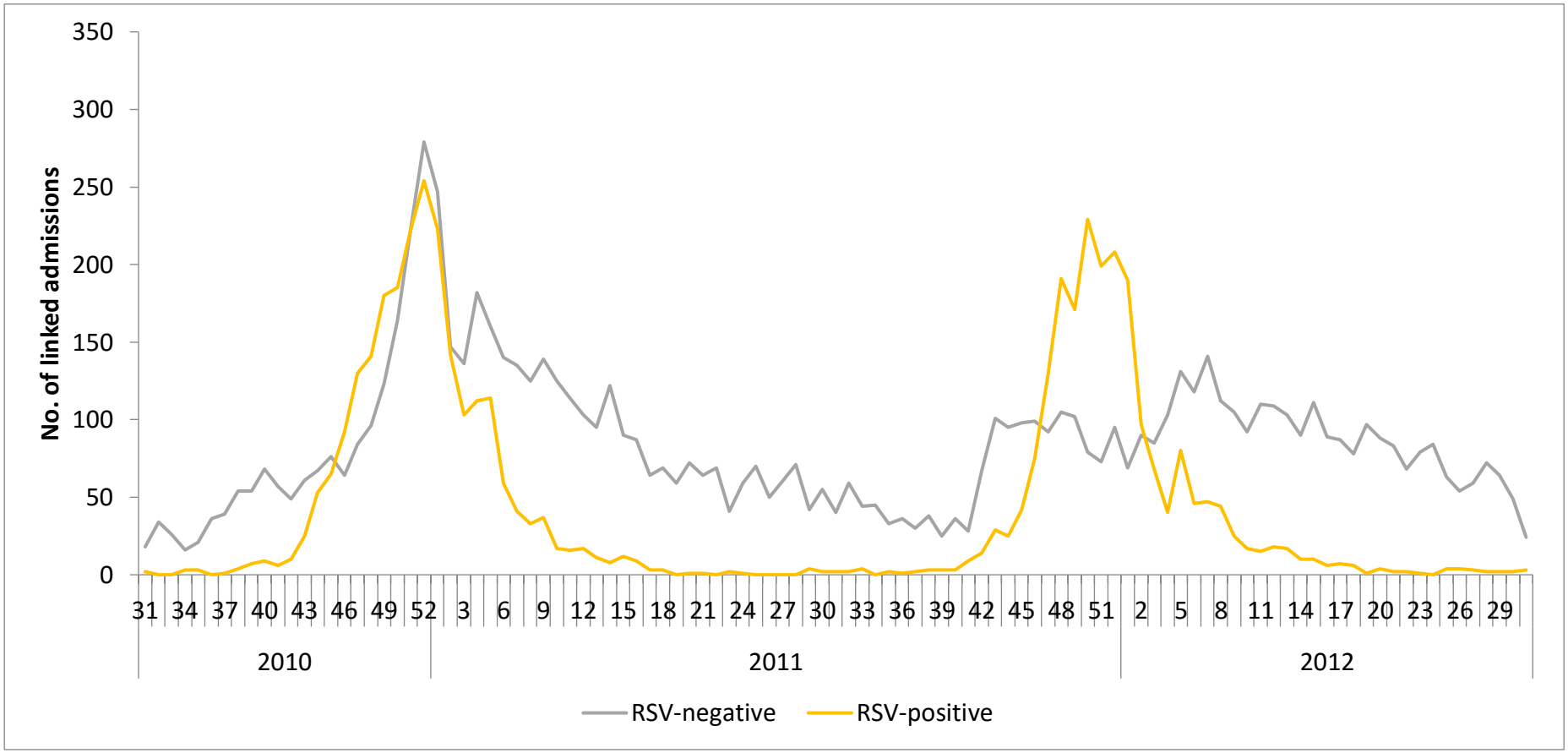

Figure 6-6. RSV-positive and RSV-negative linked admissions, by calendar week.

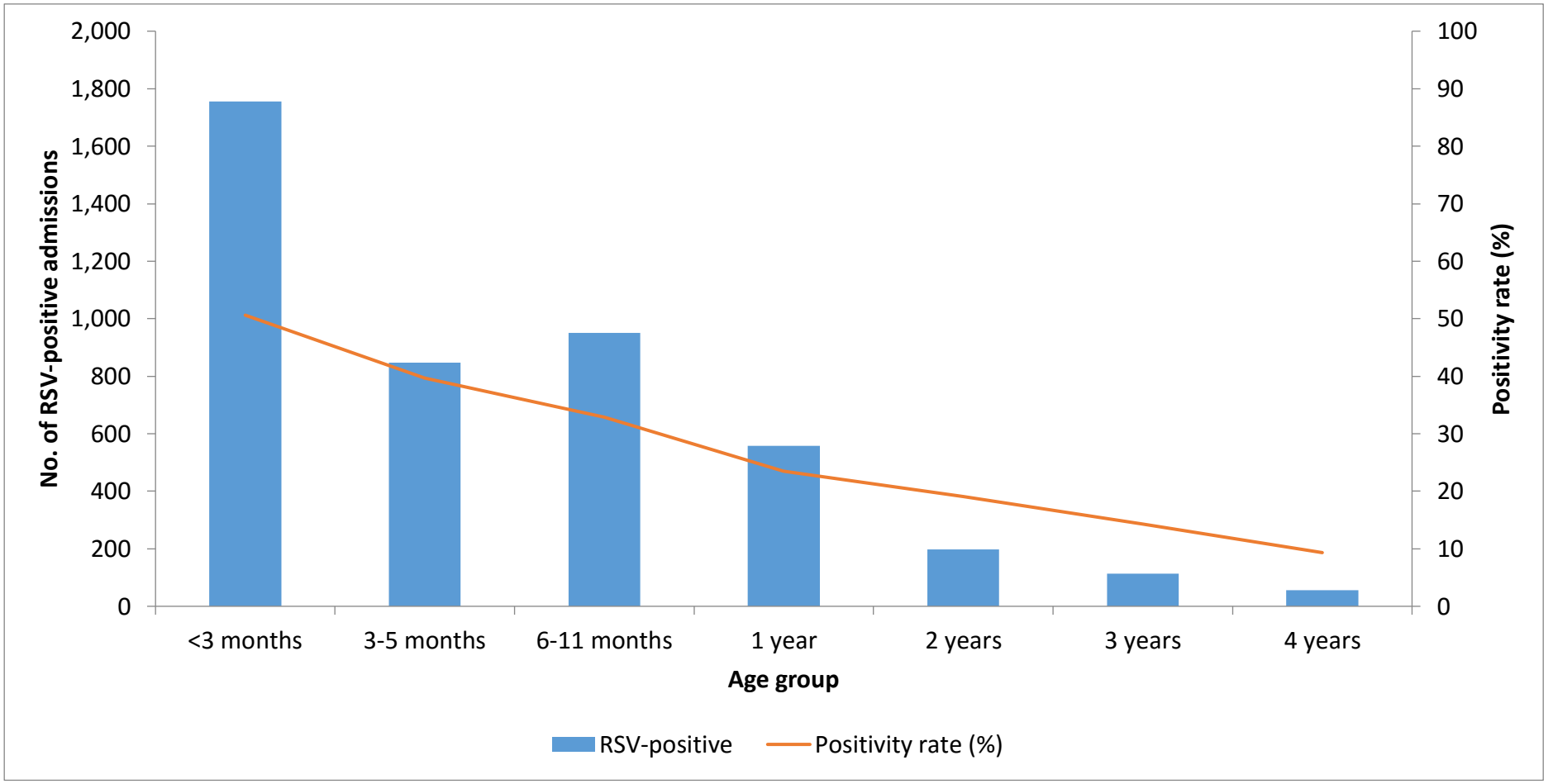

Figure 6-7. Number of linked RSV-positive admissions (columns) and positivity rate (%) (line) by age group.

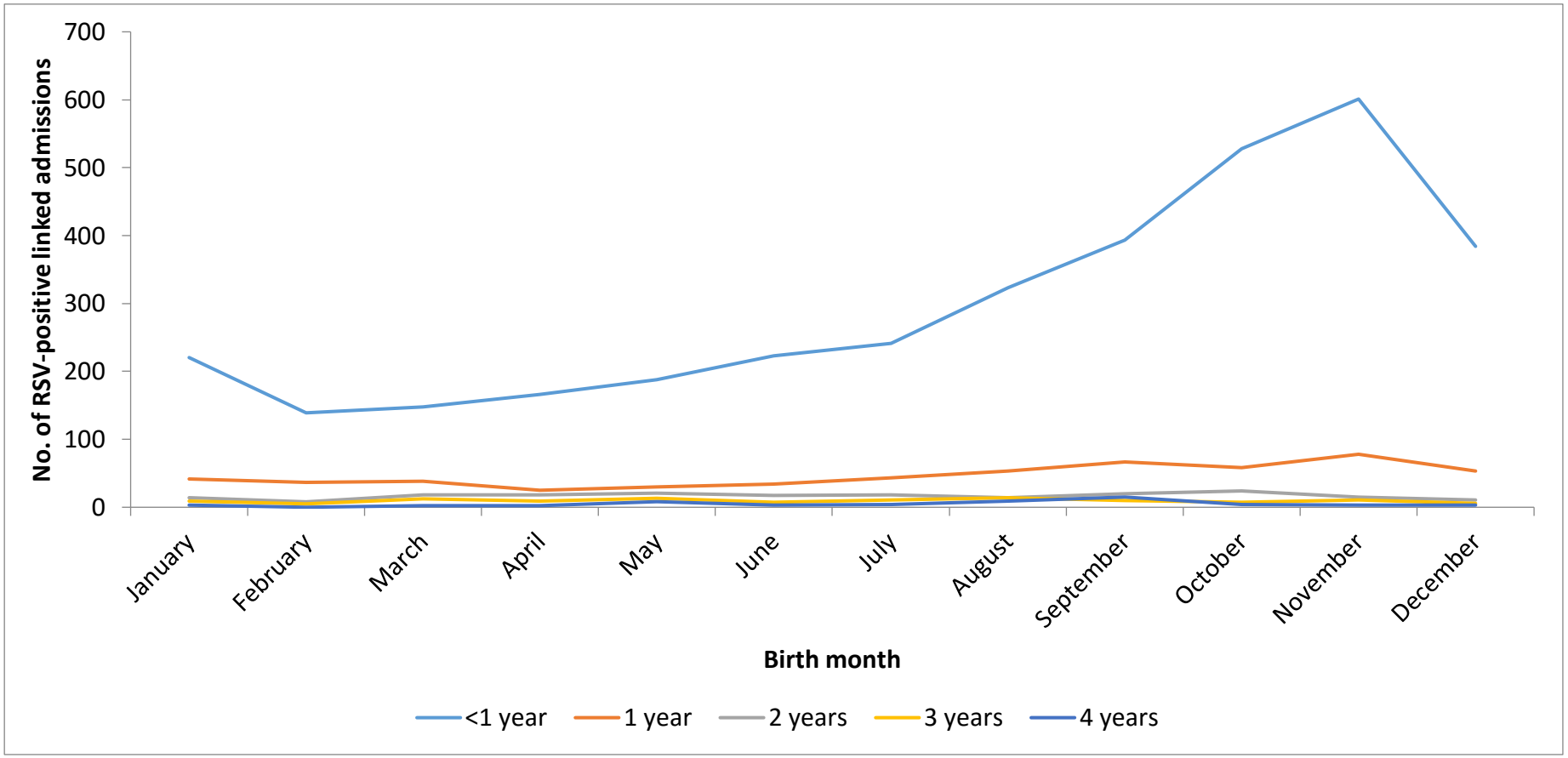

Figure 6-8. Linked RSV-positive admissions by age (in years) and birth month.

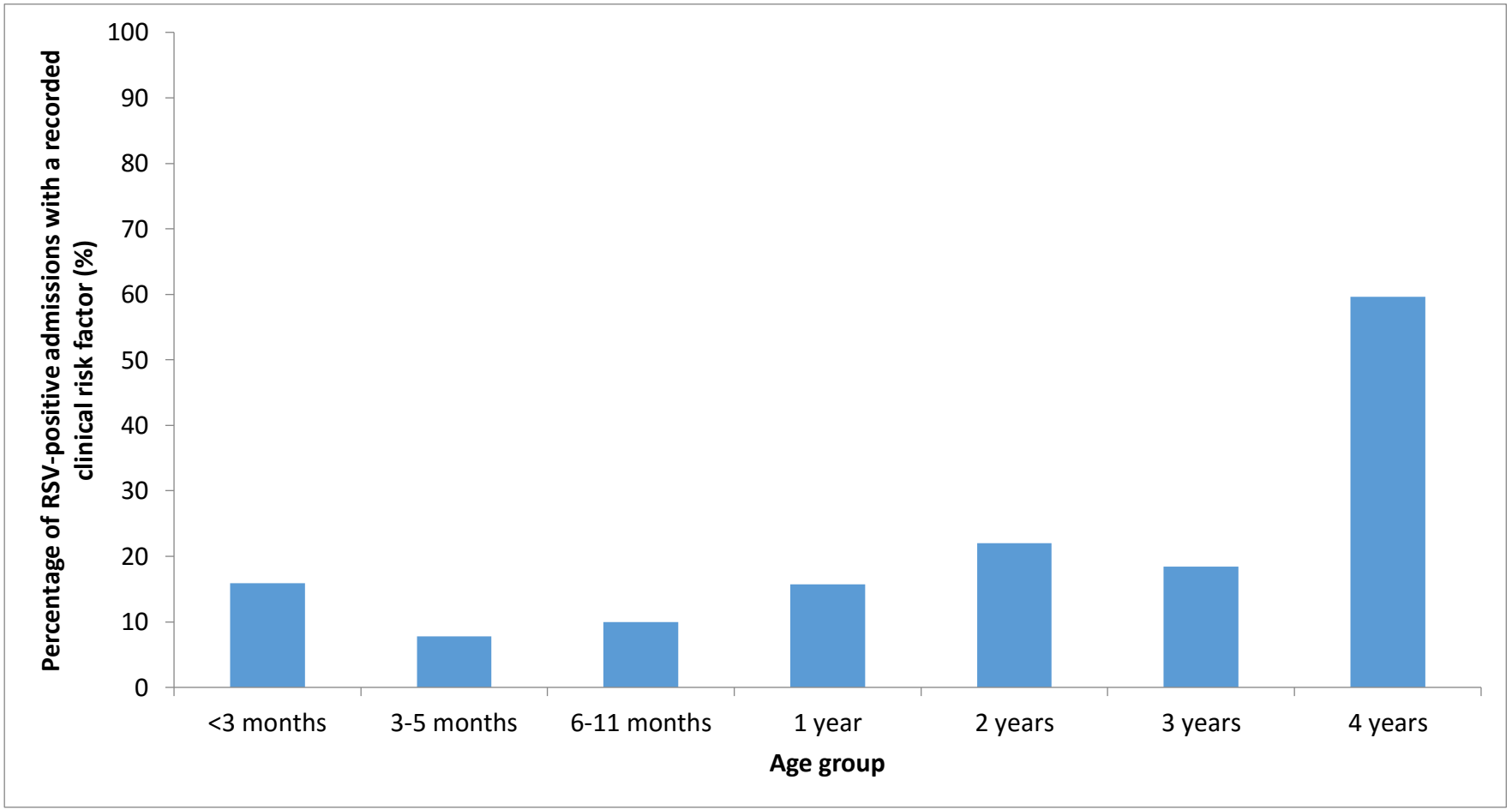

Figure 6-9. Percentage of RSV-positive admissions with a recorded clinical risk factor (%), by age group.

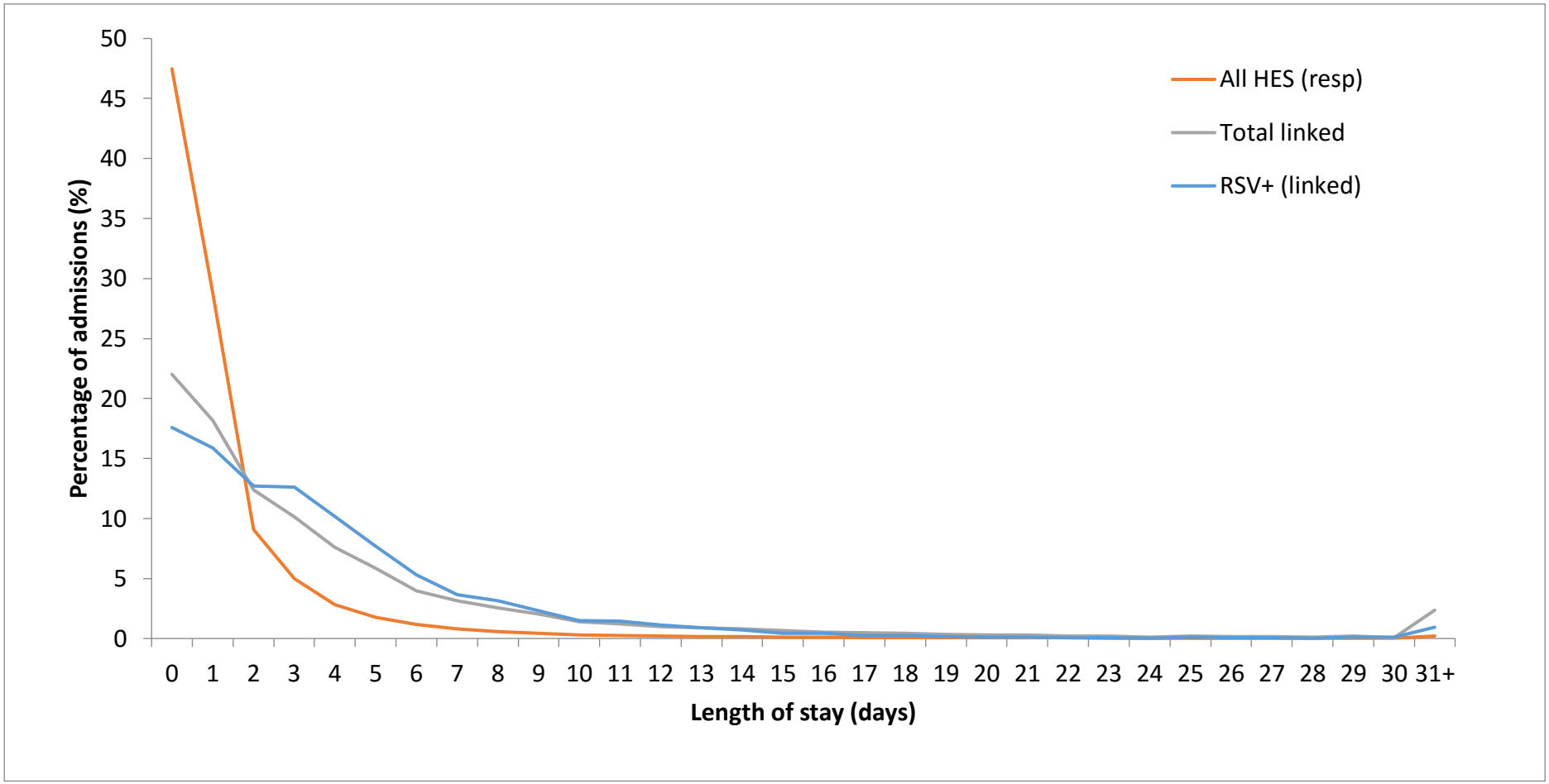

Figure 6-10. Percentage of original (unlinked) (orange), linked (grey) and RSV-positive (blue) admissions by length of stay (days).

6.5 Discussion

This is the first study in the UK using data linkage between routinely collected laboratory surveillance and administrative hospital admissions data to determine the hospital burden of RSV. The work presented in this chapter demonstrates that linkage between laboratory surveillance and hospital admissions data for RSV in England is possible, using probabilistic data linkage techniques, though assessment of linkage quality is difficult. Young age and having a known risk factor (prematurity or comorbidity) were significantly associated with linkage between HES and RDS records, indicating that these children were more likely to undergo laboratory testing to identify the causal pathogen, or more likely to have better identifiers recorded. In addition, tested (linked) children had a longer length of stay than untested (unlinked) children, suggesting that children with a longer hospital stay were more likely to undergo laboratory testing or be admitted to hospitals with better recording of identifiers on laboratory requests. Linked RSV-positive admissions were in younger children than all linked admissions or admissions in the full HES dataset. The strengths and limitations of this analysis are discussed in section 6.5.1.

Age <1 year and having a known clinical risk factor (prematurity or comorbidity) were associated with successful linkage between HES and RDS records. The total number of linked admissions, the number of RSV-positive admissions, and the RSV positivity rate all decreased with increasing age. Children with one or more known risk factors were more likely to be tested than those without known risk factors; 30% of linked admissions were in high-risk children. However, only 18% of RSV-positive admissions were in high-risk children. This proportion of RSV-positive admissions in high-risk children is similar to the proportions in the other population-based data linkage studies (in different populations) previously mentioned (114,116). Successful linkage is likely to be associated with the likelihood of testing, which suggests that these groups of children (well documented to be at higher risk of RSV infection) may more likely be tested for RSV. Furthermore, linked admissions had a longer length of stay than admissions in the full HES dataset. This suggests

that children with a longer hospital stay (an indicator of more severe disease) are more likely to undergo RSV testing and/or have better recording of identifiers which facilitates successful linkage.

My results are similar to the results of other studies using data linkage to investigate RSV-associated hospital admissions in children (114–116). The median age of RSV-associated hospital admissions in young children in my dataset was 4 months, as in the data linkage study in Ontario, Canada (116). In New South Wales (NSW), Australia, the median age of RSV-associated hospitalisations was 5 months (114). The median length of hospitalisation was also the same among data linkage studies of children admitted to hospital who have a linked, positive RSV lab result (My analysis: 3 days, NSW: 3 days, Ontario: 3 days) (114,116). 9% of RSV-associated hospital admissions in the Ontario data linkage cohort had a known clinical risk factor, compared to 18% of RSV-positive admissions in my linked dataset; however, they classified a smaller range of comorbidities as risk factors compared to my analysis (116).

6.5.1 Strengths and limitations of this analysis

This study allows the analysis of laboratory-confirmed RSV-associated hospital admissions in young children in England for the first time. The results can be used to generate estimates of RSV burden in secondary care, allowing more detailed analysis of RSV-associated hospital admissions (by age in months or weeks, and by risk group) that is required to inform vaccine policy.

However, there are a number of limitations to this analysis. The use of probabilistic linkage (as opposed to deterministic linkage) maximised the number of linked records, though linkage quality was still limited by the poor PII completeness in RDS. Only 37% of laboratory records linked to a respiratory admission that began within 7 days of the test. Completeness of PII (particularly NHS number) within RDS was low and differed significantly by laboratory; the percentage of linked laboratory records

per laboratory was directly correlated with PII completeness (particularly NHS number and postcode). If PHE the recording of NHS number in RDS in the future, this is likely to improve the number of successful links.

As there was no gold-standard dataset of true links, assessment of the linkage quality was challenging for two main reasons. Firstly, an unknown proportion of laboratory records were from primary care (see Chapter 3.1.3), and therefore would not be expected to link to a HES admission. Secondly, the RDS is a subset of laboratories in England (see Chapter 3.1.3) which has an unknown denominator of catchment populations for hospitals that submit to those laboratories – therefore, an unknown proportion of hospital admissions would not be expected to link to a RDS test.

A high number of linked records were excluded from the linked study population for having more than 7 days between the test and admission dates (Appendix 3: Figure A3-4). 59% (322/550) of these were multiple admissions in the same patients (from the multi-HES linked datasets, Appendix 3: Figure A3-2 and Figure A3-3). This is likely due to NP not applying blocking on date of admission/test, linking patients rather than admissions. Conversely, using blocking on date of test and admission could potentially increase the number of links (effectively using those dates as additional identifiers). However, not applying this blocking allowed me to investigate the timing between laboratory and hospital records to inform an appropriate cut-off, which was particularly important as linkage of these datasets had not previously been carried out. If I extended the time window to include admissions that were within 30 days of a test, this would have only resulted in an extra 550 (4%) linked admissions.

The population-based data linkage studies in other populations have used a narrow window between time of test and admission to identify RSV-associated admissions (e.g. 48 hours in Western Australia (115)). However, as viral shedding usually occurs for 3-8 days in infants and young children, I considered 7 days to be a suitable cut-off (49). Furthermore, although patients may show positive laboratory results for

RSV weeks after symptom onset (particularly those who are immunocompromised (159)), using a 7 day cut-off attempts to limit the number of nosocomial infections included (as this thesis focusses on RSV-associated hospital admissions, rather than admissions that have potentially been prolonged by nosocomial RSV infection).

A further limitation of this analysis (common to all analysis in this thesis) is that the HES extract was restricted to hospital admissions with a respiratory diagnosis. Therefore, as explained in Chapter 5.5.1, admissions with non-respiratory ICD-10 codes (such as B34.9, viral wheeze) that may also be due to RSV would not be included. In addition, as RDS only covers a subset of laboratories in England, these results are therefore an underestimate of laboratory-confirmed RSV-associated hospital admissions in England. However, I have previously demonstrated that RDS is representative of SGSS, the national laboratory surveillance system (Chapter 4). Furthermore, my results (particularly my findings by age and length of stay) are similar to the other population-based data linkage studies (114–117). This indicates that the results should be representative of laboratory-confirmed RSV-associated hospital admissions in England.

6.5.2 Implications for further analysis in this thesis

I use this linked dataset of laboratory-confirmed RSV-associated hospital admissions in Chapter 7 and 8 of this thesis, first to investigate risk factors for severe RSV-associated disease and then to estimate the total national burden of RSV in secondary care. In investigating risk factors for severe RSV-associated disease, the conclusions of my analysis need to take into consideration that children at higher risk of severe RSV infection are more likely to be tested for RSV. In estimating the total national burden of RSV in secondary care, I need to consider that these laboratory-confirmed RSV-associated hospital admissions (although likely representative of laboratory-confirmed RSV-associated hospital admissions in England) are not likely to be representative of all children with RSV-associated hospital admissions. This is particularly true for children aged 1-4 years, as these

children underwent significantly less testing than children aged <1 year. In addition, there is no RDS hospital denominator that can be derived for these linked data. Therefore, when estimating the total national burden of RSV in secondary care, I need to restrict the linked records to only those hospitals that test a relatively high percentage of respiratory admissions to maximise the representativeness of the linked admissions to all RSV-associated hospital admissions.

6.5.3 Conclusions

This work shows that linkage between laboratory surveillance and hospital admissions data for RSV in England is possible, using probabilistic data linkage techniques, though assessment of linkage quality is challenging. Improved quality of PII, particularly NHS numbers, recorded in RDS could significantly improve the quality of data linkage, improving reliability of the results.

My results demonstrate that age <1 year and having a known clinical risk factor (prematurity or comorbidity) were significantly associated with linkage between HES and RDS records, indicating that these children were more likely to undergo laboratory testing to identify the causal pathogen. Furthermore, young age (<3 months) was significantly associated with increased RSV-positivity. Therefore, as only a minority of children are tested for RSV and the linked records (particularly in children aged 1-4 years) are likely not representative of all RSV-associated hospital admissions; the analysis presented in Chapter 7 and Chapter 8 of this thesis will take these factors into consideration.

Chapter 7

Risk factors for severe RSV-associated disease

Chapter 7 Risk factors for severe RSV-associated disease

7.1 Introduction

This chapter addresses Objective 4 of this thesis: determine risk factors for severe disease (indicated by prolonged hospital stay or use of invasive ventilation) among laboratory confirmed RSV-associated hospital admissions in children younger than 5 years in England. The analysis presented in this chapter uses the laboratory-confirmed RSV-positive hospital admissions from the RDS-HES linked dataset described in Chapter 6 of this thesis.

As of August 2017, I am preparing the work presented in this chapter as a manuscript to submit for publication. This analysis is an original contribution to knowledge because:

- It is the first study of laboratory-confirmed RSV-associated hospital admissions in a UK setting to determine risk factors for severe disease among admitted children using linked laboratory and hospital data.
- I describe the distribution of clinical and demographic risk factors for severe disease (indicated by prolonged hospital stay or use of invasive ventilation) among laboratory-confirmed RSV-associated hospital admissions in England.
- My analysis confirms that premature infants and children with chronic lung disease, congenital heart disease or immunodeficiency remain at significantly increased risk of severe RSV-associated disease.

7.2 Background

RSV infection in infants and young children can lead to a broad spectrum of symptoms ranging from mild URTI to severe LRTI (section 1.3.5). In the most severe cases, children can be admitted to paediatric intensive care units (PICU) and some die as a result of their infection. It is important to identify the children most at risk of severe RSV infection in order to determine those who could benefit most from interventions that aim to either prevent infection or reduce severity of disease (23). Furthermore, one of the primary endpoints suggested by regulatory agencies for RSV vaccine trials is severe RSV-associated LRTI (section 1.3.10) (though severity is not always consistently defined) (85). Information on the children most at risk of severe RSV infection is therefore essential for economic modelling to determine the most cost-effective vaccine strategies, and to ensure that the most at-risk children can benefit from a potential future vaccine programme.

The majority of infants and young children hospitalised with RSV are born at term and are previously healthy (20–22). However, children born prematurely or those with pre-existing conditions are at increased risk of severe RSV-associated disease; more likely to be admitted to PICU or require respiratory support (23). Pre-existing conditions that are associated with a significantly higher severity and risk of death from RSV infection include those with chronic lung disease (CLD), congenital heart disease (CHD) and immunodeficiency (8,24–26). Some of these high-risk children are recommended to receive palivizumab prophylaxis (section 1.3.9), however, it is not possible to identify those who receive palivizumab as there is no national hospital prescribing database in England. As well as being important from a public health perspective, reducing disease severity is also important from an economic perspective; infants with RSV who develop LRTI have more than three times the health-care cost, on average, over the first year of life compared to infants who do not develop LRTIs (45), and patients admitted to PICU have more than four times the health-care cost, on average, compared to those on a general ward (approximately £2,597 for PICU admission compared to £535 for general admission (160)).

In the UK, small, prospective studies have been carried out which investigate predictors of RSV-associated hospital admissions in high risk infants in defined geographical areas (161,162). The largest study to investigate not just risk factors for RSV-associated hospital admissions but for severe disease (indicated by prolonged length of hospital stay) used bronchiolitis admissions as a proxy for all RSV-associated hospital admissions (110). Therefore, there is a lack of large-scale studies in the UK setting using a laboratory-confirmed endpoint which provide data on risk factors for severe RSV infection. Large scale studies can have greater generalisability, and are therefore more likely to reflect the national picture of severity. Epidemiological information on the groups most at risk of severe RSV infection is required for country-specific economic evaluations to help to determine the groups which may most benefit from current or future potential interventions (74,76) (section 1.4).

This study aims to use the unique linked dataset of laboratory and hospital data to identify risk factors for severe disease among laboratory-confirmed RSV positive hospital admissions in children younger than 5 years in England. As there is no information on PICU admissions within HES (a clear way of identifying more severe infections) I will use prolonged hospital stay and use of invasive ventilation as indicators of severe disease. A study at Royal Liverpool Children's hospital found that 98.5% of RSV-positive bronchiolitis admissions in PICU required invasive mechanical ventilation (163). Therefore, invasive ventilation is likely a reliable proxy for PICU admission in this analysis. The results of this analysis can be used in evaluations of current or future potential interventions (in conjunction with RSV burden estimates), and will enable more detailed evaluation of the groups that may benefit most from a potential future vaccine.

7.3 Methodology

7.3.1 Study population

In this chapter I use the RSV-positive hospital admissions from the RDS-HES linked dataset described in Chapter 6. This dataset contains laboratory-confirmed RSV-associated hospital admissions in children <5 years of age (at admission) from 01/08/2010 to 31/07/2012, covering two consecutive RSV seasons. Only RSV-positive RDS results successfully linked to a HES respiratory admission (within +/-7 days of the test) were included in this analysis, as described in section 6.3.4.3. From this point onwards, these admissions are referred to as *RSV-positive admissions*.

7.3.2 Defining severity outcomes

As there is no information on PICU admissions within HES, I used two indicators of severe disease: prolonged hospital stay, and requirement of invasive ventilation (including tracheostomy and tracheal intubation). 77% of RSV-positive admissions had a length of stay in hospital of 5 days or less. Therefore, I used the top quartile to define prolonged hospital stay: all admissions with a length of stay of 6 or more days were classified as having prolonged hospital stay. To identify admissions which required invasive ventilation, I used the following HES procedure codes (chosen following consultation with Sanjay Parekh, research associate at UCL Great Ormond Street Institute of Child Health, who developed a code list for indicators of higher dependency care with Padmanabhan Ramnarayan, consultant in intensive care medicine at Great Ormond Street Hospital):

- E85.1 Invasive ventilation
- E42 Tracheostomy
- X56.2 Tracheal intubation using laryngeal mask airway

Admission with any of the above codes in any of the 24 HES procedure code fields were classified as requiring invasive ventilation.

As a further marker of severity, I identified children who died in hospital during their RSV-associated hospital admission using the Discharge Method field within HES (164). However, less than 5 children in my study population died in hospital during their RSV-associated admission. Analysing these admissions would therefore put these children at risk of being identified, and I have not included death in hospital as an outcome in this analysis.

7.3.3 Defining clinical risk factors

Known clinical risk factors that I investigated in this analysis were prematurity, CLD, CHD, neurological disorder and immunodeficiency. ICD-10 codes used to identify these clinical risk factors were the same as those consistently used in this thesis, shown in Chapter 4 (Table 4-2); if a child had one or more of those ICD-10 codes in their longitudinal record of HES respiratory admissions (from 01/08/2010 to 31/07/2012), they were classified as high risk. Neither birth records (which are available in HES) nor gestational age were available in my HES extract, therefore only children with an ICD-10 code for prematurity (Table 4-2) in their longitudinal record of HES respiratory admissions were classified as premature.

7.3.4 Descriptive analysis

RSV-positive admissions, stratified according to whether they had a prolonged stay or invasive ventilation, were described by age group (<3 months, 3-5 months, 6-11 months, 1 year, 2 years, 3 years, 4 years), sex, primary diagnosis (bronchiolitis, bronchitis, pneumonia, unspecified LRTI or URTI, with remaining admissions classified as 'other'), known clinical risk factor, calendar week (defined as in section 5.3.1, and season (2010/11 and 2011/12).

7.3.5 Statistical analysis

I used a 'modified Poisson' approach to estimate the relative risk (and 95% confidence intervals) of (a) prolonged hospital stay and (b) invasive ventilation (one separate model for each) among RSV-positive hospital admissions, following the method of Zou (165) using robust error variances. This method allows the use of Poisson regression to estimate relative risk, adjusting for the overestimation of standard errors which usually occurs when Poisson regression is applied to binary data (165). I used this method to determine independent associations between each severity outcome and: age group (<3 months, 3-5 months, 6-11 months, 1-4 years), primary diagnosis, and known clinical risk factors (prematurity, CHD, CLD, neurological disorder and immunodeficiency). The variables for known clinical risk factors were treated separately, included as separate binary variables within the model. All variables were included in the initial model and then backwards stepwise regression used to remove non-significant variables (Wald test p-value >0.05).

7.4 Results

7.4.1 Overview of RSV-positive admissions

There was a total of 4,476 RSV-positive admissions in 4,011 children during the 2-year study period. 496 children had 2 admissions during the study period, and 31 had 3 or more admissions. 23% (1,046/4,476) of the RSV-positive admissions had a length of stay of 6 days or more. 4% (176/4,476) of admissions required invasive ventilation. Less than 5 (<0.1%) of these children died during their admission.

Characteristics of the study population are shown in Table 7-1. 39% (1,750/4,476) of RSV-positive admissions were in children aged <3 months. Invasive ventilation requirement decreased with increasing age: 7% (128/1,750) children <3 months old required ventilation, compared to <9% (<5/57) of children aged 4 years (Figure 7-1). However, there were very small numbers of children >1 year of age who required ventilation (*n*=24). The percentage of admissions with prolonged length of stay remained similar with age for children aged 3 months to 4 years (18-25%), but a higher percentage of children aged <3 months had prolonged hospital stay (30%, 518/1,750) (Table 7-1).

82% (3,678/4,476) of RSV-positive admissions were in children with no recorded clinical risk factor (Table 7-1). 50% (398/798) of admissions in high-risk children had prolonged length of stay, compared to 18% (648/3,678) of those with no known clinical risk factor. 9% (71/798) of admissions in high-risk children had invasive ventilation recorded, compared to 3% (105/3,678) in those with no recorded clinical risk factor. Of the admissions in children with one or more recorded clinical risk factors, admissions with prematurity recorded were most likely to require invasive ventilation (12% (8/68) of children with an ICD-10 code for prematurity required ventilation).

Although the percentage of RSV-positive admissions with a recorded clinical risk factor increased with increasing age (Chapter 6, Figure 6-8), the percentage of high-risk children with an RSV-positive admission who had prolonged hospital stay did

not substantially vary by age group (Figure 7-2). Similarly, there was no clear pattern between increasing age and requirement of mechanical ventilation among high-risk children with an RSV-positive admission (Figure 7-2).

A similar number of RSV-associated hospital admissions were in the 2010/11 season compared to the 2011/12 season (2,361 in 2010/11 compared to 2,115 in 2011/12, Table 7-1). Figure 7-3 shows RSV-positive admissions by calendar week; severe cases follow the overall seasonal pattern of RSV-associated hospital admissions.

Table 7-1. Overview of linked RSV-positive admissions, stratified by markers of severity.

	Total number of admissions	Prolonged length of stay (6 days or more) N (%)	Required invasive ventilation N (%)
Total	4,476	1,046 (23%)	176 (4%)
Sex			
Male	2,514	567 (23%)	94 (4%)
Female	1,961	479 (24%)	82 (4%)
Ratio (M:F)	1.3:1	1.2:1	1.1:1
Age			
<3 months	1,750	518 (30%)	128 (7%)
3-5 months	846	150 (18%)	14 (2%)
6-11 months	950	171 (18%)	10 (1%)
1 year	559	126 (23%)	14 (3%)
2 years	200	49 (25%)	<5[11] (<3%)
3 years	114	20 (18%)	<5 (<4%)
4 years	57	12 (21%)	<5 (<9%)
Risk status[12]			
No clinical risk factor	3,678	648 (18%)	105 (3%)
High-risk	798	398 (50%)	71 (9%)
Prematurity	*68*	*44 (65%)*	*8 (12%)*
CLD	*386*	*131 (34%)*	*25 (6%)*
CHD	*330*	*120 (36%)*	*22 (7%)*
Neurological disorder	*170*	*91 (40%)*	*6 (6%)*
Immunodeficiency	*98*	*39 (54%)*	*13 (8%)*
Primary diagnosis			
Bronchiolitis	3,066	711 (23%)	104 (3%)
Bronchitis	6	<5 (<80%)	0 (0%)
Pneumonia	220	68 (31%)	15 (7%)
URTI	327	13 (1%)	<5 (<2%)
Unspecified LRTI	302	46 (4%)	<5 (<2%)
Other	555	207 (37%)	51 (9%)
Season			
2010/11	2,361	559 (24%)	89 (4%)
2011/12	2,115	487 (23%)	87 (4%)

[11] If less than 5 admissions in a subgroup, the exact number is not specified in order to maintain anonymity of the children.

[12] 798 children had one or more of the high-risk ICD 10 codes, and is therefore included in the 'high-risk' group. The number of children with prematurity, CLD, CHD, neurological disorders and immunodeficiency totalled more than 798 as a child could have more than 1 of the comorbidities.

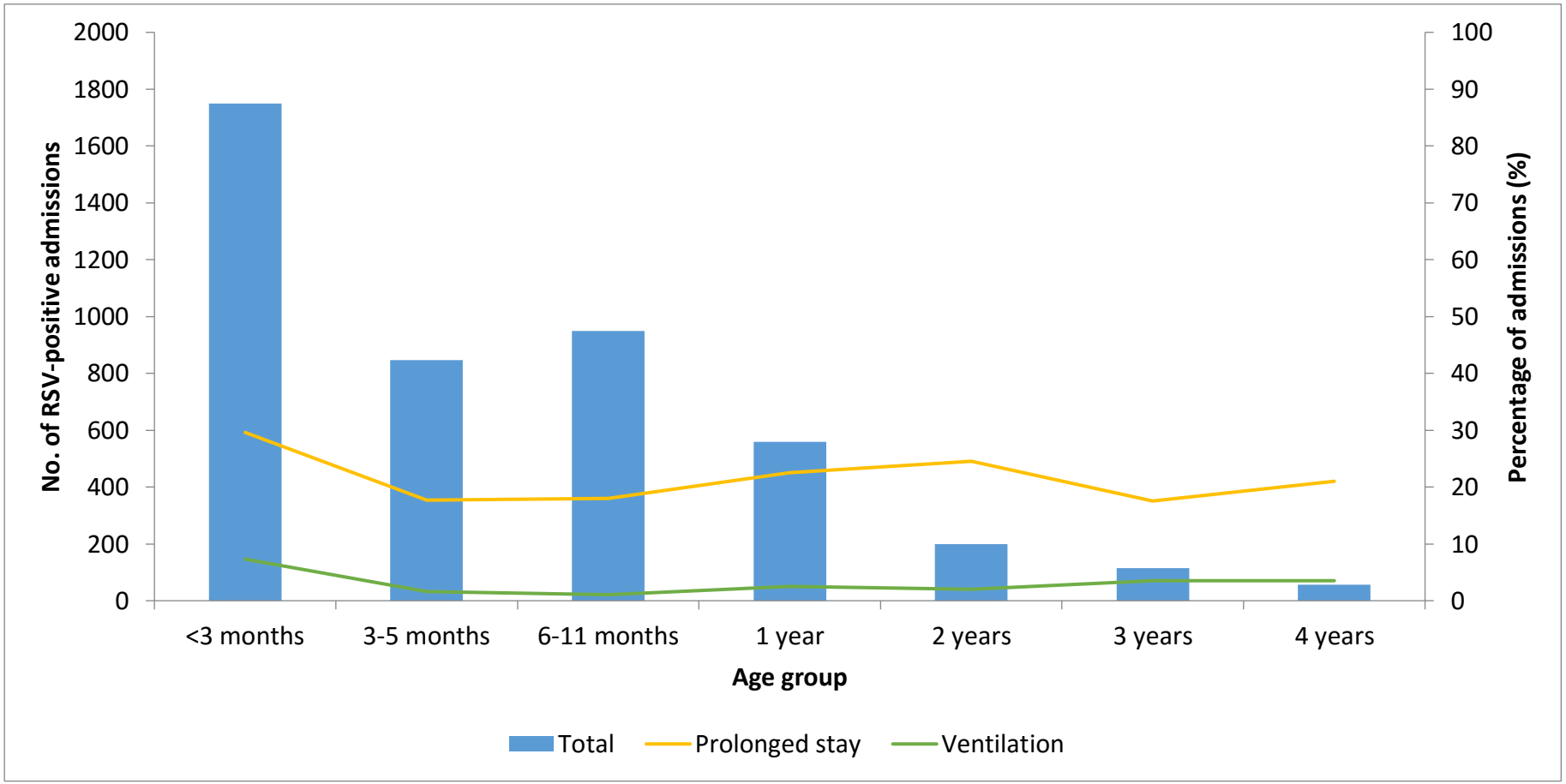

Figure 7-1. Total number of RSV-positive admissions (blue, bars), and the percentage with prolonged length of stay (orange, line) and ventilation (green, line), by age group.

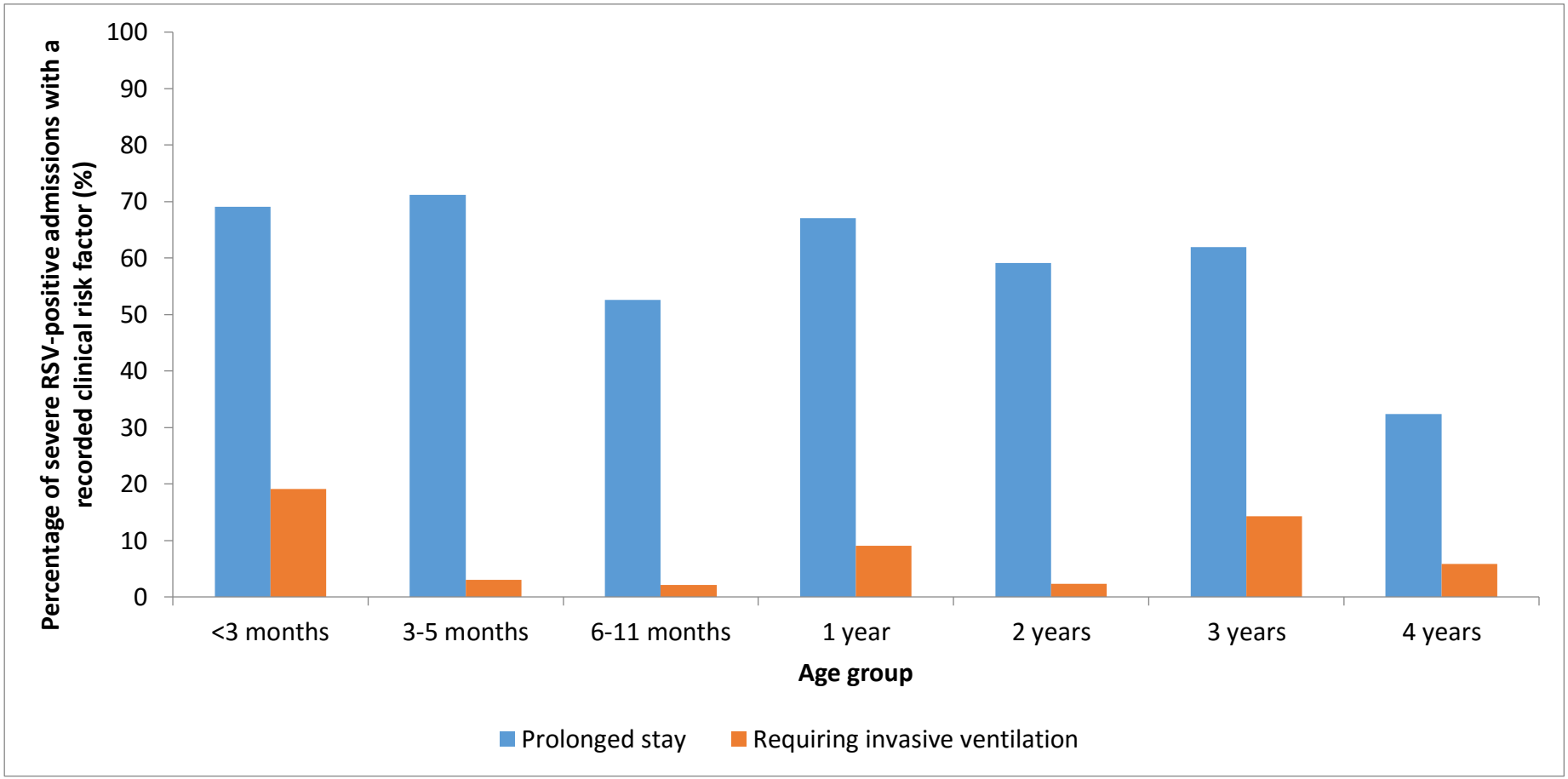

Figure 7-2. Percentage of severe RSV-positive admissions with a recorded clinical risk factor, stratified by severity outcome (prolonged stay, or invasive ventilation).

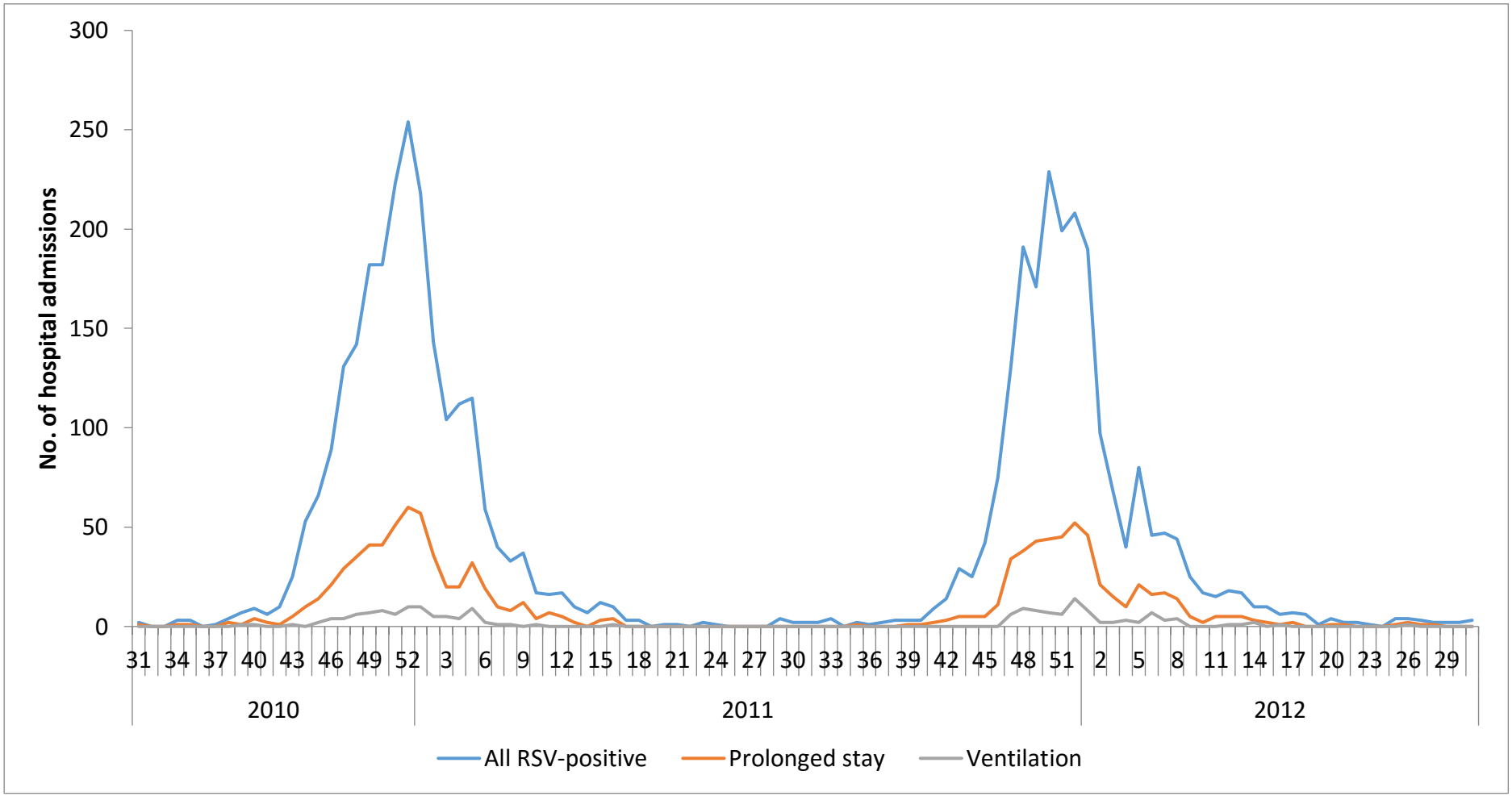

Figure 7-3. RSV-positive admissions by calendar week, stratified by severity outcome.

7.4.2 Factors associated with prolonged hospital stay

Admissions in children with known clinical risk factors, those aged <3 months, and those with a primary diagnosis of pneumonia had significantly increased risk of prolonged hospital stay (6 or more days) (Table 7-2). Admissions with CHD recorded had 2.30 (95% CI 2.02-2.61) times greater risk of prolonged hospital stay compared to those without CHD recorded. Admissions with prematurity recorded had 1.99 (95% CI 1.58-2.50) times greater risk of prolonged hospital stay compared to those without prematurity recorded.

Table 7-2. Results of final poisson regression model of relative risk of prolonged hospital stay (6 or more days) among RSV-positive hospital admissions, by key demographic and clinical characteristics.

Characteristic	Multivariate analysis relative risk (95% CI)	P-value
Age group		
<3 months	1 (ref)	-
3-5 months	0.64 (0.54-0.74)	<0.001
6-11 months	0.66 (0.56-0.77)	<0.001
1-4 years	0.73 (0.62-0.86)	<0.001
Recorded clinical risk factors		
No prematurity	1 (ref)	-
Prematurity	1.99 (1.58-2.50)	<0.001
No CLD	1 (ref)	-
. CLD	1.42 (1.23-1.66)	<0.001
No CHD	1 (ref)	-
CHD	2.30 (2.04-2.61)	<0.001
No neurological disorder	1 (ref)	-
Neurological disorder	1.58 (1.31-1.89)	<0.001
No immunodeficiency	1 (ref)	-
Immunodeficiency	1.33 (0.98-1.79)	0.066
Primary diagnosis		
Bronchiolitis	1 (ref)	-
Pneumonia	1.32 (1.04-1.66)	0.020
Unspecified LRTI	0.69 (0.52-0.92)	0.013
URTI	0.19 (0.11-0.33)	<0.001
Other	1.19 (1.03-1.37)	0.022

7.4.3 Factors associated with invasive ventilation

Admissions in children aged <3 months, or those with a diagnosis of pneumonia or 'other', had significantly increased risk of invasive ventilation (Table 7-3). Admissions with prematurity, CHD, neurological disorder or immunodeficiency recorded had higher risk of invasive ventilation, though only CHD was statistically significant (RR 2.61, 95% CI 1.69-4.03).

Table 7-3. Results of final poisson regression model of relative risk of invasive ventilation among RSV-positive hospital admissions, by key demographic and clinical characteristic.

Characteristic	Multivariate analysis relative risk (95% CI)	P-value
Age group		
<3 months	1 (ref)	-
3-5 months	0.24 (0.14-0.42)	<0.001
6-11 months	0.14 (0.07-0.26)	<0.001
1-4 years	0.22 (0.12-0.38)	<0.001
Recorded clinical risk factors		
No prematurity	1 (ref)	-
Prematurity	1.78 (0.86-3.68)	0.119
No CLD	1 (ref)	-
. CLD	0.94 (0.57-1.53)	0.793
No CHD	1 (ref)	-
CHD	2.61 (1.69-4.03)	<0.001
No neurological disorder	1 (ref)	-
Neurological disorder	1.75 (0.90-3.39)	0.096
No immunodeficiency	1 (ref)	-
Immunodeficiency	1.45 (0.59-3.56)	0.421
Primary diagnosis		
Bronchiolitis	1 (ref)	-
Pneumonia	3.65 (2.15-6.18)	<0.001
Unspecified LRTI	0.44 (0.10-1.98)	0.282
URTI	0.68 (0.26-1.80)	0.436
Other	2.47 (1.56-3.92)	<0.001

7.5 Discussion

This analysis uses the unique linked dataset of laboratory and hospital records to identify laboratory-confirmed RSV-positive admissions in children younger than 5 years and determine risk factors for severe RSV-associated disease (indicated by prolonged hospital stay or use of ventilation). 39% of RSV-positive admissions in children younger than 5 years were in children younger than 3 months. Children younger than 3 months were also at significantly increased risk of severe disease, being around five times more likely to require invasive ventilation compared to children aged 3 months to 4 years. 82% of RSV-positive admissions in children <5 years were in children with no known clinical risk factors (prematurity or comorbidity), however those with known clinical risk factors had significantly increased risk of severe disease. Of the admissions in high-risk children, those with diagnostic codes indicating prematurity or CHD had the highest relative risk of severe disease (compared to admissions without those diagnostic codes). The strengths and limitations of this analysis are discussed in section 7.5.1.

This is the first study of laboratory-confirmed RSV-associated hospital admissions using linked routinely collected data in a UK setting. My results are in line with a previous study in England that showed that 85% of infants admitted to hospital with bronchiolitis in England are born at term, with no known risk factors for severe RSV infection (110). However, my results confirm that there are clear patient groups at higher risk of severe infection: premature infants and children with CLD, CHD or neurological disorders. Although I found that children with immunodeficiency were at increased risk of prolonged hospital stay, they were not at significantly increased risk of ventilation.

My results show that infants born with prematurity recorded were nearly twice as likely to require ventilation as well as have prolonged hospital stay than those without prematurity recorded. Though this does suggest that those born prematurely are at higher risk of severe RSV-associated disease than those born at term, it should be noted that trusts with more detailed recording of comorbidities

(including prematurity) may also have more detailed and complete recording of procedures carried out – I have further described the limitations of this analysis in section 7.5.1. However, there is strong evidence for prematurity increasing the risk of severe RSV infection from other studies (166). Premature infants have an incomplete transfer of maternal antibodies, an immature immune system, and smaller, immature airways (17). The lung volume of an infant at 30 weeks gestational age is approximately three times smaller than infants born at term (17).

I found that admissions with CLD recorded were 42% more likely to have prolonged hospital stay and ventilation than those without CLD recorded. The abnormalities characteristic of children with CLD (e.g. restricted lung development, impaired alveolar development, and impaired pulmonary vascular development) predispose them to respiratory failure upon infection with RSV, and the prolonged supplementary oxygen therapy they require can cause the lung tissue to become inflamed and scarred, increasing their risk of severe LRTI (167,168). Previous studies, including in the UK, have also demonstrated that children with CLD have significantly increased length of stay compared to children without CLD (169). Previous studies have also suggested that CLD is significantly associated with the use and duration of mechanical ventilation (169). My results did not indicate that RSV-associated admissions in children with CLD were more likely to require invasive ventilation than those without CLD. However, this may be due to a power issue in my study, or difficulty in detecting children with CLD using ICD-10 codes alone.

I also found that children with CHD were 2.3 times more likely to have prolonged hospital stay and 2.6 times more likely to require invasive ventilation than those without CHD. Infants with CHD are known to be at increased risk of severe RSV infection, and to have a significantly higher mortality rate (17,26,170). The length of stay and risk of PICU admission in children with CHD varies among studies, with one study suggesting that more than half of children with CHD admitted to hospital with RSV-associated disease admitted to PICU (171). However, a systematic review of studies in the United States, Canada and Europe found that 11-20% of children with CHD required mechanical ventilation during a RSV-associated hospital admission,

which is more comparable though still higher than my results (in my analysis, 7% of children with CHD required invasive ventilation) (171). CHD increases risk of severe RSV infection by limiting cardiac output and oxygen delivery, increasing the risk of respiratory or cardiac failure and prompting the use of ventilator support (172). As infants with CHD can deteriorate rapidly after infection with RSV, these infants are recommended to be admitted to hospital for monitoring and observation following the onset of symptoms (172); this explains their increased risk of prolonged hospital stay and use of invasive ventilation shown in my results.

My results show that children with neurological disorders or immunodeficiency were at increased risk of prolonged hospital stay, but not at significantly increased risk of invasive ventilation (compared to infants without those conditions). Several studies have demonstrated that children with neurological disorders have increased risk of severe RSV infection, with increased morbidity and mortality (17,24). The increased risk of severe disease in these children is due to their reduced ability to cough (important in removing secretions from the airways) and the high prevalence of gastro-oesophageal reflux and swallowing dysfunction, which leads to aspiration (173). Patients with severe combined immunodeficiency syndrome (SCID), acquired immune deficiency syndrome, undergoing organ or hematopoietic stem cell transplantation or receiving immunosuppressive therapy are highly susceptible to severe infections, particularly those caused by viruses, including RSV (17). Among previous studies reporting on RSV infections in immunocompromised children, there appears to be considerable variation in rates of admission depending on the specific causes of the compromised immune system (174).

My results support previous studies which demonstrate that the burden of RSV in high-risk children remains high and these high-risk children are still at increased risk of severe disease (78). Though some of these high-risk children are recommended for palivizumab prophylaxis (section 1.3.9), which can reduce the severity of infection, there are no national studies of what proportion of eligible children receive palivizumab, nor the effectiveness of palivizumab in the extremely high-risk population who are eligible to receive it in the UK. Any future vaccination

programme therefore needs to ensure that vulnerable infants who are at high risk of severe RSV-associated disease are protected. The increased risk of severe RSV infection in premature infants raises questions about the timing of potential maternal vaccination in pregnancy; although vaccination late in pregnancy would ensure that antibodies last as long as possible in the infant, it should be ensured that premature babies are protected. Evaluation of a potential future universal vaccine programme should consider the impact on these high-risk children who are more likely to develop severe RSV-associated disease.

7.5.1 Strengths and limitations of this analysis

This study is the first to use linked routinely collected laboratory and hospital records to investigate risk factors for severe RSV-associated disease in a UK setting. Using a large number of laboratory-confirmed RSV admissions with national coverage improves the representativeness of the results compared to small-scale studies in single centres, and improves reliability of the results compared to, for example, studies using bronchiolitis admissions as a proxy for RSV-associated admissions. Furthermore, as outlined above, my results are comparable with the findings of previous studies.

However, further to the limitations discussed in Chapter 6.5.1 relating to the linked dataset, there are a number of specific limitations of this analysis. Firstly, I could only identify codes for prematurity and comorbidities that were recorded in the longitudinal HES record of respiratory admissions, as my HES extract only included admissions with respiratory diagnosis codes (ICD-10 Chapter X, see section 3.3). HES clinical coding guidance states that all relevant co-morbidities must be recorded for each admission (and reference cannot be made to previous admissions) (175). However, the accuracy of clinical coding varies significantly by trust, and there is consistent underreporting of comorbidities (176). Therefore, my analysis likely underestimates the number of children with comorbidities. I am particularly likely to have missed coding for premature infants, as I did not have access to birth

records (and therefore gestational age). Only 2% of my RSV-associated admissions had an ICD-10 code indicating prematurity, whereas a previous birth cohort study using HES estimated that 4.7% of bronchiolitis admissions were in infants born prematurely (110). This under-recording may have led me to underestimate (or overestimate) the risk of severe disease in these high-risk children. To overcome the potential under-recording of known clinical risk factors in future work (and enable the total size of high-risk groups to be calculated – allowing a denominator of high-risk children to be used to calculate the overall risk of severe RSV-associated admission by risk group), a birth cohort within HES could be created (110) – this would include the birth record and therefore have more complete recording of comorbidities.

Secondly, some of the high-risk children in my study population will have been recommended for palivizumab prophylaxis. However, there is no national data on hospital prescribing in England, and gestational age was not available in the HES extract that was available for this study. Therefore, I was unable to determine which children were recommended for palivizumab or whether palivizumab was administered or not. I can therefore not distinguish whether my findings that high-risk children remain at increased risk of severe RSV-associated disease are because palivizumab wasn't recommended to the high-risk children in my study population, whether it wasn't administered to all (or a proportion of) recommended children in my study population, or whether it was administered but was not effective in reducing the severity of disease in the high-risk infants in my study population to that in non-high-risk infants.

Finally, a further limitation of this analysis is that an indicator for PICU admission is not included in HES. I used prolonged hospital stay (less than a quarter of RSV-associated admissions had a hospital stay of 6 days or more) and use of invasive ventilation as proxies for severe disease. Invasive ventilation will only be used in severe cases and is therefore a reliable proxy for PICU admission (as explained in section 7.2, one study at Royal Liverpool Children's hospital found that 98.5% of RSV-positive bronchiolitis admissions in PICU required invasive mechanical

ventilation (163)). However, some very high-risk children may have been kept in hospital longer as a precaution, which may not correlate with the severity of their disease. In addition, as mentioned above, it is possible that children admitted to hospitals which have more detailed recording of comorbidities are more likely to have more detailed and complete recorded of procedures, which could increase the strength of association between these risk factors and the measured outcomes. To overcome the limitation of not having PICU admission recorded, and to allow more information of severe RSV-associated admissions requiring PICU admission to be analysed, linking to the Paediatric Intensive Care Audit Network (PICANet) (177) should be carried out.

7.5.2 Implications for further analysis in this thesis

The analysis presented in this Chapter shows that severity of RSV-associated hospital admissions differs significantly by patient characteristics, with young infants and those with known clinical risk factors at significantly increased risk of prolonged hospital stay and invasive ventilation. This difference in severity needs to be taken into account in the analysis presented in Chapter 8 of this thesis, when estimating the burden of RSV-associated admissions using the linked data. I will therefore use the linked data to not only estimate the total number of admissions in England that are due to RSV (stratified by key patient and clinical characteristics), but estimate the total number of bed days due to these RSV-associated admissions, and examine how they differ by risk group. This will enable more detailed evaluation of the groups that may benefit most from a potential future vaccine.

7.5.3 Conclusions

Previously, I have shown that young age and having a known risk factor (prematurity or comorbidity) were significantly associated with successful linkage, indicating that these children were more likely to undergo laboratory testing to identify the causal pathogen (Chapter 6). The analysis presented in this Chapter demonstrates that, among the tested children that were RSV-positive, the youngest children (<3 months old) and those with known clinical risk factors had a higher risk of severe RSV-associated disease. Evaluation of a potential future universal vaccine programme should consider the impact on these high-risk children who are more likely to develop more severe disease.

Chapter 8

Estimating the secondary care burden of RSV using linked laboratory surveillance and hospital admissions data

Chapter 8 Estimating the secondary care burden of RSV using linked laboratory surveillance and hospital admissions data

8.1 Introduction

This chapter addresses Objective 5 of this thesis: use the linked dataset to estimate the total national burden of RSV-associated hospital admissions in children younger than five years of age in England (stratified by age, primary diagnosis, calendar week and risk group). I have presented the work in this chapter (in its preliminary stages) at the 10th International Respiratory Syncytial Virus Symposium (Patagonia, Argentina) at the Public Health Science 2016 conference (Cardiff, Wales). The abstract of this latter presentation was published in The Lancet (a copy included in Appendix 4):

> *RM Reeves, P Hardelid, R Gilbert, N Panagiotopoulos, M Minaji, RG Pebody (2016) Use of linked laboratory surveillance and hospital data to estimate the burden of admissions for respiratory syncytial virus infection in children younger than 5 years in England. The Lancet, Volume 388, S99*

As of August 2017, I am preparing the work presented in this chapter as a manuscript to submit for publication. This analysis is an original contribution to knowledge because:

- I develop a novel method of using linked laboratory surveillance and hospital admissions data to develop a predictive model for RSV-positivity, and apply this predictive model to the unlinked HES data to estimate which respiratory admissions (that did not link to a laboratory test) were due to RSV.
- I produce estimates of RSV-associated admissions and bed days, stratified by age in weeks, birth month, calendar week and risk group.
- I estimate that approximately 55% of RSV-associated bed days and 45% of RSV-associated admissions in infants were in those younger than 3 months.

8.2 Background

Recent advances in the understanding of RSV biology and innovations in immunogen design have resulted in a number of potential RSV vaccines currently in Phase 2 and 3 clinical trials (section 1.3.10) (84). An effective vaccine is estimated to be available within the next 5-10 years (85). As explained previously in this thesis, there are a number of key gaps in knowledge of the burden of RSV on the health service that must be answered to provide a robust evidence base to guide vaccine policy decisions (section 1.4). One of these key gaps in knowledge of RSV burden is precise, detailed estimates of RSV-associated hospital admissions by age and risk group. However, a number of factors complicate the production of these estimates; only a minority of children undergo laboratory testing to confirm RSV as the causal pathogen of infection (section 4.2), the majority of respiratory admissions are coded in HES as having unspecified cause (section 4.3), and routinely collected hospital admissions data does not contain information on laboratory testing (section 4.3).

Time-series modelling has been used previously to estimate the national secondary care burden of RSV, and I have presented estimates of the burden of RSV in secondary care in England using this technique in Chapter 5 of this thesis. However, as these estimates are based on aggregate data they are limited in the detail they can provide – for example, my analysis groups RSV-associated admissions in children <5 years old into those in children aged <6 months, aged 6-11 months and aged 1-4 years, by primary diagnosis. However, estimates of the burden of RSV in secondary care also need to include stratification by risk group, as I have demonstrated the importance of known clinical risk factors in the increased severity of RSV-associated disease. In addition, with proposed potential vaccine strategies including maternal vaccination (to protect newborns) or a direct vaccination of infants (potentially young infants aged <6 months, or older infants, section 1.3.10.2), it is important to have estimates of RSV burden stratified at least by month of age to use in vaccine impact studies. Furthermore, with my previous analysis demonstrating the importance of the timing of birth in relation to RSV

season in the risk of RSV-positivity (section 4.2), burden estimates should also consider including month of birth as well as age in months.

Worldwide, few studies have utilised population-based data linkage to describe the aetiology or pathogen-specific burden of acute lower respiratory infections (115). In this thesis, I have demonstrated that linkage of routinely collected laboratory surveillance and hospital admissions data allows laboratory-confirmed RSV-associated hospital admissions to be described in detail (by key demographic and clinical characteristics). Furthermore, as this linked dataset contains information on both RSV-positive and RSV-negative hospital admissions, I have developed a novel methodology of using this linked data to predict RSV-positivity among linked admissions. Using this predictive model, I can then estimate the total national burden of RSV-associated admissions, and produce stratified estimates of the number of RSV-associated admissions by age and risk group (as well as other demographic and clinical characteristics), in sufficient detail to be used in vaccine impact studies.

In this chapter, I aim to:

1. Define the HES denominator population of the linked data (as RDS covers a subset of laboratories and therefore only a subset of HES providers, section 8.3.2).
2. Use this subset of linked admissions (from denominator providers only) to generate a predictive model for RSV-positivity among linked admissions.
3. Apply the predictive model to the unlinked (i.e. untested) HES respiratory admissions to predict which admissions were RSV-associated.
4. Generate stratified estimates of the total annual number of RSV-hospital admissions (and associated bed days) in England by key patient and clinical characteristics.

Due to the substantial differences in testing and clinical presentation between the age groups <1 years and 1-4 years (significantly more testing was carried out in the <1 year age group (i.e. there were more linked admissions in children <1 year old, and the RSV-positivity rate was significantly lower in children 1-4 years, section 6.4), I decided to carry out separate analysis for admissions in children aged <1 year and those in children aged 1-4 years. However, it was not possible to generate a good-fitting predictive model for admissions in children aged 1-4 years (see Appendix 5). This is most likely because the older tested children constituted a small, unrepresentative subgroup of all RSV-associated admissions in this age group. An average of 22% of respiratory admissions in children aged <1 year (in HES providers considered to be part of the RDS denominator population, section 8.4) linked to an RDS test, compared to only 4% in children aged 4 years (Appendix 5).

As a consequence of this, the analysis presented in this chapter estimates the burden of RSV-associated admissions in children <1 year of age only. Baseline epidemiology data in this young age group is most crucial in determining whether an indirect (i.e. maternal) or direct immunisation strategy is the most effective strategy to protect the young infants (<6 months) most at risk of RSV-associated admissions and more severe infection. Therefore, focusing only on the <1 year age group in this analysis still allows me to achieve the main aim of this thesis (Chapter 2). More intensive testing in some hospitals already reporting to RDS could allow sufficient data to be included in a RDS-HES linked dataset to investigate the burden of RSV-associated admissions in children aged 1-4 years. This could also be extended to infants aged <1 year to validate the results that I present in this thesis. Alternatively, a single- or multi-site prospective study (collecting detailed information, including laboratory testing, on respiratory admissions in children aged 1-4 years, such as in Leicestershire (178)) could be a successful approach to investigate the burden of RSV-associated admissions in children aged 1-4 years.

8.3 Methodology

8.3.1 Data extraction

In this chapter I use RSV-positive and RSV-negative hospital admissions in children <1 year of age only (at admission) from the RDS-HES linked dataset described in Chapter 6. This dataset contains laboratory-confirmed RSV-associated hospital admissions from 01/08/2010 to 31/07/2012, covering two consecutive RSV seasons. Only admissions with an RDS record within +/-7 days of the admission were included in this analysis, as described in section 6.3.4.3.

8.3.2 Population used for modelling

The RDS dataset collected RSV-positive and RSV-negative test results from a subset of major laboratories in England (section 3.2.3). There is not a clear denominator of hospitals that submit to RDS laboratories and this is the first study to look at RSV testing practices in English hospitals. Therefore, in order to determine a suitable denominator for this analysis, I first explored the percentage of respiratory admissions in each NHS trust (identified using the provider code 'procode3' in HES, section 3.3.3) that linked to an RDS test (in the final linked study population).

In the study population used for the predictive modelling in this chapter, I only included data from NHS trusts which linked to >50 RDS records and tested >10% of their total respiratory admissions in infants younger than 1 year for RSV. I generated a predictive model for RSV-positivity using a random 2/3 sample of this study population, and used the remaining 1/3 as the test sample.

8.3.3 Statistical analysis

Multivariable logistic regression was used to determine factors predicting the probability of testing positive for RSV among tested children. The outcome variable was whether the linked admission was positive for RSV or not. The independent variables investigated as potential predictors of an RSV-positive test among tested children were: age (continuous variable), any diagnosis of bronchiolitis (indicator variable), any diagnosis of unspecified LRTI (indicator variable), any diagnosis of pneumonia (indicator variable), any diagnosis of URTI (indicator variable), ICD-10 code indicating RSV as cause of disease (indicator variable) and risk group (indicator variable). All potential predictors were included in the initial model, and then backwards stepwise regression used to remove non-significant variables (likelihood ratio test $p>0.05$). To adjust for the seasonality of RSV circulation, one sine and one cosine function was introduced into the model, following the methodology of Stolwijk (179) and Edwards (180). I constructed receiver operator characteristic (ROC) curves by plotting the true-positive rate (sensitivity) against the false-positive rate (1-specificity). The model was then validated using the test sample, and the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) calculated to explore predictive accuracy of the model (Appendix 6: Table A6-2). A predicted probability cut-off of 0.5 (PPV 79%, 95% CI 77-81%, Appendix 6) was chosen in order to maximise both sensitivity and specificity (Appendix 6: Table A6-1). I assessed the fit of the final model using the test sample (comparing predicted and true positive admissions by calendar week, age in months, and primary diagnosis).

8.3.4 Estimating the secondary care burden of RSV in England

I applied the final logistic regression model (to predict RSV-positivity) to the whole HES extract of respiratory admissions in children <1 year in England, including the admissions which were included in the model fitting process. A cumulative frequency curve of predicted probabilities is included in Appendix 6. All admissions

with predicted RSV-positivity equal to or higher than the cut-off probability predicted from the model were classified as RSV-associated admissions. 95% CIs for the predicted probabilities were calculated by first generating the CIs for the linear predictors and then converting them to probabilities[13].

I described RSV-associated admissions and the associated number of hospital bed days of these admissions by calendar week, age, birth month, and risk group. Calendar weeks were defined as in section 5.3. Admissions with a length of hospital stay of <1 day were counted as accounting for 0.5 bed days. Admission rates were calculated using ONS mid-year population estimates for children <1 year old in England (140,141); an average of the ONS mid-year population estimates for 2010 and 2011 were used for the 2010/11 season an average of the ONS mid-year population estimates for 2011 and 2012 were used for the 2011/12 season (140,141).

[13] http://www.stata.com/support/faqs/statistics/prediction-confidence-intervals/

8.4 Results

8.4.1 Overview of population used for modelling

I included data from 24 out of 178 potential NHS trusts recorded in my HES extract with >10% of respiratory admissions linked to an RDS test (and >50 linked admissions) (Table 8-1). This resulted in final dataset of 6,758 linked admissions in infants <1 year old (used to generate the predictive model) - this included 80% of all linked admissions in infants <1 year old.

Of the 6,758 linked admissions, 44% (*n*=2,947) were positive for RSV. 49% (*n*=1, 445) of RSV-positive linked admissions were in infants aged <3 months, which was the age group with the highest RSV positivity rate (53% of linked admissions were RSV-positive) (Table 8-2). 81% (*n*=2, 375) of RSV-positive linked admissions were in infants with a primary diagnosis of bronchiolitis. 85% (*n*=2,502) of RSV-positive linked admissions had no known risk factor (comorbidity or prematurity) for hospital admission.

There were 3,221 linked admissions in the 2010/11 season and 3,573 in the 2011/12 season (Table 8-2, Figure 8-1). In the 2010-11 season, there was a peak in admissions during week 51 (*n*=142) and a peak in positivity-rate during week 50 of 70% (107/158). In the 2011-12 season, there was a peak in admissions during week 50 (*n*=178) and a peak in positivity-rate also during week 50 of 83% (178/215).

Table 8-1. NHS providers included in the study population used for modelling.

Provider code	Provider name	Linked admissions	Total HES admissions	% tested
RR1	Heart of England NHS Foundation Trust	2,582	1,099	43%
RX1	Nottingham University Hospitals NHS Trust	1,461	836	57%
RA7	University Hospitals Bristol NHS Foundation Trust	1,399	680	49%
RM2	University Hospital of South Manchester NHS Foundation Trust	779	321	41%
RWE	University Hospitals of Leicester NHS Trust	1,924	684	36%
RNJ/ R1H	Barts and The London NHS Trust	870	246	25%
RW3	Central Manchester University Hospitals NHS Foundation Trust	1,967	490	25%
RD1	Royal United Hospital Bath NHS Trust	852	207	24%
RQX	Homerton University Hospital NHS Foundation Trust	602	132	22%
RAE	Bradford Teaching Hospitals NHS Foundation Trust	1,240	296	24%
RTD	The Newcastle Upon Tyne Hospitals NHS Foundation Trust	1,781	223	13%
RR8	Leeds Teaching Hospitals NHS Trust	1,947	271	14%
RHM	Southampton University Hospitals NHS Trust	1,722	211	12%
REF	Royal Cornwall Hospitals NHS Trust	1,294	213	16%
RNH	Newham University Hospital NHS Trust	346	56	16%
RR7	Gateshead Health NHS Foundation Trust	653	98	15%
RM3	Salford Royal NHS Foundation Trust	543	64	12%
RWJ	Stockport NHS Foundation Trust	1,205	156	13%
RNQ	Kettering General Hospital NHS Foundation Trust	904	100	11%
RGT	Cambridge University Hospitals NHS Foundation Trust	540	68	13%
RCB	York Teaching Hospital NHS Foundation TRUST	835	89	11%
RMP	Tameside Hospital NHS Foundation Trust	831	96	12%
RVW	North Tees and Hartlepool NHS Foundation Trust	1,040	122	11%
Total		27,317	6,758	

Table 8-2. Characteristics of the linked dataset used to generate the predictive model.

	Total N (%)	RSV-positive N (%)	RSV positivity rate (%)
Total	6,758	2,947	45%
Male	4,011 (59%)	1,672 (57%)	42%
Female	2,747 (41%)	1,275 (43%)	46%
Ratio (M:F)	1.5:1	1.3:1	
Age (months)			
<3	2,744 (41%)	1,445 (49%)	53%
3-5	1,701 (25%)	700 (24%)	41%
6-11	2,313 (34%)	802 (27%)	35%
Risk group			
No risk	5,070 (75%)	2,502 (85%)	49%
Risk factor	1,688 (25%)	445 (15%)	26%
Primary diagnosis			
Bronchiolitis	4,047 (60%)	2,375 (81%)	59%
Bronchitis	31 (<1%)	6 (<1%)	19%
Pneumonia	290 (4%)	72 (2%)	25%
URTI	817 (12%)	139 (5%)	17%
Unspecified LRTI	299 (4%)	61 (2%)	20%
Other	1,274 (19%)	24 (10%)	23%
Season			
2010/11	3,221 (48%)	1,380 (47%)	43%
2011/12	3,537 (52%)	1,567 (53%)	44%

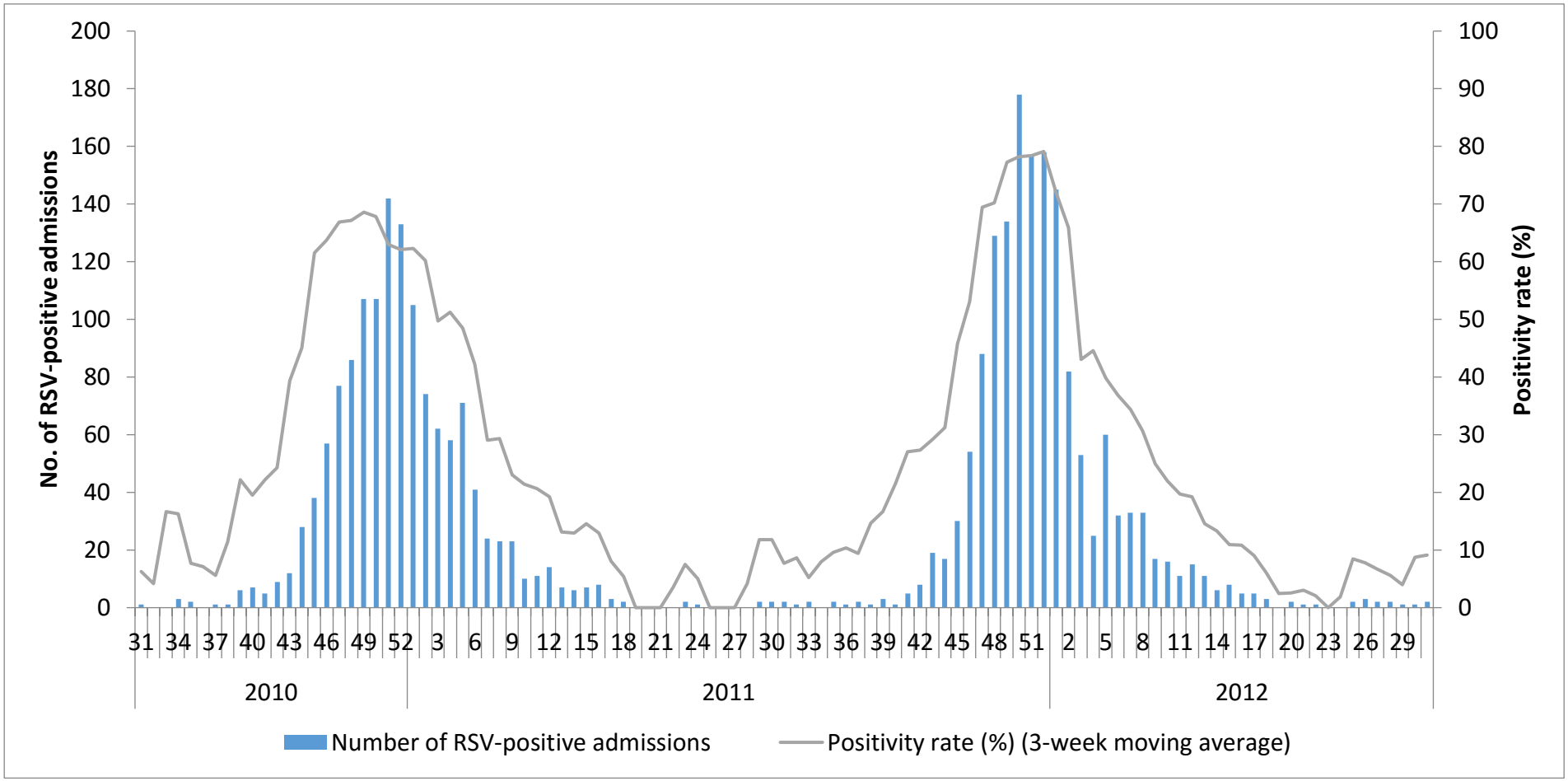

Figure 8-1. Number of RSV-positive linked admissions and RSV-positivity rate (no. of RSV-positive/total linked, as a 3-week moving average), by calendar week.

8.4.2 Predicting RSV-positivity in linked admissions

The variables included in the final logistic regression model included any diagnosis of bronchiolitis, any diagnosis of unspecified LRTI, RSV-specific diagnosis code, any code indicating high risk status, age, and the cyclical function of calendar week. The equation of the final model was as follows:

$$\log\left[\frac{p}{1-p}\right] = -2.55 + 1.31 * bronchiolitis + 0.54 * unspecifiedLRTI + 2.55 * RSVcode + 0.19 * sex - 1.14 * riskgroup - 0.52 * age - 0.57 * sin(\frac{2\pi t}{52}) + 1.78 * cos(\frac{2\pi t}{52})$$

Infants with a diagnosis of bronchiolitis, unspecified LRTI or with an RSV-specific code had higher odds of RSV-positivity (OR=3.70 (95% CI 3.03-4.51), OR=1.72 (95% CI 1.19-2.50, and OR=12.77 (95% CI 10.06-16.20), respectively). Infants with a known risk factor (i.e. comorbidity or prematurity) had reduced odds of RSV-positivity (OR=0.32, 95% CI 0.26-0.39). RSV-positivity was significantly associated with calendar week and age.

8.4.2.1 Assessing model fit

The area under the ROC curve of 0.9 indicated that the final model had good predictive accuracy (Appendix 6: Figure A6-1). The PPV of the final model was 79% (95% CI 77-81%) and the NPV 86% (95% CI 84-87%) (Appendix 6). The sensitivity and specificity of the model were 82% (95% CI 79-84%) and 84% (95% CI 81-86%), respectively (Appendix 6).

The model was a good fit for the seasonality and age distribution of RSV-associated admissions; however, the model slightly overestimated the number of RSV-positive admissions in infants aged 1 and 2 months (Figure 8-2, Figure 8-3). The model over-predicted bronchiolitis and underestimated the number of RSV-associated pneumonia and URTI admissions (Figure 8-4).

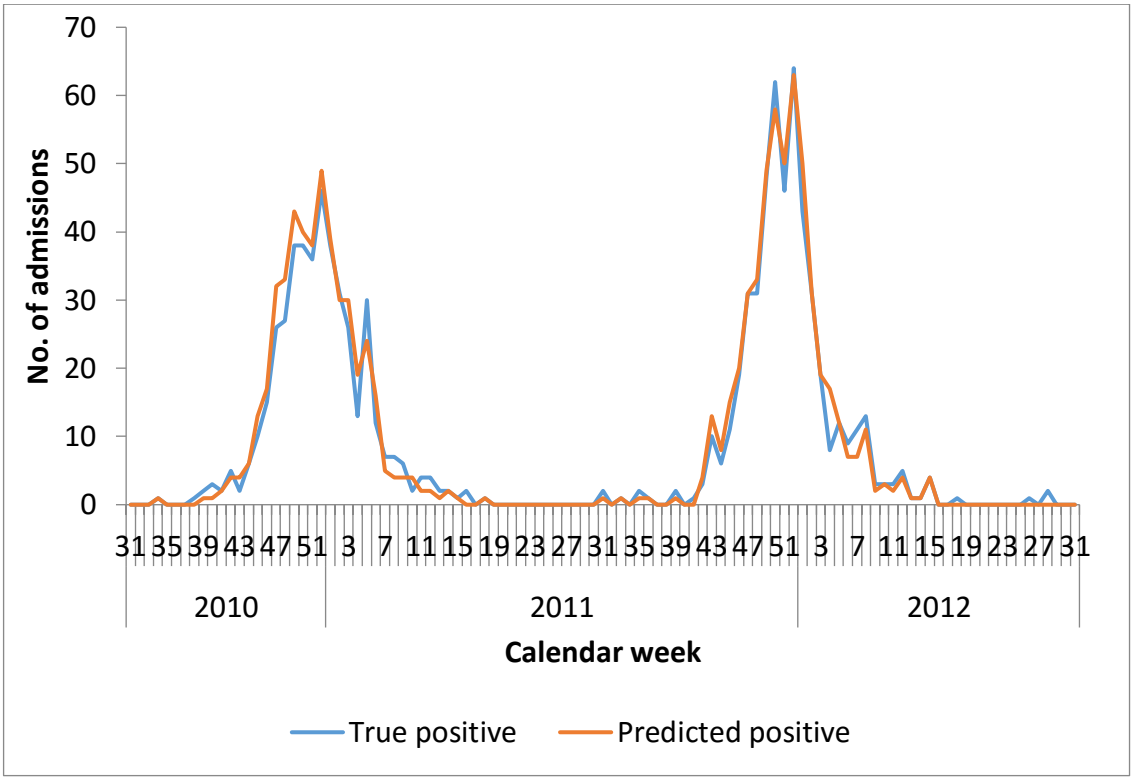

Figure 8-2. Number of true RSV-positive admissions and the number predicted due to RSV by the final model, in the test sample, by week.

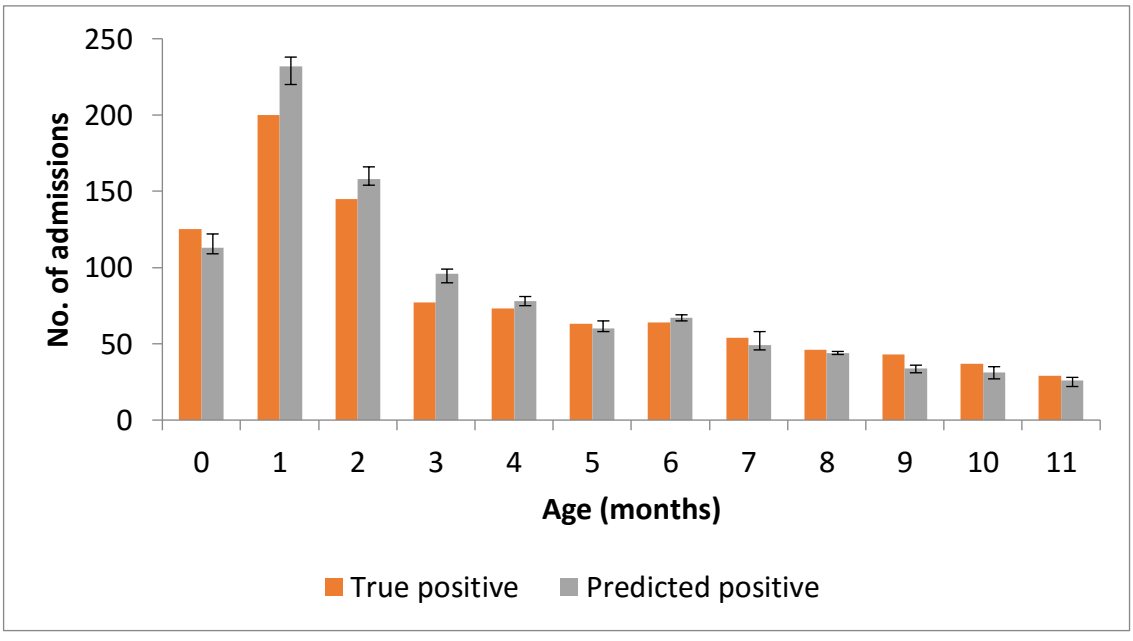

Figure 8-3. Number of true RSV-positive admissions and the number predicted due to RSV by the model, in the test sample, by age in months.

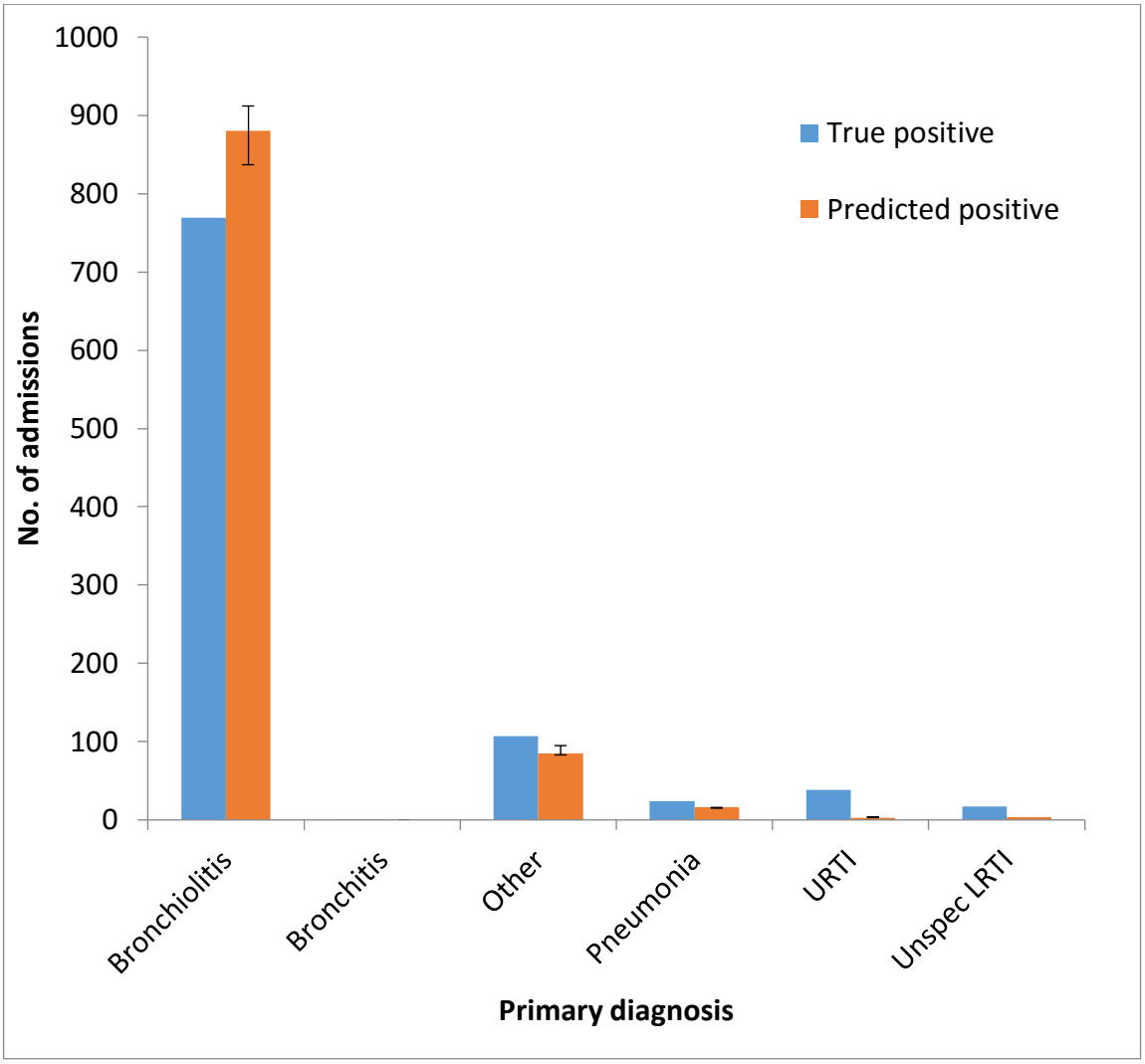

Figure 8-4. Number of true RSV-positive admissions and the number predicted due to RSV by the model, in the test sample, by age in months.

8.4.3 Estimating the secondary care burden of RSV in infants in England

The predictive model estimated a total of 40,717 (95% CI 38,472- 44,056) RSV-associated admissions in infants <1 year of age in England during the 2-year period from 01/08/2010 to 31/07/2012; an annual average of 20,359 (95% CI 19,236 - 22,028) RSV-associated admissions. The average annual rate of RSV-associated admissions was 29.63 (95% CI 27.99-32.06) per 1,000 infants <1 year of age (Table 8-3). I estimate a total annual average of 57,907 (95% CI 55,391-61,637) bed days due to RSV-associated admissions in infants <1 year of age each year in England (Table 8-4). 31% (6,368/20,359) of RSV-associated admissions were for <1 day, accounting for 5% (6,368/57,907) of RSV-associated bed days.

The number of estimated RSV-associated admissions by calendar week is shown in Figure 8-5. During both seasons there was a peak in estimated RSV-associated admissions in week 52 (2010: *n*=2,281, 95% CI 2,057 – 2,498. 2011: *n*=2,637, 95% CI 2,348 - 2,848), coinciding with the peak in laboratory reports of RSV-positive respiratory samples from SGSS (section 4.2) (64). The highest number of RSV-associated admissions as a percentage of all weekly respiratory admissions during the 2010/11 season occurred in week 49 in 2010 (66%, 95% CI 61 – 73%) and during the 2011/12 season occurred in week 52 in 2011 (70%, 95% CI 62 –76%).

A total of 74% (15,126/20,359) of our estimated RSV-associated admissions were in infants younger than 6 months, accounting for 80% (46,124/57,907) of the annual bed days due to RSV (Table 8-4). 55% (31,987/57,907) of bed days and 45% (9,130/20,359) of RSV-associated admissions were in infants younger than 3 months. The annual number of RSV-associated admissions peaked at age 6 weeks (*n*=1,019, 95% CI 974-1,052), then declined with increasing age (Figure 8-6). RSV-associated admissions also peaked in infants born in September (*n*=2,730, 95% CI 2,612-2,861), October (*n*=3,252, 95% CI 3,148-3,384) and November (*n*=2,716, 95% CI 2,624-2,944) (Figure 8-7). Only 5% (944/20,359) of the RSV-associated admissions were in infants with an ICD-10 code indicating high-risk status, but these accounted for 21% (12,160/57,907) of bed days.

Table 8-3. Estimated admission rates of RSV-associated hospital admissions (per 1,000 infants), stratified by primary diagnosis and season.

Primary diagnosis	Estimated admission rate of RSV-associated hospital admissions (per 1,000 infants) (95% CI)		
	2010/11	2011/12	Average
Bronchiolitis	28.58 (27.00-30.03)	26.82 (25.44-28.18)	27.70 (26.22-29.11)
Bronchitis	0.08 (0.08-0.08)	0.07 (0.07-0.08)	0.07 (0.07-0.08)
Other	1.34 (1.29-1.56)	1.22 (1.14-1.28)	1.28 (1.22-1.50)
Pneumonia	0.24 (0.22-1.56)	0.17 (0.16-0.18)	0.21 (0.20-0.87)
URTI	0.22 (0.20-0.26)	0.18 (0.16-0.23)	0.20 (0.18-0.25)
Unspec LRTI	0.20 (0.13-0.27)	0.13 (0.07-0.86)	0.17 (0.10-0.56)
Total	30.66 (28.94-33.13)	28.60 (27.04-30.98)	29.63 (27.99-32.06)

58-75% of bronchiolitis admissions were estimated to be due to RSV, with the percentage of RSV-associated admissions decreasing with increasing age (Figure 8-8). 5-15% of pneumonia admissions, 12-27% of bronchitis admissions, and 1-8% of unspecified LRTI admissions were estimated to be due to RSV.

Table 8-4. Total estimated annual number of hospital admissions and bed days due to RSV in infants <1 year old in England - stratified by age, risk group, birth month and primary diagnosis.

	Total annual admissions (95% CI)		Total annual bed days (95% CI)	
	N (95% CI)	%	*N* (95% CI)	%
Total	20,359 (19,236 - 22,028)	-	57,907 (55,391-61,637)	-
Sex				
Male	11,725 (11,022-12,630)	58%	32,177 (30,816-34,119)	56%
Female	8,632 (8,213-9,395)	42%	25,727 (24,573-27,515)	44%
Ratio (M:F)	1.3:1	-	1.3:1	-
Age (months)				
<1	1,772 (1,712-1,913)	9%	10,529 (10,088-11,483)	18%
1	4,174 (3,995-4,312)	21%	12,729 (12,297-13,112)	22%
2	3,184 (3,073-3,371)	16%	8,729 (8,510-9,118)	15%
3	2,323 (2,201-2,451)	11%	5,745 (5,539-5,979)	10%
4	2,013 (1,895-2,174)	10%	4,809 (4,481-5,052)	8%
5	1,661 (1,563-1,791)	8%	3,584 (3,412-3,851)	6%
6	1,359 (1,269-1,502	7%	2,819 (2,647-3,069)	5%
7	1,121 (1,051-1,274)	6%	2,601 (2,484-2,827)	4%
8	911 (836-1,051)	4%	2,053 (1,948-2,286)	4%
9	771 (698-897)	4%	1,759 (1,636-1,971)	3%
10	580 (528-693)	3%	1,508 (1,425-1,679)	2%
11	493 (419-603)	2%	1,045 (924-1,211)	2%
Risk group				
No risk	19,415 (18,318-20,961)	95%	45,747 (43,819-48,456)	79%
Risk factor	944 (919-1,0671)	5%	12,160 (11,572-13,181)	21%
Birth month				
January	896 (790-1,042)	4%	3,018 (2,842-3,348)	5%
February	694 (634-805)	3%	1,985 (1,772-2,241)	3%
March	831 (770-986)	4%	2,123 (1,980-2,414)	4%
April	996 (912-1,147)	5%	2,237 (2,080-2,507)	4%
May	1,270 (1,192-1,406)	6%	2,937 (2,668-3,176)	5%
June	1,434 (1,353-1,546)	7%	3,001 (2,867-3,164)	5%
July	1,764 (1,695-1,913)	9%	4,288 (4,171-4,668)	7%
August	2,134 (2,009-2,225)	10%	5,357 (5,114-5,682)	9%
September	2,730 (2,612-2,861)	13%	7,198 (6,974-7,436)	12%
October	3,252 (3,148-3,384)	16%	10,060 (9,701-10,351)	17%
November	2,716 (2,624-2,944)	13%	9,609 (9,430-10,194)	17%
December	1,643 (1,501-1,771)	8%	6,095 (5,792-6,455)	12%

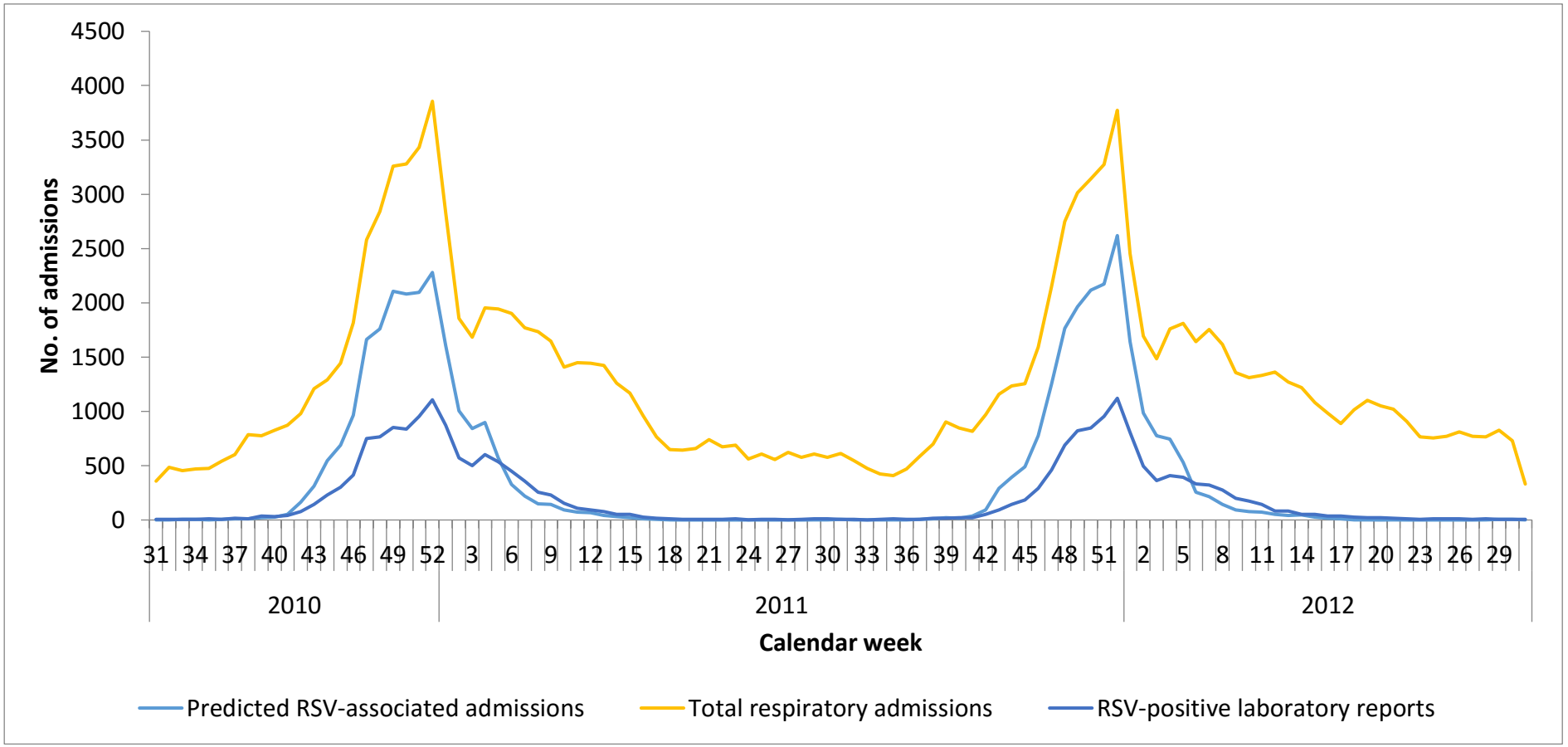

Figure 8-5. Estimated RSV-associated hospital admissions in infants <1 year old in England, by calendar week. RSV-positive laboratory reports from SGSS in children <5 years shown to illustrate timing of RSV circulation (previously reported (139)**).**

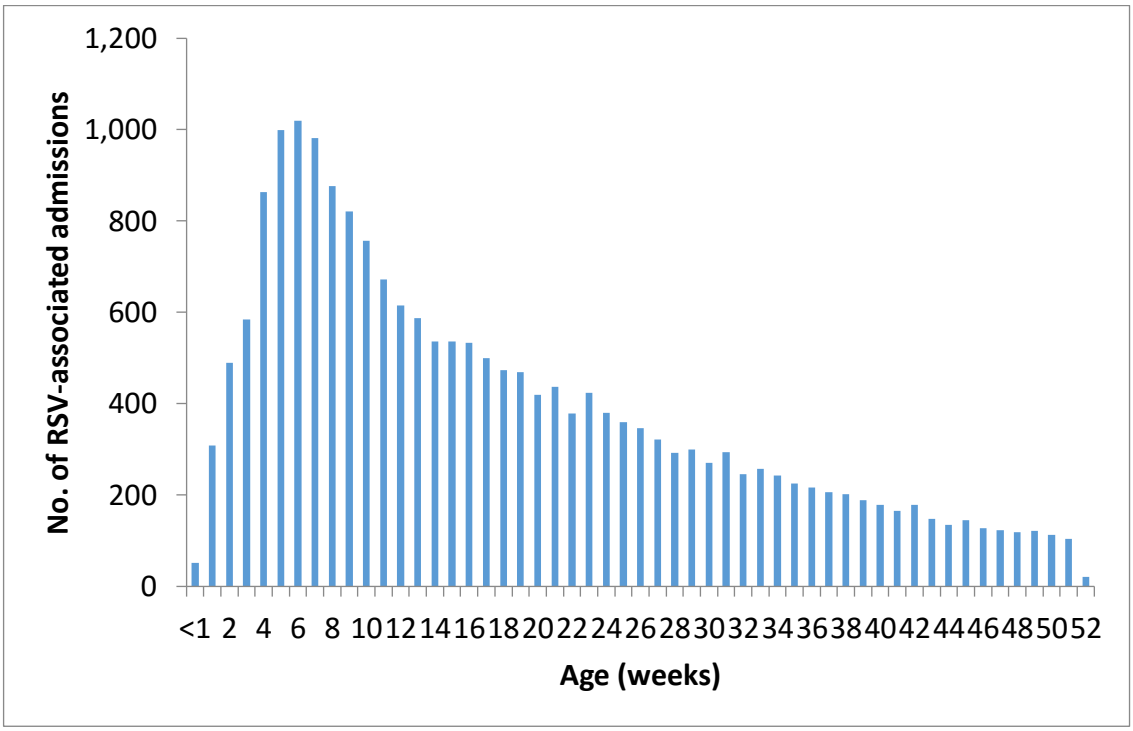

Figure 8-6. Annual estimated number of RSV-associated admissions, by age in weeks.

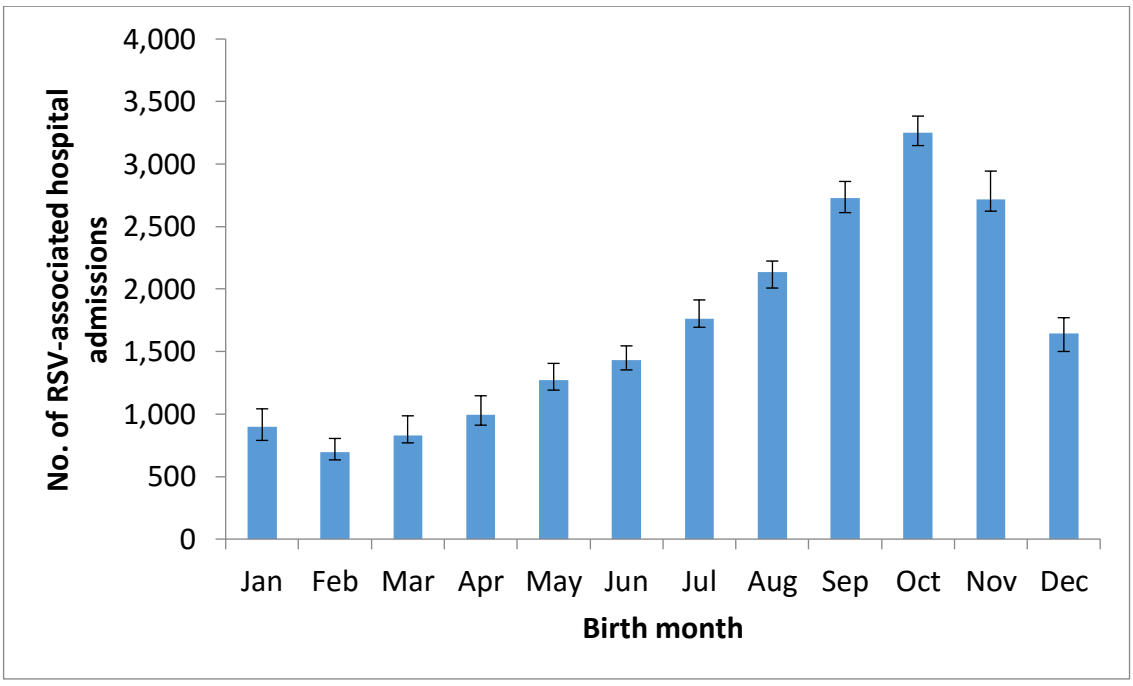

Figure 8-7. Annual estimated number of RSV-associated admissions, by birth month.

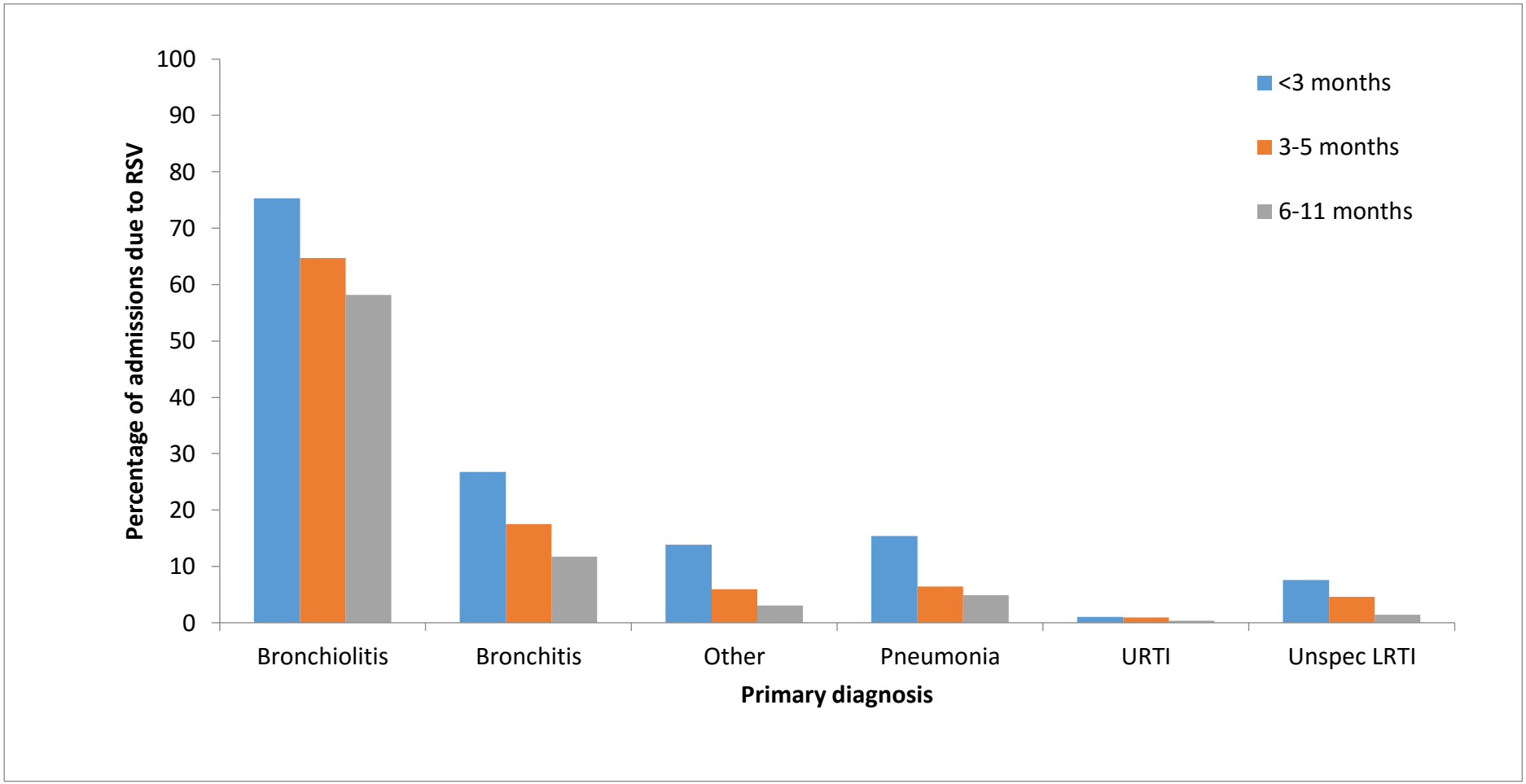

Figure 8-8. Percentage of admissions estimated to be due to RSV per primary diagnosis, stratified by age group.

8.5 Discussion

This analysis is the first to use linked laboratory and hospital records to estimate the secondary burden of RSV in England. Using the unique RDS-HES linked dataset, I estimate a total annual average of 20,359 (95% CI 19,236-22,028) RSV-associated admissions in infants younger than 1 year of age in England during the two-year period from mid-2010 to mid-2012. These admissions accounted for approximately 57,907 (95% CI 55,391-61,637) annual bed days. Approximately 55% of RSV-associated bed days and 45% of RSV-associated admissions were in infants younger than 3 months. There was a peak in RSV-associated admissions in infants aged 6 weeks, as well as in infants born in September, October and November. Only 5% of RSV-associated admissions were in high-risk infants, but these infants accounted for 21% of the estimated bed days.

This estimate of annual RSV-associated admissions is similar to - though around 13% lower than - my previous estimate of 23,310 (95% CI 21,816-25,738) RSV-associated admissions in infants younger than 1 year in England using time-series modelling, which is the only other estimate in England covering the same time period (Chapter 5) (139). This difference could be caused by my model under-predicting admissions with a diagnosis of unspecified LRTI, pneumonia or URTI (as there were less linked RSV-positive admissions with these diagnoses), or could be due this being a more accurate prediction. However, as this analysis only includes respiratory admissions and does not capture admissions with non-respiratory ICD-10 codes which may also be due to RSV (e.g. R06.2 - Wheezing) (as for all this analysis in this thesis), I believe that these results should therefore be considered a minimum estimate of the true burden of RSV-associated admissions in infants younger than 1 year in England.

That this analysis estimated a large burden of RSV-associated admissions in infants younger than 3 months of age is consistent with previous studies in Western countries (181). A future maternal immunisation strategy that protects this group therefore has the potential to significantly reduce the secondary care burden of RSV in England. However, vaccine coverage in pregnant women in England is low,

ranging from 45% for influenza to 60% for pertussis (94). I have undertaken crude analysis that demonstrates the minimum potential effectiveness of potential vaccine strategy scenarios in terms of hospital admissions and bed days prevented – this analysis is included in Appendix 7. This crude analysis uses my estimates generated in this chapter to explore the minimum potential effectiveness of maternal vaccine strategies with 70% efficacy and 45% coverage (similar to influenza maternal vaccine uptake) or 60% coverage (similar to pertussis vaccine uptake), and direct vaccination of infants aged 4-6 months or 6-11 months with 90% coverage. My results indicate that a combination approach of maternal immunization (targeted to infants born at the beginning of or during RSV season, i.e. September, October, November, December) as well as a vaccine targeted to older infants may be the most effective intervention (Appendix 7).

In the analysis presented in this chapter, infants with risk factors for severe RSV-associated illness accounted for one-fifth of bed days but only 5% of admissions. This is comparable to previous studies in the UK which estimate that 2-8% of RSV-associated admissions are in high-risk children (110,118,161). It also highlights the increased severity of RSV-associated illness in these children with known clinical risk factors. An extension of guidelines for palivizumab prophylaxis to cover a wider range of high-risk children - such as those with rare diseases including congenital and acquired immunodeficiency or neurological disorders – therefore has the potential to prevent (or reduce severity of) these admissions (182). However, due to the high cost of palivizumab and the requirement of multiple injections throughout the RSV season, a universal vaccine approach (i.e. not targeted to those with comorbidities) is likely to be a more cost-effective strategy.

8.5.1 Strengths and limitations of this analysis

This analysis is the first to use linked laboratory and hospital records to estimate the number of RSV hospital admissions in infants in England. The benefit of using the RDS dataset is that both RSV-positive and negative laboratory records can be

analysed, which has allowed me to develop a novel method of using the linked data to generate a predictive model of RSV-positive hospital admissions. My predictive model had good fit, with high sensitivity and specificity, though slightly overestimated the number of RSV-positive admissions in infants aged 1 and 2 months, over-predicted bronchiolitis, and underestimated the number of RSV-associated pneumonia and URTI admissions (likely due to the large number of RSV-associated bronchiolitis admissions in my dataset). By applying this model to the unlinked HES data (i.e. all respiratory admissions, regardless of whether they were tested for RSV or not), I was able to generate more detailed estimates of the secondary care burden of RSV compared to my estimates using time-series modelling, with narrower age groupings and analysis of clinical risk factors. Following my analysis of severity of RSV-associated admissions in Chapter 7 of this thesis, in this analysis I estimated not just the number of RSV-associated admissions but also the number of RSV-associated bed days by key patient and clinical characteristics.

There are several limitations to this analysis, further to the specific limitations of the data linkage presented in Chapter 6 of this thesis. Firstly, it was not possible for us to fully ascertain the success of the linkage methodology as it is not possible to disentangle missed links due to missing patient identifiable information from missing links due to the laboratory test being carried out in primary care (RDS records laboratory tests from primary and secondary care, but lacks complete information on which setting the sample was from). Secondly, the HES dataset I had access to only included patients admitted to hospital with an ICD-10 code belonging to Chapter X: Diseases of the Respiratory System. Therefore, patients with non-respiratory codes are not included and my analysis should be considered an underestimate of the true secondary care burden of RSV. In addition, only two years of data were linked, limiting the sample size. Finally, the laboratory surveillance data only covered a subset of laboratories in England, with no known denominator population of hospitals that submit to these laboratories. However, I limited the population used for predictive modelling to only trusts that tested >10%

of their respiratory admissions in order to estimate the denominator population of hospitals.

8.5.2 Conclusions

I have demonstrated that population-based data linkage between laboratory surveillance and hospital admissions data for RSV in England facilitates the estimation of the total national secondary care burden of RSV. That these estimates are similar to previous studies using statistical modelling techniques offers validation of my methodology. Furthermore, this methodology allows the secondary care burden of RSV to be examined in much more detail than has previously been achieved. My methodology has the potential to be utilised for other pathogens, particularly in similar instances where only a minority of cases undergo laboratory testing to identify the causal pathogen. My results can be used in cost-effectiveness evaluations (section 1.3.10.3) to identify optimal target populations for a potential future RSV vaccine.

Chapter 9

Discussion and conclusions

Chapter 9 Discussion and conclusions

9.1 Summary of findings

The overall aim of this thesis was to estimate the burden of RSV in infants and young children in secondary care in England using routinely collected laboratory surveillance and hospital admissions data. I have produced estimates of the secondary care burden of RSV by two different methods – time-series modelling, and a novel data linkage method – and identified risk factors for severe RSV-associated disease. My results provide important epidemiological data that highlights key target populations for interventions, such as a potential vaccine programme, and my estimates can be used in vaccine impact studies and cost-effectiveness evaluations to drive policy decisions.

The main findings of this thesis can be summarised as follows:

- Using time-series modelling, I estimate that there was an annual average of 33,500 (95% CI: 30,400-38,500) hospital admissions for RSV-associated RTIs in children <5 years of age. 84% (95% CI: 81-91%) of these RSV-associated admissions were coded as LRTIs. 82% (95% CI: 79-87%) of hospital admissions for bronchiolitis in children aged <6 months could be attributed to RSV.
- My novel data linkage methodology produced a similar estimate of RSV-associated hospital admissions in children <1 year of age to the time-series modelling estimate (20,400, 95% CI 19,200-22,000, vs 23,300, 95% CI 21,800-25,700, per year, respectively) – this offers validation of the data linkage results, which is particularly important due to the difficulties in assessing data linkage quality.
- In the two-year period from mid-2010 to mid-2012, approximately 55% of RSV-associated bed days and 45% of RSV-associated admissions in children younger than 1 year were in infants younger than 3 months, with a peak in RSV-associated admissions in infants 6 weeks old. These young infants (<3

months old) were also at increased risk of severe disease requiring prolonged hospital stay and the use of mechanical ventilation.

- There was a peak in RSV-associated admissions and RSV-positivity in infants born in September, October and November – highlighting the importance of birth around the beginning of RSV season in the risk of RSV-associated disease.
- Only 5% of estimated RSV-associated admissions were in infants with known clinical risk factors, but these infants accounted for 21% of the estimated bed days. Infants with known clinical risk factors were at significantly increased risk of more severe disease requiring prolonged hospital stay and the use of mechanical ventilation.

In the respective chapters of this thesis, I have discussed how each of these findings support (or, are supported by) findings of previous studies, and how these findings contribute to the existing knowledge base. The implications of these findings for future RSV research and potential vaccine strategy are discussed below.

9.2 Implications of these findings

My results highlight that key target populations for potential interventions, such as a future vaccine programme, are:

a) Infants aged <3 months – these children have the highest risk of RSV-positivity, the highest number of RSV-associated hospital admissions (with a peak at 1 month of age), and are at increased risk of severe RSV-associated disease.
b) Infants born at the beginning of RSV season (September, October and November).
c) Children with known clinical risk factors – these children do not account for a high percentage of admissions, but account for a disproportionate amount

of RSV-associated bed days and have significantly increased severity of disease.

The increased risk of RSV-associated hospital admission and severe RSV-associated disease has previously been demonstrated in infants aged <6 months, and these infants are considered a high priority for vaccine development (section 1.3.10.2). However, my results highlight the importance of targeting interventions to the youngest infants aged <3 months to protect those most at risk of severe disease. Infants aged <3 months may be most successfully targeted via maternal immunisation than a direct immunisation strategy, as these very young infants may not respond adequately to vaccination and are at risk of vaccine-enhanced disease (section 1.3.10.2). A maternal vaccine would therefore need to elicit immunity in the newborn for at least 3 months to protect the most at-risk young infants. The Novovax RSV F Vaccine Phase 2 clinical trial in healthy women of child bearing age demonstrated significantly elevated antibody levels for 3 months following immunisation (88). However, the results of the recently completed Phase 2 trial in third-trimester pregnant women are not yet published (183); these results will indicate whether the vaccine is capable of eliciting protection for the whole of the high-risk 3-month period in newborn infants. Alternatively, vaccination of siblings or other household contacts could be an effective alternate strategy of providing protection for very young infants (section 1.3.10.2).

The crude analysis of potential vaccine strategy that I have carried out using my estimates of RSV secondary care burden from the linked data (included in Appendix 7), hypothesises that a maternal vaccine programme targeted at pregnant women due to give birth around the beginning of RSV season (i.e. September, October, November, December) may be the most effective strategy in reducing RSV-associated admissions and associated bed days in this young at-risk group. However, sufficient coverage would likely need to exceed that of the current influenza maternal vaccine coverage (approximately 45%). I also demonstrate that there is a significant burden of RSV-associated admissions in older infants and children – my time-series modelling analysis estimated that 52% (95% CI 45-63%) of

RSV-associated admissions in children <5 years per year occurred in children aged 6 months to 4 years. Therefore, a combination approach of a maternal vaccine (as described above), as well as direct immunisation of older infants and/or young children, could potentially reduce the burden of RSV-associated admissions infants and young children most effectively. However, if large numbers of children would need to be vaccinated within a short space of time, this could pose a logistical challenge. Though my crude vaccine impact analysis considers variation in vaccine coverage and effectiveness, it does not consider transmission dynamics and indirect effects of immunising this group. Vaccination of older infants and young children is likely to have an indirect protection effect on very young infants by reducing virus transmission. Dynamic models of potential vaccine strategy which include the potential indirect effects are therefore required in order to accurately estimate the impact of a future vaccine, and should consider more detailed analysis of this combination approach that I hypothesise could be most effective. The baseline epidemiological data that I have produced in this thesis can be used in these models.

Certain children with known clinical risk factors are recommended for Palivizumab prophylaxis: preterm infants with chronic lung disease or congenital heart disease, children under the age of 24 months with severe combined immunodeficiency syndrome (SCID), or long term ventilated (LTV) children aged less than 12 months at the start of RSV season (or aged less than 24 months with an additional comorbidity) (section 1.3.9). My analysis potentially supports the extension of palivizumab guidelines to include children with neurological disorders, acquired immunodeficiency or those receiving immunosuppressive therapy, in order to prevent the more severe (and more expensive) RSV-associated admissions in these at-risk groups. In my study, 40% of laboratory-confirmed RSV-associated hospital admissions during which invasive ventilation was used (an indicator of PICU admission) were in children with a known clinical risk factor (Chapter 7). Therefore, if palivizumab was effective in reducing severity of RSV infection in these high-risk children, an extension of palivizumab guidelines to protect these children could significantly reduce the number of RSV-associated PICU admissions. However,

economic models would need to be undertaken to determine the cost-effectiveness of this extension, particularly as palivizumab is expensive and requires multiple injections throughout the RSV season (section 1.3.9). In addition, targeting this small group of children would still not significantly reduce the overall burden of RSV in secondary care. Nonetheless, evaluation of a potential future universal vaccine programme should consider the impact on these high-risk children who are more likely to develop more severe disease.

9.2.1 Use of administrative data for RSV research

My PhD demonstrates the advantages of utilising routinely collected datasets in generating epidemiological data on the secondary-care burden of RSV. While the laboratory surveillance and hospital admissions datasets used in this thesis were not collected for research purposes and have notable limitations, they are information-rich datasets which I have shown to be valuable for advancing knowledge of RSV epidemiology and secondary care burden.

9.2.1.1 Use of linked laboratory and hospital data for RSV research

I have shown that routinely collected laboratory surveillance data or hospital admissions data on their own are not sufficient in providing detailed information on the secondary care burden of RSV in infants and young children. However, my PhD demonstrates the value of linking these datasets to identify laboratory-confirmed RSV-associated hospital admissions. I have utilised data linkage between laboratory surveillance and hospital admissions data to develop a predictive model for RSV-associated admissions, which has allowed me to generate an individual-level dataset of hospital admissions that are confirmed to be or have a high probability of being RSV-associated.

A key contribution of my work is that I have estimated both the number of RSV-associated hospital admissions and the number of RSV-associated hospital bed days,

stratified by patient and clinical characteristics. My results show the pattern of RSV-associated admissions by key characteristics such as age in weeks, risk group and diagnosis. The data on RSV-associated hospital admissions that I have generated is essential for predictive modelling studies to determine optimal vaccine strategy, and as baseline data to inform the measurement of the actual impact of a potential future vaccine programme. In addition, applying my methodology to more recent and current data in England would allow monitoring of trends in RSV-associated hospital admissions over time.

9.3 Strengths and limitations of this work

In this thesis, I have produced detailed estimates of the secondary care burden of RSV by two different methods: time-series modelling, and a novel data linkage method. For the first time in England, I have used linked laboratory surveillance and hospital admissions data; this has allowed me to describe laboratory-confirmed hospital admissions by patient and clinical characteristics and develop a predictive model for RSV-positive hospital admissions. The similarity of my results from these two methods suggests that both estimates are valid.

I have highlighted the limitations of my analysis in the discussion sections of each results chapter. The main limitations of my work can be summarised as follows:

- This analysis only covers admission with diagnosis code from Chapter X of the ICD-10 (that is, ICD-10 codes J00-J99); therefore, admissions with diagnostic codes from other ICD-10 chapters which may also be used for RSV – (e.g. R06.0-R06.8 (Abnormalities of breathing) and B34.9 (or R06.2) (Wheeze) – are not included in the dataset. There are, for example, approximately 1,300 admissions with a primary diagnosis of R06.2 (Wheeze) in children <1 year in England each year (according to NHS Digital annual reports (184)), some of which will be due to RSV. Therefore, my results are

likely to be a slight underestimate of the true burden of RSV in secondary care.

- I could only identify codes for prematurity and comorbidities that were recorded in the longitudinal HES record of respiratory admissions, as my HES extract only included admissions with respiratory diagnosis codes – therefore, I am likely to have underestimated the number of children with comorbidities in my analysis.
- The data linkage carried out in this thesis was limited by the PII available in the RDS dataset (held by PHE). Only 37% of laboratory records linked to a respiratory admission that began within 7 days of the test. Completeness of NHS number and postcode –the two most discriminatory PII variables available – was very low in some laboratories, which significantly reduced the number of successful linked records from those laboratories. Furthermore, assessment of linkage quality was challenging, as an unknown proportion of laboratory records were from primary care (or A&E) and therefore would not be expected to link to a HES admission, and (as the RDS is a subset of laboratories in England which has an unknown denominator of hospitals that submit to those laboratories) an unknown proportion of HES admissions would not be expected to link to a RDS test.
- I was unable to estimate the secondary care burden of RSV in children aged 1-4 using the linked data due to the low proportion of laboratory testing in this age group. This lack of testing limited the development of a suitable predictive model for RSV-positivity in children aged 1-4 years. This was likely because the linked admissions were not representative of all respiratory admissions in this age group.

To overcome the limitations of the data linkage and assessment of linkage quality, the use of the +/-7 day cut off (between admission and test) for linked records offers some validation, reducing the potential for false matches (due to a false match being unlikely to have an admission within +/- 7 days of the RSV test). In addition, when developing the predictive model using the linked data, I specified a denominator of HES providers based on those with a high percentage of linked

records – this attempted to adjust for the low linkage in some areas due to poor RDS PII, only including providers which have linked (tested) records more representative of all their respiratory admissions. Furthermore, I compared the characteristics of admissions in the original, unlinked dataset of HES admissions to the admissions in the linked dataset, to identify any potential biases.

However, as highlighted in the relevant discussion sections of each results chapter, my results are similar to previous studies of RSV-associated hospital admissions in infants and young children – particularly the pattern of admissions by age (Chapter 6), and the risk factors for severe RSV-associated admissions (Chapter 7). Furthermore, my estimates of RSV burden using linked data are similar to my estimates using time-series modelling, which offers additional validation of my results. I have identified future work to utilise my results and overcome the limitations of my methodology, which I have described below.

9.4 Recommended future research

The key future research projects following on from the analysis in this thesis are the cost-effectiveness models to identify optimal target populations for a future licenced RSV vaccine. The data that I have generated in this thesis (RSV-associated admissions and bed days by sex, calendar week, age in months, known clinical risk factors, birth month, diagnosis, and severity of disease) is essential data to be included in these cost-effectiveness models (section 1.3.10.3). Vaccines strategies also need to consider the primary care and community burden of RSV, however this data is very limited for RSV compared to secondary care data (45). The Royal College of General Practitioners (RCGP) primary care database includes data from the RCGP flu surveillance system, which covers around 1.5% of the English population and reports the weekly consultations for influenza-like illness and other acute respiratory illnesses (185). This surveillance scheme also includes laboratory testing of respiratory swabs; therefore, utilisation of this dataset could allow the

primary care burden of RSV in infants and young children to be analysed (186). Alternatively, the Clinical Practice Research Datalink (CPRD) contains anonymised primary care data and can readily be linked to HES (though has no laboratory test results recorded) (187); utilising this linked CPRD-HES database, potentially in conjunction with my predictive model for RSV-positive hospital admissions, could also provide an opportunity to evaluate the primary care burden of RSV. In addition, as well as considering potential vaccination strategies, investigation into alternative interventions such as improving hygienic measures such as handwashing and reducing transmission (for example, among siblings, or by health visitors and community midwives) should also be carried out (42).

To overcome some of the limitations of my analysis, such as the reliance on recorded ICD-10 codes to identify high-risk children, linkage of the RDS-HES dataset to ONS birth records and NHS birth notification records should be carried out to allow more complete identification of at-risk children (particularly if birth weight and gestational age were included). Alternatively, a birth cohort with much more complete information on birth weight and gestational age could be created within HES by linking mother's delivery and baby's birth records (188). Linkage to the Paediatric Intensive Care Audit Network (PICANet) (177) would allow further analysis of severe RSV-associated admissions requiring PICU admission, and access to national hospital prescribing data would enable identification of those children who receive palivizumab prophylaxis in order to determine its use and effectiveness in a real-world setting.

Future work should also focus on the improvement of linkage success between RDS and HES. To facilitate this, I recommend that PHE improves the PII quality of RDS. One method of achieving this could be to change the record submission system for RDS, so that records cannot be submitted without complete PII – this would significantly improve the linkage quality and greatly increase the number of successfully linked records. A further recommendation for PHE regarding the RDS dataset would be the addition of a well-completed variable indicating whether the sample for laboratory testing was taken in primary or secondary care; this would

enable significantly better assessment of linkage quality to be undertaken. My methodology should be continuously applied to more recent and current data in England to allow monitoring of trends in RSV-associated hospital admissions over time. In addition, maternal and child vaccination records linked to the RDS-HES linked dataset would facilitate evaluation of a future vaccination programme.

Finally, I was unable to estimate the secondary care burden of RSV in children aged 1-4 years using the linked data, due to the lack of laboratory testing in this age group limiting the development of a suitable predictive model for RSV-positivity. More intensive testing in some hospitals already reporting to RDS could allow sufficient data to be included in a RDS-HES linked dataset to investigate the burden of RSV-associated admissions in children aged 1-4 years. This could also be extended to infants aged <1 year to validate the results that I present in this thesis. Alternatively, a single- or multi-site prospective study (collecting detailed information, including laboratory testing, on respiratory admissions in children aged 1-4 years, such as in Leicestershire (178)) could be a successful approach to investigate the burden of RSV-associated admissions in children aged 1-4 years.

In summary, I recommend that the following future work be carried out:

- Vaccine cost-effectiveness models to identify optimal vaccine strategy, utilising the detailed epidemiological data that I have produced.
- PHE to improve recording of PII (particularly NHS number) within RDS, to improve linkage success between RDS and HES, and to facilitate the inclusion of a primary/secondary care source indicator within RDS.
- Ongoing application of my methodology to more recent and current data in England to allow monitoring of trends in RSV-associated hospital admissions over time (including analysis of all HES admissions, not just those with a respiratory ICD-10 code). If an RSV vaccine programme is introduced in the future, having recent and continually updated estimates would allow the real-world impact of a vaccine to be calculated.

- More detailed estimates of the secondary care burden of RSV in children aged 1-4 (potentially by increased testing in specific hospitals which report to RDS, or using prospective studies of respiratory admissions in these children), as well as studies to investigate the primary care and community burden of RSV in infants and young children in England.

9.5 Concluding remarks

In this thesis, I have utilised routinely collected laboratory surveillance and hospital admissions data to produce estimates of the secondary care burden of RSV by two different methods – time-series modelling, and a novel data linkage method – and identify risk factors for severe RSV-associated disease. My results provide detailed epidemiological data that highlights key target populations for interventions: infants younger than 3 months, infants born around the beginning of RSV season, and infants and young children with known clinical risk factors. With a number of RSV vaccines now in late stage clinical trials, my results can be used in vaccine impact studies to drive policy decisions regarding a potential future vaccine programme. My work also provides a methodological foundation for future studies of the secondary care burden of RSV using linked routinely collected datasets.

References

1. Blount RE, Morris JA, Savage RE. Recovery of cytopathogenic agent from chimpanzees with coryza. Proc Soc Exp Biol Med. 1956;92(3):544–9.

2. Collins PL, Graham BS. Viral and host factors in human respiratory syncytial virus pathogenesis. J Virol. 2008;82(5):2040–55.

3. Collins PL, Fearns R, Graham BS. Respiratory syncytial virus: virology, reverse genetics, and pathogenesis of disease. Curr Top Microbiol Immunol. 2013;372:3–38.

4. Dudas RA, Karron RA. Respiratory syncytial virus vaccines. Clin Microbiol Rev. 1998;11(3):430–9.

5. Enders G. Paramyxoviruses. Medical Microbiology. University of Texas Medical Branch at Galveston; 1996.

6. Ogra PL. Respiratory syncytial virus: the virus, the disease and the immune response. Paediatr Respir Rev. 2004;5 Suppl A:S119-26.

7. McNamara PS, Smyth RL. The pathogenesis of respiratory syncytial virus disease in childhood. Br Med Bull. 2002;61(1):13–28.

8. Public Health England. Respiratory syncytial virus: the green book, chapter 27a [Internet]. 2013 [cited 2017 Feb 26]. Available from: https://www.gov.uk/government/publications/respiratory-syncytial-virus-the-green-book-chapter-27a

9. McLellan JS, Ray WC, Peeples ME. Structure and function of respiratory syncytial virus surface glycoproteins. Curr Top Microbiol Immunol. 2013;372:83–104.

10. Nair H, Nokes DJ, Gessner BD, Dherani M, Madhi SA, Singleton RJ, et al. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. Lancet. 2010;375(9725):1545–55.

11. McNamara PS, Flanagan BF, Selby AM, Hart CA, Smyth RL. Pro- and anti-inflammatory responses in respiratory syncytial virus bronchiolitis. Eur Respir J. 2003;23(1).

12. Glezen WP, Taber LH, Frank AL, Kasel JA. Risk of primary infection and reinfection with respiratory syncytial virus. Am J Dis Child. 1986;140(6):543–6.

13. Ohuma EO, Okiro EA, Ochola R, Sande CJ, Cane PA, Medley GF, et al. The Natural History of Respiratory Syncytial Virus in a Birth Cohort: The Influence of Age and Previous Infection on Reinfection and Disease. Am J Epidemiol. 2012;176(9):794–802.

14. Shi T, McAllister DA, O'Brien KL, Simoes EAF, Madhi SA, Gessner BD, et al. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. Lancet. 2017;390:946–58.

15. Resch B. Burden of respiratory syncytial virus infection in young children. World J Clin Pediatr. 2012;1(3):8–12.

16. Collins CL, Pollard AJ. Respiratory Syncytial Virus Infections in Children and Adults. J Infect. 2002;45(1):10–7.

17. Sommer C. Risk Factors for Severe Respiratory Syncytial Virus Lower Respiratory Tract Infection. Open Microbiol J. 2011;5(1):144–54.

18. Paul WE, Zhu J. How are T(H)2-type immune responses initiated and

amplified? Nat Rev Immunol. 2010;10(4):225–35.

19. Pinto RA, Arredondo SM, Bono MR, Gaggero AA, Díaz P V. T helper 1/T helper 2 cytokine imbalance in respiratory syncytial virus infection is associated with increased endogenous plasma cortisol. Pediatrics. 2006;117(5):e878-86.

20. Hall CB, Weinberg GA, Iwane MK, Blumkin AK, Edwards KM, Staat MA, et al. The burden of respiratory syncytial virus infection in young children. N Engl J Med. 2009;360(6):588–98.

21. Cilla G, Sarasua A, Montes M, Arostegui N, Vicente D, Pérez-Yarza E, et al. Risk factors for hospitalization due to respiratory syncytial virus infection among infants in the Basque Country, Spain. Epidemiol Infect. 2006;134(3):506–13.

22. García CG, Bhore R, Soriano-Fallas A, Trost M, Chason R, Ramilo O, et al. Risk factors in children hospitalized with RSV bronchiolitis versus non-RSV bronchiolitis. Pediatrics. 2010;126(6):e1453-60.

23. Pockett RD, Campbell D, Carroll S, Rajoriya F, Adlard N. A comparison of healthcare resource use for rotavirus and RSV between vulnerable children with co-morbidities and healthy children: a case control study. J Med Econ. 2013;16(4):560–5.

24. Thorburn K. Pre-existing disease is associated with a significantly higher risk of death in severe respiratory syncytial virus infection. Arch Dis Child. 2009;94(2):99–103.

25. Horn SD, Smout RJ. Effect of prematurity on respiratory syncytial virus hospital resource use and outcomes. J Pediatr. 2003;143(5 Suppl):S133-41.

26. MacDonald NE, Hall CB, Suffin SC, Alexson C, Harris PJ, Manning JA. Respiratory Syncytial Viral Infection in Infants with Congenital Heart Disease. N Engl J Med. 1982;307(7):397–400.

27. Kristensen K, Hjuler T, Ravn H, Simões EAF, Stensballe LG. Chronic diseases, chromosomal abnormalities, and congenital malformations as risk factors for respiratory syncytial virus hospitalization: a population-based cohort study. Clin Infect Dis. 2012;54(6):810–7.

28. White LJ, Waris M, Cane PA, Nokes DJ, Medley GF. The transmission dynamics of groups A and B human respiratory syncytial virus (hRSV) in England & Wales and Finland: seasonality and cross-protection. Epidemiol Infect. 2005;133(2):279–89.

29. Stensballe LG, Devasundaram JK, Simoes EA. Respiratory syncytial virus epidemics: the ups and downs of a seasonal virus. Pediatr Infect Dis J. 2003 Mar;22(2 Suppl):S21-32.

30. Bloom-Feshbach K, Alonso WJ, Charu V, Tamerius J, Simonsen L, Miller MA, et al. Latitudinal variations in seasonal activity of influenza and respiratory syncytial virus (RSV): a global comparative review. PLoS One. 2013;8(2):e54445.

31. Seemungal T, Harper-Owen R, Bhowmik A, Moric I, Sanderson G, Message S, et al. Respiratory Viruses, Symptoms, and Inflammatory Markers in Acute Exacerbations and Stable Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2001;164(9):1618–23.

32. Hussey GD, Apolles P, Arendse Z, Yeates J, Robertson A, Swingler G, et al. Respiratory syncytial virus infection in children hospitalised with acute lower respiratory tract infection. S Afr Med J. 2000;90(5):509–12.

33. Madhi SA, Venter M, Madhi A, Petersen MK, Klugman KP. Differing manifestations of respiratory syncytial virus-associated severe lower respiratory tract infections in human immunodeficiency virus type 1-infected and uninfected children. Pediatr Infect Dis J. 2001;20(2):164–70.

34. Falsey AR, Formica MA, Hennessey PA, Criddle MM, Sullender WM, Walsh EE. Detection of Respiratory Syncytial Virus in Adults with Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2006;173(6):639–43.

35. Zlateva KT, Vijgen L, Dekeersmaeker N, Naranjo C, Van Ranst M. Subgroup prevalence and genotype circulation patterns of human respiratory syncytial virus in Belgium during ten successive epidemic seasons. J Clin Microbiol. 2007;45(9):3022–30.

36. Public Health England. Respiratory virus circulation, England and Wales [Internet]. 2017 [cited 2017 Jun 24]. Available from: https://www.gov.uk/government/publications/respiratory-virus-circulation-england-and-wales

37. Andabaka T, Nickerson JW, Rojas-Reyes MX, Rueda JD, Bacic Vrca V, Barsic B. Monoclonal antibody for reducing the risk of respiratory syncytial virus infection in children. Cochrane database Syst Rev. 2013;4:CD006602.

38. Bitko V, Musiyenko A, Barik S. Viral infection of the lungs through the eye. J Virol. 2007;81(2):783–90.

39. Domachowske JB, Rosenberg HF. Respiratory syncytial virus infection: immune response, immunopathogenesis, and treatment. Clin Microbiol Rev. 1999;12(2):298–309.

40. Ruuskanen O, Ogra PL. Respiratory syncytial virus. Curr Probl Pediatr. 1993;23(2):50–79.

41. Drysdale SB, Green CA, Sande CJ. Best practice in the prevention and management of paediatric respiratory syncytial virus infection. Ther Adv Infect Dis. 2016;3(2):63–71.

42. Jefferson T, Del Mar CB, Dooley L, Ferroni E, Al-Ansary LA, Bawazeer GA, et al.

Physical interventions to interrupt or reduce the spread of respiratory viruses. Cochrane Database Syst Rev. 2011;7:CD006207.

43. Munywoki PK, Koech DC, Agoti CN, Lewa C, Cane PA, Medley GF, et al. The source of respiratory syncytial virus infection in infants: a household cohort study in rural Kenya. J Infect Dis. 2014;209(11):1685–92.

44. Poletti P, Merler S, Ajelli M, Manfredi P, Munywoki PK, Nokes D, et al. Evaluating vaccination strategies for reducing infant respiratory syncytial virus infection in low-income settings. BMC Med. 2015;13:49.

45. Drysdale SB, Sande CJ, Green CA, Pollard AJ. RSV vaccine use – the missing data. Expert Rev Vaccines. 2016;15(2):149–52.

46. Scott PD, Ochola R, Ngama M, Okiro EA, James Nokes D, Medley GF, et al. Molecular analysis of respiratory syncytial virus reinfections in infants from coastal Kenya. J Infect Dis. 2006;193(1):59–67.

47. Hall CB, Walsh EE, Long CE, Schnabel KC. Immunity to and frequency of reinfection with respiratory syncytial virus. J Infect Dis. 1991;163(4):693–8.

48. Lambert L, Sagfors AM, Openshaw PJM, Culley FJ. Immunity to RSV in Early-Life. Front Immunol. 2014;5:466.

49. Eiland LS. Respiratory syncytial virus: diagnosis, treatment and prevention. J Pediatr Pharmacol Ther. 2009;14(2):75–85.

50. Bawage SS, Tiwari PM, Pillai S, Dennis V, Singh SR. Recent advances in diagnosis, prevention, and treatment of human respiratory syncytial virus. Adv Virol. 2013;2013:595768.

51. Green CA, Yeates D, Goldacre A, Sande C, Parslow RC, McShane P, et al. Admission to hospital for bronchiolitis in England: trends over five decades,

geographical variation and association with perinatal characteristics and subsequent asthma. Arch Dis Child. 2015;0:1–7.

52. Nair H, Verma VR, Theodoratou E, Zgaga L, Huda T, Simões EAF, et al. An evaluation of the emerging interventions against Respiratory Syncytial Virus (RSV)-associated acute lower respiratory infections in children. BMC Public Health. 2011;11 Suppl 3:S30.

53. Singleton RJ, Bulkow LR, Miernyk K, DeByle C, Pruitt L, Hummel KB, et al. Viral respiratory infections in hospitalized and community control children in Alaska. J Med Virol. 2010;82(7):1282–90.

54. Ruuskanen O, Lahti E, Jennings LC, Murdoch DR, Yuen K, Rello J. Viral pneumonia. Lancet (London, England). 2011;377(9773):1264–75.

55. Rhedin S, Lindstrand A, Hjelmgren A, Ryd-Rinder M, Öhrmalm L, Tolfvenstam T, et al. Respiratory viruses associated with community-acquired pneumonia in children: matched case–control study. Thorax. 2015;70(9):847–53.

56. Turner C, Turner P, Cararra V, Eh Lwe N, Watthanaworawit W, Day NP, et al. A high burden of respiratory syncytial virus associated pneumonia in children less than two years of age in a South East Asian refugee population. PLoS One. 2012;7(11):e50100.

57. Cunningham S, Nair H, Campbell H. Deciphering clinical phenotypes in acute viral lower respiratory tract infection: Bronchiolitis is not an island. Thorax. 2016;71(8):679–80.

58. Church NR, Anas NG, Hall CB, Brooks JG. Respiratory syncytial virus-related apnea in infants. Demographics and outcome. Am J Dis Child. 1984;138(3):247–50.

59. Bruhn FW, Mokrohisky ST, McIntosh K. Apnea associated with respiratory

syncytial virus infection in young infants. J Pediatr. 1977;90(3):382–6.

60. Behrendt CE, Decker MD, Burch DJ, Watson PH. International variation in the management of infants hospitalized with respiratory syncytial virus. International RSV Study Group. Eur J Pediatr. 1998;157(3):215–20.

61. Gill PJ, Goldacre MJ, Mant D, Heneghan C, Thomson A, Seagroatt V, et al. Increase in emergency admissions to hospital for children aged under 15 in England, 1999-2010: national database analysis. Arch Dis Child. 2013;98(5):328–34.

62. Collins PL, Melero JA. Progress in understanding and controlling respiratory syncytial virus: still crazy after all these years. Virus Res. 2011;162(1–2):80–99.

63. Wu P, Hartert T V. Evidence for a causal relationship between respiratory syncytial virus infection and asthma. Expert Rev Anti Infect Ther. 2011;9(9):731–45.

64. Reeves RM, Hardelid P, Gilbert R, Ellis J, Zhao H, Donati M, et al. Epidemiology of laboratory-confirmed respiratory syncytial virus infection in young children in England, 2010–2014: the importance of birth month. Epidemiol Infect. 2016;144(10):2049–56.

65. Adams O, Weis J, Jasinska K, Vogel M, Tenenbaum T. Comparison of human metapneumovirus, respiratory syncytial virus and Rhinovirus respiratory tract infections in young children admitted to hospital. J Med Virol. 2014;87(2):275–80.

66. Henrickson KJ, Hall CB. Diagnostic assays for respiratory syncytial virus disease. Pediatr Infect Dis J. 2007;26(11 Suppl):S36-40.

67. Popow-Kraupp T, Aberle JH. Diagnosis of respiratory syncytial virus infection.

Open Microbiol J. 2011;5:128–34.

68. Mahony JB. Detection of respiratory viruses by molecular methods. Clin Microbiol Rev. 2008;21(4):716–47.

69. Thornton H V, Blair PS, Lovering AM, Muir P, Hay AD. Clinical presentation and microbiological diagnosis in paediatric respiratory tract infection: a systematic review. Br J Gen Pract. 2015;65(631):e69-81.

70. Murray J, Saxena S, Sharland M. Preventing severe respiratory syncytial virus disease: passive, active immunisation and new antivirals. Arch Dis Child. 2014;99(5):469–73.

71. Higgins D, Trujillo C, Keech C. Advances in RSV vaccine research and development - A global agenda. Vaccine. 2016;34(26):2870–5.

72. Carande EJ, Pollard AJ, Drysdale SB. Management of Respiratory Syncytial Virus Bronchiolitis: 2015 Survey of Members of the European Society for Paediatric Infectious Diseases. Can J Infect Dis Med Microbiol = J Can des Mal Infect la Microbiol medicale. 2016;9139537.

73. Johnson S, Oliver C, Prince GA, Hemming VG, Pfarr DS, Wang SC, et al. Development of a humanized monoclonal antibody (MEDI-493) with potent in vitro and in vivo activity against respiratory syncytial virus. J Infect Dis. 1997;176(5):1215–24.

74. Joint Committee on Vaccination and Immunisation (JCVI). Joint Committee on Vaccination and Immunisation Statement on immunisation for Respiratory Syncytial Virus [Internet]. 2012. Available from: http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@ab/documents/digitalasset/dh_120395.pdf

75. Thomas M, Bedford-Russell A, Sharland M. Hospitalisation for RSV infection in

ex-preterm infants-implications for use of RSV immune globulin. Arch Dis Child. 2000;83(2):122–7.

76. Wang D, Bayliss S, Meads C. Palivizumab for immunoprophylaxis of respiratory syncytial virus (RSV) bronchiolitis in high-risk infants and young children: a systematic review and additional economic modelling of subgroup analyses. Health Technol Assess. 2011;15(5):iii–iv, 1-124.

77. IMpact-RSV Study Group. Palivizumab, a humanized respiratory syncytial virus monoclonal antibody, reduces hospitalization from respiratory syncytial virus infection in high-risk infants. The IMpact-RSV Study Group. Pediatrics. 1998;102(3 Pt 1):531–7.

78. Resch B, Egger B, Kurath-Koller S, Urlesberger B. Respiratory syncytial virus hospitalizations in infants of 28 weeks gestational age and less in the palivizumab era. Int J Infect Dis. 2017;57:50–3.

79. Kim HW, Canchola JG, Brandt CD, Pyles G, Chanock RM, Jensen K, et al. Respiratory syncytial virus disease in infants despite prior administration of antigenic inactivated vaccine. Am J Epidemiol. 1969;89(4):422–34.

80. Murphy BR, Walsh EE. Formalin-Inactivated Respiratory Syncytial Virus Vaccine Induces Antibodies to the Fusion Glycoprotein That Are Deficient in Fusion-Inhibiting Activity. J Clin Microbiol. 1988;1595–7.

81. Fulginiti VA, Eller JJ, Sieber OF, Joyner JW, Minamitani M, Meiklejohn G. Respiratory virus immunization. I. A field trial of two inactivated respiratory virus vaccines; an aqueous trivalent parainfluenza virus vaccine and an alum-precipitated respiratory syncytial virus vaccine. Am J Epidemiol. 1969;89(4):435–48.

82. Kapikian AZ, Mitchell RH, Chanock RM, Shvedoff RA, Stewart CE. An epidemiologic study of altered clinical reactivity to respiratory syncytial (RS)

virus infection in children previously vaccinated with an inactivated RS virus vaccine. Am J Epidemiol. 1969;89(4):405–21.

83. Acosta PL, Caballero MT, Polack FP. Brief History and Characterization of Enhanced Respiratory Syncytial Virus Disease. Clin Vaccine Immunol. 2015;23(3):189–95.

84. PATH. RSV Vaccine Snapshot - PATH Vaccine Resource Library [Internet]. 2017. Available from: http://www.path.org/publications/files/CVIA_rsv_snapshot_final.pdf

85. Modjarrad K, Giersing B, Kaslow DC, Smith PG, Moorthy VS. WHO consultation on Respiratory Syncytial Virus Vaccine Development Report from a World Health Organization Meeting held on 23-24 March 2015. Vaccine. 2016;34(2):190–7.

86. Bont L, Checchia PA, Fauroux B, Figueras-Aloy J, Manzoni P, Paes B, et al. Defining the Epidemiology and Burden of Severe Respiratory Syncytial Virus Infection Among Infants and Children in Western Countries. Infect Dis Ther. 2016;5(3):271–98.

87. Anderson LJ, Dormitzer PR, Nokes DJ, Rappuoli R, Roca A, Graham BS. Strategic priorities for respiratory syncytial virus (RSV) vaccine development. Vaccine. 2013;31 Suppl 2:B209-15.

88. August A, Glenn GM, Kpamegan E, Hickman SP, Jani D, Lu H, et al. A Phase 2 randomized, observer-blind, placebo-controlled, dose-ranging trial of aluminum-adjuvanted respiratory syncytial virus F particle vaccine formulations in healthy women of childbearing age. Vaccine. 2017;35(30):3749–59.

89. Novavax Inc. Novavax; Part 2: A Deep Dive Into Phase 3 Of Its RSV Vaccine For Older Adults [Internet]. 2016. [cited 2017 Jul 8]. Available from:

https://seekingalpha.com/article/4007281-novavax-part-2-deep-dive-phase-3-rsv-vaccine-older-adults

90. Bavarian Nordic. Bavarian Nordic announces positive date from ongoing Phase 2 study investigating a universal RSV vaccine [Internet]. [cited 2017 Jul 8]. Available from: http://www.bavarian-nordic.com/investor/news/news.aspx?news=5247

91. Anderson LJ. Respiratory syncytial virus vaccine development. Semin Immunol. 2013;25(2):160–71.

92. Neuzil KM. Progress toward a Respiratory Syncytial Virus Vaccine. Clin Vaccine Immunol. 2016;23(3):186–8.

93. Haynes LM. Progress and challenges in RSV prophylaxis and vaccine development. J Infect Dis. 2013;208 Suppl(suppl_3):S177-83.

94. McQuaid F, Jones C, Stevens Z, Plumb J, Hughes R, Bedford H, et al. Factors influencing women's attitudes towards antenatal vaccines, group B Streptococcus and clinical trial participation in pregnancy: an online survey. BMJ Open. 2016;6(4):e010790.

95. Jorquera PA, Anderson L, Tripp RA. Understanding respiratory syncytial virus (RSV) vaccine development and aspects of disease pathogenesis. Expert Rev Vaccines. 2016;15(2):173–87.

96. Ericsson CD, Steffen R, Ess SM, Szucs TD. Economic Evaluation of Immunization Strategies. Clin Infect Dis. 2002;35(3):294–7.

97. Lugnér AK, Mylius SD, Wallinga J. Dynamic versus static models in cost-effectiveness analyses of anti-viral drug therapy to mitigate an influenza pandemic. Health Econ. 2009;19(5):n/a-n/a.

98. Cromer D, van Hoek AJ, Newall AT, Pollard AJ, Jit M. Burden of paediatric respiratory syncytial virus disease and potential effect of different immunisation strategies: a modelling and cost-effectiveness analysis for England. Lancet Public Heal. 2017;2(8):e367–74.

99. Baguelin M, Flasche S, Camacho A, Demiris N, Miller E, Edmunds WJ. Assessing Optimal Target Populations for Influenza Vaccination Programmes: An Evidence Synthesis and Modelling Study. Leung GM, editor. PLoS Med. 2013;10(10):e1001527.

100. Hodgson D, Baguelin M, van Leeuwen E, Panovska-Griffiths J, Ramsay M, Pebody R, et al. Effect of mass paediatric influenza vaccination on existing influenza vaccination programmes in England and Wales: a modelling and cost-effectiveness analysis. Lancet Public Heal. 2017;2(2):e74–81.

101. Campbell H, Bont L, Nair H. Respiratory syncytial virus (RSV) disease - new data needed to guide future policy. J Glob Health. 2015;5(2):20101.

102. Bont L, Baraldi E, Fauroux B, Greenough A, Heikkinen T, Manzoni P, et al. RSV-Still More Questions Than Answers. Pediatr Infect Dis J. 2014;33(11):1177–9.

103. Lloyd PC, May L, Hoffman D, Riegelman R, Simonsen L. The effect of birth month on the risk of respiratory syncytial virus hospitalization in the first year of life in the United States. Pediatr Infect Dis J. 2014;33(6):e135-40.

104. Houben ML, Bont L, Wilbrink B, Belderbos ME, Kimpen JLL, Visser GHA, et al. Clinical prediction rule for RSV bronchiolitis in healthy newborns: prognostic birth cohort study. Pediatrics. 2011;127(1):35–41.

105. Holberg CJ, Wright AL, Martinez FD, Ray CG, Taussig LM, Lebowitz MD. Risk factors for respiratory syncytial virus-associated lower respiratory illnesses in the first year of life. Am J Epidemiol. 1991;133(11):1135–51.

106. Figueras-Aloy J, Carbonell-Estrany X, Quero-Jiménez J, Fernández-Colomer B, Guzmán-Cabañas J, Echaniz-Urcelay I, et al. FLIP-2 Study: risk factors linked to respiratory syncytial virus infection requiring hospitalization in premature infants born in Spain at a gestational age of 32 to 35 weeks. Pediatr Infect Dis J. 2008;27(9):788–93.

107. ReSViNET. About - ReSViNET [Internet]. [cited 2017 Sep 14]. Available from: http://www.resvinet.org/about1.html

108. RESCEU. RESCEU – REspiratory Syncytial virus Consortium in EUrope [Internet]. [cited 2017 Sep 14]. Available from: http://resc-eu.org/

109. Jutte DP, Roos LL, Brownell MD. Administrative Record Linkage as a Tool for Public Health Research. Annu Rev Public Health. 2011;32(1):91–108.

110. Murray J, Bottle A, Sharland M, Modi N, Aylin P, Majeed A, et al. Risk factors for hospital admission with RSV bronchiolitis in England: a population-based birth cohort study. PLoS One. 2014;9(2):e89186.

111. Pitman RJ, Melegaro A, Gelb D, Siddiqui MR, Gay NJ, Edmunds WJ. Assessing the burden of influenza and other respiratory infections in England and Wales. J Infect. 2007;54(6):530–8.

112. Cromer D, van Hoek AJ, Jit M, Edmunds WJ, Fleming D, Miller E. The burden of influenza in England by age and clinical risk group: a statistical analysis to inform vaccine policy. J Infect. 2014;68(4):363–71.

113. Taylor S, Taylor RJ, Lustig RL, Schuck-Paim C, Haguinet F, Webb DJ, et al. Modelling estimates of the burden of respiratory syncytial virus infection in children in the UK. BMJ Open. 2016;6(6):e009337.

114. Homaira N, Oei J-L, Mallitt K-A, Abdel-Latif ME, Hilder L, Bajuk B, et al. High burden of RSV hospitalization in very young children: a data linkage study.

Epidemiol Infect. 2015;1–10.

115. Moore HC, de Klerk N, Keil AD, Smith DW, Blyth CC, Richmond P, et al. Use of data linkage to investigate the aetiology of acute lower respiratory infection hospitalisations in children. J Paediatr Child Health. 2012;48(6):520–8.

116. Pisesky A, Benchimol EI, Wong CA, Hui C, Crowe M, Belair M-A, et al. Incidence of Hospitalization for Respiratory Syncytial Virus Infection amongst Children in Ontario, Canada: A Population-Based Study Using Validated Health Administrative Data. PLoS One. 2016;11(3):e0150416.

117. Stensballe LG. An epidemiological study of respiratory syncytial virus associated hospitalizations in Denmark. Respir Res. 2002;3 Suppl 1:S34-9.

118. Müller-Pebody B, Edmunds WJ, Zambon MC, Gay NJ, Crowcroft NS. Contribution of RSV to bronchiolitis and pneumonia-associated hospitalizations in English children, April 1995-March 1998. Epidemiol Infect. 2002;129(1):99–106.

119. Zhao H, Green H, Lackenby A, Donati M, Ellis J, Thompson C, et al. A new laboratory-based surveillance system (Respiratory DataMart System) for influenza and other respiratory viruses in England: results and experience from 2009 to 2012. Euro Surveill. 2014;19(3).

120. Bradley CJ, Penberthy L, Devers KJ, Holden DJ. Health services research and data linkages: issues, methods, and directions for the future. Health Serv Res. 2010;45(5 Pt 2):1468–88.

121. Harron K, Goldstein H, Wade A, Muller-Pebody B, Parslow R, Gilbert R. Linkage, evaluation and analysis of national electronic healthcare data: application to providing enhanced blood-stream infection surveillance in paediatric intensive care. PLoS One. 2013;8(12):e85278.

122. Atchison CJ, Lopman BA, Harris CJ, Tam CC, Iturriza Gómara M, Gray JJ. Clinical laboratory practices for the detection of rotavirus in England and Wales: can surveillance based on routine laboratory testing data be used to evaluate the impact of vaccination? Euro Surveill. 2009;14(20).

123. Harron K, Mok Q, Parslow R, Muller-Pebody B, Gilbert R, Ramnarayan P. Risk of bloodstream infection in children admitted to paediatric intensive care units in England and Wales following emergency inter-hospital transfer. Intensive Care Med. 2014;40(12):1916–23.

124. Pearson A, Chronias A, Murray M. Voluntary and mandatory surveillance for methicillin-resistant Staphylococcus aureus (MRSA) and methicillin-susceptible S. aureus (MSSA) bacteraemia in England. J Antimicrob Chemother. 2009;64 Suppl 1(Supplement 1):i11-7.

125. Jaro MA. Probabilistic linkage of large public health data files. Stat Med. 14(5–7):491–8.

126. Public Health England. Respiratory infections: laboratory reports 2016 [Internet]. 2016 [cited 2017 Jan 10]. Available from: https://www.gov.uk/government/publications/respiratory-infections-laboratory-reports-2016

127. Public Health England; Weekly national flu reports - GOV.UK [Internet]. 2017 [cited 2017 Jul 26]. Available from: https://www.gov.uk/government/statistics/weekly-national-flu-reports

128. NHS Digital. Hospital Episode Statistics [Internet]. NHS Digital; [cited 2016 Dec 23]. Available from: http://content.digital.nhs.uk/hes

129. NHS Digital. Payment by Results data [Internet]. NHS Digital; [cited 2016 Dec 22]. Available from: http://content.digital.nhs.uk/PbR

130. NHS Digital. How do we collect and process HES data? [Internet]. [cited 2017 Jan 10]. Available from: http://content.digital.nhs.uk/article/1824/How-do-we-collect-and-process-HES-data

131. Hagger-Johnson G, Harron K, Fleming T, Gilbert R, Goldstein H, Landy R, et al. Data linkage errors in hospital administrative data when applying a pseudonymisation algorithm to paediatric intensive care records. BMJ Open. 2015;5(8):e008118.

132. NHS Digital. The processing cycle and HES data quality [Internet]. [cited 2017 Jan 10]. Available from: http://content.digital.nhs.uk/article/1825/The-processing-cycle-and-HES-data-quality

133. World Health Organisation; ICD-10 Version:2010 [Internet]. 2017 [cited 2017 Jul 26]. Available from: http://apps.who.int/classifications/icd10/browse/2010/en

134. McGuiness CB, Boron ML, Saunders B, Edelman L, Kumar VR, Rabon-Stith KM. Respiratory syncytial virus surveillance in the United States, 2007-2012: results from a national surveillance system. Pediatr Infect Dis J. 2014;33(6):589–94.

135. Mullins JA, Lamonte AC, Bresee JS, Anderson LJ. Substantial variability in community respiratory syncytial virus season timing. Pediatr Infect Dis J. 2003;22(10):857–62.

136. Mytton OT, Rutter PD, Donaldson LJ. Influenza A(H1N1)pdm09 in England, 2009 to 2011: a greater burden of severe illness in the year after the pandemic than in the pandemic year. Eurosurveillance. 2012;17(14):3.

137. Alvarez AE, Marson FA de L, Bertuzzo CS, Arns CW, Ribeiro JD. Epidemiological and genetic characteristics associated with the severity of acute viral bronchiolitis by respiratory syncytial virus. J Pediatr (Rio J).

2013;89(6):531–43.

138. PATH. RSV Vaccine Snapshot - PATH Vaccine Resource Library [Internet]. 2015 [cited 2016 Jan 7]. Available from: http://www.path.org/vaccineresources/details.php?i=1562

139. Reeves RM, Hardelid P, Gilbert R, Warburton F, Ellis J, Pebody RG. Estimating the burden of respiratory syncytial virus (RSV) on respiratory hospital admissions in children less than five years of age in England, 2007-2012. Influenza Other Respi Viruses. 2017;11(2):122–9.

140. Office for National Statistics. Population Estimates for UK, England and Wales, Scotland and Northern Ireland. 2011.

141. Office for National Statistics. Population Estimates for UK, England and Wales, Scotland and Northern Ireland, Mid-2011 and Mid-2012. 2013.

142. Cooper DL, Smith GE, Edmunds WJ, Joseph C, Gerard E, George RC. The contribution of respiratory pathogens to the seasonality of NHS Direct calls. J Infect. 2007;55(3):240–8.

143. Matias G, Taylor R, Haguinet F, Schuck-Paim C, Lustig R, Shinde V. Estimates of mortality attributable to influenza and RSV in the United States during 1997-2009 by influenza type or subtype, age, cause of death, and risk status. Influenza Other Respi Viruses. 2014;8(5):507–15.

144. Fleming DM, Taylor RJ, Lustig RL, Schuck-Paim C, Haguinet F, Webb DJ, et al. Modelling estimates of the burden of Respiratory Syncytial virus infection in adults and the elderly in the United Kingdom. BMC Infect Dis. 2015;15:443.

145. Simoes EAF. Environmental and demographic risk factors for respiratory syncytial virus lower respiratory tract disease. J Pediatr. 2003;143(5 Suppl):S118-26.

146. Ajayi-Obe EK, Coen PG, Handa R, Hawrami K, Aitken C, McIntosh EDG, et al. Influenza A and respiratory syncytial virus hospital burden in young children in East London. Epidemiol Infect. 2008;136(8):1046–58.

147. Iwane MK, Edwards KM, Szilagyi PG, Walker FJ, Griffin MR, Weinberg GA, et al. Population-based surveillance for hospitalizations associated with respiratory syncytial virus, influenza virus, and parainfluenza viruses among young children. Pediatrics. 2004;113(6):1758–64.

148. Simpson MD, Kieke BA, Sundaram ME, McClure DL, Meece JK, Sifakis F, et al. Incidence of Medically Attended Respiratory Syncytial Virus and Influenza Illnesses in Children 6-59 Months Old During Four Seasons. Open forum Infect Dis. 2016;3(2):ofw081.

149. Dusetzina S, Tyree S, Meyer A, Al. E. Linking Data for Health Services Research: A Framework and Instructional Guide [Internet]. In: An Overview of Record Linkage Methods. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014. p. Available from: https://www.ncbi.nlm.nih.gov/books.

150. Harron K, Goldstein H, Dibben C. Methodological Developments in Data Linkage. John Wiley & Sons, Ltd; 2016.

151. Harron K. Introduction to Data Linkage [Internet]. 2016 [cited 2017 Sep 12]. Available from: https://adrn.ac.uk/media/174200/data_linkage_katieharron_2016.pdf

152. Dusetzina SB, Tyree S, Meyer A-M, Meyer A, Green L, Carpenter WR. Linking Data for Health Services Research: A Framework and Instructional Guide [Internet]. In: An Overview of Record Linkage Methods. Agency for Healthcare Research and Quality (US); 2014.

153. NHS Digital; Methodology for creation of the HES Patient ID (HESID) [Internet]. 2014 [cited 2017 Jul 30]. Available from:

http://content.digital.nhs.uk/media/1370/HES-Hospital-Episode-Statistics-Replacement-of-the-HES-patient-ID/pdf/HESID_Methodology.pdf

154. Sayers A, Ben-Shlomo Y, Blom AW, Steele F. Probabilistic record linkage. Int J Epidemiol. 2016;45(3):954–64.

155. Grannis SJ, Overhage JM, Hui S, McDonald CJ. Analysis of a probabilistic record linkage technique without human review. AMIA . Annu Symp proceedings AMIA Symp. 2003;2003:259–63.

156. DuVall SL, Kerber RA, Thomas A. Extending the Fellegi-Sunter probabilistic record linkage method for approximate field comparators. J Biomed Inform. 2010;43(1):24–30.

157. Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data via the EM Algorithm. J R Stat Soc Ser B. 1977;39:1–38.

158. Market Research Society (MRS). Postcode Format [Internet]. [cited 2017 Apr 18]. Available from: https://www.mrs.org.uk/pdf/postcodeformat.pdf

159. Abels S, Nadal D, Stroehle A, Bossart W. Reliable detection of respiratory syncytial virus infection in children for adequate hospital infection control management. J Clin Microbiol. 2001;39(9):3135–9.

160. Department of Health. NHS reference costs 2009-2010 [Internet]. 2010 [cited 2017 Jul 27]. Available from: https://www.gov.uk/government/publications/nhs-reference-costs-2009-2010

161. Deshpande SA, Northern V. The clinical and health economic burden of respiratory syncytial virus disease among children under 2 years of age in a defined geographical area. Arch Dis Child. 2003;88(12):1065–9.

162. Clark SJ, Beresford MW, Subhedar N V, Shaw NJ. Respiratory syncytial virus infection in high risk infants and the potential impact of prophylaxis in a United Kingdom cohort. Arch Dis Child. 2000;83(4):313–6.

163. Thorburn K. Pre-existing disease is associated with a significantly higher risk of death in severe respiratory syncytial virus infection. Arch Dis Child. 2008;94(2):99–103.

164. NHS Digital. HES data dictionary [Internet]. 2017 [cited 2017 Jul 27]. Available from: http://content.digital.nhs.uk/hesdatadictionary

165. Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol. 2004;159(7):702–6.

166. Shi T, Balsells E, Wastnedge E, Singleton R, Rasmussen ZA, Zar HJ, et al. Risk factors for respiratory syncytial virus associated with acute lower respiratory infection in children under five years: Systematic review and meta–analysis. J Glob Health. 2015;5(2):20416.

167. Carpenter TC, Stenmark KR. Predisposition of infants with chronic lung disease to respiratory syncytial virus-induced respiratory failure: a vascular hypothesis. Pediatr Infect Dis J. 2004;23(Supplement):S33–40.

168. Wang D, Cummins C, Bayliss S, Sandercock J, Burls A. Immunoprophylaxis against respiratory syncytial virus (RSV) with palivizumab in children: a systematic review and economic evaluation. Health Technol Assess. 2008;12(36):iii, ix–x, 1-86.

169. Paes B, Fauroux B, Figueras-Aloy J, Bont L, Checchia PA, Simões EAF, et al. Defining the Risk and Associated Morbidity and Mortality of Severe Respiratory Syncytial Virus Infection Among Infants with Chronic Lung Disease. Infect Dis Ther. 2016;5(4):453–71.

170. Jung JW. Respiratory syncytial virus infection in children with congenital heart disease: global data and interim results of Korean RSV-CHD survey. Korean J Pediatr. 2011;54(5):192–6.

171. Checchia PA, Paes B, Bont L, Manzoni P, Simões EAF, Fauroux B, et al. Defining the Risk and Associated Morbidity and Mortality of Severe Respiratory Syncytial Virus Infection Among Infants with Congenital Heart Disease. Infect Dis Ther. 2017;6(1):37–56.

172. Fixler DE. Respiratory syncytial virus infection in children with congenital heart disease: A review. Pediatr Cardiol. 1996;17(3):163–8.

173. Resch B, Manzoni P, Lanari M. Severe respiratory syncytial virus (RSV) infection in infants with neuromuscular diseases and immune deficiency syndromes. Paediatr Respir Rev. 2009;10(3):148–53.

174. Manzoni P, Figueras-Aloy J, Simões EAF, Checchia PA, Fauroux B, Bont L, et al. Defining the Incidence and Associated Morbidity and Mortality of Severe Respiratory Syncytial Virus Infection Among Children with Chronic Diseases. Infect Dis Ther. 2017;6(3):282–411.

175. Royal College of Surgeons. Clinical coding and your data [Internet]. 2015 [cited 2017 Aug 1]. Available from: https://www.rcseng.ac.uk/-/media/files/rcs/standards-and-research/standards-and-policy/audit-outcomes-and-data/clinical-coding-and-your-data-updated-2016.pdf?la=en

176. Capita. The quality of clinical coding in the NHS: Payment by Results data assurance framework [Internet]. 2014. Available from: http://www.chks.co.uk/userfiles/files/The_quality_of_clinical_coding_in_the_NHS.pdf

177. Paediatric Intensive Care Audit Network (PICANet); About PICANet [Internet]. 2017 [cited 2017 Jul 30]. Available from: http://www.picanet.org.uk/About/

178. Nicholson KG, McNally T, Silverman M, Simons P, Stockton JD, Zambon MC. Rates of hospitalisation for influenza, respiratory syncytial virus and human metapneumovirus among infants and young children. Vaccine. 2006;24(1):102–8.

179. Stolwijk AM, Straatman H, Zielhuis GA. Studying seasonality by using sine and cosine functions in regression analysis. J Epidemiol Community Health. 1999;53(4):235–8.

180. Edwards JH. The recognition and estimation of cyclic trends. Ann Hum Genet. 1961;25:83–7.

181. Bont L, Checchia PA, Fauroux B, Figueras-Aloy J, Manzoni P, Paes B, et al. Defining the Epidemiology and Burden of Severe Respiratory Syncytial Virus Infection Among Infants and Children in Western Countries. Infect Dis Ther. 2016;5(3):271–98.

182. Resch B. Respiratory Syncytial Virus Infection in High-risk Infants - an Update on Palivizumab Prophylaxis. Open Microbiol J. 2014;8:71–7.

183. Novavax Inc. RSV F Vaccine Maternal Immunization Study in Healthy Third-trimester Pregnant Women. ClinicalTrials.gov [Internet]. 2017 [cited 2017 Jul 25]. Available from: https://clinicaltrials.gov/ct2/show/NCT02247726?term=Novavax&rank=4

184. NHS Digital; Publications Calendar: April 2011 - March 2012 [Internet]. NHS Digital, 1 Trevelyan Square, Boar Lane, Leeds, LS1 6AE, United Kingdom; [cited 2017 Jul 30]. Available from: http://content.digital.nhs.uk/article/1816/Publications-Calendar-April-2011---March-2012

185. Royal College of General Practitioners. Research and Surveillance Centre [Internet]. 2017 [cited 2017 Jul 30]. Available from:

http://www.rcgp.org.uk/clinical-and-research/our-programmes/research-and-surveillance-centre.aspx

186. Taylor S, Taylor RJ, Lustig RL, Schuck-Paim C, Haguinet F, Webb DJ, et al. Modelling estimates of the burden of respiratory syncytial virus infection in children in the UK. BMJ Open. 2016;6(6):e009337.

187. Clinical Practice Research Datalink; CPRD Linked Data [Internet]. 2017 [cited 2017 Jul 30]. Available from: https://www.cprd.com/dataAccess/linkeddata.asp

188. Harron K, Gilbert R, Cromwell D, van der Meulen J, Wisner K, Verdoux H. Linking Data for Mothers and Babies in De-Identified Electronic Health Data. Gebhardt S, editor. PLoS One. 2016;11(10):e0164667.

Appendices

Appendix 1

Epidemiology and Infection publication

Appendix 2

Influenza and Other Respiratory Viruses publication

Accepted: 2 December 2016

DOI: 10.1111/irv.12443

ORIGINAL ARTICLE

WILEY

Estimating the burden of respiratory syncytial virus (RSV) on respiratory hospital admissions in children less than five years of age in England, 2007-2012

Rachel Melanie Reeves[1,2,3] | Pia Hardelid[1,2] | Ruth Gilbert[1,2] | Fiona Warburton[4] | Joanna Ellis[5] | Richard G. Pebody[3]

[1]Farr Institute of Health Informatics Research, London, UK

[2]Institute of Child Health, University College London, London, UK

[3]Respiratory Diseases Department, Public Health England, Colindale, London, UK

[4]Statistics and Modelling Economics Department, Public Health England, Colindale, London, UK

[5]Virus Reference Department, Public Health England, Colindale, London, UK

Correspondence
Rachel Melanie Reeves, Respiratory Diseases Department, Public Health England, London, UK.
Email: Rachel.Reeves@phe.gov.uk

Background: Respiratory syncytial virus (RSV) is a leading cause of hospital admission in young children. With several RSV vaccines candidates undergoing clinical trials, recent estimates of RSV burden are required to provide a baseline for vaccine impact studies.

Objectives: To estimate the number of RSV-associated hospital admissions in children aged <5 years in England over a 5-year period from 2007 using ecological time series modelling of national hospital administrative data.

Patients/Methods: Multiple linear regression modelling of weekly time series of laboratory surveillance data and Hospital Episode Statistics (HES) data was used to estimate the number of hospital admissions due to major respiratory pathogens including RSV in children <5 years of age in England from mid-2007 to mid-2012, stratified by age group (<6 months, 6-11 months, 1-4 years) and primary diagnosis: bronchiolitis, pneumonia, unspecified lower respiratory tract infection (LRTI), bronchitis and upper respiratory tract infection (URTI).

Results: On average, 33 561 (95% confidence interval 30 429-38 489) RSV-associated hospital admissions in children <5 years of age occurred annually from 2007 to 2012. Average annual admission rates were 35.1 (95% CI: 32.9-38.9) per 1000 children aged <1 year and 5.31 (95% CI: 4.5-6.6) per 1000 children aged 1-4 years. About 84% (95% CI: 81-91%) of RSV-associated admissions were for LRTI. The diagnosis-specific burden of RSV-associated admissions differed significantly by age group.

Conclusions: RSV remains a significant cause of hospital admissions in young children in England. Individual-level analysis of RSV-associated admissions is required to fully describe the burden by age and risk group and identify optimal prevention strategies.

KEYWORDS

bronchiolitis, bronchitis, child, England, hospital admissions, pneumonia, respiratory syncytial virus

1 | BACKGROUND

Respiratory syncytial virus (RSV) is a major cause of respiratory tract infections (RTI) worldwide.[1] In older children and adults, RSV infection often leads to mild upper respiratory tract infection (URTI). However, in infants and young children, RSV is an important cause of severe respiratory infection, particularly bronchiolitis, which may require hospital admission.[2] With a number of RSV vaccine candidates now in phase 2 and 3 clinical trials, it is essential to have accurate estimates of the hospital burden of RSV by age and risk group in order to determine the potential benefits of a future vaccine programme.[3]

Calculating the national burden of disease due to RSV is not straightforward. A reliable diagnosis of RSV infection relies on the detection of RSV in respiratory secretions, but only a minority of children hospitalised with an acute respiratory infection will undergo laboratory testing to identify the causal pathogen.[4] The vast majority of respiratory infections are therefore recorded in hospital admission data under non-specific diagnosis such as unspecified pneumonia or bronchiolitis. Hospital admission data alone can therefore not be used to accurately calculate the burden of RSV in secondary care.

Respiratory pathogens have varying temporal patterns which can be observed using laboratory surveillance data. Statistical models which utilise the seasonal variation in laboratory reports by pathogen can be constructed to attribute hospital admissions to different viruses.[4,5] This method of estimating the hospital burden of RSV has previously been used in the UK; however, the most recent study only considers data up to 2009.[4–6] Surveillance of respiratory viruses including RSV has been strengthened in England since the 2009 influenza A (H1N1) pandemic, with more widespread use of laboratory confirmation for respiratory viruses with PCR methods, and a greater degree of reporting to national surveillance schemes.[7] In addition, a recent study in the UK demonstrates that hospital admission due to bronchiolitis is increasing over time.[8] These developments emphasise the need for more recent estimates of RSV-associated hospital admissions as previous estimates may not reflect the current burden of disease.

The aim of this work was to estimate the number of hospital admissions attributable to RSV in children <5 years of age in England in the period from mid-2007 to mid-2012 using ecological time series modelling of national laboratory surveillance and hospital administrative data.

2 | METHODS

2.1 | Data sources

2.1.1 | National laboratory reports

The Second Generation Surveillance System (SGSS)—formerly LabBase2—is a microbiology laboratory surveillance database at the Centre for Infectious Disease Surveillance and Control (CIDSC) at Public Health England (PHE), the national centre responsible for infectious disease surveillance, prevention and control. Positive test results for microorganisms recorded in SGSS are collected from local pathology systems transmitted from PHE, NHS and private microbiology laboratories in England.[9] All clinically significant microorganisms should be reported, although no guidelines for the judgement of clinical significance are defined. Weekly laboratory reports of samples positive for RSV, influenza A, influenza B, rhinovirus, parainfluenza, human metapneumovirus (hMPV), adenovirus, *Streptococcus pneumoniae*, *Mycoplasma pneumoniae* and *Haemophilus influenza* in children <5 years of age (at the time of the sample) in England from calendar week 27 (mid-July) in 2007 to calendar week 28 (mid-July) in 2012 were extracted. If individuals had multiple positive samples for the same virus within a two-week period, only one record was included in this analysis—this was to avoid multiple tests in the same individual within the same infection episode being counted as separate infections.

2.1.2 | Hospital Episode Statistics (HES)

The Hospital Episode Statistics (HES) admitted patient care database, held by the Health and Social Care Information Centre (HSCIC), contains routinely collected data on all admissions to all NHS hospitals in England. Records include clinical, geographical and administrative information including admission and discharge dates, on every patient. In this analysis, an admission refers to a single HES spell—from admission to discharge in one hospital.

Diagnoses are recorded in HES using International Classification of Diseases 10th Revision (ICD-10) codes, with up to 20 diagnosis codes allowed per HES episode. Each admission is allocated a primary diagnosis which is the main reason for the length of stay in hospital. All admissions with a primary diagnosis of ICD-10 codes for bronchiolitis (ICD-10 J21), pneumonia (J12-18), unspecified lower respiratory tract infection (LRTI) (J22), bronchitis (J20) and upper respiratory tract infection (URTI) (J00-06) in children <5 years of age (at admission) in England from calendar week 27 in 2007 to calendar week 26 in 2012 were included in the study. Calendar weeks were defined as blocks of 7 days beginning on 1st January each year, with week 52 allowed to have more than 7 days. Only the primary diagnosis was included to avoid double counting of admissions which may have two or more of these diagnoses. HES data were not available for 2013 onwards. The weekly number of hospital admissions was stratified into three groups: <6 months, 6-11 months and 1-4 years.

2.2 | Statistical analysis

In this study, we use the observed temporal variation in weekly laboratory reports of potential causative pathogens to estimate the number of RTI hospital admissions that could be attributed to RSV, building on methods applied in previous modelling studies.[4,10,11] Separate models were developed for each primary diagnosis, using the weekly number of hospital admissions in children <5 years of age in England for each respective diagnosis as the dependent variable. For each diagnosis, separate models were constructed by age group (<6 months, 6-11 months, 1-4 years).

Multiple linear regression models were used to estimate the number of hospital admissions due to RSV from HES data coded as acute bronchiolitis, pneumonia, unspecified LRTI, bronchitis and URTI. These models have been used in similar studies.[4,10,12,13] All models used the weekly number of laboratory-confirmed episodes in children <5 years of age in England for the following pathogens as the independent variables: RSV, influenza A, influenza B, rhinovirus, parainfluenza, human metapneumovirus (hMPV) and adenovirus. The weekly number of laboratory-confirmed episodes of *Streptococcus pneumoniae*, *Mycoplasma pneumoniae* and *Haemophilus influenza* in children <5 years of age in England was also included as independent variables in the models for pneumonia and unspecified LRTI. All variables were first included in the models, then those with negative coefficients removed (in order of decreasing significance) due to biological implausibility (pathogens cannot cause a negative number of hospital admissions), followed by those that did not contribute significantly to the model (*F*-test $P>.05$). Interactions between all pathogens in the final model were investigated ($P\leq.01$ was considered significant) due to the potential for co-circulation of pathogens. Interactions between pathogens in the final model and an indicator variable taking the value 0 for the pre-pandemic period (before week 20 2009) and 1 for the pandemic and post-pandemic period were investigated to account for potential changes in testing practice following the 2009 influenza pandemic ($P\leq.01$ was considered significant).

Final estimates of the number of hospital admissions attributable to each pathogen (including 95% confidence intervals (CIs)) were calculated by multiplying the coefficient from the final model by the total number of weekly laboratory-confirmed episodes for each relevant pathogen. The total number of hospital admissions for all children <5 years old was calculated as the sum of the number of hospital admissions calculated in each age group for each pathogen.

Admission rates were calculated using ONS mid-year population estimates for England by age group (<1 year, 1-4 years).[14,15] The average of ONS mid-year population estimates for 2007 and 2008 was used as the denominator for the 2007/2008 epidemiological year and the average of ONS mid-year population estimates for 2008 and 2009 used for the 2008/2009 epidemiological year, etc.

3 | RESULTS

3.1 | Seasonality of laboratory reports and hospital admissions

The temporal variation in laboratory reports by pathogen is shown in Figure 1. The temporal trends in hospital admissions for children <5 years old in England with a primary diagnosis of bronchiolitis, pneumonia, bronchitis, unspecified LRTI or URTI are shown in Figure 2. Hospital admissions with a primary diagnosis of bronchiolitis were markedly seasonal and mirror the pattern of laboratory-confirmed RSV infections. Hospital admissions with a primary diagnosis of pneumonia had a very similar seasonal pattern to hospital admissions with a primary diagnosis of unspecified LRTI, with peaks also occurring at the same time as the peaks in laboratory-confirmed RSV infections in SGSS each year (Figure 2).

3.2 | Estimated RSV-associated hospital admissions

3.2.1 | All RTI admissions

Of the 121 968 hospital admissions with a primary diagnosis of bronchiolitis, pneumonia, bronchitis, unspecified LRTI or URTI that occurred annually in children <5 years of age from mid-2007 to mid-2012, we estimate that 28% (33 561/121 968, 95% CI: 25-32%) were due to RSV. Of RSV-associated RTI hospital admissions, 48% (16 202/33 561, 95% CI: 46-52%) were in children <6 months of age, 21% (7108/33 561, 95% CI: 19-25%) were in children 6-11 months of age, and 31% (10 251/33 561, 95% CI: 26-38%) were in children aged 1-4 years (Table 1). The majority, 84% (28 111/33 561, 95% CI: 81-91%), of RSV-associated RTI hospital admissions were for LRTI. About 65% (21 418/33 561, 95% CI: 62-70%) of RSV-associated hospital admissions were coded as bronchiolitis (Tables 1 and 2).

On average, the estimated admission rate of any RSV-associated RTI hospital admission was 35.1 (95% CI: 32.9-38.9) per 1000 children <1 year of age and 5.31 (95% CI: 4.5-6.6) per 1000 children 1-4 years of age, per epidemiological year (Table 3). Estimated rates of RSV-associated hospital admissions were, on average, higher in <1-year-olds for all diagnoses except unspecified LRTI, where admission rates in 1- to 4-year-olds were higher (Table 3). All LRTI admission rates increased in both age groups over time, peaking in 2010/2011 (the RSV season following the 2009 influenza A (H1N1) pandemic, which was an intense seasonal influenza season dominated by A/H1N1pdm09) and with a slight decrease in 2011/2012 for all diagnoses except pneumonia. Admissions for URTI decreased over the study period, particularly in the 1-4 years of age group which saw a 70% decrease from 2.9 (95% CI: 2.5-3.4) admissions per 1000 children in 2007/2008 to 0.9 (95% CI: 0.3-2.1) per 1000 children in 2011/2012. The percentage of weekly hospital admissions attributed to RSV varied by calendar week for all primary diagnoses and age groups (Figure 3).

3.2.2 | LRTI admissions: bronchiolitis, pneumonia, bronchitis and unspecified LRTI

Overall, 53% (28 111/53 514, 95% CI: 51-57%) of LRTI admissions in children <5 years of age could be attributed to RSV, and 54% (15 260/28 111, 95% CI: 52-58%) of RSV-associated LRTIs were in infants <6 months old (Table 1). Of RSV-associated LRTI admissions, 78% (21 830/28 111, 95% CI: 74-83%) were coded as bronchiolitis, 13% (3766/28 111, 95% CI: 12-15%) were coded as unspecified LRTI, 8% (2346/28 111, 95% CI: 7-9%) were coded as pneumonia, and <1% (169/28 111) were coded as bronchitis.

There was an annual average of 27 969 hospital admissions with a primary diagnosis of bronchiolitis in children <5 years of age in England from 2007 to 2012. We estimate that approximately 78% (21 830/27 969, 95% CI: 75-83%) of these were due to RSV (Tables 1 and 2). There were differences by age, with approximately 82%

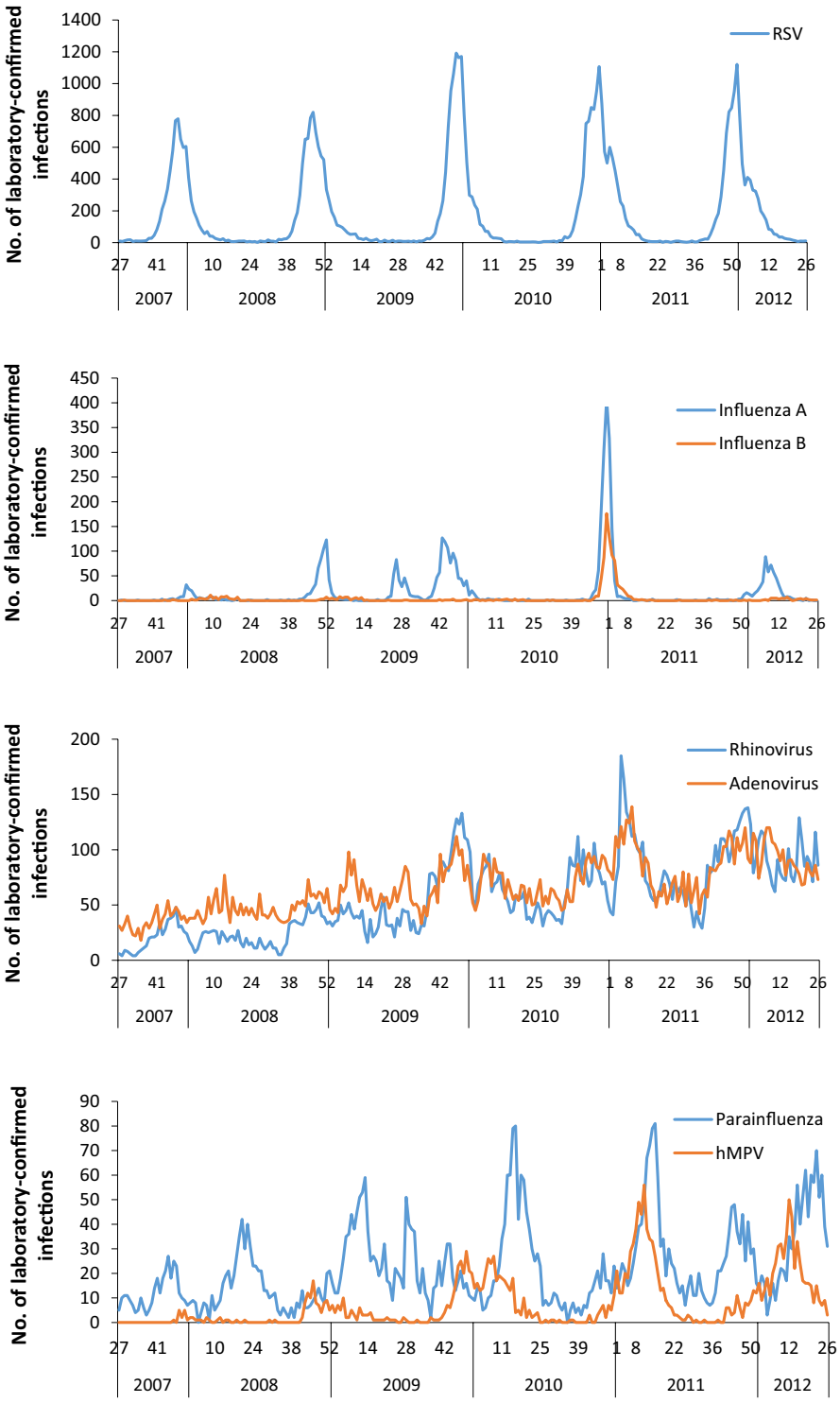

FIGURE 1 Weekly number of laboratory-confirmed cases of major respiratory viruses recorded in SGSS for children <5 y of age, over time

(14 962/18 246, 95% CI: 79-87%) of all bronchiolitis admissions in children aged <6 months attributable to RSV, 70% (5319/7652, 95% CI: 66-75%) of bronchiolitis admissions in children aged 6-11 months attributable to RSV and 75% (1549/2071, 95% CI: 71-79%) of bronchiolitis admissions in children aged 1-4 years attributable to RSV (Table 2). During the period from calendar week 46 to week 2, over 90% of bronchiolitis admissions in children aged <6 months and over 80% in children aged 6-11 months and 1-4 years were attributable to RSV each week (Figure 3). The other explanatory pathogens for bronchiolitis admissions were parainfluenza, hMPV and rhinovirus, with differences in causal pathogens by age (see Supplementary Data).

There was an annual average of 9537 hospital admissions with a primary diagnosis of pneumonia in children <5 years of age in England from 2007 to 2012. We estimate that approximately 25% (2346/9537, 95% CI: 22-27%) of these were due to RSV (Table 1). There were differences by age, with approximately 26% (1923/7503,

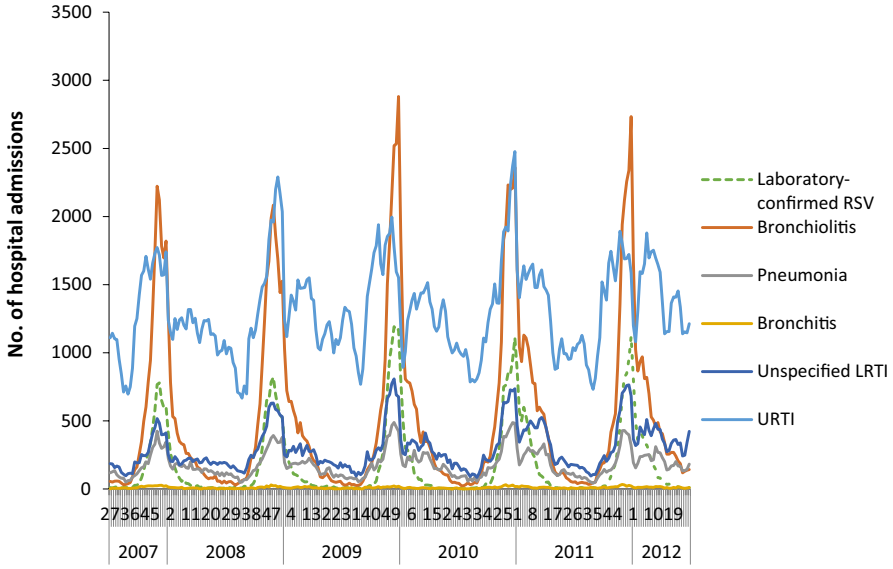

FIGURE 2 Weekly number of hospital admissions in children <5 y of age in England with any diagnosis of URTI, bronchiolitis, pneumonia, bronchitis or unspecified LRTI. Weekly number of laboratory-confirmed RSV infections in children <5 y of age in England (SGSS) shown in orange

95% CI: 23-28%) of pneumonia admissions in children aged 1-4 years attributable to RSV compared to 19% in children aged <6 months (138/739, 95% CI: 14-28%). The other main explanatory pathogens for hospital admissions with a primary diagnosis of pneumonia in children <5 years of age were *S. pneumoniae*, hMPV, parainfluenza, influenza A and rhinovirus, with differences in causal pathogens by age.

There was an annual average of 547 hospital admissions with a primary diagnosis of acute bronchitis in children <5 years of age in England from 2007 to 2012. We estimate that approximately 31% (169/547, 95% CI: 23-42%) of these were due to RSV (Table 1). There were differences by age, with approximately 56% (89/159, 95% CI: 48-71%) of bronchitis hospital admissions in children <6 months of age attributed to RSV, compared to only 14% (39/276, 95% CI: 8-20%) in children aged 1-4 years (Table 2). The other explanatory pathogens for hospital admissions with a primary diagnosis of bronchitis were parainfluenza and adenovirus in children aged 1-4 years only.

There was an annual average of 15 461 hospital admissions with a primary diagnosis of unspecified LRTI in children <5 years of age in England from 2007 to 2012. We estimate that approximately 24%

TABLE 1 Average annual number of hospital admissions in children <5 y of age in England: total per primary diagnosis and number estimated to be attributed to RSV by the final models, stratified by age group (<6 mo, 6-11 mo, 1-4 y)

	Annual number of hospital admissions (95% CI)			
Primary diagnosis	**<6 mo**	**6-11 mo**	**1-4 y**	**Total**
Bronchiolitis				
Total	18 246	7652	2071	27 969
RSV-associated	14 962 (14 396-15 942)	5319 (5066-5775)	1549 (1467-1631)	21 830 (20 929-23 348)
Pneumonia				
Total	739	1294	7503	9537
RSV-associated	138 (101-158)	285 (245-324)	1923 (1 718-2 128)	2346 (2064-2610)
Bronchitis				
Total	159	113	276	547
RSV-associated	89 (76-113)	41 (29-62)	39 (23-55)	169 (128-230)
Unspecified LRTI				
Total	1224	2546	11 690	15 461
RSV-associated	71 (37-104)	429 (322-619)	3266 (3007-3525)	3766 (3366-4248)
URTI				
Total	8868	12 737	46 850	68 455
RSV-associated	942 (802-1082)	1034 (742-1559)	3474 (2398-5412)	5450 (3222-8053)
Total	29 236	24 342	68 390	121 968
Total RSV-associated	16 202 (15 412-17 399)	7108 (6404-8339)	10 251 (8613-12 751)	33 561(30 429-38 489)

TABLE 2 Total RTI hospital admissions estimated to be due to RSV by the final models, stratified by age group (<6 mo, 6-11 mo, 1-4 y) and primary diagnosis, as a percentage of the total hospital admissions for the respective primary diagnosis

	Percentage of annual hospital admissions (per primary diagnosis) attributed to RSV (95% CI)			
Primary diagnosis	**<6 mo**	**6-11 mo**	**1-4 y**	**Total**
Bronchiolitis	82% (79-87%)	70% (66-75%)	75% (71-79%)	78% (75-83%)
Pneumonia	19% (14-28%)	22% (19-25%)	26% (23-28%)	25% (22-27%)
Bronchitis	56% (48-71%)	37% (28-55%)	14% (8-20%)	31% (23-42%)
Unspecified LRTI	6% (3-8%)	17% (13-24%)	28% (26-30%)	24% (22-27%)
URTI	11% (9-12%)	8% (6-12%)	7% (5-12%)	8% (5-12%)

(3766/15 461, 95% CI: 22-27%) of these were due to RSV. There were differences by age, with approximately 28% (3266/11 690, 95% CI: 26-30%) of unspecified LRTI hospital admissions in children aged 1-4 years of age attributed to RSV, compared to only approximately 6% (71/1224, 95% CI: 3-8%) in children <6 months of age (Table 2). The other main explanatory pathogens for hospital admissions with a primary diagnosis of unspecified LRTI in children <5 years of age were adenovirus, *S. pneumoniae*, hMPV, parainfluenza, rhinovirus and influenza A, with differences in causal pathogens by age.

3.2.3 | URTI admissions

From 2007 to 2012, there was an annual average of 68 455 hospital admissions with a primary diagnosis of URTI in children <5 years of age in England. We estimate that approximately 8% (5450/68 455, 95% CI: 5-12%) of these were due to RSV (Table 1). There were no significant differences in the percentage of URTI admissions attributable to RSV by age group, although a slightly higher percentage of admissions in children <6 months of age were attributed to RSV (11%, 942/8868, 95% CI: 9-12%). The other explanatory pathogens for hospital admissions with a primary diagnosis of URTI were influenza A, parainfluenza and adenovirus.

The number of hospital admissions attributed to each pathogen in each final model is shown in Table S1.

4 | DISCUSSION

Every year, approximately 33 561 (95% CI: 30 429-38 489) hospital admissions for RTIs in children <5 years of age from 2007 to 2012 could be attributed to RSV. This represents annual RSV-associated RTI admission rates of 35.1 (95% CI: 32.9-38.9) per 1000 children <1 year of age and 5.31 (95% CI: 4.5-6.6) per 1000 children 1-4 years of age. The vast majority of these RSV-associated admissions (84%, 95% CI: 81-91%) were coded as LRTIs. Of the RSV-associated LRTIs, nearly half were in children aged <6 months (48%, 95% CI: 46-52%). Approximately 82% (95% CI: 79-87%) of hospital admissions for bronchiolitis in children aged <6 months could be attributed to RSV.

Our study is the first to determine burden of RSV-associated respiratory hospital admissions in children in England according to primary diagnosis and age group. Our study uses smaller age groupings than previous studies to more precisely reflect the differences in RSV-associated illness by age. Our study is also the first to estimate

TABLE 3 Estimated admission rates of RSV-associated hospital admissions per 1000 children <5 y of age in England.

Estimated admission rate of RSV-associated hospital admissions (per 1000) (95% CI)							
	Age group	**2007/8**	**2008/9**	**2009/10**	**2010/11**	**2011/12**	**Average**
URTI	<1 y	3.00 (2.50-3.50)	3.41 (2.84-3.41)	2.73 (2.03-4.00)	3.00 (2.23-4.39)	2.59 (1.92-3.79)	2.95 (2.31-3.93)
	1-4 y	2.94 (2.49-3.38)	3.29 (2.79-3.29)	0.95 (0.35-2.27)	1.04 (0.38-2.48)	0.90 (0.33-2.14)	1.82 (1.27-2.81)
Bronchiolitis	<1 y	27.1 (26.2-28.1)	30.8 (29.8-31.9)	30.5 (29.2-33.4)	33.6 (32.1-36.7)	29.0 (27.7-31.7)	30.2 (29.0-32.3)
	1-4 y	0.64 (0.60-0.67)	0.72 (0.68-0.76)	0.87 (0.82-0.91)	0.94 (0.89-0.99)	0.82 (0.77-0.86)	0.90 (0.75-0.84)
Unspecified LRTI	<1 y	0.40 (0.24-0.57)	0.46 (0.27-0.65)	0.93 (0.70-1.36)	1.03 (0.78-1.50)	0.89 (0.67-1.29)	0.74 (0.53-1.07)
	1-4 y	1.34 (1.24-1.45)	1.52 (1.40-1.64)	1.82 (1.68-1.97)	1.99 (1.83-2.15)	1.72 (1.58-1.86)	1.68 (1.55-1.81)
Pneumonia	<1 y	0.54 (0.45-0.63)	0.62 (0.52-0.72)	0.65 (0.53-0.84)	0.72 (0.58-0.93)	2.69 (2.28-3.16)	1.04 (0.87-1.26)
	1-4 y	0.79 (0.71-0.88)	0.89 (0.81-0.99)	1.07 (0.96-1.19)	1.17 (1.05-1.30)	1.01 (0.91-1.12)	0.99 (0.88-1.09)
Bronchitis	<1 y	0.21 (0.18-0.24)	0.24 (0.20-0.27)	0.17 (0.13-0.26)	0.19 (0.15-0.29)	0.17 (0.13-0.25)	0.20 (0.16-0.26)
	1-4 y	0.02 (0.01-0.02)	0.02 (0.01-0.03)	0.02 (0.01-0.03)	0.02 (0.01-0.03)	0.02 (0.01-0.03)	0.02 (0.01-0.03)
Total	<1 y	31.3 (29.5-33.0)	35.6 (33.6-37.6)	35.0 (32.6-39.8)	38.5 (35.9-43.8)	35.5 (32.7-40.1)	35.1 (32.9-38.9)
	1-4 y	5.73 (5.05-6.40)	6.43 (5.67-7.20)	4.74 (3.82-6.37)	5.17 (4.17-6.95)	4.47 (3.61-6.01)	5.31 (4.46-6.59)

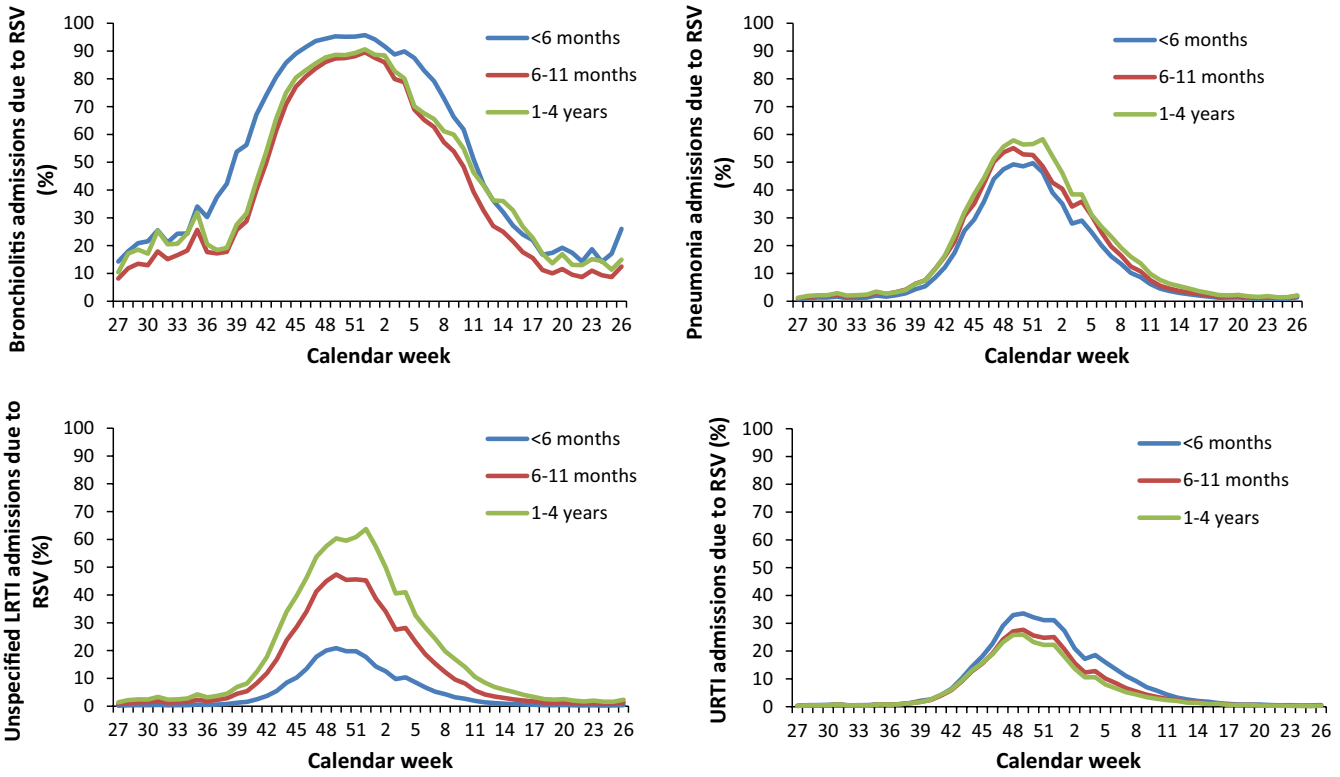

FIGURE 3 Percentage of total weekly hospital admissions due to RSV for each primary diagnosis, by age group

the burden of RSV in England in the post-2009 influenza A (H1N1) pandemic era. However, the methods used here do have potential limitations. Firstly, our models assume that the temporal variation in laboratory reports of the causative agents is an accurate representation of their relative incidence over time. The results will be biased if this is not the case (ie due to seasonal changes in laboratory testing or reporting). Secondly, there may be other reasons for temporal variations that are not accounted for in this analysis (eg meteorological variables), and it is possible that some hospital admissions could be attributed to other pathogens with similar temporal patterns to those that we have considered (eg other bacterial pathogens).[10] Thirdly, it is possible that age-related differences in testing practices may have impacted our quantitative estimates—particularly those for influenza-associated admissions. However, that our results demonstrate a significantly higher burden of RSV-associated hospital admissions in young children compared to other respiratory pathogens, including influenza, is consistent with the results of previous studies.[16–18] Finally, as we restricted hospital admissions to those with a primary diagnosis of URTI or LRTI, our results are likely to be an underestimate of the true burden of RSV-associated hospital admissions.

Previous estimates of RSV burden in secondary care in England range from 26 500 to 29 160 RSV-associated hospital admissions per year in children <5 years of age—all lower than our estimate of 33 561 (95% CI: 30 429-38 489) RSV-associated admissions, which is probably due to those studies considering earlier time periods.[4–6,10] Our study found a steady, general increase in the admission rates of RSV-associated LRTI hospital admissions from 2007/2008 to 2011/2012, peaking during the 2010/2011 season. There are a number of potential explanations for this increase over time. Our models are sensitive to the number of positive laboratory tests; therefore, our results could be affected by a change in the relative sensitivity or specificity of the assays for the different pathogens over time, although the majority of laboratories have been using RT-PCR for most of the study period. However, from 1999 to 2010, there has been an increase of 28% in emergency hospital admissions in children <5 years of age, particularly admissions for acute illness, and the annual number of hospital admissions due to bronchiolitis in young children in the UK has also increased sevenfold between 1979 and 2011.[8,19] This evidence suggests that the increase is not due to an increase in the severity of infection, nor the virulence of RSV in the age group, because paediatric intensive care admission rates have changed little from 2004 to 2012.[8] Instead, these trends are likely due to a general increase in hospital admission, potentially due to hospital admission thresholds being lowered (particularly in the younger infants) or failure to manage these acute illnesses in the community care setting.[20]

Our study found a high burden of RSV-associated hospital admissions in children <6 months of age, a group well documented as being at high risk of RSV-associated hospital admission.[21] However, our study also found a high number of RSV-associated pneumonia and unspecified LRTI admissions in children aged 1-4 years. RSV was the pathogen associated with the highest number of admissions for all types of LRTI in all age groups compared to the other pathogens, except for pneumonia and unspecified LRTI in children aged <6 months

and 6-11 months. It is therefore possible that RSV-associated hospital admissions are more likely to be coded as bronchiolitis in young infants, but as pneumonia or unspecified LRTI in children older than one year.

As this analysis is at the population level, it is not possible to investigate individual-level risk factors for RSV-associated hospital admission or additional outcomes indicating the severity of illness such as length of stay or admission to paediatric intensive care. Individual-level analysis of RSV-associated hospital admission could be achieved through linkage of laboratory surveillance data and hospital admissions data, as demonstrated in Western Australia.[22]

In conclusion, RSV is a significant cause of hospital admissions for LRTI in children <5 years of age in England. Large-scale individual-level analysis of RSV-associated hospital admissions in young children in secondary care in England is required to compare the severity of cases by age. This work provides a baseline for vaccine impact studies if such vaccines are introduced into the UK immunisation programme in the future.

ACKNOWLEDGEMENTS

Rachel M. Reeves is supported by awards to establish the Farr Institute of Health Informatics Research at UCLPartners, from the Medical Research Council, Arthritis Research UK, British Heart Foundation, Cancer Research UK, Chief Scientist Office, Economic and Social Research Council, Engineering and Physical Sciences Research Council, National Institute for Health Research, National Institute for Social Care and Health Research, and Wellcome Trust (grant MR/K006584/1).

REFERENCES

1. Hall CB, Weinberg GA, Iwane MK, et al. The burden of respiratory syncytial virus infection in young children. *N Engl J Med*. 2009;360:588–598.
2. Nair H, Nokes DJ, Gessner BD, et al. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. *Lancet*. 2010;375:1545–1555.
3. PATH. RSV Vaccine Snapshot - PATH Vaccine Resource Library [Internet]. 2016. http://www.path.org/vaccineresources/details.php?i=1562. Accessed January 3, 2017.
4. Pitman RJ, Melegaro A, Gelb D, Siddiqui MR, Gay NJ, Edmunds WJ. Assessing the burden of influenza and other respiratory infections in England and Wales. *J Infect*. 2007;54:530–538.
5. Cromer D, van Hoek AJ, Jit M, Edmunds WJ, Fleming D, Miller E. The burden of influenza in England by age and clinical risk group: a statistical analysis to inform vaccine policy. *J Infect*. 2014;68:363–371.
6. Taylor S, Taylor RJ, Lustig RL, et al. Modelling estimates of the burden of respiratory syncytial virus infection in children in the UK. *BMJ Open*. 2016;6:e009337.
7. Zhao H, Green H, Lackenby A, et al. A new laboratory-based surveillance system (Respiratory DataMart System) for influenza and other respiratory viruses in England: results and experience from 2009 to 2012. *Euro Surveill* 2014;19:pii: 20808.
8. Green CA, Yeates D, Goldacre A, et al. Admission to hospital for bronchiolitis in England: trends over five decades, geographical variation and association with perinatal characteristics and subsequent asthma. *Arch Dis Child* 2016;101:140–146.
9. Public Health England. Sources of UK flu data: influenza surveillance in the UK - Detailed guidance [Internet]. 2014. https://www.gov.uk/guidance/sources-of-uk-flu-data-influenza-surveillance-in-the-uk. Accessed January 3, 2017.
10. Müller-Pebody B, Edmunds WJ, Zambon MC, Gay NJ, Crowcroft NS. Contribution of RSV to bronchiolitis and pneumonia-associated hospitalizations in English children, April 1995-March 1998. *Epidemiol Infect*. 2002;129:99–106.
11. Cooper DL, Smith GE, Edmunds WJ, Joseph C, Gerard E, George RC. The contribution of respiratory pathogens to the seasonality of NHS Direct calls. *J Infect*. 2007;55:240–248.
12. Matias G, Taylor R, Haguinet F, Schuck-Paim C, Lustig R, Shinde V. Estimates of mortality attributable to influenza and RSV in the United States during 1997-2009 by influenza type or subtype, age, cause of death, and risk status. *Influenza Other Respi Viruses*. 2014;8:507–515.
13. Fleming DM, Taylor RJ, Lustig RL, et al. Modelling estimates of the burden of Respiratory Syncytial virus infection in adults and the elderly in the United Kingdom. *BMC Infect Dis*. 2015;15:443.
14. Office for National Statistics. Population Estimates for UK, England and Wales, Scotland and Northern Ireland. 2011.
15. Office for National Statistics. Population Estimates for UK, England and Wales, Scotland and Northern Ireland, Mid-2011 and Mid-2012. 2013.
16. Ajayi-Obe EK, Coen PG, Handa R, et al. Influenza A and respiratory syncytial virus hospital burden in young children in East London. *Epidemiol Infect*. 2008;136:1046–1058.
17. Iwane MK, Edwards KM, Szilagyi PG, et al. Population-based surveillance for hospitalizations associated with respiratory syncytial virus, influenza virus, and parainfluenza viruses among young children. *Pediatrics*. 2004;113:1758–1764.
18. Simpson MD, Kieke BA, Sundaram ME, et al. Incidence of medically attended respiratory syncytial virus and influenza illnesses in children 6-59 months old during four seasons. *Open forum Infect Dis*. 2016;3:ofw081.
19. Gill PJ, Goldacre MJ, Mant D, et al. Increase in emergency admissions to hospital for children aged under 15 in England, 1999-2010: national database analysis. *Arch Dis Child*. 2013;98:328–334.
20. Public Health England. Respiratory syncytial virus: the green book, chapter 27a [Internet]. 2013. https://www.gov.uk/government/publications/respiratory-syncytial-virus-the-green-book-chapter-27a. Accessed January 3, 2017.
21. Simoes EAF. Environmental and demographic risk factors for respiratory syncytial virus lower respiratory tract disease. *J Pediatr*. 2003;143(5 Suppl):S118–S126.
22. Moore HC, de Klerk N, Keil AD, et al. Use of data linkage to investigate the aetiology of acute lower respiratory infection hospitalisations in children. *J Paediatr Child Health*. 2012;48:520–528.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Reeves RM, Hardelid P, Gilbert R, Warburton F, Ellis J, and Pebody RG. (2017), Estimating the burden of respiratory syncytial virus (RSV) on respiratory hospital admissions in children less than five years of age in England, 2007-2012. *Influenza Other Respir Viruses* 11, 122–129. doi: 10.1111/irv.12443

Appendix 3

Data Linkage – Supplementary Material

Table A3-1. The error rates (%) of patient identifiers in HES and RDS (calculated using the training dataset (section 6.3.4.2). Calculated by Nikolaos Panagiotopoulos (PHE).

Identifier	Error rates (%)	
	HES	RDS
NHS number	0.30	1.02
Sex	<0.01	<0.01
DOB (day)	0.03	0.07
DOB (month)	<0.01	0.03
DOB (year)	<0.01	<0.01
Post code prefix	4.03	4.01
Post code suffix	7.19	6.97

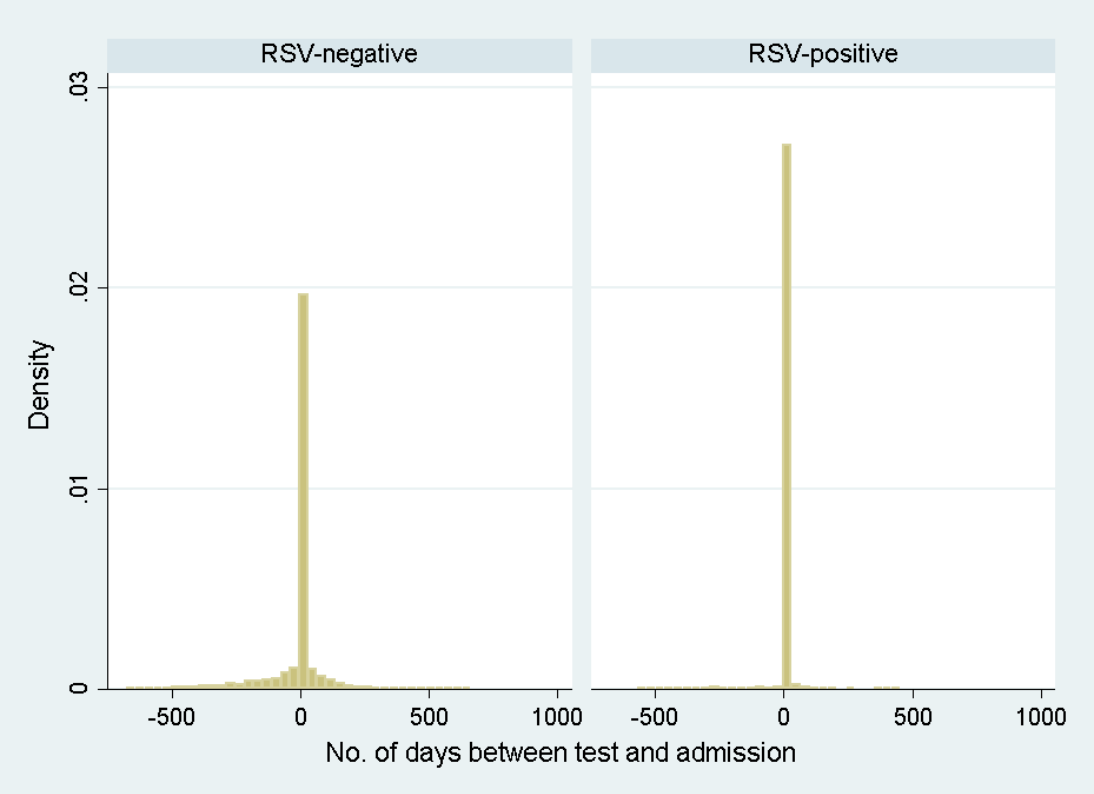

Figure A3-1. Time between test and admissions dates (days) for linked record in the Single-HES dataset.

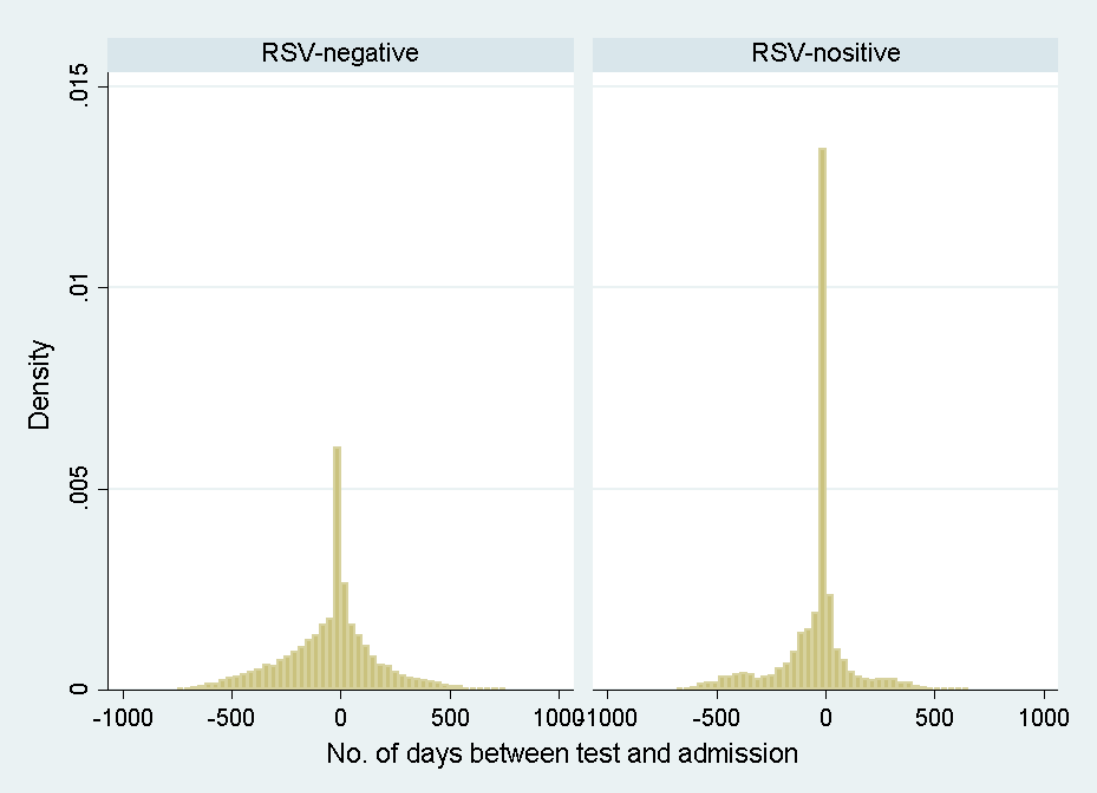

Figure A3-2. Time between test and admissions dates (days) for linked record in the Multi-HES (no error) dataset.

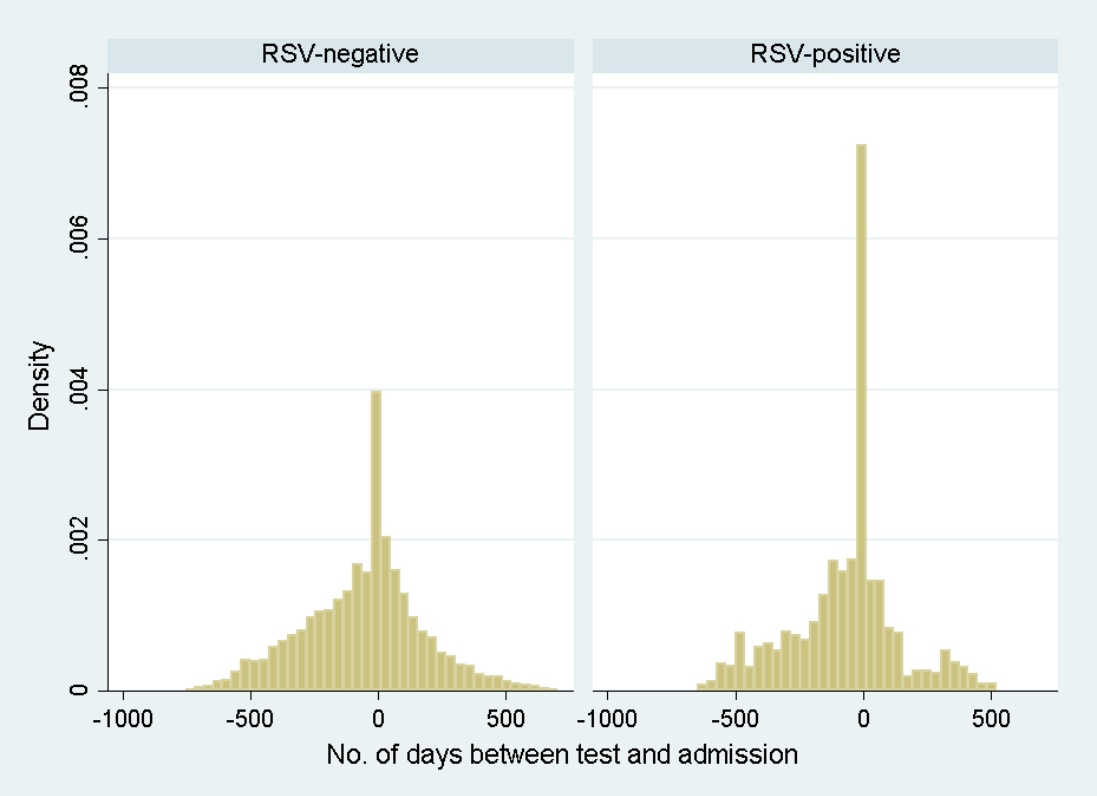

Figure A3-3. Time between test and admissions dates (days) for linked record in the Multi-HES (with error) dataset.

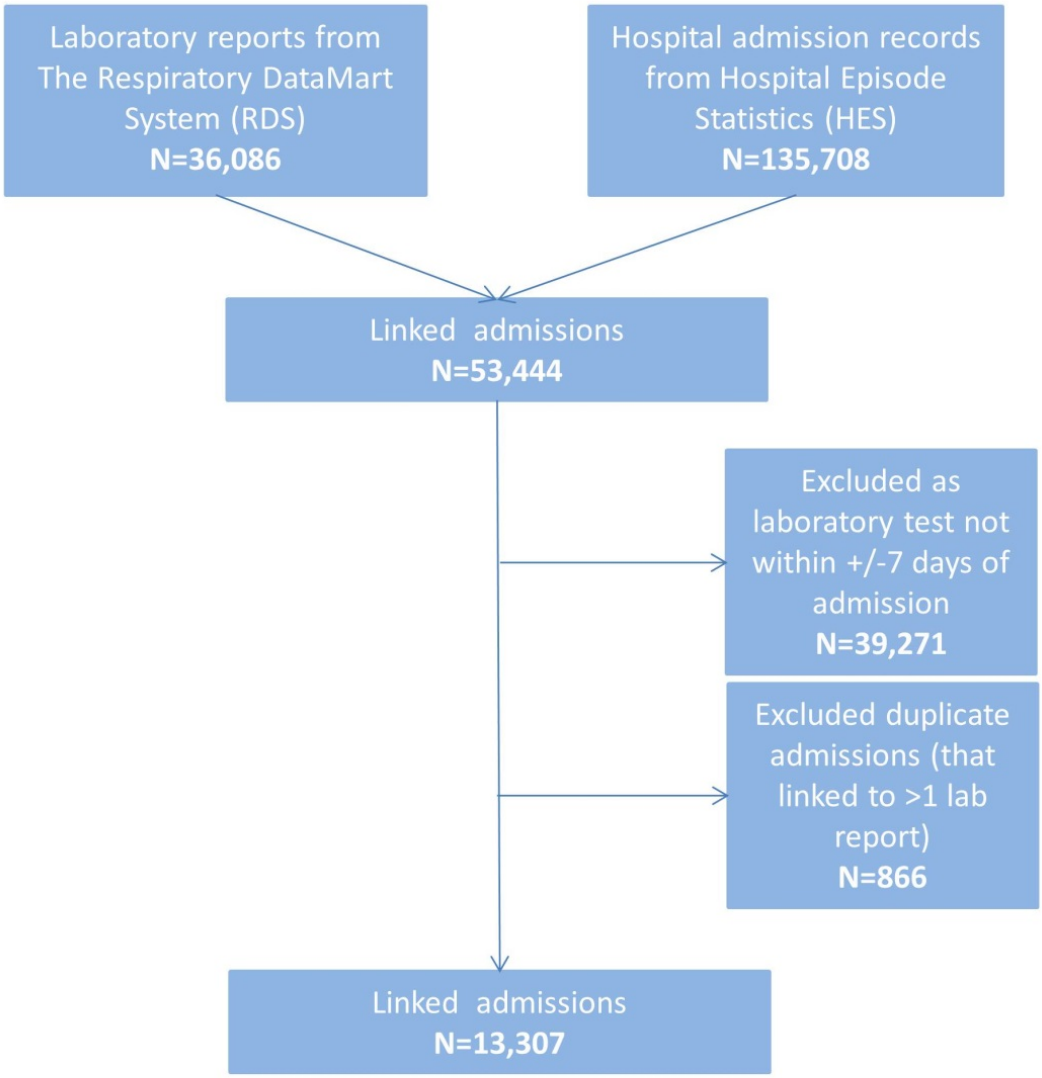

Figure A3-4. Inclusion/exclusion criteria for the linked RDS-HES study population.

Appendix 4

The Lancet publication

Appendix 5

Estimating the secondary care burden of RSV in children aged 1-4 years using linked data

Generating a predictive model for RSV-positivity in children aged 1-4

36% (*n*=4,811) of linked admissions in the RDS-HES linked dataset were in children aged 1-4. Even in the 24 providers with the highest percentage of testing being carried out (i.e. the highest percentage of admissions which linked to an RDS RSV test, Chapter 8: Table 8-1) only 4% of respiratory admissions linked to an RDS test. Only 3 providers tested >10% of their respiratory admissions (with a total of 252 linked admissions from these providers), and testing among the remaining providers ranged from 0-6%.

I attempted to generate a predictive model for RSV-positivity among linked admissions in children aged 1-4 years, following the same methodology as in Chapter 8 for admissions in infants aged <1 year. I also investigated whether using primary diagnosis instead of any diagnosis, and whether using the number of clinical risk factors instead of a binary variable for clinical risk group improved the model fit, but it did not. Using backwards stepwise regression to remove non-significant (p<0.05) variables from the model, the best fitting model had the outcome of RSV-positivity and the following predictors: any diagnosis of unspecified LRTI, any code indicating RSV as the cause of disease, age, and the cyclical function of calendar week (as in Chapter 8). I used a 2/3 sample of the data to generate the model, and a 1/3 sample to test the model fit. This model had relatively low sensitivity and specificity (Table A5-1), poor fit by calendar week (Figure A5-1) and age (Figure A5-2, and a ROC curve of 0.7825.

This relatively poor fit is likely due to the low number of linked admissions as well as testing bias in children aged 1-4 years (and therefore the non-generalisability of the results to non-tested, i.e. unlinked, admissions). Therefore, due to the low number of linked admissions, and the resulting difficulty in developing a good and well-fitting predictive model for RSV-positivity in this age group, I have not used this method to estimate the total secondary care burden of RSV in children aged 1-4 years in this thesis.

Table A5-1. Sensitivity and specificity of the final model with different probability cut-off points.

Probability cut-off	Specificity (%)	Sensitivity (%)	Sum
0.01	27	99	126
0.1	46	97	143
0.2	59	87	146
0.3	64	70	134
0.4	79	66	145
0.5	84	53	137
0.6	94	48	142
0.7	97	36	133
0.8	98	23	131
0.9	99	17	116

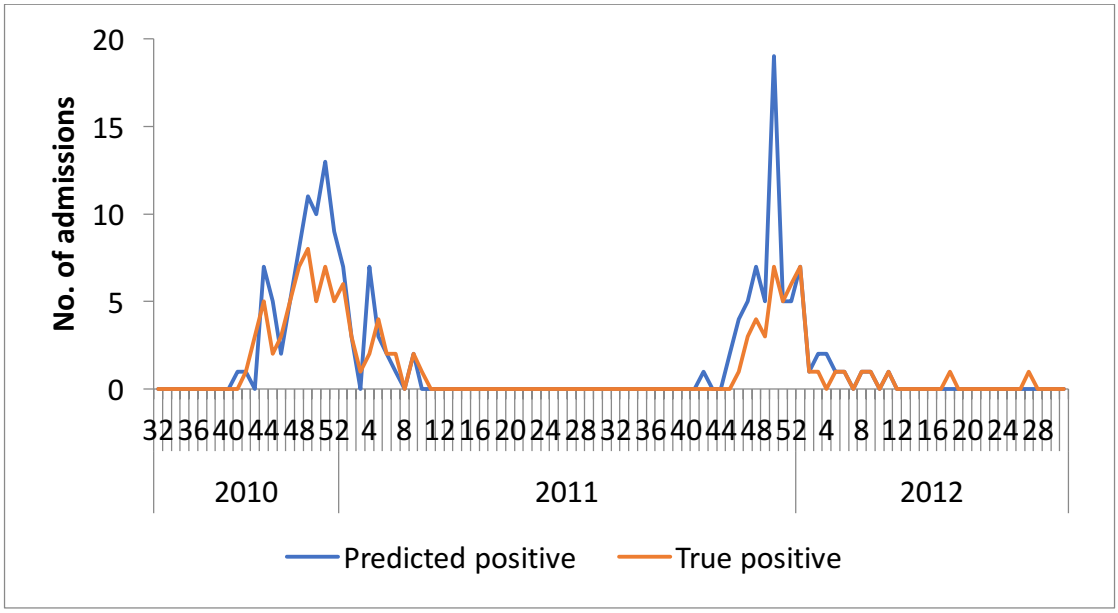

Figure A5-1. True RSV-positive admissions and predicted RSV-positive admissions from the best-fitting model of RSV-positivity in children aged 1-4 years, by calendar week.

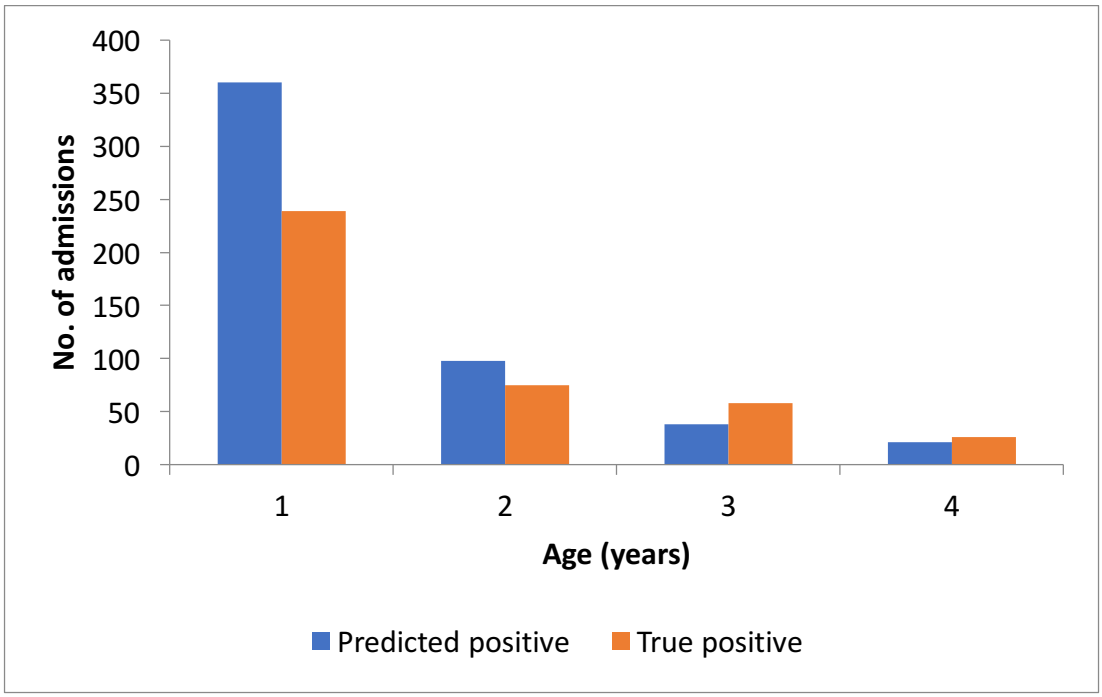

Figure A5-2. True RSV-positive admissions and predicted RSV-positive admissions from the best-fitting model of RSV-positivity in children aged 1-4 years, by age in years.

Appendix 6

Estimating the secondary care burden of RSV in infants aged <1 year using linked data – Supplementary Material

Table A6-1. Sensitivity and specificity of the final model with different probability cut-off points.

Probability cut-off	Specificity (%)	Sensitivity (%)	Sum
0.01	6	100	106
0.1	42	99	141
0.2	55	96	151
0.3	67	92	159
0.4	79	86	165
0.5	84	82	166
0.6	88	74	162
0.7	95	51	146
0.8	96	45	141
0.9	98	35	133
0.99	100	0	100

Table A6-2 shows the number of true RSV-positive and RSV–negative compared to the number of predicted RSV-positive and RSV–negative in the test sample. The PPV of the final model was 79% (95% CI 77-81%) and the NPV 86% (95% CI 84-87%). The sensitivity and specificity of the model were 82% (95% CI 79-84%) and 84% (95% CI 81-86%), respectively.

Table A6-2. The number of true RSV-positive and RSV-negative in the test sample compared to the number predicted by the final logistic regression model.

		True RSV-status (laboratory test)		
		Positive	**Negative**	**Total**
Predicted RSV-status (model)	**Positive (95% CI)**	784 (758-806)	204 (182-236)	988 (940-1,042)
	Negative (95% CI)	172 (150-198)	1,055 (1,023-1,077)	1,227 (1,173-1,275)
	Total	956	1,259	2,215

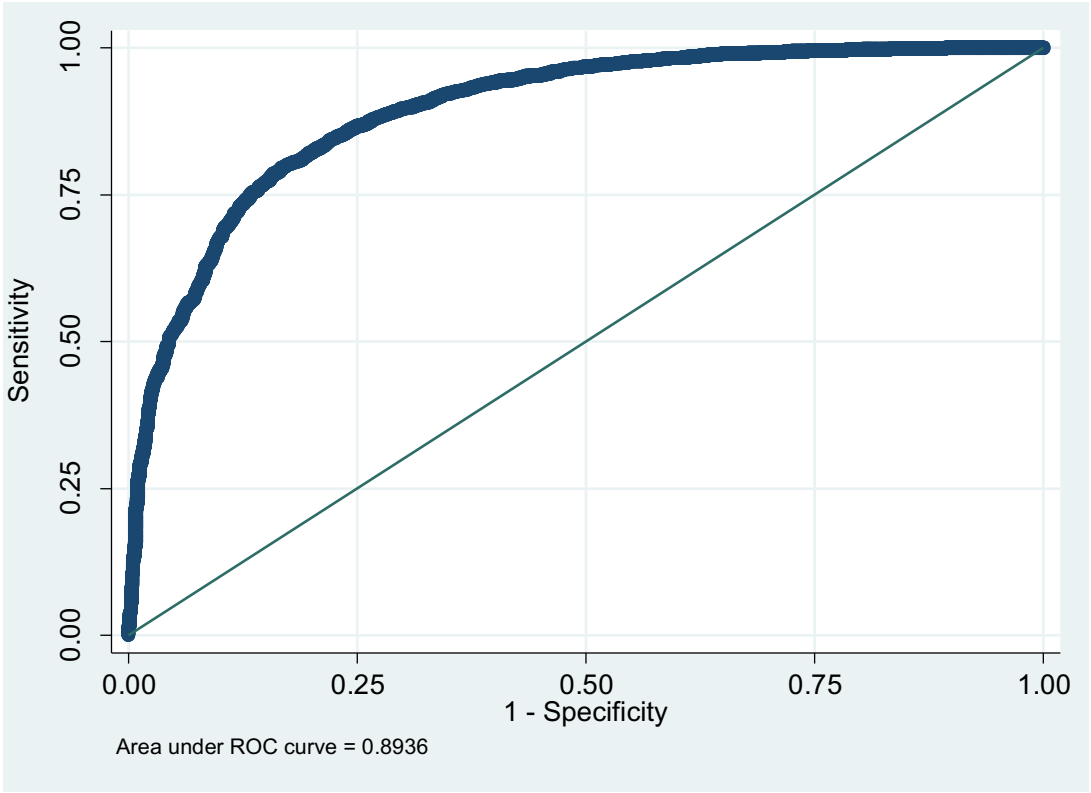

Figure A6-1. Receiver-operator curve (ROC) for the final logistic regression model.

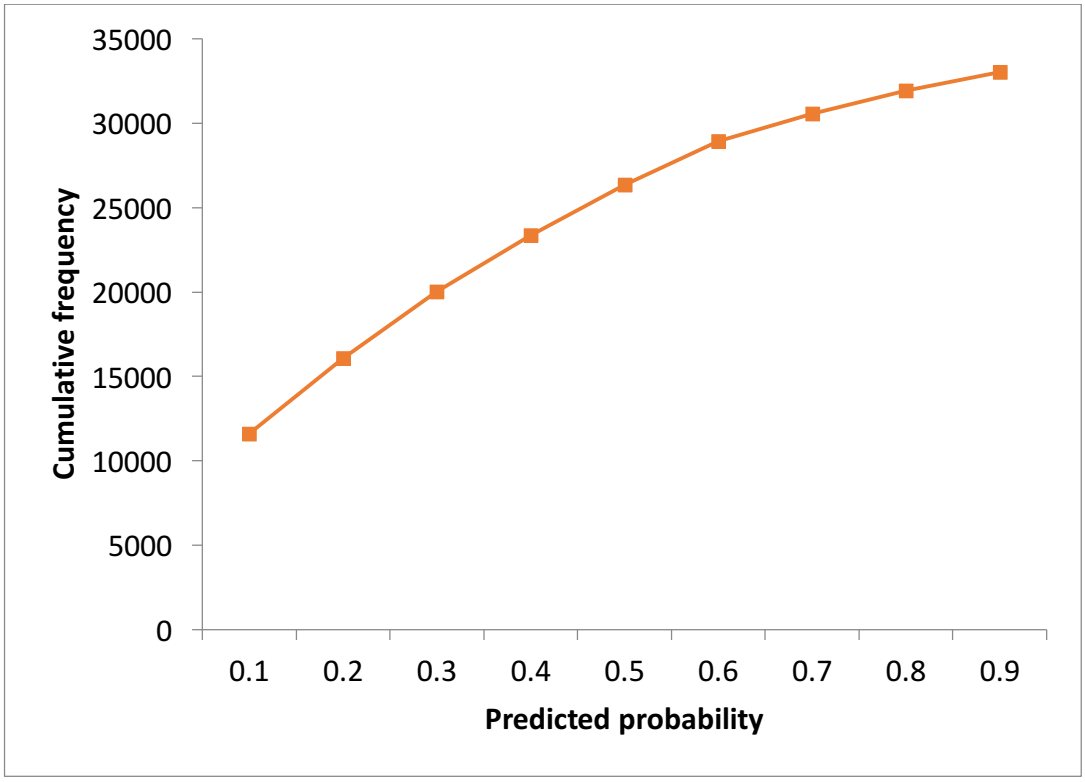

Figure A6-2. Cumulative frequency curve of predicted probabilities from final model.

Appendix 7

Crude vaccine impact analysis

Crude vaccine impact analysis

I have carried out crude analysis of the minimum estimated admissions and bed days potentially prevented by different vaccination strategies. For this analysis, I have used the estimates calculated from the linked data (Chapter 8 of this thesis) – see Figure A7-1 and Figure A7-2.

I have investigated 4 potential scenarios:

a) Scenario 1: maternal vaccine to protect infants aged 0-2 months of age, with 45% coverage (similar to influenza maternal vaccine uptake[1]) and 70% efficacy.
b) Scenario 2: maternal vaccine to protect infants aged 0-2 months of age, with 60% coverage (similar to pertussis vaccine uptake[2]) and 70% efficacy.
c) Scenario 3: direct vaccination of children aged 4-6 months, with 90% coverage and 70% efficacy.
d) Scenario 4: direct vaccination of children aged 6-11 months, with 90% coverage and 70% efficacy.

I also investigated combination strategies of scenarios 1 + 3, 1 + 4, 2 + 3 and 2 + 4.

My results indicate that the most effective vaccination strategy could be scenario 2 (for infants born in September, October, November and December) in combination with scenario 4.

[1] http://ecdc.europa.eu/en/publications/Publications/Seasonal-influenza-vaccination-Europe-2012-13.pdf

[2] http://bmjopen.bmj.com/content/6/4/e010790

Table A7-1. Number of RSV-associated admissions in children <1 year, by age and birth month, as estimated using the linked data (Chapter 8 of this thesis).

		Age (months)											
		0	1	2	3	4	5	6	7	8	9	10	11
Birth Month	January	203	135	34	3	0	0	0	1	8	93	263	158
	February	45	38	6	0	1	1	2	8	92	297	172	35
	March	13	11	0	0	2	3	11	111	356	265	53	10
	April	2	1	1	1	2	13	137	418	325	84	11	3
	May	1	1	2	3	22	180	506	433	101	16	7	2
	June	1	4	4	15	205	568	491	118	21	9	1	0
	July	1	5	23	228	693	627	157	25	7	1	0	0
	August	3	46	294	791	731	218	45	7	1	0	0	0
	September	35	413	1074	878	282	38	9	2	0	0	1	1
	October	219	1410	1234	310	61	13	3	0	1	0	1	3
	November	673	1485	412	75	12	2	0	0	1	1	4	53
	December	578	628	103	21	5	1	1	0	1	7	70	230

Table A7-2. Number of RSV-associated bed days in children <1 year, by age and birth month, as estimated using the linked data (Chapter 8 of this thesis).

		Age (months)											
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Birth Month	**January**	1119	505	161	7	0	0	0	3	56	186	670	311
	February	438	158	13	0	2	1	3	35	245	626	368	99
	March	81	53	0	0	5	27	76	305	758	558	201	60
	April	10	7	1	3	7	36	312	939	649	234	38	4
	May	4	1	11	20	371	425	915	856	250	58	20	7
	June	14	13	92	56	441	1051	911	322	70	30	1	0
	July	1	13	216	572	1583	1345	421	111	18	10	0	0
	August	130	378	872	1811	1474	524	143	24	2	0	0	0
	September	388	1102	2737	2116	691	128	27	6	0	0	2	1
	October	1681	4248	3182	699	196	41	8	0	1	0	1	4
	November	3637	4364	1078	384	30	5	0	0	4	1	13	95
	December	3029	1889	368	78	10	2	3	0	1	56	196	464

Table A7-3. Potential number of RSV-associated admissions and bed days in children <1 year prevented by the 4 vaccine scenarios.

Target population	Coverage	Target birth months	Potential admissions prevented	Potential bed days prevented
Infants 0-2 months (maternal vaccine)	45%†	Nov + Dec	1222	4525
		Oct + Nov + Dec	2123	7394
		Sept + Oct + Nov + Dec	2602	8726
		Aug + Sept + Oct + Nov	2298	7495
		Sept + Oct + Nov	2190	7061
Infants 0-2 months (maternal vaccine)	60%*	Nov + Dec	1629	6033
		Oct + Nov + Dec	2831	9859
		Sept + Oct + Nov + Dec	3470	11634
		Aug + Sept + Oct + Nov	3064	9993
		Sept + Oct + Nov	2920	9414
Infants aged 4-6 months	90%	All	3170	7063
Infants aged 6-11 months	90%	All	3296	7423

Table A7-4. Potential number of RSV-associated admissions and bed days in children <1 year prevented by a combination of strategy 3 (direct vaccination of infants aged 4-6 months) and a maternal vaccine.

Target population	Coverage	Target birth months	Potential admissions prevented	Potential bed days prevented
Infants 0-2 months (maternal vaccine)	45%	Nov + Dec	4392	11587
		Oct + Nov + Dec	5294	14457
		Sept + Oct + Nov + Dec	5773	15788
		Aug + Sept + Oct + Nov	5468	14558
		Sept + Oct + Nov	5360	14124
Infants 0-2 months (maternal vaccine)	60%	Nov + Dec	4799	13095
		Oct + Nov + Dec	6001	16922
		Sept + Oct + Nov + Dec	6640	18697
		Aug + Sept + Oct + Nov	6234	17056
		Sept + Oct + Nov	6090	16477

Table A7-5. Potential number of RSV-associated admissions and bed days in children <1 year prevented by a combination of strategy 4 (direct vaccination of infants aged 6-11 months) and a maternal vaccine.

Target population	Coverage	Target birth months	Potential admissions prevented	Potential bed days prevented
Infants 0-2 months (maternal vaccine)	45%	Nov + Dec	4518	11947
		Oct + Nov + Dec	5420	14817
		Sept + Oct + Nov + Dec	5899	16148
		Aug + Sept + Oct + Nov	5594	14918
		Sept + Oct + Nov	5486	14484
Infants 0-2 months (maternal vaccine)	60%	Nov + Dec	4925	13455
		Oct + Nov + Dec	6127	17282
		Sept + Oct + Nov + Dec	6766	19057
		Aug + Sept + Oct + Nov	6360	17416
		Sept + Oct + Nov	6216	16837

Appendix 8

Conference presentations on the work in this PhD

Conference presentations on the work in this PhD

- Using electronic health records to evaluate the burden of RSV. *Meeting: Vaccine evaluation, implementation and policy, LSHTM, London. February 2017 (Oral).*
- Use of linked laboratory surveillance and hospital data to estimate the burden of admissions for respiratory syncytial virus infection in children younger than 5 years in England. *The Lancet Public Health Science Conference, Cardiff. November 2016 (Poster).*
- The burden of RSV-associated hospital admissions in children <5 years old in England: A comparison of estimates using time-series modelling and data linkage methods. *10th International Respiratory Syncytial Virus Symposium, Argentina. September 2016 (Poster).*
- Using probabilistically linked data to investigate the burden of respiratory syncytial virus (RSV) in children <5 years of age on secondary care in England. *2016 International Population Data Linkage Conference, Swansea. August 2016 (Oral).*
- Epidemiological evaluation of infants and young children with laboratory-confirmed Respiratory Syncytial Virus infection in England, 2010-2014. *1st International Meeting on Respiratory Pathogens, Singapore. September 2015 (Oral).*
- Describing the epidemiology of Respiratory Syncytial Virus in infants and young children in England using a routine laboratory database. *The Farr Institute International Conference, St Andrews. August 2015 (Oral).*

Additional meetings attended

- 3rd ReSViNET High-Level Expert Meeting "RSV – The Next Steps". *Amsterdam, March 2017.*