Imperial College London

The clinical burden of respiratory syncytial virus (RSV) bronchiolitis among infants in the United Kingdom (UK)

Submitted for the Degree of Doctor of Philosophy

Joanna Catherine Murray

Department of Primary Care and Public Health

Imperial College London

Statement of Contributions

I, Joanna Catherine Murray, declare that the work presented in this thesis is my own. Where information has been derived from other sources this has been indicated and appropriately referenced.

Acknowledgements

I have completed this thesis thanks to the help and support of many people who I wish to sincerely thank.

First and foremost, I would like to thank Sonia Saxena and Alex Bottle for the fantastic supervision, support and guidance they have generously provided throughout my PhD.

I also thank Azeem Majeed, Neena Modi, the Neonatal Data Analysis Unit team and the Dr Foster Unit at Imperial College London for providing access to data and for all their help and expertise. I owe a big thank you to all my colleagues and friends at Imperial College London, especially Liz Koshy for all her support.

I would like to thank the Medicines for Neonates Investigator group for funding my studentship and fees through the National Institute for Health Research “Medicines for Neonates” programme grant for applied research (RP-PG-0707-10010) held by Neena Modi.

I thank all my friends, “the sons” and especially my family for all their love, support and encouragement throughout the last few years. Without them this would not have been possible.

Finally I would like to thank and lovingly dedicate this work to Andrew Mastrantone, who as always has been there for me every step of the way.

Funding

This PhD presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research Programme (Grant Reference Number RP-PG-0707-10010). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Outputs from this PhD

Publications relating to this work

Murray J, Sharland M, Bottle A, Curcin V, Majeed A, Modi N & Saxena S (2013) Incidence, natural history and management of bronchiolitis in infants: a population-based birth cohort study in UK general practice (Under review - July 2013).

Murray J, Bottle A, Sharland M, Modi N, Aylin P, Majeed A & Saxena S (2013) "Changes in the severity and age of RSV bronchiolitis hospital admission among infants in England: a population-based birth cohort study" (Under review - July 2013).

Murray J, Saxena S, Modi N, Majeed A, Aylin P & Bottle A (2012) Quality of routine hospital birth records and the feasibility of their use for creating birth cohorts. *Journal of Public Health* doi:10.1093/pubmed/fds077 (Epub ahead of print).

Murray J, Modi N, Sharland M & Saxena S (2011) BMJ online rapid response letter - Management of bronchiolitis in primary care – the need for an improved evidence base. http://www.bmj.com/content/342/bmj.d1714.full/reply#bmj_el_257426

Conference presentations and prizes arising from this work

Poster presentation at the Society for Academic Primary Care Annual Scientific Meeting 2012

Poster & oral presentation at the Royal College of Paediatrics and Child Health Conference 2012

Oral presentation at the Neonatal Society Autumn Meeting 2011

Prize for best oral presentation at Imperial College School of Public Health PhD Symposium 2011

Oral presentation & highly commended young researcher award at the Society for Academic Primary Care (SAPC) Annual Regional Meeting Cambridge 2011

Poster presentation at the European Society for Paediatric Infectious Diseases (ESPID) The Hague 2011

Prize plenary presentation at the Society for Academic Primary Care Annual Scientific Meeting Norwich 2010

Other outputs during this time

Koshy E, Murray J, Bottle A, Aylin P, Sharland M, Majeed A & Saxena S (2012) Significantly increasing hospital admissions for acute throat infections among children in England: is this related to tonsillectomy rates? *Archives of Disease in Childhood* (Epub ahead of print)

Murray J, Majeed A, Saleem-Khan M, Lee J T & Nelson P (2011) Use of the NHS Choices website for primary care consultations: results from online and general practice surveys. *Journal of the Royal Society of Medicine (JRSM) Short Reports* **2**: 56 DOI 10.1258

Gajria C, Birger R, Murray J, Rice A & Majeed A. (2011) Primary Care Records can be used to identify patients with neuropathic pain. *Informatics in Primary Care* **19**(2): 83-90

Koshy E, Murray J, Saxena S, Bottle A & Sharland M (2010) Postscript letter: Authors' response. *Thorax* doi:10.1136/thx.2010.154914

Koshy E, Murray J, Saxena S, Bottle A & Sharland M. (2010) Impact of the seven-valent pneumococcal conjugate vaccination (PCV7) on childhood hospital admissions for bacterial pneumonia and empyema in England: national time-trends study. *Thorax* **65**(9): 770-4

Murray J, Saxena S, Millett C, Curcin V, de Lusignan S & Majeed A. (2010) Reductions in coronary heart disease risk factors by ethnic group in South-West London: ten year longitudinal study (1998 – 2007) *Family Practice* **27** (4): 430-8

Keith T, Saxena S, Murray J & Sharland M. (2010) The positive and negative effects of restricting antimicrobial prescribing in children. *Current Opinion in Infectious Disease* **23** (3): 242-8

Verma A, Birger R, Bhatt H, Murray J, Millett C, Saxena S, Banarsee R, Gnani S & Majeed A (2010) Ethnic disparities in diabetes management: a 10-year population-based repeated cross-sectional study in UK primary care. *Journal of Public Health* **32**(2): 250-8

Murray J, Saxena S, Sharland M, Bottle A & Majeed A (2009) BMJ online rapid response letter - Long term outcomes following admission for RSV bronchiolitis in children. http://www.bmj.com/cgi/eletters/338/jun11_1/b1935

Abstract

Background and Aim: Studies of the epidemiology of respiratory syncytial virus (RSV) bronchiolitis to date have been small; focused only on the hospital setting and selective groups of high-risk infants such as those born preterm; lacked long-term follow-up and few have been based on data from the United Kingdom (UK). Hence, the existing evidence base was likely to underestimate the scale and impact of RSV on health in UK infants. The aim of this thesis was to provide better estimates of the wider clinical burden of RSV bronchiolitis among infants presenting across primary and secondary care settings in the UK.

Methods: The clinical spectrum of bronchiolitis illness across different healthcare settings was examined using routine data from electronic health records to develop longitudinal, population-based cohorts with follow-up from birth through early childhood. Databases examined in this thesis included Hospital Episodes Statistics (HES), the General Practice Research Database (GPRD) and the National Neonatal Research Database (NNRD).

Results: The estimated bronchiolitis admission rate in NHS hospitals in England was 24.2 admissions per 1000 infants aged less than 1 year (95% CI 23.7 to 24.8) with a median length of stay of 1 day (IQR 0 to 3 days) at a median age of 120 days (IQR 61 to 209 days). 15% of infants admitted with bronchiolitis were born preterm (47.3 per 1000 infants) and 24% had at least one known clinical risk factor for severe RSV infection. Cystic fibrosis, cerebral palsy and Down's syndrome also increase an infant's risk of bronchiolitis admission. The bronchiolitis consultation rate in UK general practice was 58.1 per 1000 infants (95% CI 56.5 to 59.8) at a mean age of 5.5 months (SD=3.2). Using a broader bronchiolitis case definition the estimated consultation rate was 206.7 per 1000 infants (95% CI 204.0 to 209.6). 36% of bronchiolitis consultations resulted in a prescription, of which 28% were for antibiotics and 27% for beta agonists, despite no evidence to support their use. Bronchiolitis in infancy is a predictor of subsequent hospital admissions and general practice consultations for asthma and wheezing in early childhood.

Conclusions: The clinical burden of RSV bronchiolitis across healthcare settings in the UK is greater than previously estimated. Between 4% and 21% of infants have a bronchiolitis GP consultation in their first year. 2.4% of the national birth cohort are admitted to hospital with bronchiolitis in the first year of life, most of whom are born at term with no known clinical risk factors for severe RSV infection. This thesis has identified new groups of infants who may be at increased risk of severe RSV disease including those with cystic fibrosis, cerebral palsy and Down's syndrome. These findings highlight the need to prioritise development of new approaches for the prevention and treatment of RSV infection and have important implications for clinical training and management of bronchiolitis across healthcare settings.

Table of Contents

Statement of Contributions 1
Acknowledgements 2
Outputs from this PhD 3
Abstract 5
Table of Contents 6
List of Figures 11
List of Boxes 11
List of Tables 11
Abbreviations 12
1.0 Introduction 13
1.1 Childhood respiratory illness 14
1.2 Respiratory syncytial virus (RSV) infection 14
Clinical symptoms 14
Virology 14
Immunity 14
Global burden 15
UK burden 15
1.3 Bronchiolitis 16
Incidence 16
1.4 Diagnosis and management of bronchiolitis in the UK 17
Diagnosis 17
Treatment 18
1.5 Risk factors for bronchiolitis 19
1.6 Prevention 20
1.6.1 Hygiene 20
1.6.2 Passive immunotherapy 21
1.6.3 Active vaccination 22
1.7 Impact on respiratory health in early childhood 23
1.8 Summary of chapter 24
Key findings from literature review justifying this work 25
2.0 Aim and objectives 26
2.1 Summary of priorities identified for further research: 26
2.2 Rationale for thesis 27
2.3 Hypothesis 28
2.4 Research questions 28

2.5 Aim 30
2.6 Objectives 30
3.0 Use of routinely collected data for child health research 31
3.1 Problems in conducting child health research 31
3.1.1 Experimental study design 32
3.1.2 Non-experimental study designs 33
3.2 Use of routinely collected electronic patient health records 33
3.2.1 Ethical considerations and consent for long-term follow-up 34
3.2.2 Potential uses of child health records for research 35
3.3 Linkage with other health records 35
3.3.1 Examples of linkage between birth and child health records to date 36
3.4 Database linkage 37
3.4.1 Advantages and limitations of using linked health records for research 38
3.5 Summary of chapter 40
4.0 Overview of health record databases in the UK 41
4.1 Introduction to child health records in the UK 41
4.2 Hospital Episode Statistics (HES) 42
4.2.1 Overview of dataset 42
4.2.2 Data collection and purpose 43
4.2.3 HES maternity records and baby tails 43
4.3 The General Practice Research Database (GPRD/CPRD) 45
4.3.1 Overview of dataset 45
4.3.2 Dataset uses 45
4.4 The National Neonatal Research Database (NNRD) 46
4.4.1 Overview of dataset 46
4.4.2 Data collection and uses 46
4.5 Research governance and ethical approval 48
4.5.1 HES data access and approval 48
4.5.2 GPRD data access and approval 49
4.5.3 NNRD data access and approval 49
4.6 Summary of chapter 49
5.0 Methods: Creating birth cohorts using child health records 50
5.1 RSV-associated admission rates 50
5.2 Creating a birth cohort using HES records 52
5.2.1 Linkage across time to produce a HES cohort 52
5.2.2 Processing and methodological issues using HES maternity data 52

5.2.3 Coverage and completeness of recording 54
5.2.4 Comparison with data from ONS birth registrations 56
5.2.5 Approach for minimising the effect of variable data completion between hospitals 57
5.2.6 Explaining these findings 59
5.2.7 Limitations of this work 59
5.2.8 Summary 60
5.2.9 Birth cohort study 60
5.2.10 Statistical analyses 63
5.3 Creating a birth cohort using records from the GPRD 64
5.4 Creating birth cohorts with follow-up into early childhood 70
5.4.1 Developing a cohort with longer follow-up after admission to hospital with bronchiolitis, using HES 70
5.4.2 Developing a cohort with longer follow-up after a bronchiolitis consultation in infancy, using the GPRD 72
5.5 Summary of chapter 74
6.0 Results – Study 1: Bronchiolitis hospital admissions 75
6.1 Introduction 75
6.2 Results 76
6.2.1 Trends in RSV-associated admissions 76
6.2.2 Birth cohort study 77
6.3 Summary of chapter findings 83
7.0 Results - Study 2: Bronchiolitis in the community setting 84
7.1 Introduction 84
7.2 Results 85
7.2.1 Cohort characteristics 85
7.2.2 Bronchiolitis consultations 85
7.2.3 Consultations in the first year of life 87
7.2.4 Consultations before and after bronchiolitis 88
7.2.5 Prescriptions 88
7.3 Summary of chapter findings 89
8.0 Results – Study 3: Impact of bronchiolitis on respiratory health in early childhood 90
8.1 Introduction 90
8.2 Results 91
8.2.1 HES cohort study with follow-up to age 5 years 91
8.2.2 GPRD cohort study with follow-up to age 7 years 94

8.3 Summary of chapter findings 96
9.0 Preliminary developments and progress in linking neonatal care and subsequent hospital admission records 97
9.1 Background 97
9.2 Research governance and applying for ethical approval 98
9.3 Proposed linkage methodology 98
9.4 Progress to date and on-going work 101
9.5 Summary of chapter 105
10.0 Discussion 106
10.1 Overall summary of thesis findings 106
10.2 Discussion of study 1: Bronchiolitis hospital admissions 108
10.2.1 Main findings 108
10.2.2 Strengths and weaknesses 108
10.2.3 Findings in relation to other studies 109
10.2.4 Implications and future research 112
10.2.5 Conclusion 113
10.3 Discussion of study 2: Bronchiolitis in the community setting 114
10.3.1 Main findings 114
10.3.2 Strengths and weaknesses 114
10.3.3 Findings in relation to other studies 116
10.3.4 Implications and future research 117
10.3.5 Conclusion 118
10.4 Discussion of study 3: Impact of bronchiolitis on respiratory health in childhood 119
10.4.1 Main findings 119
10.4.2 Strengths and weaknesses 120
10.4.3 Findings in relation to other studies 122
10.4.4 Implications and future research 123
10.4.5 Conclusion 124
10.5 Methodological strengths and weaknesses 125
Strengths 125
Weaknesses 125
Chance, bias and confounding 126
10.6 Implications for policy and practice 128
Implications for health care policy and the development of clinical guidance 128
Implications for clinical management and service provision 129
10.7 Remaining research questions and future work 130

10.7.1 Improve understanding of the typical care pathway for infants with RSV 130
10.7.2 Prevention and treatment of RSV bronchiolitis 132
10.7.3 Improve understanding of the long-term impact of RSV infection on airways 132
10.7.4 The future of using routine data for child health research 132
10.8 Overall thesis conclusion 134
11.0 Appendices 135
Appendix 1: Palivizumab prescribing data database from the NHS Information Centre . 135
Appendix 2: Published methodology paper 136
References 146

List of Figures

Figure 1: Pyramid of care for infants with RSV bronchiolitis. 29
Figure 2: Trends in RSV-associated hospital admissions per 1000 live births, from 1997/98 to 2010/11. 77
Figure 3: Age of bronchiolitis admission in months, for admissions coded as RSV or unspecified bronchiolitis. 79
Figure 4: Incidence of bronchiolitis hospital admissions among birth cohort, by Strategic Health Authority (SHA) of admission. 82
Figure 5: Age in months at first diagnosis of bronchiolitis in a GP consultation. 86
Figure 6: Seasonality of acute bronchiolitis consultations in general practices across the UK. 86
Figure 7: Subsequent respiratory admissions in early childhood (up to age 5 years) among infants in high- and low- risk groups, comparing those with and without a history of bronchiolitis admission during childhood. 93
Figure 8: Pyramid summarising estimates of RSV bronchiolitis across healthcare settings, derived from findings in this thesis. 107

List of Boxes

Box 1: Key information collected in HES maternity records. 44
Box 2: ICD-10 codes and CCS groupings used to identify at-risk groups. 62
Box 3: Probabilistic linkage process. 100
Box 4: HES data capture and transfer pathway. 102

List of Tables

Table 1: Completeness of recording of baby tail fields in HES birth records (2005/06 to 2009/010) 55
Table 2: Comparison of live birth recording in HES with Office for National Statistics reporting (2007/08). 56
Table 3: Comparison of maternity characteristics between hospitals with high and low completeness of recording in their HES birth admission records†, from financial year 2007/08. 58
Table 4: GPRD codes for acute bronchiolitis. 65
Table 5: GPRD code list for wider bronchiolitis definition of consultations in first year of life. 66
Table 6: Coded diagnosis groups for consultations in the first year of life 69
Table 7: Characteristics of birth cohort and infants admitted with bronchiolitis. 78
Table 8: The relative risk of a bronchiolitis hospital admission among infants in different risk groups. 81
Table 9: The most common diagnoses recorded during consultations among infants in the cohort, comparing those with and without a bronchiolitis consultation during their first year of life. 87
Table 10: Risk of subsequent respiratory hospital admissions in first five years of life, following a bronchiolitis admission in infancy. 92
Table 11: Risk of subsequent GP consultations for asthma or wheeze in the first 7 years of life, comparing children with a history of bronchiolitis consultation to those without. 95

Abbreviations

A&E	Accident and Emergency
ALSPAC	Avon Longitudinal Study of Parents and Children
BMJ	British Medical Journal
CCS	Clinical Classification System
CI	Confidence Interval
CPRD	Clinical Practice Research Datalink
CQC	Care Quality Commission
DFI	Dr Foster Intelligence
DFU	Dr Foster Unit
ED	Emergency Department
GP	General Practitioner
GPRD	General Practice Research Database
HES	Hospital Episode Statistics
HMPV	Human metapneumovirus
HPA	Health Protection Agency
HTA	Health Technology Assessment
ICU	Intensive Care Unit
JCVI	Joint Committee on Vaccination and Immunisation
LRTI	Lower Respiratory Tract Infection
LTV	Long Term Ventilated
MHRA	Medicines and Healthcare Regulatory Authority
MRC	Medical Research Council
NDAU	Neonatal Data Analysis Unit
NEJM	New England Journal of Medicine
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NICU	Neonatal Intensive Care Unit
NIGB	National Information Governance Board for Health and Social Care
NIHR	National Institute of Health Research
NNRD	National Neonatal Research Database
ONS	Office for National Statistics
PCA	Prescribing Cost Analysis
QALY	Quality Adjusted Life Year
RCT	Randomised Controlled Trial
RTI	Respiratory Tract Infection
RSV	Respiratory Syncytial Virus
SCID	Severe Combined Immunodeficiency Syndrome
SIGN	Scottish Intercollegiate Guidelines Network
UK	United Kingdom
USA	United States of America

1.0 Introduction

Chapter overview

This chapter explores the existing evidence base relating to respiratory syncytial virus (RSV) bronchiolitis among infants. It begins by describing in general the significant burden of respiratory illness among children, followed by detailed information relating specifically to RSV infection in infancy. The burden of RSV illness among infants in the wider global context is outlined before the incidence, diagnosis and management of bronchiolitis in the United Kingdom (UK) are described in detail. This is followed by an exploration of existing preventative regimes and the possible future potential of an active vaccination against RSV infection. The longer-term impact of RSV infection in infancy is then considered, examining the evidence for any association with subsequent respiratory morbidity in childhood and into early adulthood. Finally a summary of what is known and the remaining gaps in our knowledge is presented, providing the rationale behind the central topic of this thesis.

Literature review methodology

A systematic approach to reviewing the existing evidence base was adopted. The Cochrane Library, EMBASE, MEDLINE (OVID) and PUBMED databases were all searched for studies published in English, from January 1st 1990 to December 31st 2012. Searches were carried out using varying combinations of the following search terms: (''Respiratory Syncytial Virus, Human"[Mesh]" OR "Respiratory Syncytial Virus Infections"[Mesh]), ("Bronchiolitis"[Mesh] OR "Bronchiolitis, Viral"[Mesh]), (''Risk'' OR "Risk Factors"[Mesh]), ("Primary Health Care"[Mesh] OR "General Practice"[Mesh]), "Patient Admission"[Mesh]", “Community Health Services”[Mesh], Asthma"[Mesh], ”Wheez*”.

Reference lists from selected papers and recent review articles were also examined along with published conference abstracts. In addition to searching for published journal articles, grey literature was also searched to ensure a wider collection of relevant documents were identified, such as government reports, policy documents etc.

1.1 Childhood respiratory illness

The fourth Millennium Development Goal aims to reduce mortality children under 5 years old by two-thirds by 2015.[2] Despite advances in healthcare and availability of treatments, respiratory tract infections (RTIs) remain the leading cause of death in young children worldwide and present a major hindrance in achieving this target.[3-5] Acute respiratory conditions including infectious illness are a major cause of childhood morbidity and remain important causes of health service use[6;7] accounting for around 50% of consultations in primary care and an estimated 15% of hospital admissions in developed countries.[8;9] Lower respiratory tract infections (LRTIs) are particularly common in infants worldwide,[10] the most common cause of which is RSV, which is therefore considered the most important respiratory pathogen in childhood.[11]

1.2 Respiratory syncytial virus (RSV) infection

Clinical symptoms

Most children will have been infected with RSV by 2 years of age.[12] RSV infection will usually result in mild upper respiratory tract illness with nasal congestion, cough, rhinorrhoea and fever.[13] Around 40% of primary RSV infections will result in LRTI, principally bronchiolitis and also pneumonia.[13;14]

RSV infection results in inflammation and oedema of the bronchiolar wall, mucus build up clogging the airways and necrosis of the respiratory epithelium.[15;16] It is this damage which is thought may have longer-term effects on lung function in childhood and possibly into adulthood.[13;17]

Virology

RSV is a negative-strand, non-segmented RNA pneumovirus.[16] There are two RSV subtypes which co-circulate, RSV A is more common than type B and is thought to cause more clinically significant illness.[13;18] In temperate climates RSV tend to circulate during winter months, from October through to March,[13;19;20] while in tropical climates outbreaks tend to occur during the rainy season.[21;22]

Immunity

Viral, host and environmental factors contribute to the pathogenesis of RSV, with the extent of their contributions varying considerably between infected individuals.[16] Previous infection with RSV has been shown to only confer partial immunity to RSV, meaning individuals can be repeatedly infected with the same or different RSV strains throughout life.[23] Reinfection tends to be less severe in childhood after a first illness[13] and becomes progressively less

severe in early adult life but in later life, RSV infection presents a major cause of community-acquired pneumonia among the elderly.[24;25]

The underlying immune response to RSV infection remains a vast area for research since it is still poorly understood. It was previously thought that an overly aggressive, hyper-responsive immune response to RSV infection resulted in severe illness but more recent evidence suggests recovery is dependent on the innate immune response.[25] It is unclear whether reinfection can occur because infants fail to generate sufficient antibody titres or if these wane too rapidly after infection.[13]

Despite RSV long being a priority area for vaccine development and control,[26] the difficulties in determining how RSV interacts with our immune system have resulted in it being the only organism of the three main pathogens that cause death from RTI, *Haemophilus influenzae* and *Streptococcus pneumoniae* being the other two, for which no vaccine is currently available.[3;4]

Global burden

RSV is the leading cause of acute LRTIs and admissions to hospital worldwide.[3;21] A systematic review and meta-analysis published in the Lancet in 2010, estimated the global burden of acute LRTI due to RSV in young children, reporting an estimated 33.8 million new episodes of RSV LRTI worldwide, among children under 5 years in 2005.[21] The study found that around 10%, or at least 3.4 million of these RSV episodes were severe enough to require hospital admission.[21] RSV-associated mortality among infants in developed countries is relatively low, with studies in the United States of America (USA) for example, estimating that RSV was responsible for between 100 and 180 pneumonia related deaths among children between 1993 and 1995.[14;27] Globally however, Nair et al report an estimated 66,000 to 199,000 children under 5 years die each year due to RSV LRTI, 99% of these occurring in developing countries.[21]

UK burden

RSV infection is the leading cause of hospitalisation in children under 1 year of age,[28] with typical rates of 28 per 1000 children under 1 year admitted in England.[29;30] The incidence decreases in older children as a study from the Health Protection Agency (HPA) found using data from the late 1990s, the mean annual incidence of hospital admissions attributable to RSV was 1.3 per 1000 children aged 1 to 4 years.[30] A small, regional study carried out in a single strategic health authority in England also used data from this period and reported hospital admission rates for RSV illness were 24.4 per 1000 infants under 1 year and for bronchiolitis admissions the rate was 30.8 per 1000 infants under 1 year.[31] The RSV-

attributed death rate in infants aged one to 12 months has been estimated to be 8.4 per 100,000 population in the UK.[32]

1.3 Bronchiolitis

Acute bronchiolitis is the most common LRTI in infants, estimated to affect around 10% of children under 1 year.[21;31;33-35] It most commonly presents in infants aged 3 to 6 months.[36] It is usually a mild, self-limiting illness usually lasting between 3 and 7 days, but in some infants may be more severe, requiring hospital admission. Infants typically present with symptoms such as cough, breathing difficulties, poor feeding, irritability and apnoea. The diagnosis is based on these clinical symptoms combined with wheeze and/or crepitations/crackles on auscultation.[36] Bronchiolitis is typically caused by RSV in around 75% of cases,[37] with peak incidence during the winter months.[38;39]

Other bronchiolitis cases can be caused by infection with a range of different viruses including parainfluenza virus, influenza virus, adenovirus, rhinovirus or human metapneumovirus (HMPV). The incidence of dual infection with more than one virus is between 5% and 10% among infants with severe bronchiolitis.[40] Bacterial co-infections with community acquired organisms occurs in around 22% of infants hospitalised with bronchiolitis, who also required significantly more ventilator support than those without bacterial co-infection.[40;41] There is also evidence the dual infection with RSV and HMPV may be associated with a more severe course of illness and the requirement for ventilation.[40;42] There is also an increased risk of more severe bronchiolitis illness among infants with dual infection of RSV and human bocavirus.[43] Rhinovirus infection is thought to cause a milder form of bronchiolitis than RSV.[43]

Incidence

An estimated 2 to 3% of all infants living in developed countries are admitted to hospital with bronchiolitis caused by RSV.[21;31;34;44] In the UK, bronchiolitis admission rates in infants under 1 year have previously been estimated from small regional studies to be 30.8 per 1000[31] and 31.2 per 1000 in the USA.[10] A more recent study in Texas reported higher bronchiolitis admission rates of 5.5% among children under 2 years,[45] with the higher incidence estimate likely to be due to the wider inclusion of children aged 12 to 24 months in their study, coupled with variation in the clinical definition of bronchiolitis used in the USA compared with that used in the UK.[13] To date, no national studies have reported the disease burden at a population level in the UK.

Early findings from The Houston Family Study in the USA published in 1986, a birth cohort of 125 infants with community follow-up for the first year of life, estimated the rate of RSV-confirmed lower respiratory tract disease to be 21.6 per 100 infants under 1 year.[46] Large

prospective birth cohorts are needed to determine the current clinical burden of bronchiolitis in the community, but these are expensive and not practical. Previous studies have focused on identifying risk factors for hospitalisation with severe disease but there are no reliable community-based estimates of the incidence of bronchiolitis in the UK. With widespread use of electronic patient records in general practices across the UK[47], it could now be possible to estimate the frequency of acute bronchiolitis consultations in the community.

1.4 Diagnosis and management of bronchiolitis in the UK

The Scottish Intercollegiate Guidelines Network (SIGN) clinical guideline on bronchiolitis in children, published in 2006, is the main source of information for bronchiolitis management in the UK.[36] The guideline provides evidence based recommendations on prevention, diagnosis and treatment of bronchiolitis, grading each recommendation on the strength of the evidence on which it is based, ranging from 1 for high quality meta-analyses of randomised controlled trials (RCTs) to 4 where only anecdotal evidence from expert opinion is available.[36]

Little is known about the presentation, natural history and management of bronchiolitis in the community and the SIGN guideline lacks sufficient evidence from and applicable to management of bronchiolitis in general practice.[36] The guidelines recommends further research in needed to examine the prevalence of bronchiolitis in primary care and current treatment practice in the community and risk profiling to identify those who may have future respiratory symptoms.[36]

Diagnosis

RTI consultations are considered the 'bread and butter' of daily practice in UK primary care.[48] Despite this, it can be challenging for General Practitioners (GPs), presented with an infant with respiratory symptoms, to differentiate their diagnosis between bronchiolitis, bacterial respiratory tract infections and wheezy bronchitis/early asthma. The diagnosis of bronchiolitis is clinical, based on history and physical examination, with symptoms including dry, wheezy cough, nasal discharge, fever and fine inspiratory crackles or high pitched expiratory wheeze.[36;49-52]

Although virological testing for RSV is not routine practice and is not required for the clinical diagnosis of bronchiolitis to be made, it can be of benefit when admitting infants to hospital. For example a case-control study has shown this can reduce length of stay, antibiotic use and subsequent microbiological testing.[53;54] Other studies have also reported that a

definitive viral diagnosis reduces unnecessary interventions.[55-57] Rapid antigen testing has been shown to be a useful method for detecting RSV mediated disease in very young infants, with good specificity but decreasing diagnostic sensitivity in infants older than 3 months.[55;58;59]

Treatment

Evidence from systematic reviews suggests no available treatments shorten the natural course or provide clinically relevant improvements in bronchiolitis symptoms.[33;60-65] Since there is no universally effective drug treatment, the mainstay of management for infants with RSV bronchiolitis is supportive care and a small proportion of more severely affected infants are treated in intensive care.[33;36;61;62]

A Cochrane systematic review into the use of antiviral ribavirin for RSV infection, found that many of the studies reviewed, including RCTs, either had excluded children with significant comorbidities, had no placebo control group, included very small numbers, had too short follow-up or no blinding to treatment allocation, meaning the quality of evidence was quite poor, potentially biased and had questionable generalisability.[36;60;63;66-68] Similarly, although many studies are underpowered, no current evidence supports the use of inhaled beta 2 agonist bronchodilators or nebulised ipratropium for the treatment of infants with acute bronchiolitis.[69;70] RCTs of anti-inflammatory treatments for RSV infection have also proved to be unsuccessful, as neither inhaled nor oral systemic corticosteroids have demonstrated any reductions in duration of symptoms of length of bronchiolitis hospital admission.[62;71-73]

A multicentre, double-blind RCT of nebulised epinephrine was published in the New England Journal of Medicine (NEJM) in 2003, reporting that nebulized epinephrine did not significantly reduce the length of time until an infant admitted to hospital with bronchiolitis was ready for discharge.[74] However, in more recent years, some evidence for the effectiveness and superiority of adrenaline for short-term outcomes among outpatients with bronchiolitis has emerged.[75;76] This systematic review and meta-analysis published in the British Medical Journal (BMJ), also reports evidence from a single study suggesting benefits of combined epinephrine and dexamethasone steroid. A Cochrane review published in 2008 has also shown that nebulised 3% saline may be effective in reducing length of stay and clinical severity among infants hospitalised with bronchiolitis.[77]

For infants hospitalised with bronchiolitis, nasal suction is recommended in those who exhibit respiratory distress due to nasal blockage and nasogastric feeding should be considered in those infants who cannot maintain oral intake or hydration.[36] Although there is very little published evidence for its use, the expert clinical consensus is that supplemental oxygen

should be given to infants with severe respiratory distress, cyanosis or oxygen saturation level ≤92%.[36] A Cochrane systematic review investigating the use of chest physiotherapy to treat bronchiolitis among infants without comorbidities, found no evidence that this improves clinical severity or reduces length of hospital stay.[78-80]

1.5 Risk factors for bronchiolitis

Children known to be at high risk of developing severe RSV infection include those with chronic lung disease,[81-83] haemo-dynamically significant congenital heart disease,[81;84-86] immunodeficiency, low birth weight[87] and those born preterm (approximately 8% of newborns in the UK).[3;22;44;88-94] This is particularly concerning given the reported increases in preterm birth rates in developed countries over recent years, combined with improved survival of preterm infants, meaning that the population of at-risk infants is increasing.[31;36;95] Evidence from the US, published in the NEJM in 2009, suggests most children (66%) admitted to hospital with RSV infection have no co-existing medical conditions or characteristics that significantly identify them as being at greater risk of severe disease.[44]

Other factors may also be associated with RSV infection, for example deprivation,[96] residential crowding,[97] tobacco smoke exposure,[98-102] maternal smoking during pregnancy,[103;104] air pollution and traffic exposure,[105] lack of breast-feeding,[106;107] multiple gestation,[108] elective caesarean delivery,[109] winter birth and discharge from NICU between September and December.[110]

Although there is a wealth of evidence regarding the association between some known risk factors and severe bronchiolitis, the association between other potential risk factors and bronchiolitis remains unclear. Furthermore, the effect of potential risk factors on disease *severity* needs to be determined. There remains a poor evidence base of prognostic factors for hospital admission due to RSV infection.[111] A 2008 Health Technology Assessment (HTA) identified further large observational studies are needed to estimate the mortality rate among children known to be at highest risk of RSV infection such as those with chronic lung disease, congenital heart disease and ex-premature infants.[111] The specific effects of premature birth and low birth weight on risk of bronchiolitis hospitalisation have yet to be determined, as recommended by Wang et al, future research ought to, "*address major uncertainties among patient subgroups, including preterm infants of different gestational ages*".[111]

A more recent update of the HTA examining Palivizumab use for immunoprophylaxis of RSV bronchiolitis in high-risk infants and young infants, was published in 2011.[112] This report explicitly recommended that larger, better powered studies are needed to derive better estimates of risk factor effect sizes. In particular, improved estimates are needed to determine the effect size of parental education, multiple births, family history of atopy, lack of breastfeeding and gestational age.[112] This information would help to improve Palivizumab cost-effectiveness estimates.

A complex interplay between host and viral factors is thought to determine the severity of RSV infection.[92] Despite the known clinical risk factors for RSV bronchiolitis hospital admission described above, the decision as to whether bronchiolitis should be treated in hospital or in the community is a difficult one.[113] Whether an infant is hypoxic is commonly cited by clinicians as a factor noted when considering whether to admit an infant with bronchiolitis, yet the use of pulse oximetry for this purpose has questionable benefits.[22;113;114] Other studies have suggested that oxygen saturations below 92% are the strongest predictor of bronchiolitis hospital admission.[113;115]

One UK study reported on the development and use of a simple clinical risk scoring system to aid decision making in busy Emergency Departments.[113] The sensitivity and specificity of the tool varied according to risk factors included, but the study reported that the optimal 5 risks to include in the scoring system were: symptom duration of less than 5 days, respiratory rate of ≥50 breaths per minute, heart rate of ≥155 beat per minute, oxygen saturation less than 97% and age at presentation of less than 18 weeks. Further validation of the tool is required, among a much larger sample of infants, but this provides some positive development in the quest to improve the reliability of diagnosis and management of severe RSV bronchiolitis.

1.6 Prevention

1.6.1 Hygiene

Interruption of RSV transmission can be achieved through conservative prevention strategies such as reducing exposure to potential environmental risk factors such as passive smoking and avoiding overcrowded areas during the peak winter transmission season.[116] The primary focus for reducing nosocomial RSV transmission, particularly in hospital settings, is encouraging good hand-washing techniques or glove and gown use among healthcare professionals and those in contact with infected individuals.[53;116;117] This is

particularly important to prevent nosocomial RSV outbreaks among vulnerable, high-risk infants in neonatal units.[118]

1.6.2 Passive immunotherapy

Prophylactic passive immunotherapy with Palivizumab (Synagis®, MedImmune), a humanised mouse monoclonal antibody against RSV infection, is licensed for use in some countries for those most at high risk of RSV infection.[13;119] Palivizumab only provides short-term protection against RSV as it is a passive immunisation, not a vaccine.[120;121] Two large randomised, double blind, placebo controlled trials have shown palivizumab to be both safe and effective in reducing RSV hospitalisation rates and serious complications among high-risk infants.[122;123] Infants aged less than or equal to 35 weeks gestational age or those with bronchopulmonary dysplasia (n=1502) were randomized to receive 5 injections of palivizumab (15 mg/kg) or equivalent volume of placebo every 30 days. Overall, those receiving palivizumab (n=500) had a 55% reduction in RSV hospital admissions compared to the placebo group (n=1002).[122]

Palivizumab is at present the only passive immunisation licensed in the UK for the prevention of serious LRTI requiring hospitalisation caused by RSV. However, it is very expensive (the estimated cost for a single dose of palivizumab for an infant aged 6 months, weighing 7.5kg, is £1023.11[112] so the total cost for 5 doses over the winter season is just over £5000 per patient).

The Joint Committee for Vaccination and Immunisation (JCVI) RSV subgroup in the UK recently revised guidance for the use of palivizumab, though there was some debate over the limited evidence base on which these recommendations have been made.[124] Currently the JCVI state that palivizumab is only cost-effective and recommended for use in the following sub-groups of infants at most risk of severe disease:[111;112;124]

- Children < 2 years requiring treatment for chronic lung disease within the last 6 months
- Children < 2 years with haemodynamically significant congenital heart disease
- Children born at 35 weeks of gestation or less and less than 6 months of age at the onset of the RSV season.
- Children who have severe combined immunodeficiency syndrome (SCID), until immune reconstituted.
- All long term ventilated (LTV) children less than 12 months at the start of the RSV season and LTV children aged less than 24 months with additional co-pathology (heart disease/intrinsic lung disease as reflected by oxygen dependency).

Experts from the American Academy of Pediatrics have recently highlighted the insufficiency of current evidence for the benefits of prophylaxis particularly in immuno-compromised infants (such as solid organ or hematopoietic stem cell transplant recipients, HIV-infected infants and children with other primary and secondary immune deficiencies).[125] Further experimental studies are currently evaluating the use of palivizumab in infants with cystic fibrosis and immunosuppressed bone marrow transplant recipients.[13;116]

A HTA published in May 2011, considered population subgroups with different combinations of risk factors for whom the use of palivizumab may be cost-effective. The study found that at a willingness-to-pay threshold of £30,000 per quality adjusted life year (QALY), prophylaxis with palivizumab is only cost-effective among the subgroups of children with no chronic lung disease or congenital heart disease *if* they have at least two other risk factors besides gestational age at birth (including siblings at school, being male, smoking exposure, household overcrowding or multiple births).[126]

Adding to the continuing debate over precisely which infants are most at risk of severe RSV disease, a NEJM study published in 2009 reported that many children with RSV infection are previously healthy, so targeting only high risk infants with prophylaxis will be limited in its effects on the total RSV disease burden.[44]

Another prophylactic drug was also under development, motavizumab, which is also a humanised monoclonal IgG antibody which attacks the RSV F protein and was developed to achieve greater binding affinity and virus neutralising activity than palivizumab[127] A phase III, double-blinded, randomised, multinational trial was completed in 2006, comparing palivizumab and motavizumab, using the same inclusion and exclusion criteria as the original IMpact trial.[127;128] The study found motavizumab was non-inferior to palivizumab, with a 26% reduction in RSV hospital admissions compared with palivizumab. However, recipients of motavizumab were found to be at increased risk of skin reactions compared with those receiving palivizumab. Consequently, the USA's Food and Drug Administration (FDA) did not license use of motavizumab, expressing concerns over its safety regarding the increased hypersensitivity reactions.[129;130] MedImmune who developed motavizumab, have discontinued development of its use for prophylaxis against RSV but are maintaining research into its possible use for treatment of RSV infection.[127;131]

1.6.3 Active vaccination

Ideally either vaccination in early infancy or a maternal vaccine is needed to prevent the large burden of RSV-associated illness in early infancy. However, RSV presents a

significant challenge because of the difficulty in eliciting immunity in such young infants, combined with the inadequate attenuation of live viral vaccine candidates.[13;132;133] Furthermore, in this age group waning maternal antibody levels are not sufficient enough to protect against disease and can decrease vaccine immunogenicity.[134] There are also considerable safety concerns about inactivated vaccine candidates following the failure of a formalin-inactivated RSV vaccine which was trialled in the 1960s and resulted in two infant deaths and more severe RSV disease in those who had received the vaccine.[13;135;136] The potential impact of an RSV vaccine in the Netherlands has been modelled with the results suggesting it would be a cost-effective intervention, but in the absence of clinical trial data many assumptions about the vaccine characteristics had to be made.[137]

1.7 Impact on respiratory health in early childhood

In addition to the acute effects of RSV bronchiolitis, evidence suggests RSV may be associated with short and longer-term respiratory complications in childhood, including breathing difficulties and cardiovascular abnormalities.[138;139] Other studies have shown that RSV infection may have a long-term effect on the respiratory tract, causing recurrent wheezing, asthma and reactive airways disease in early childhood.[14;40;140-152] A further study of a prospective cohort in Sweden, showed RSV bronchiolitis is associated with an increased prevalence of allergic asthma persisting at age 18 years, but this was a small study with follow-up of only 46 RSV cases and 92 matched controls.[141] There is a growing body of evidence suggesting severe RSV infection in infancy may predict asthma and exacerbations of allergic airway disease in adulthood.[153;154] In addition, the long-term as well as acute effects of RSV have been shown to be associated with decreased quality of life and increased healthcare utilisation.[103;155-157]

However, the 2008 HTA concluded that there are still major uncertainties about the effects of RSV infection on long term respiratory health and mortality among infants in the UK.[111] For example, conflicting evidence has shown RSV infection in infancy to be a risk factor for reduced lung function in early adult life but not for atopy or asthma.[158] Other research has examined the role of genetic versus functional predisposition to chronic respiratory morbidity following RSV infection.[159] Recent studies have hypothesised that the disease phenotype may be more important than the causative agent, in predicting future asthma and wheeze.[160] This is supported by evidence from immunological studies which have shown the inflammatory response in acute bronchiolitis is quite different from that observed in asthma.[161]

It is not possible to determine from observational studies alone, whether RSV infection *causes* recurrent wheeze or asthma or is an indicator of pre-existing pulmonary vulnerability.[162] The double-blind, placebo controlled MAKI trial has provided new evidence of the association between RSV infection and wheezing in early childhood, as otherwise healthy preterm infants who received palivizumab prophylaxis were found to have a 61% relative reduction in risk of wheezing in the first year of life compared with those not receiving the passive immunisations.[162;163] Another prospective multi-centre trial found that palivizumab decreases the relative risk of pre-school recurrent wheezing by 80% among non-atopic children but had no impact on children with history of atopy.[164] This finding implies that in the absence of any genetic predisposition to atopy, RSV infection may have an important causative role in wheezing pathogenesis.[164] However, the study focused on preterm born infants only, so may not be generalisable to all children.

Emerging evidence has also shown that the inception of asthma may be associated with the persistence of latent RSV residing in bone marrow, where resident stromal cells which provide an immunologically privileged sanctuary, harbour the virus and allow persistence of infection.[165]

1.8 Summary of chapter

In this chapter current knowledge about the burden of RSV bronchiolitis has been summarised, from the wider global context to its current management and importance in the UK. Known risk factors have been described and current gaps in our existing knowledge of RSV epidemiology have been identified. Since there is no recommended treatment which shortens the natural course of infection and passive immunotherapy is only cost-effective in those at very severe risk, there is a real need to improve our understanding and the current evidence base for what constitutes optimal management of bronchiolitis in both primary and secondary care settings in the UK. Large, population-based observational studies have been recommended to derive better estimates of risk factor effect sizes. This is important information to inform future policy if an effective active vaccination candidate is found.

Key findings from literature review justifying this work

1) RSV infection is the leading cause of childhood LRTI with an estimated 33.8 million new episodes of RSV LRTI worldwide each year, among children under 5 years. It is therefore a priority area for development of novel prevention strategies and treatments.

2) There is a paucity of evidence surrounding the burden of bronchiolitis illness in the community, with no UK studies reporting the incidence in primary care settings.

3) In developed countries including the UK, small regional studies have estimated that 2 to 3% of all infants are admitted to hospital with RSV bronchiolitis. No recent, large, observational studies have estimated the clinical burden of RSV bronchiolitis at a population scale.

4) There is no known treatment to shorten the natural course of RSV bronchiolitis hence the mainstay of intervention is passive immunotherapy.

5) In the UK passive immunotherapy is only considered cost-effective and recommended for use in a small sub-group of high-risk infants.

6) Preterm birth, immunodeficiency, chronic lung disease and congenital heart disease increase the risk of hospital admission with severe RSV infection. However, evidence from the US suggests around two thirds of infants admitted to hospital are previously healthy with none of the known clinical risk factors.

From the literature review I conclude that large observational studies are needed to improve risk factor effect estimates. Policy makers need reliable estimates to model the risks and benefits and predict the impact of any emerging preventive interventions such as a novel passive or active immunisation against RSV infection. Since the wider impact of RSV disease in the community is not clear, larger studies with long-term follow-up are needed to clarify the impact of RSV bronchiolitis in infancy on subsequent respiratory health in the UK.

2.0 Aim and objectives

Chapter Overview

This chapter outlines the overall aims of this thesis. It begins by summarising the gaps in our existing knowledge of the clinical burden of RSV bronchiolitis and the opportunities that the availability of routine electronic health records could present, in enabling the development of large, observational cohorts with long-term follow-up to facilitate child health research. This forms the rationale for the methodology used and research topic that forms the focus of this thesis. The hypothesis being tested and research questions are then presented, followed by the overall aim and specific objectives of this thesis.

2.1 Summary of priorities identified for further research:

- Identify the key clinical risk factors that predict hospital admission with severe RSV bronchiolitis in the UK.
- Determine which of these pre-disposing factors are most important in relation to disease severity.
- Determine what proportion of infants with RSV bronchiolitis is born at term and has no known clinical risk factors.
- Address major uncertainties among patient subgroups – in particular, determine the specific impact of gestational age at birth on risk of RSV bronchiolitis hospitalisation.
- Estimate the incidence of RSV bronchiolitis in the community.
- Determine the impact of RSV bronchiolitis in infancy on longer-term respiratory health.

2.2 Rationale for thesis

It is well established that RSV infection is the leading cause of childhood LRTI worldwide and is a significant cause of morbidity particularly in infancy.[21] With birth rates in the UK rising and improved survival of infants born preterm, the population at risk of severe RSV disease is increasing.[88;166] This in combination with the lack of effective treatment for RSV bronchiolitis and expensive passive immunotherapy only being considered cost-effective in a minority of extremely high risk infants, raises the importance of improving the existing evidence base from which we can determine how best to manage affected infants.

Presently there are no national estimates of the incidence of bronchiolitis in the community and even less is known about how it is being managed in primary care and whether GPs are following existing guidance. Existing research has focused on the more severe end of the disease spectrum, but even research examining the risk of hospital admission with RSV bronchiolitis has been limited to small, regional studies, many from individual hospitals and few based on UK data. To date, no large, observational studies have estimated the clinical burden of RSV bronchiolitis at a population scale. Such studies are needed to improve risk factor effect estimates.

Longer-term follow-up of affected infants is required, to address current uncertainty surrounding the impact of RSV bronchiolitis in infancy on subsequent respiratory health. However, prospective follow-up and large cohort studies, particularly in child populations, are often precluded by cost and practicability.

Databases of routinely-collected electronic health records, particularly here in the UK where everyone has universal access to healthcare, can provide a solution to some of these issues which have previously hindered progress in this field of research. Observational studies designed using these data will provide representative, population-level data that are highly generalisable, relate directly to real clinical practice and are sufficiently powered to detect small effect sizes. Hence this justifies the use of such databases as the approach I have chosen to explore the specific research questions in this thesis.

The gaps in our current knowledge of the epidemiology of RSV bronchiolitis in the UK have not only been identified by my own review of the existing literature evidence, but these areas for further research have also been highlighted in current UK guidance for bronchiolitis management, as well as recommended in the National Institute for Health Research (NIHR) Health Technology Assessments on this topic. This provides further justification that it is essential that we identify which infants are at high risk and quantify the wider, clinical burden of RSV bronchiolitis across healthcare settings in the UK, in order to ensure appropriate

distribution of resources and optimal management of affected infants. This information could be useful for developing an electronic risk prediction tool to aid clinical decision making when managing infants with bronchiolitis.

The findings of this thesis will be of vital importance in the event that a successful, safe and effective RSV vaccination candidate is identified. Health policy makers in the UK will need this information to help them to determine the target population for an emerging vaccine and to model the potential risks and benefits to predict its impact. This research will also provide more detailed clinical evidence from across healthcare settings in the UK, to ensure existing recommendations for the use of palivizumab passive immunotherapy are targeting infants at greatest risk of severe RSV infection. In this thesis I seek to add to and improve the existing evidence base on which recommendations for the prevention and clinical management of RSV bronchiolitis can be made.

2.3 Hypothesis

Existing evidence underestimates the wider clinical burden of RSV bronchiolitis among infants presenting across healthcare settings in the UK.

2.4 Research questions

What is the clinical burden of RSV bronchiolitis among infants in the UK?

To describe the total clinical burden of disease, it is necessary to examine the wider clinical spectrum of bronchiolitis illness across different healthcare domains (Figure 1).

1) What are the risk factors for admission to hospital with RSV bronchiolitis and how do admission rates vary between infants with comorbidities, preterm and term born infants?

2) What is the incidence of acute bronchiolitis among infants presenting to general practitioners and how is it being managed in the community?

3) What is the long-term impact of RSV bronchiolitis on subsequent respiratory health in early childhood?

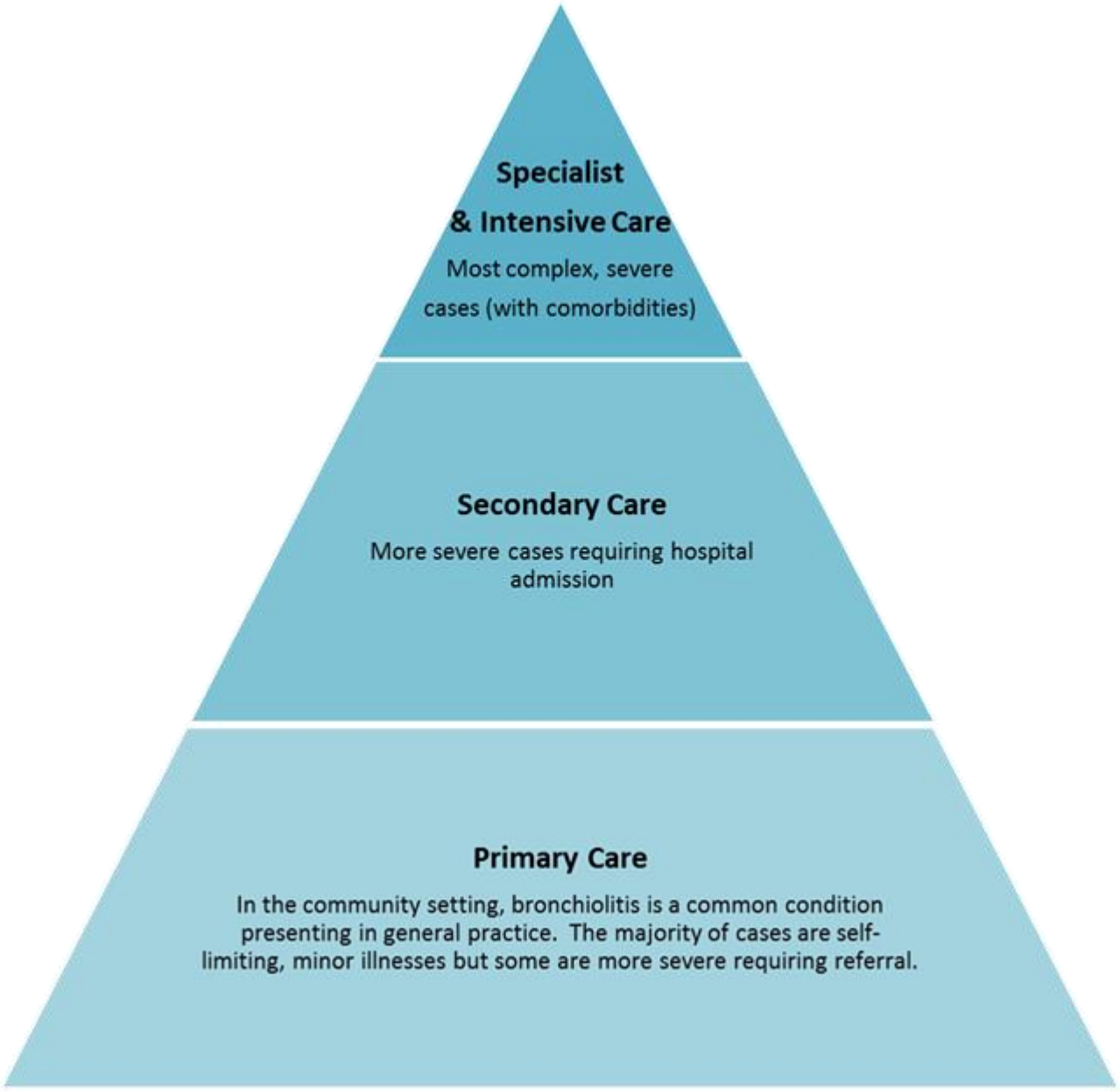


Figure 1: Pyramid of care for infants with RSV bronchiolitis.

2.5 Aim

To improve estimates of the wider clinical burden of RSV bronchiolitis among infants in the UK.

2.6 Objectives

1. To describe the clinical burden of RSV bronchiolitis among infants admitted to National Health Service (NHS) hospitals in England, comparing admission rates between infants with and without clinical risk factors for the condition (Study 1 – Using linked HES hospital records to develop a nationally representative birth cohort).

2. To estimate the incidence of RSV bronchiolitis among infants presenting in primary care and to describe how it is being managed by general practitioners, in relation to national guidance (Study 2 Using GPRD patient records longitudinally linked across time to develop a nationally representative cohort).

3. To determine the longer-term effects of RSV bronchiolitis on subsequent respiratory health in early childhood (Study 3 - GPRD and HES long-term follow-up studies).

3.0 Use of routinely collected data for child health research

Chapter overview

In this chapter the practical issues and ethical challenges that arise when conducting child health research are described in detail. The use of electronic patient records for observational research in children and linkage between such databases is discussed. This chapter provides important background and rationale for the methodological approach applied in this thesis.

3.1 Problems in conducting child health research

Events in early childhood have a significant impact on health and development throughout life.[167;168] Infant health is therefore a very important area of research, where the impact of science on clinical care has long-lasting effects into adult life. Multidisciplinary research is required to improve the scientific basis of strategies to prevent, diagnose and treat diseases affecting the health and well-being of children. Conducting clinical research in child populations can be logistically difficult, expensive, time consuming and ethically challenging.[169] Large epidemiological studies are a crucial tool for enhancing the evidence base used to inform clinical practice. Unfortunately, assessing health outcomes in large populations of children has been severely limited by costs and complexity of acquiring data.[170] Alternative methods for investigating child health over longer periods of time are required. Routinely collected electronic patient records could provide an invaluable data resource for clinical research among children.

Undertaking clinical research studies in children presents distinct challenges to researchers, in what is a rapidly evolving, dynamic field.[171;172] Yet it is widely recognised that research into the determinants, diagnosis, treatment and prevention of childhood disease is of vital importance,[171] not least because clinical decisions need to be justified with valid up-to-date evidence.[173] In the field of paediatrics this can be challenging, given the dearth of clinical trials addressing issues relevant to child health.[174;175] Closer integration and translation of child health research into core National Health Service (NHS) activities is required.[176] Over recent years steps have been taken to address these issues in the UK and utilise the research opportunities presented by a NHS providing universal coverage. We now have research support organisations such as Clinical Research Networks and the necessary infrastructure to facilitate improvements in paediatric research.[174] The Medicines for

Children Research Network (MCRN) reported a total of 75 studies of medicines in children in the UK on their portfolio, in 2008,[177] 7 of which were registered studies open or in set-up with their clinical trials unit. The production of the British National Formulary for Children (BNFC) was also a landmark event in improving medicines for children. The Royal College of Paediatrics and Child Health (RCPCH) are implementing initiatives such as 'Turning the Tide' to try to strengthen child health research by increasing exposure and involvement among undergraduate medical students through to paediatric trainees and consultants.[178] Despite these improvements, the difficulties of establishing clinical studies among child populations remain. Here, in the context of paediatric research, I will evaluate alternative study designs and highlight how more efficient use of routine data might improve the quality and range of research produced in the UK.

3.1.1 Experimental study design

Clinical trials are the preferred study design for investigating the effects of a particular intervention and comparing with existing alternatives or placebo. The obvious advantage of such experimental studies is that the assignment of an exposure or treatment to individuals is controlled by the investigator. Even distribution of potential unknown confounders can be maintained through randomisation and participants can be blinded as to which intervention they receive to reduce bias. Randomised controlled trials (RCTs) can therefore provide strong empirical evidence of the efficacy and safety of therapeutic interventions in a clinical setting.

However, conducting clinical trials in child populations presents distinct ethical challenges. There have been several tragic examples where infant deaths or disease may have been prevented had more in-depth studies been conducted. These include the effects of thalidomide on development of the fetus, the use of propofol as a sedative resulting in at least 15 deaths of critically ill children[179-181] and the association between grey baby syndrome and use of the antibiotic chloramphenicol in neonates.[181;182]

In addition to valid ethical anxieties, the logistical considerations, intensive resources and financial requirements of establishing trials in children are extensive and often prohibit long-term follow-up of patients enrolled in trials. Costs of conducting trials among children are greater than among healthy adults[180] and consequently most medicines used in children, particularly among neonates, have not been subjected to the same level of clinical evaluation as in adults.[183] Recruitment of children to RCTs can be difficult with complex issues in acquiring consent. Consequently, where trials in child populations are established, sample sizes are frequently too small; one review reported that around 50% of published trials had sample sizes below 40.[174;184] Limited sample sizes means trials are frequently

under powered for examining side effects and have limited generalisability. Following the marketing of a drug, funding from pharmaceutical companies may be less forthcoming. Indeed, many phase IV trials cannot recruit large enough populations to sufficiently monitor side-effects associated with long-term use, especially as very large numbers of individuals need to be exposed to a drug in order to detect reactions that occur with a low probability.[185] So findings from artificial trial settings among highly selected individuals may not always translate into sustainable, safe interventions in routine clinical practice.

3.1.2 Non-experimental study designs

Whilst RCTs are considered the gold standard study design, it is also important to consider situations where alternative approaches to clinical research may be more appropriate. Large observational studies such as case-control, case-series, ecological and cohort designs are a crucial tool in expanding the clinical evidence base, particularly for child health. Observational studies can provide more generalisable findings than experimental studies, as included individuals tend to be more representative of the general population since there are not the strict exclusion criteria required for clinical trials, eliminating the inherent healthy volunteer bias present in most RCTs. Observational studies can also provide data on real-life effectiveness and complications, which may differ from findings from artificial trial settings with strict participant selection criteria. Sufficiently large sample sizes and long-term follow-up are more easily achieved in observational studies such as EPICure (UK) and EPIPAGE (France); longitudinal population-based birth cohorts of extremely preterm babies born in the mid-1990s.[186] Large prospective cohort studies such as these with long-term follow-up over many years are rare because they are so expensive and time consuming. Even with large cohort studies, teasing out the burden of many conditions at a national population level, especially very rare disorders, is still a challenge. Assignment of individuals is not easily controlled by researchers conducting observational studies, so the strength of evidence provided can be limited by the potential for biases due to unknown confounders. Non-experimental studies are usually less expensive than RCTs but can similarly be limited by attrition difficulties.[187;188] Observational research can provide the firm hypothesis base and information for sample size calculations required for subsequent experimental studies and as such are an invaluable clinical research tool, providing scientific justification for clinical trials.

3.2 Use of routinely collected electronic patient health records

A new challenge in paediatric research is to better utilise evolving healthcare technologies to facilitate faster dissemination of population-level interventions into the public domain.[170] This should certainly be possible in the UK which is considered to be at the forefront of research

involving secondary use of healthcare data, as the NHS provides a unique source of national data due to its universal coverage.[189] In light of the current financial climate and widely anticipated cutbacks to funding for research, universities and the NHS, attention is turning towards the utilisation of existing data sources for health services and clinical research. Data from routinely collected electronic patient health records are increasingly being used to address the methodological difficulties of carrying out research in populations that are typically challenging to study, due to significant ethical concerns and vulnerability, such as patients with very rare diseases, pregnant women and children. Already secondary use of routinely collected data from children has significantly contributed to improvements in child health research in the UK.[6;190-194]

For example, real-time data has been used to alert trusts of safety issues such as higher than expected child mortality following cardiac surgery.[195] The fundamental advantages of using such routine databases are the minimal costs involved and magnitude of information they can provide, often at a national level. These are factors which have previously severely limited evaluation of health outcomes over long periods of time, in large geographically defined populations.[170] Where survival *is* reported in neonatal studies for example, data are often from specialised units and are therefore subject to inherent selection bias resulting from different criteria for admissions, treatment or referrals.[95] Routine data permit investigation of long-term outcomes which is so rarely feasible in RCTs, thereby providing invaluable information about real-world implementation in populations outside trial inclusion criteria.

3.2.1 Ethical considerations and consent for long-term follow-up

The inherent difficulties surrounding anonymity and obtaining consent when using secondary data are well established[189;196] and it has been suggested that current UK research regulations can impede follow-up in multicentre studies.[197-199] Difficulties in acquiring data on long-term outcomes were highlighted by the Bristol Royal Infirmary inquiry in 2001.[197] The requirement for an individual's consent to confirm they are alive and monitor survival often precludes effective long-term follow-up for prospective studies due to the ethical and practical complications of tracing participants who change address or die.[197] A series of BMJ editorials highlighted some important issues surrounding the use of identifiable personal data for medical research purposes.[189;196;200] This series stressed the need for regulating bodies to accept that current UK laws permit the secondary use of data without consent or full anonymisation, provided the likely public benefit is demonstrably proportionate to the risks of identification and subsequent distress this may cause.[189]

The National Information Governance Body (NIGB) Ethics and Confidentiality Committee determines when obtaining individual consent may not be practicable, for example when undertaking large population studies of survival, such as among children with chronic conditions.[197] Hansson's BMJ article rightly argues the need for a wider view of autonomy in epidemiological research.[201] Systematically obtaining individual signed consent for sharing patient identifiable information is difficult and can sometimes hamper research.[201-204] though for particular research questions it may be wholly necessary to obtain individual/parental consent .[198]

3.2.2 Potential uses of child health records for research

Information on birth characteristics, such as gestational age and birth weight, is needed for many epidemiological and health services research studies examining short- and longer-term clinical outcomes.[205] Detailed information on early birth exposures is routinely captured in hospitals throughout England and can provide a rich electronic source of clinical information about the health status of individuals at birth, as well as for measuring the quality of maternity care. Secondary use of such data involves comparatively minimal costs in comparison to bespoke prospective observational studies. Large epidemiological datasets are particularly useful for infant health research where life-long assessment of outcomes is required, but all too often precluded by reason of cost, methodological difficulty and practicability. Consequently, much infant research is selective, poorly generalisable and focused on short-term outcomes. The secondary uses of data from electronic health records cannot replace nor precisely replicate the value of data collected for prospective studies, but the potential benefits for clinical and health services research and impact on patient care are well established.[206-209] These data can facilitate benchmarking and comparison of outcomes between different hospital sites or geographical areas. Cohorts based on these data, particularly when combined with novel techniques such as Mendelian randomisation, can provide representative population-level information that is highly generalisable, has power to detect small effect sizes and relates directly to real clinical practice. Such cohorts have the potential to answer an extensive range of research questions that require longitudinal follow-up, from examining long-term health outcomes and healthcare utilisation following preterm birth, to identifying risk factors for childhood hospital admission with influenza or respiratory syncytial virus infection.

3.3 Linkage with other health records

In health systems with universal coverage, the value of clinical databases for research can be significantly enhanced by linking different data sources, creating the potential to develop

cradle-to-grave datasets for the whole population. Probabilistic or deterministic linkage methods can be used to match records for the same individual from different databases, using unique personal identifiers such as NHS number. Algorithms can be developed to assign individual unique identifiers using variables such as NHS number, date of birth and postcode. Linkage between datasets can provide validation of recording, coding and completeness of data, help to develop more robust clinical case definitions and provide information on events that happen outside healthcare settings.[210] For example, linking to Office for National Statistics (ONS) mortality data in England yields information on cause of death and out-of-hospital deaths.

Successful matching of records is dependent on the quality of unique identifier variables available,[189] but linkage methods can correctly identify a high proportion of matched records even with limited availability of personal identifiers.[211] Large scale population-based linkage presents its own complex ethical challenges of consent, confidentiality and secure data storage that are best addressed by robust information governance systems and engaging patients and the public about how their records are to be used.[189;212]

3.3.1 Examples of linkage between birth and child health records to date

Non-UK

The effectiveness of routine health record linkage in *adults* has been demonstrated in Australia and Canada,[210;212-214] where it has improved both data quality and utility.[210] Birth records have been successfully linked to hospital discharge data in Australia, with matching rates of 99%.[215] In several regions of the United States, data from birth certificates have been linked to hospital discharge records to examine maternal outcomes.[216;217] In Scandinavian countries, the assignment of unique personal identification numbers permits linkage between civil registration systems, enabling the development of population-based cohorts. These have facilitated a broad array of epidemiological studies, including research in Denmark investigating the impact of place of birth and familial risk factors on risk of autism[218] and examining the association between prenatal exposure to high levels of coffee and attention-deficit hyperactivity disorder in childhood.[219]

UK

In Wales, the Secure Anonymised Information Linkage (SAIL) databank brings together anonymised person-based electronic health and social care data, which are now also being

combined to establish an anonymised Wales wide Electronic Cohort for Children (WECC).[220;221] This databank has successfully been used to assist with identifying potential clinical trial participants from primary care data.[220;222] The Scottish Health Informatics Programme (SHIP) is an example of a complex database of linked electronic patient records, providing health and social care information from birth through to death.[223] To date, SHIP data have primarily been used to conduct pharmacovigilance and diabetes epidemiology research.[223;224] Another well-established system in the UK is the Oxford Record Linkage Study covering over 10 million records for around 5 million people since 1963.[225] This database has been used to carry out longitudinal research studies identifying maternal and perinatal risk factors for a range of conditions such as inflammatory bowel disease[226], asthma [227] and coeliac disease.[228]

Administrative birth data are captured in England in both hospital discharge records and also birth registration information collected by the ONS. Use of information contained within the English HES delivery and birth records for research has been limited. A previous feasibility study showed high rates of linkage between HES maternity and ONS birth records.[229] In addition to these national sources of birth data, the NHS Numbers for Babies (NN4B) Service, introduced in 2002, collects a small dataset containing some key fields that are not recorded in birth registrations, such as gestational age.[230] This NN4B service ensures every baby is allocated a unique NHS number shortly after they are born. Linkage between NN4B and HES birth records has indicated that improvements in the quality and completeness of HES maternity data are needed.[230;231] HES maternity data have also been used to validate birth information recorded in the Millennium Cohort Study, a longitudinal observational cohort of nearly 19,000 babies born across the UK.[232]

3.4 Database linkage

With uncertainty surrounding the future of the Connecting for Health summary care record, it is imperative that we maximise the research potential of routinely collected patient data. Indeed previous reports have highlighted the need for initiatives to support linkage between different databases, recommending a “federated structure of data sources” rather than individual databases.[233] Although the quality of clinical databases can vary,[234] the intrinsic value of their use for research purposes can be significantly enhanced by linking different sources. Furthermore, comparison between datasets can provide validation of clinical recording, coding and completeness of data. As well as the quality assurance benefits of linking datasets, additional information on events that happen outside the healthcare setting

of data capture can be provided;[210] for example linking to Office of National Statistics (ONS) mortality data to obtain vital information on out-of-hospital deaths.

Probabilistic matching techniques have been developed to link health service data since the 1960s[235] and have been considered an acceptable method for conducting population based health research for some time,[212] made possible by significant advances in computer technologies. Field and colleagues report that although issues surrounding data ownership require some consideration, the advantages of routine linkage of health record databases are readily apparent and their effectiveness has already been demonstrated in countries such as Australia and Canada.[210;213] Their study demonstrates the benefits of linkage between oncology databases and suggests routine combining of data from such sources can improve both data quality and utility.[210] The value of linking individual records from disease registries and other healthcare databases for epidemiological and health services research has also been demonstrated in the UK. For example, the Scottish Medical Record Linkage System and the Oxford Record Linkage Studies described earlier.[225;236]

3.4.1 Advantages and limitations of using linked health records for research

Databases of linked medical records could be useful for many purposes: as public health tools for conducting epidemiological research, to provide long-term data on patient care pathways, to evaluate interventions and also for health services planning and management. In comparison to conducting large prospective studies with several years' follow-up, linking existing patient health records is considerably less expensive and time consuming and when using national data findings are highly generalisable. In addition, the development of very large cohorts with high statistical power permits the detection of even small effect sizes and can overcome some of the limitations of cross-sectional data such as under ascertainment of cases.[237] For widely used medications, large routine datasets can provide useful information on side effects and safety, which under-powered clinical trials fail to sufficiently measure. Specifically for epidemiological research among groups that are conventionally challenging to study, linkage between databases represents a potentially invaluable tool. This is especially true in the field of neonatal medicine as with improved neonatal care, survival rates among preterm babies are increasing so longer-term follow-up following discharge from neonatal intensive care units (ICUs) is imperative.

There are some inevitable limitations of linking electronic patient records for research purposes which must be considered. Using secondary data sources is not always ideal as researchers have no control over the information available since the data are often collected for an unrelated purpose. In many clinical trials and observational studies, losing patients during follow-up due to death or moving residence can be a problem. Utilising linked patient

health records should make it easier to track individuals for example between different hospitals or trusts within the UK. A considerable limitation of using electronic records for research purposes is variation in the accuracy of clinical coding. However, in recent years the quality of coding in many national datasets has improved[238-240] and they are increasingly being used for clinical research.

Possible methodological limitations of linking healthcare records include the degree to which data can be anonymised without losing vital information and the successful matching of records is dependent on the quality of unique identifier variables.[189] However, provided methods are tailored to the datasets being used and the context of how data were collected is considered, linkage methods can correctly identify a high proportion of matched records even with limited quality or availability of identifiers.[211] Furthermore, efficient methods of validating probabilistic record linkage have been widely reported.[241] Large-scale population-based linkage presents its own complex ethical challenges as well as considerable difficulties with respect to consent.[189;212] The key to making linked resources useful and accessible is ensuring confidentiality and secure data storage.[189;212]

Record linkage is not a viable solution to answer *all* research questions. It will be necessary in some studies to obtain individual/parental consent.[198] Equally, observational study designs are not appropriate for all paediatric research and are no substitute for rigorous RCTs of interventions. However, record linkage can supplement clinical trials by providing information on long-term outcomes. This is especially important in paediatric research where longitudinal data are so infrequently available. To facilitate linkage between existing databases of patient records, cooperation between data owners will be vital as is the requirement for rigorous, research governance. Educating and engaging with the public is essential to provide reassurance about the confidentiality of personal medical data used for research purposes, especially in a society where personal data are commonly shared among commercial organisations.

3.5 Summary of chapter

High quality paediatric research is essential to provide a firm evidence base for clinicians, to improve patient care and develop better diagnostic tools and treatments for children throughout the UK and worldwide.[174] The UK is in a unique position with many healthcare settings routinely collecting and storing detailed information on child health. Existing data sources can be used to improve clinical trial design as well as for monitoring safety following the marketing of new interventions. If discernibly high standards of patient confidentiality and data security can be guaranteed and linkage methods validated, utilising existing resources for research purposes should be encouraged. Among populations that can be so challenging to study, this represents a cost-effective alternative source of data for longitudinal observational research. Linking routinely collected electronic health records from different databases could help to address the current paucity of information on long-term health outcomes among children, which is essential to improve the translation of research into clinical practice.

4.0 Overview of health record databases in the UK

Chapter overview

Having discussed in the previous chapter the use of electronic health records for child health research, this chapter presents more detailed contextual information about some specific databases of routinely collected healthcare records in the UK. The research governance procedures and ethical approvals required to use these data are explained and the access and approvals I have to permit use of these particular databases are outlined. These data sources will be used to try to answer the research questions I have developed and to test the hypothesis on which this thesis is based (chapter 2.0).

4.1 Introduction to child health records in the UK

The UK is in a unique position among developed countries, because of the universal access to healthcare provided by the NHS, unlike more fragmented healthcare systems such as in the USA. The NHS is also more technologically advanced than healthcare systems in other developed countries, with most areas of healthcare using electronic systems to store health records. As a result, a wealth of patient information is stored across a range of healthcare databases, which can provide valuable data for health services and clinical research, particularly among populations which are typically difficult to study such as pregnant women and children.

Patient information is recorded whenever and wherever an individual accesses healthcare, from primary care settings such as visiting a GP to secondary and tertiary care settings where inpatient and outpatient records are collected. Administrative, clinical, diagnostic, therapeutic and microbiological information from patients are all collected in different ways and recorded in different systems across different healthcare settings within the NHS. In the absence of a universal care record which could contain all this information in one system, to be shared and accessed by all health professionals, the current fragmented approach means there is much duplication of effort and data collection. There are therefore many different regional and national databases where these health data are stored.

The literature from observational studies described in chapter 1, were predominantly from small scale, regional studies. National patient datasets provide a means of estimating the

burden of illness at a much larger, population scale. Not only for studies of ecological design, but the methods chapter will also describe how I piloted the use of these data to develop cohorts of children, with longitudinal follow-up over time, from birth.

In this thesis I describe studies developed using secondary and primary care data. National hospital admission data for England are recorded in the Hospital Episode Statistics (HES) database and the General Practice Research Database (GPRD) contains data from a nationally representative sample of 600 general practices in the UK. The National Neonatal Research Database (NNRD) contains neonatal records from almost all neonatal units across the UK.

4.2 Hospital Episode Statistics (HES)

4.2.1 Overview of dataset

HES is a national administrative database, electronically recording all admissions to NHS hospitals in England since 1989.[242;243] HES contains a wide range of data including coded clinical diagnoses, procedures, geographical information on where patients are treated and demographic information such as patient age, sex, socio-economic status and ethnicity (www.hesonline.nhs.uk). In addition to inpatient records, outpatient attendance data are available since 2003, with more than 40 million new records added annually and more recently A&E data have also been collected since 2007.[242] Information on deaths is also available in HES, as records are linked to mortality data collected by the Office for National Statistics.

The basic unit recorded in HES is the finished consultant episode, which covers the period of time during which a patient is under the care of a given consultant.[242] A 'spell' or admission is defined as the continuous period of time spent as a patient within one hospital,[243] from admission to discharge or transfer to another provider. Spells may therefore include more than one consultant episode, for example if a patient is transferred from a general medical ward to a specialist or intensive care unit.[243]

HES data are divided into financial years from April 1st in a given year to March 31st in the following year. The International Classification of Diseases system version 10 (ICD-10) codes are used to record the main reason for admission or 'primary diagnosis', with up to 13 secondary diagnoses.[244;245] procedures and interventions are coded using the 4th revision of the Office of Population, Censuses and Surveys: Classification of interventions and procedures (OPCS-4).[242;245] The dataset has changed over time, with older versions of

these clinical classification systems being used in earlier years, for example ICD-9 codes were used until 1995 when use of the tenth version was introduced.

4.2.2 Data collection and purpose

The mechanisms used to collect HES data have changed over time, as changes to the structure and organisation of the NHS have occurred. Currently, providers of healthcare such as NHS Trusts, submit data to the National Programme for IT's Secondary Uses Service (SUS) (previously to the NHS-Wide Clearing Service).[242] To support patient care and inform commissioners of that care, healthcare providers collect clinical and administrative information locally. These data are submitted to the SUS, which make this information available to commissioners and also copy it into a database. At set times throughout the year, SUS send an extract of their database to HES.[242] This national HES data warehouse for England is run by the NHS Information Centre for Health and Social Care. Data within SUS can continue to be updated and changed, whereas data in HES is fixed as it was on the date that the extract was taken, so there can be discrepancies between them. HES data are then cleaned, validated, new fields are derived and quality checks are carried out.[242;245]

There have been some concerns regarding the accuracy of diagnostic coding in HES and variation between hospitals, but coding has been improving over time.[240;246-248] HES is the data source for a wide range of healthcare analysis, for research organisations, the NHS and the government.[242] HES data have been used extensively for research, measuring quality of healthcare delivery, examining disease time trends, variations in practice and to make international comparisons.[6;249-255]

4.2.3 HES maternity records and baby tails

Maternity data in HES consists of delivery records (for mothers) and birth records (for babies). These contain additional data fields known as the "baby tail", which include information such as birth weight and gestational age.[245;256;257] As soon as a mother has given birth, her hospital admission record changes from a general inpatient admission record to a maternity record and is updated as such before it is submitted to HES.[242] HES contains two types of maternity record, the delivery and the birth record, both of which contain a "baby tail" of an additional 19 fields.[242] The delivery record is for the mother and contains the same information as a general record plus an additional baby tail with information about the delivery.[242] The birth record is for the baby and also contains general record information plus the baby tail, which contains identical information to the corresponding baby tail in the mother's delivery record.[242] Diagnoses and procedures recorded in the birth record refer to the baby and, conversely, diagnoses and procedures in the delivery record refer to the

mother.[242] For multiple births, separate tails for each baby will appear in the mother's delivery record, but each birth record will contain only the individual baby's relevant tail.[242] HES data are collected from births in NHS and non-NHS hospitals and at home, although information from births outside NHS hospitals is often incomplete.[242] A list of some of the maternity data collected within HES can be found in Box 1: Key information collected in HES maternity records.

.

Box 1: Key information collected in HES maternity records.

- Method of onset of labour
- Place and date of delivery
- Miscarriages and ectopic pregnancies
- Live and still births
- Deliveries and births with complications
- Multiple deliveries
- Length of gestation
- Birth weight
- Neonatal unit admissions and level of care
- Antenatal and postnatal length of stay

4.3 The General Practice Research Database (GPRD) now part of the Clinical Practice Research Datalink (CPRD)

4.3.1 Overview of dataset

The General Practice Research Database (GPRD) database contains computerised clinical records from about 5 million active patients, 12 million patients overall, from 600 primary care practices across the UK and is a nationally representative sample of around 8% of the UK population.[258] From 29th March 2012, GPRD is now part of the Clinical Practice Research Datalink (CPRD), the new observational data and interventional research service for the English NHS, funded by the Medicines and Healthcare Regulatory Agency (MHRA) and the NIHR.[258;259] The CPRD aims to maximise the way anonymised clinical data from the NHS can be used for observational research, using linkage to integrate data from primary care, secondary care and disease registries, with the aim of facilitating research that is beneficial to improving public health.[259]

It is estimated that over 98% of the UK population are registered with a GP, almost all of whom use computerised record systems.[260] Databases such as the GPRD are therefore a rich source of longitudinal patient data. Information collected from GP records includes demographics, diagnoses, symptoms, preventative care, prescriptions and referrals.[260] In the UK, a standardised hierarchical classification system of READ codes is used to record medical information in patient records.[261] The Oxford Medical Information System (OXMIS) is another clinical coding system that was previously used by GPs contributing to the GPRD.

4.3.2 Dataset uses

The C/GPRD is the largest and most comprehensive source of primary care data in the UK and has been widely used for research, from pharmacovigilance to risk score development.[258;262] It has been particularly useful for child health research, for example for studies examining the association between antiepileptic drugs and child mortality,[263] antibiotic use and the risk of community-associated methicillin-resistant *Staphylococcus aureus* (MRSA)in children[264] and quantifying the risk of convulsions in children who received the monovalent H1N1 and trivalent influenza vaccines.[265] The GPRD was also used to examine any possible association between the mumps, measles and rubella (MMR) vaccine and autism, with the findings showing no association between use of the vaccine and autism.[262;266;267]

4.4 The National Neonatal Research Database (NNRD)

4.4.1 Overview of dataset

In 2003 a Department of Health review[268] led to the reorganisation of neonatal services in the NHS into clinical networks with coordinated care services and shared management.[170;269] As a result, a new system of electronic neonatal records was developed and is now in use by the majority of neonatal intensive care units (NICUs) across the UK (approximately 165 units).[170;270;271] The Neonatal Data Analysis Unit (NDAU) is an independent academic unit at Imperial College London, established in 2007 to analyse data held on electronic neonatal records. The NDAU obtain a standard core clinical dataset from NICU admission to discharge, which was developed by the British Association of Perinatal Medicine (BAPM)[272] and is held on a common NHS platform, neonatal.net. This database is now known as the NNRD and is held at the NDAU. The NDAU undertake clinical audits, monitoring, service evaluations, administrative reports and health services research. The NDAU Steering Board currently includes representation from Imperial College London, the National Perinatal Epidemiology Unit, BAPM and the new-born charity BLISS.

Included in the dataset are data from ten audit questions which form the National Neonatal Audit Programme (NNAP) which was set up by the RCPCH in January 2006, with the aim of informing good clinical practice in aspects of neonatal care by auditing national standards.[273] The dataset also incorporates the mandatory National Critical Care minimum dataset. In 2012, the UK National Neonatal Collaborative was established, comprising all NHS trusts that contribute data to the NNRD, to support UK neonatal units in utilising clinical data to improve patient care and outcomes and to support service delivery and evaluation and the NNAP.[271;273]

4.4.2 Data collection and uses

The data held in the NNRD are entered by neonatal clinical staff through a system of drop-down boxes, incorporating entry-level logic and range checks, with minimal free text entry.[271] In general, data are of a high quality as they are used daily for clinical care, generating discharge summaries and to support commissioning and network management. Babies are frequently transferred between neonatal units providing different levels of care, so these electronic records provide instant access to full clinical details by a receiving hospital when a baby is transferred. Since 2011, these data have also provided the basis for Payment by Results (PbR) for newborn services.

At present, records from around 60,000 babies admitted to NHS neonatal units across the UK are entered onto the electronic neonatal data system each year, which is managed by Clevermed, an authorised NHS hosting company.[274]

The neonatal database is unique internationally as data are collected in a standardised format during the course of clinical care. The database provides information from a large geographically defined population which can be used to facilitate applied neonatal research.[271;275]

Data are entered onto the neonatal.net platform by clinical staff on neonatal wards, so these are actively recorded, integrated records of clinical care which are used to produce final discharge summaries. A range of data fields are collected, including static, episodic and daily items. Static items are recorded once, usually when a baby is initially admitted and include items such as NHS number, name, gestational age and birth weight. Episodic items are recorded at each episode of care and include information such as diagnoses, infections and admission and discharge times. Daily items are recorded on each day of care and include information on care received such as respiratory support or enteral feeding.

These neonatal records are securely uploaded and stored on an NHS server by Clevermed Ltd. For all UK units who have provided the required Caldicott Guardian permission, these records are then encrypted and securely transferred via NHS connection to the NDAU where they are stored on an NHS server. The NDAU then examine the data, particularly completeness of recording and carry out logical checks for consistency and realistic ranges of values.[276] Some data cleaning and processing is also carried out, such as converting date variables, removing duplicate records and assigning numeric equivalents to string items or to free text items that occur frequently.

4.5 Research governance and ethical approval

Aggregated, anonymised data from some of these routine healthcare databases are freely available and accessible to all online. Where use of individual and/or identifiable patient data for research is required, ethical approval is needed. The National Information Governance Board for Health and Social Care (NIGB) is an independent statutory body responsible for the improvement, monitoring and promotion of information governance in health and social care in the UK.[277] Their role is to advise and ensure the lawful and ethical use of patient information, both for individual patient benefit and for the greater public good. Section 251 of the NHS Act of 2006 (previously Section 60 of the Health & Social Care Act 2001), allows identifiable patient information to be used without individual consent, in specific circumstances where anonymised information may not be sufficient and obtaining patient consent may not be practicable.[278] Researchers must apply to the Ethics and Confidentiality Committee (ECC) which oversees this aspect of the NIGB's remit.

Confidentiality

Data are defined as confidential if they may be personally identifiable. Confidential information is subject to the provisions of the Data Protection Act 1998.

4.5.1 HES data access and approval

The Dr Foster Unit (DFU) at Imperial College London is within the Department of Primary Care and Public Health and is principally funded by a research grant from Dr Foster Intelligence (DFI), an independent healthcare information company and joint venture with the NHS Information Centre. The DFU holds HES data consisting of inpatient, outpatient and A&E data for England which is deemed confidential as individual level information such as postcode, NHS number and date of birth are held, from which it may be possible to identify individuals. The unit has permission from the NIGB under Section 251 of the NHS Act 2006 (formerly Section 60 approval from the Patient Information Advisory Group) to hold such confidential data and has ethical approval to use them for research and measuring quality of delivery of healthcare, from the South East Ethics Research Committee.

Through the DFU I have been granted access to HES data, in accordance with their non-disclosure and confidentiality agreements and standard terms of data protection.

4.5.2 GPRD data access and approval

I submitted a protocol outlining the study proposal described in the next chapter, which was approved by the Independent Scientific Advisory Committee for the MHRA who review all research proposals for the use of the GPRD. Access to the data was granted free of charge under the previous Medical Research Council (MRC) license scheme with the GPRD.

4.5.3 NNRD data access and approval

In 2010 the NDAU was granted National Research Ethics approval to establish the NNRD.[279] As part of the Medicines for Neonates programme of work (of which this PhD research forms one component) we applied to the NIGB for approval to link the NNRD managed by the NDAU to national HES data at an individual level.

Provisional approval was received in December 2010, subject to clarification and amendments to some aspects of the application, such as providing further information to parents about the potential uses of their baby's data and to facilitate a mechanism to permit dissent to be included. Final NIGB approval was granted in January 2012. Following this, an application to receive HES data on site at the NDAU, with personal identifiers to permit linkage to the NNRD, was submitted to the NHS Information Centre in February 2012. This application was approved and a HES data extract provided to the NDAU at the end of August 2012.

4.6 Summary of chapter

This chapter has provided an overview of the electronic health record databases that I have access to and will be used to answer the research questions in this thesis, outlined earlier in chapter 2.0. Specifically, HES, G/CPRD and the NNRD databases have been described, in terms of the history and purpose of each dataset, the types of information they contain and what these data are used for. This chapter also presents important research governance information regarding the access and approvals I have for using these data for research.

5.0 Methods: Creating birth cohorts using child health records

Chapter overview

In this chapter the processes used to create birth cohorts using routine healthcare records are described. First, I explain how RSV-associated admission rates were examined. Then I describe in detail the novel application I developed of using HES data to create a cohort, by linking individual birth records to subsequent hospital admission records and the feasibility of doing this. The coverage and completeness of recording in HES birth records is discussed and their validity tested by comparing to birth registration data collected by the ONS. An approach for minimising the effect of variable data completion between hospitals is then presented, with sensitivity analyses to determine whether this is a valid solution. I then describe how GP records were obtained and can also be linked over time to develop a cohort. Finally I describe how cohorts of children born in 2003 were developed using both the GPRD and HES datasets, to examine the impact of an episode of bronchiolitis on subsequent consultations and admissions for further respiratory diagnoses.

5.1 RSV-associated admission rates

In addition to developing a birth cohort of infants born in English hospitals and examining which factors increase their risk of bronchiolitis admission, I decided it would be useful to obtain some background information about underlying admission rates for RSV-associated illness among infants, over time. This provides important contextual information as secular trends in RTI admissions and changing thresholds for childhood admissions may impact on RSV admission rates. Examining trends over time will provide useful data to determine the robustness of the bronchiolitis admission rates I calculated.

All emergency admissions to hospital among infants under 1 year, from 1997/98 to 2010/11, with a primary diagnosis of RSV-associated illness were identified using the HES database. This time period was chosen since these were the earliest and latest years for which these admissions data were available. Following consultation with general practitioners, a paediatric infectious disease consultant and clinicians at the Health Protection Agency, a list of ICD-10 codes for RSV-associated illnesses was identified and these were categorised into the following groups:

Specific RSV diagnosis

J205 Acute bronchitis due to RSV

J210 Acute bronchiolitis due to RSV

J121 Respiratory Syncytial Virus pneumonia

Unspecified bronchiolitis diagnosis

J209 Acute bronchitis unspecified

J219 Acute bronchiolitis unspecified

Unspecified viral pneumonia diagnosis

J128 Viral pneumonia unspecified

J129 Bronchopneumonia unspecified

J180 Lobar pneumonia unspecified

J189 Pneumonia unspecified

J22 Unspecified acute lower respiratory infection

Annual admission rates were then calculated by dividing the total the number of admissions in each of the three groups in a given year, by the estimated number of live births in that year, using mid-year population estimates obtained from the ONS.

5.2 Creating a birth cohort using HES records

A birth cohort is a longitudinal study design where individuals who are born within a defined period of time and location are followed up over time. Among the group the incidence of one or more outcome is measured and information on exposures to different factors is recorded across the time period. In this thesis I will be using routine data to carry out epidemiological research by developing birth cohorts of individual patient records linked across time.

5.2.1 Linkage across time to produce a HES cohort

The processing of HES birth records described for the first time here can be used to develop cohorts of children with the potential for long-term follow-up. Using a unique personal identifier, individual birth records can be linked across time to subsequent admission records, so that longitudinal information on severe episodes of illness requiring hospitalisation can be provided. The unique identifier used in HES ("HESID") is derived by matching records for the same patient using a combination of NHS number and local patient identifier, plus the patient's date of birth, sex and postcode.[242]

As well as the completeness of recording and quality of coding, it is important when linking records across time to consider the issue of cohort attrition, because although an individual may be born in an NHS hospital in England and therefore included in a HES birth cohort, migration, deaths and admissions outside of NHS hospitals in England are difficult to identify without individual patient consent and follow-up. I will now outline the key data processing and methodological issues required to use HES birth records to develop a birth cohort.

5.2.2 Processing and methodological issues using HES maternity data

Identifying birth episodes

The first step in developing a birth cohort study using HES was identifying all individual birth episodes within a given year. To identify birth episodes, I used the "admimeth" variable which contains a code recording how the patient was admitted to hospital.[245] This field was used to select all records with an admission method coded as 82 (other: babies born in health care provider), or 83 (other: babies born outside the health care provider, except when born at home as intended).

Removing duplicate birth episodes

The next step was identifying any duplicate birth episodes (a typical but limited problem with administrative data) using the unique personal identifier ("HESID"). I looked specifically at the first episode within the given spell (i.e. the initial birth episode itself) using the "epiorder"

field which contains the number of the episode within the current spell.[245] In the dataset from financial year 2007/08, 5600 individuals were identified with more than one birth episode recorded, of which 5545 had two and 55 had more than two birth episodes recorded. Individuals with >2 birth episodes were excluded from the cohort because it would be almost impossible to decipher which birth record ought to be retained when comparing more than two duplicates. Many records from individuals with two birth episodes were identical matches (2750 individuals), having two exact duplicate birth episodes with identical information recorded in all fields. Where this was the case, one of the identical records was deleted.

Among the remaining individuals with two birth episodes, the records did not contain matching information. It is very difficult to determine which of two birth records with the same unique identifier is likely to be the true record of that individual's birth. In previous studies using HES and similar databases, the common approach has been to retain only the first observation recorded. Upon observation, it became clear that the basic demographic information for each pair of birth episodes tended to be common to both and for most it was only the diagnostic fields that differed. Consequently, records for individuals with duplicate birth episodes were compared using recording of *diagnostic information* and only the birth episode with the most diagnostic information recorded (number of non-empty diagnosis fields) was retained in the cohort. It was not possible to determine whether this approach is better than randomly retaining one birth episode. However the record with the most diagnostic information might be the more accurate of the two because it is common for hospitals to resubmit data after they have carried out all their diagnosis coding more thoroughly (their first submissions may contain only very basic diagnostic information). Therefore, records can appear as duplicates if hospitals fail to report them as re-submissions to the Secondary Uses Service (SUS, the NHS data repository). Consequently, for an individual with two birth episodes, the one with more diagnostic information may be the more accurate record. Handling duplicate records is a common, challenging problem when using routine data such as HES and requires careful consideration of the study context.

Summarising information from additional episodes in a birth spell

Babies can have more than one episode of care within their birth admission, for example if a baby receives specialist care from a different consultant, is transferred between hospitals or is admitted to a neonatal unit. These additional episodes occur within the same birth spell but, where the initial birth event would have an "epiorder" value of 1, subsequent episodes in the birth spell have an "epiorder" value >1. To facilitate one-to-many linkage to subsequent hospital admission records, it is easiest to develop a dataset consisting of one observation (or row) per individual. 7410 individuals in our 2007/08 birth cohort had >1 episode in their

birth spell. To simplify the dataset to consist of only one birth record per patient, key information such as any premature or congenital anomaly diagnoses was summarised using flags. These were then linked back into the original birth episode and subsequent episodes in the birth spell were dropped.

Cleaning variables

A range of exclusion criteria were developed to clean key variable fields and examine the quality of coding. The Care Quality Commission (CQC) conducted a review exploring quality indicator specifications, used to assess the quality of HES maternity data from 2009/10.[280;281] The criteria identified within the CQC review[280] and HES inpatient cleaning rules[282] were combined and applied to the HES birth fields to ensure suspicious data and invalid records were removed. In addition, any stillbirths that were recorded were identified (using the birth status and discharge method fields) and removed from the final cohort of live births.

Identical baby tail information recorded for each baby can be found in their mother's delivery record. Maternity systems can record up to 9 birth tails for each delivery, allowing information from multiple births to appear in the mother's delivery record. In some instances, I found baby's information was not recorded in the first field of a given variable. For example, if a baby was the second twin, their gestational age at birth ("gestat") may have appeared in the second field ("gestat_2") with the first field blank ("gestat_1") because in the mother's delivery record this would contain the first twin's gestation. To simplify analyses, I condensed individual birth records so that only one field for each variable existed. So for the above example, I transferred the relevant information for the gestation variable from "gestat_2" into "gestat_1" and then removed all additional fields for that variable (i.e. "gestat_2" to "gestat_9").

5.2.3 Coverage and completeness of recording

Each year, the NHS Information Centre publish quality reports examining coverage of the data submitted by NHS hospitals to HES.[256;283] These reports have repeatedly found that for many hospitals completeness of mandatory fields relating to births is often inadequate.[256;283] I examined the completeness of recording for each baby tail field over five recent years and also compared the numbers with ONS birth registrations (see 5.2.4). Recording of most fields was fairly consistent between years until 2008/09 when overall completeness of recording slightly improved (Table 1). The values presented differ from those published by HES online because of the extensive data cleaning processes I applied, which resulted in removal of some invalid birth records (such as those with an invalid date of birth). Among the 2008/09 dataset, birth weight and gestational age were recorded for 69% and 65% of births respectively.

The completeness of recording of information in hospital birth records is highly variable between hospitals but is generally improving. The proportion of missing or unknown data in key birth record fields has been decreasing annually, such as gestational age (from 46.2% in 2005/06 to 34.6% in 2008/09) and birth weight (from 43.9% in 2005/06 to 31.3% in 2008/09).

Table 1: Completeness of recording of baby tail fields in HES birth records (2005/06 to 2009/010).

Baby tail fields in HES birth records (field name)	% missing or unknown 2005/06	% missing or unknown 2006/07	% missing or unknown 2007/08	% missing or unknown 2008/09	% missing or unknown 2009/10	P value
Anaesthetic given during labour/delivery (delpren)	41.9	41.8	44.8	29.6	16.5	<0.001
Anaesthetic given post-labour/delivery (delposn)	48.1	46.0	49.6	34.8	21.6	<0.001
Antenatal days of stay (antedur) (derived field)	0.2	0.2	0.2	0.1	0.1	<0.001
Baby's age in days (neodur) (derived field)	0.1	0.0	0.0	0.0	0.0	<0.001
Birth order (birorder)	33.9	36.9	39.4	24.9	13.7	<0.001
Birth weight (birweit)	43.9	47.1	50.1	31.3	16.9	<0.001
Delivery place change reason (delchang)	45.2	45.8	47.4	34.6	21.7	<0.001
Delivery method (delmeth)	35.1	35.8	44.3	30.6	14.9	<0.001
Delivery place (actual) (delplace)	44.0	46.8	57.0	41.3	17.9	<0.001
Delivery place (intended) (delinten)	41.3	42.7	43.6	30.2	14.9	<0.001
First antenatal assessment date (anasdate)	41.7	44.3	44.6	34.7	20.4	<0.001
Gestation (weeks) at first antenatal assessment (anagest)	54.5	63.9	55.2	45.6	28.3	<0.001
Length of gestation (gestat)	46.2	54.2	48.0	34.6	18.1	<0.001
Birth status (birstat)	43.9	47.0	48.0	32.9	16.2	<0.001
Labour / delivery onset method (delonset)	36.2	37.7	41.1	25.5	11.5	<0.001
Mother's age at delivery (matage)	42.4	43.3	43.0	34.5	30.5	<0.001
Neonatal level of care (neocare)	16.1	16.0	17.1	18.4	12.4	<0.001
Number of babies (numbaby)	31.8	33.3	36.1	23.6	11.9	<0.001
Resuscitation method (biresus)	44.2	45.3	48.0	34.2	21.1	<0.001
Status of person conducting delivery (delstat)	38.9	42.6	48.4	33.9	19.1	<0.001
Total number of births*	554521	566749	575493	589684	603786	-

*Total number of births in cohort after removal of duplicate episodes and completion of data cleaning processes as described earlier.

5.2.4 Comparison with data from ONS birth registrations

Overall, the HES cohort captured 87% of all live births recorded by the ONS in England during the time period (Table 2). There are various explanations for this difference. An estimated 97% of all live births in England occur in NHS hospitals. Less than 1% of births annually occur in non-NHS establishments and between 2-3% occur at home, according to ONS birth registration data. Those occurring in private hospitals or at home may not be recorded in HES. In addition, because HES contains discharge records, some births which occurred within the cohort year but were discharged from hospital after this period would not be captured in our dataset. Most of the births that are not captured within this HES cohort may have been excluded as a result of the quality criteria I applied to remove records with invalid information reported in key fields, or due to the incomplete coverage issues described earlier.

Table 2: Comparison of live birth recording in HES with Office for National Statistics reporting (2007/08).

Month & year	Number of live births reported in ONS data	Number of live births in created HES cohort data	Percentage of births reported by ONS that were recorded in the HES cohort
Apr-07	51456	44841	87.1
May-07	55863	49011	87.7
Jun-07	53664	46971	87.5
Jul-07	57158	49614	86.8
Aug-07	57874	50048	86.5
Sep-07	57189	49575	86.7
Oct-07	57092	49843	87.3
Nov-07	54195	46666	86.1
Dec-07	54993	47712	86.8
Jan-08	56621	49364	87.2
Feb-08	52414	45963	87.7
Mar-08	55919	45885	82.1
Total for year	664438	575493	86.6

5.2.5 Approach for minimising the effect of variable data completion between hospitals

Using a cut-off of 90% completeness of recording for key birth fields, hospitals with highly complete data were compared with the others in terms of hospital characteristics from available data, to assess how representative of all of England a birth cohort formed from babies at high-coding hospitals would be.

In the 2007/08 dataset, 55% of records had no birth weight recorded, 48% were missing gestational age, 43% had no maternal age recorded and 17% had no indication of whether the baby received any specialist neonatal care, with considerable variation in the completeness of information recorded in birth records between NHS providers. Some hospitals with many births had recorded gestational age and birth weight in less than 10% of birth records, whilst other similarly large hospitals recorded this information for >90% of births. There appears to be no correlation between the number of births occurring in a hospital and the completeness of recording of these prematurity indicators. Therefore, depending on study purpose and the exposure and outcome measures of interest, I suggest selecting birth records only from hospitals with the highest completeness of recording. This was tested by creating a 2007/08 birth cohort where I selected only birth records from hospitals where ≥ 90% of their birth records contained complete recording of key variables, birth weight and gestational age. The resulting cohort included 296,618 babies born at 71 hospitals across England.

Table 3 shows a comparison of characteristics of included (n=71) and excluded (n=85) hospitals (hospitals in the same NHS trust have been grouped together, as requested by journal peer reviewers). The mean number of births, maternity beds and access to neonatal intensive care were similar among hospitals with high and low completeness of recording of birth record information. There were no statistically significant differences between the two groups for any of the characteristics compared. This suggests that a birth cohort derived from these 71 hospitals with high levels of completeness of recording would be representative of the whole country.

Table 3: Comparison of maternity characteristics between hospitals with high and low completeness of recording in their HES birth admission records†, from financial year 2007/08.

Hospital maternity factors	Hospitals with *high* completeness (n=71)	Hospitals with *low* completeness (n=85)	P value
Mean number of births in year (SD)	3957 (2011)	3465 (1997)	0.1287
Mean number of maternity beds available (SD)	55.1 (30.3)	55.3 (26.4)	0.9649
Mean number of maternity beds reported to be occupied (SD)	35.4 (21.4)	35.2 (18.0)	0.9495
Number with access to on site neonatal intensive care (%)	52 (73)	68 (80)	0.3022
Mean number of available beds in neonatal intensive care unit (SD)	10.6 (11.7)	11.4 (10.9)	0.6595
Mean maternal age (% birth records missing this data)	28.9 (18.4)	29.0 (70.1)	0.3964
Proportion of births per hospital in most deprived Carstairs Deprivation Score quintile (% birth records missing this data)	0.472 (69.4)	0.435 (56.2)	0.6406
Proportion of births per hospital of non white British ethnicity (% birth records missing this data)	0.527 (7.1)	0.564 (5.1)	0.6438

† Low completeness of recording was defined as hospitals where <90% of their birth admission records contained complete recording of birth weight and gestational age.

* SD = standard deviation

5.2.6 Explaining these findings

The Healthcare Commission's review of maternity services found that even among larger, well-respected maternity units, information technology was poor and data collection systems inadequate.[280] One of the key recommendations from this review was that, "[hospitals] must ensure that maternity units are equipped with appropriate IT systems that comply with Connecting for Health, enabling completion of mandated national data sets and the provision of accurate and systematic data on outcomes and management information on which to plan, commission and manage the resources required for maternity care".[280] Coverage of hospital deliveries was estimated to be around 73% between 2001/02 and 2005/06, but for home deliveries the figure was just 14%.[242] As a result of these quality concerns, additional rigorous validation was carried out by the NHS Information Centre on HES maternity data, including intensive cleaning of systematic coding errors.[242] Some of the quality and coverage issues specifically affecting HES maternity data include:

- Stand-alone maternity systems in around 20 hospitals are not linked to their patient administration system (PAS), from which HES data are obtained (via the Secondary Uses Service).
- Some hospitals submit a significantly higher number of delivery episodes than birth episodes.
- Some hospitals return data on the number of birth episodes or delivery episodes but not both.
- Transfer of maternity information between systems leaves scope for data errors and shortfalls.
- Stillbirths are not reliably recorded in every hospital and are not allocated an NHS number.[229;232;256]
- Lack of a priori definitions for data variables resulting in inconsistencies in data entry.
- Use of aggregate or categorised fields rather than raw data.

5.2.7 Limitations of this work

It was not possible to determine how accurately the identifier used to deterministically link records across time is allocated to unique individuals, although it is known that the NHS Information Centre now use an improved, complex algorithm to derive "HESID", based on a combination of other identifiers as described earlier. The comparison of maternity characteristics between hospitals with high and low completeness of recording, was limited to specific fields for which we had access to information from all hospitals, such as their mean number of births, maternity beds and specialist neonatal care facilities. However, I am

confident that the comparison of these key hospital factors between hospitals with high and low completeness of recording of birth information, provides sufficient evidence that these do not significantly differ.

5.2.8 Summary

Detailed clinical information is routinely captured in hospital records in the HES database. Despite some complex data processing requirements, HES records provide a powerful source of national baseline data on birth episodes that can be used for population scale epidemiological research. These records contain valuable clinical information on prematurity, birth weight, treatments, complications and comorbidities. However, the completeness of birth information recorded in English hospitals is variable. Where key birth information such as gestational age and birth weight are missing, it may be preferable to select data only from hospitals with high levels of completeness of recording; the analysis here suggests that the results may be generalisable to all hospitals. Careful consideration of the context in which information in these records is collected and how this will be used is required. HES data can be used to develop longitudinal cohorts by linking individual birth records to subsequent hospital admission records, providing information on longer-term health outcomes among infants.

5.2.9 Birth cohort study

This 2007/08 cohort was designed to examine research question 1 and objective 1 of this thesis. As has been described in detail earlier, a birth cohort for all infants born in English NHS hospitals and discharged during a twelve month period (from 1st April 2007 to 31st March 2008) was created, using HES ID to link birth and subsequent admission records up to an infant's first birthday. Births from 2007/08 were selected since this was the most recent and complete year of HES data available to me that would permit one year of follow-up from birth, in which to examine hospital admissions in each individual's first year of life. This cohort included only records from live births and excluded infants born in hospitals (85/156) with poor recording (<90% complete) of key indicators (birth weight and gestational age), to enable us to group infants into term and preterm categories.[257] Sensitivity analyses described earlier showed no statistically significant differences between included and excluded hospitals, so a study using only those with high completeness of recording of gestational age and birth weight should be generalisable.[257] Pre-existing linkage of these records to ONS mortality records allowed us to identify any deaths among the birth cohort during their first year of life, including out-of-hospital deaths.

Outcome and exposure measures

Information from the detailed review of current literature (described in chapter 1) was used to decide which prior exposures were clinically relevant to examine their association with a bronchiolitis hospital admission. Diagnostic information recorded in individual birth records and any subsequent hospital admission records in the first year of life, were used to flag whether infants had any of these exposures/potential risk factors for severe RSV infection (Box 2), using specific ICD-10 codes or larger subgroups of codes based on the Agency for Healthcare Quality and Research's Clinical Classification System (CCS).[284] CCS is a categorisation scheme which aggregates individual ICD-10 codes into broader diagnosis groups to facilitate reporting and statistical analyses.[284] Only risk factors that were clinically diagnosed were examined, since information on other environmental risk factors such as exposure to passive smoking, is not available in these medical records.

Infants were considered preterm if their gestational age at birth was less than 37 weeks (Box 2), in accordance with the WHO definition of premature birth.[1] Infants missing gestational age were assumed to be born at full term. The reasons justifying this decision were three-fold. Firstly because sensitivity analyses revealed this group of infants with unknown gestational age had a similar median length of stay at birth and similarly low neonatal unit admission rates to those with a gestational age of ≥37 weeks. Secondly, since infants born preterm are typically less healthy and may require clinical interventions, their gestational age at birth is perhaps more likely to be recorded that those born at term, who may be discharged quickly. It is plausible therefore that the majority of infants missing gestational age are more likely to have been born at term than preterm infants requiring further hospital care. The final justification for considering infants with missing gestational age to be born at term, was that the proportion of preterm births identified in the cohort was 7.5%, which is very similar to what one would expect based on ONS birth registration data which reports that 1 in 13 babies in England and Wales are born preterm.[90;91;285]

Infants admitted as an emergency with a primary diagnosis of acute bronchiolitis were identified using the 'J21' ICD-10 codes ('J210' - acute bronchiolitis due to respiratory syncytial virus, 'J218' - acute bronchiolitis due to other specified organisms and 'J219' - acute bronchiolitis, unspecified). Most bronchiolitis admissions were coded with unspecific aetiology and approximately a third were coded as being due to RSV. All bronchiolitis codes were therefore subsequently grouped into a single "RSV bronchiolitis" category, since RSV is the commonest cause of bronchiolitis, but specific microbiological diagnoses are poorly coded in HES and testing for RSV is not standard clinical practice for bronchiolitis management. Bronchiolitis admissions in children over 1 year were excluded because of the uncertain nature of this clinical diagnosis in older infants. Infants were defined as all

individuals under 1 year of age. Age at bronchiolitis admission was examined and the median length of stay (LOS) for bronchiolitis admissions was calculated to provide a proxy measure of severity of illness.

Box 2: ICD-10 codes and CCS groupings used to identify at-risk groups.

Prematurity

Babies were categorised as born prematurely if their gestational age at birth was <37 weeks.[1]

If a birth record had no gestational age recorded, i.e. premature status was unknown, then they were assumed to be not premature (justified on the basis that infants in the unknown group had similarly low ICU admission rates and short length of stay at birth, to infants in the group known to be born at term.)

Immunodeficiency: CCS group 57 – Immunity disorders (This includes ICD-10 codes D80, D81, D82, D83, D84 and D89 which includes diagnoses such as hypogammaglobulinemia and severe combined immunodeficiency.)

Cystic fibrosis: CCS group 56 – Cystic fibrosis (This includes ICD-10 codes under E84.)

Chronic lung disease: ICD-10 codes P27 – Chronic respiratory disease originating in the perinatal period and P28 – Other chronic respiratory diseases originating in the perinatal period.

Congenital heart diseases: CCS group 213 (This includes ICD-10 codes Q20, Q21, Q22, Q23, Q24, Q25, Q26, Q27, Q28.)

Nervous system congenital anomalies: CCS group 216 (This includes ICD-10 codes Q00 to Q07 which incorporates conditions such as spina bifida, anencephaly and other congenital malformations of the nervous system.)

Other congenital anomalies & perinatal conditions: CCS groups 224 and 217 (This includes a broad range of congenital anomalies and perinatal conditions with ICD-10 P- and Q- codes, excluding those included within other definitions listed above, such as codes for chronic lung disease.)

Down's syndrome: ICD-10 code Q90

Cerebral palsy: ICD-10 code G80

5.2.10 Statistical analyses

The absolute risk of a bronchiolitis admission was calculated among infants with and without risk factors for severe RSV infection, including preterm infants and infants with a clinical diagnosis of cystic fibrosis, congenital heart disease, nervous system congenital anomalies, other congenital anomalies, immunodeficiency, cerebral palsy or Down's syndrome (case definitions and codes in Box 2). Infants without a particular risk factor condition were considered "healthy". Associated 95% confidence intervals (CI) were calculated using Poisson approximation. The relative risk (RR) of a bronchiolitis admission was calculated, with associated 95% CI for infants in each individual risk group, by comparing with the baseline group of infants without the particular risk factor. Infants could belong to more than 1 of these risk groups, so this potential confounding was controlled for using Poisson multiple regression models to calculate the adjusted RR of bronchiolitis admission for infants in each risk group. Confounders controlled for in these analyses were selected a priori, based on a detailed literature review of risk factors for hospital admission with bronchiolitis. In addition to each of the diagnostic categories described earlier (see box 2) candidate covariates in each model included sex, month of birth, deprivation (using Carstairs score quintile which is based on patient postcode), age at bronchiolitis admission and multiple birth. Covariates were selected for inclusion in the multivariable model if the p value was below the 0.05 significance level. All the covariates listed were found to be significant and retained in the final model.

Two-way interactions of variables deemed to be of clinical interest were explored (following consultation with clinicians), including between sex and deprivation (measured in Carstairs quintiles). No interactions were found to be statistically significant. Diagnostic plots of residuals against fitted values were produced and Pearson goodness of fit tests carried out to assess the validity of the model. These showed no evidence of outliers or overdispersion.

Clustering of observations within hospitals can lead to confidence intervals that are too narrow and to coefficient bias. Hierarchical modelling techniques were initially conducted to adjust for clustering by hospital, but this resulted in intra-class correlation coefficients (ICC) with values less than 5% (a standard threshold above which adjustment for clustering is recommended) and similar coefficients and standard errors to those obtained without this clustering adjustment. Hence, it was decided that the modelling would not include adjustment for clustering by hospital. Indeed, different methods of adjustment are based on different assumptions and there is often disagreement about which methods are most appropriate to use in different contexts.[286]

Further work could involve comparison of the baby tail fields from birth records, with the corresponding maternal delivery record and baby tail, to determine the extent to which the information recorded in these correlates. Linkage to other sources of data on potential confounders (for example environmental risk factors such as exposure to passive smoking) would strengthen the analyses as these could then be controlled for in the modelling process.

Chi Squared tests were used to test for significant differences between proportions and Mann-Whitney tests were used to compare median values for non-normal data. Data were analysed using the SAS 9.2 software package (SAS Institute, Cary, North Carolina, USA).

5.3 Creating a birth cohort using records from the GPRD

Data for this study were obtained from the General Practice Research Database (GPRD), a computerised database of anonymised longitudinal medical records from primary care in the UK (described in more detail earlier in chapter 4.3). The approach described below for creating a birth cohort, was used to address the second and third research questions and corresponding second and third objectives of this thesis:

Objective 2 – To estimate the incidence of RSV bronchiolitis among infants presenting in primary care and to describe how it is being managed by general practitioners, in relation to national guidance (Study 2 – Creating a nationally representative birth cohort using GPRD patient records).

Objective 3 – To determine the longer-term effects of RSV bronchiolitis on subsequent respiratory health in early childhood (Study 3 - GPRD and HES long-term follow-up studies).

Research question 2 – What is the incidence of acute bronchiolitis among infants presenting to general practitioners and how is it being managed in the community?

Research question 3 – What is the long-term impact of RSV bronchiolitis on subsequent respiratory health in early childhood?

Description of study cohort

82861 infants born in 2003 were identified, registered across 600 GPRD general practices in the UK (these were all "Up To Standard" practices for which the data are deemed to be of suitable research quality).[258] Complete electronic primary care records for the first year of life were obtained for each of these infants, to create a retrospective cohort from birth to age

1 year, using a scientific workflow infrastructure.[287] These workflow infrastructures have been developed specifically to assist with the challenges of mining large databases of primary care records, particularly to ensure reproducibility of methods used.[287] Scientific workflows are a method for representing data processing operations, recording all steps taken in the cleaning of data, coding definitions used and variables selected for inclusion in final datasets, thereby making analysis of large databases more manageable and scalable.[287] Individuals who died or transferred to a different general practice before their first birthday were then excluded (n=1885), so the birth cohort with complete follow-up for their first year of life consisted of 80976 infants.

Outcome and exposure measures

Following consultation with GPs and a consultant paediatrician, a diagnosis code list was developed to identify consultations where a diagnosis of acute bronchiolitis was made (Table 4).

Table 4: GPRD codes for acute bronchiolitis.

GPRD Medcode	Description
1019	Acute bronchiolitis
18451	Acute bronchiolitis due to respiratory syncytial virus
17917	Acute bronchiolitis NOS
66228	Acute bronchiolitis due to other specified organisms

This dataset did not contain the free text components of GP records, so the definition of a bronchiolitis consultation was based solely on the clinical diagnostic codes recorded in the infants' medical records.

For comparison I also examined the impact of expanding our conservative definition of acute bronchiolitis, to include consultations in the first year of life where other diagnosis codes were recorded which are also likely to represent infants with bronchiolitis, who may have been given a less specific or symptomatic diagnosis code. For this wider definition, consultations in the first year of life with the following diagnosis codes were identified, as deemed appropriate after detailed discussion with the GPs and paediatricians I consulted (Table 5):

Table 5: GPRD code list for wider bronchiolitis definition of consultations in first year of life.

GPRD Medcode	Description
1019	Acute bronchiolitis
18451	Acute bronchiolitis due to respiratory syncytial virus
17917	Acute bronchiolitis NOS
66228	Acute bronchiolitis due to other specified organisms
92	Cough
1273	C/O cough
292	Chesty cough
68	Chest infection
2581	Chest infection NOS
2891	Wheezing symptom
173	Wheezing
4836	Nocturnal cough/wheeze
5861	O/E – expiratory wheeze
2210	[D] Wheezing
74908	Viral wheeze
83478	Viral induced wheeze
78	Asthma
81	Asthma monitoring
10043	Asthma annual review
13176	Asthma follow-up
7378	Asthma management plan given
185	Acute exacerbation of asthma
719	H/O: asthma
5515	Seen in asthma clinic
13064	Asthma severity
19539	Asthma monitoring check done
16070	Asthma NOS
10274	Asthma medication review
13173	Asthma not disturbing sleep
232	Asthma attack
1555	Bronchial asthma
19520	Asthma treatment compliance satisfactory
26501	Asthma never causes daytime symptoms
38143	Asthma never disturbs sleep
13174	Asthma not limiting activities
4442	Asthma unspecified
26504	Asthma never restricts exercise
11022	Asthma trigger
10318	Suspected asthma
16785	Asthma control step 1
16667	Asthma control step 2
18224	Asthma control step 3

26861	Asthma sometimes restricts exercise
233	Severe asthma attack
26503	Asthma causes daytime symptoms most days
7416	Asthma disturbing sleep
24884	Asthma causes daytime symptoms 1 to 2 times per week
7229	Asthma prophylactic medication used
42824	Asthma daytime symptoms
31225	Asthma causes daytime symptoms 1 to 2 times per month
1208	Childhood asthma
3018	Mild asthma
13065	Moderate asthma
25181	Asthma restricts exercise
13175	Asthma disturbs sleep frequently
18223	Step down change in asthma management plan
5138	Patient in asthma study
19167	Asthma monitoring by nurse
25791	Asthma clinical management plan
45073	Intrinsic asthma NOS
98185	Asthma control test
18323	Intrinsic asthma with asthma attack
24506	Further asthma - drug prevent.
30382	Asthma monitoring admin NOS
19519	Asthma treatment compliance unsatisfactory
10487	Asthma - currently active
41554	Asthma monitor offer default
31167	Asthma night-time symptoms
3458	Occasional asthma
30815	Asthma causing night waking
5267	Intrinsic asthma
38144	Asthma limits walking up hills or stairs
11370	Asthma confirmed
9552	Change in asthma management plan
3665	Late onset asthma
20886	Asthma control step 4
38146	Asthma disturbs sleep weekly
18763	Referral to asthma clinic
20860	Asthma control step 5
25796	Mixed asthma
45782	Extrinsic asthma NOS
23481	Asthma – cardiac
27926	Extrinsic asthma with status asthmaticus
29325	Intrinsic asthma without status asthmaticus
30458	Asthma monitoring by doctor
4892	Status asthmaticus NOS
8355	Asthma monitored
47337	Asthma accident and emergency attendance since last visit

6707	Extrinsic asthma with asthma attack
39570	Asthma causes night symptoms 1 to 2 times per month
3366	Severe asthma
26506	Asthma severely restricts exercise
41020	Absent from work or school due to asthma
12987	Late-onset asthma
8335	Asthma attack NOS
24479	Emergency asthma admission since last appointment
14777	Extrinsic asthma without status asthmaticus

The age of an infant at the time of the bronchiolitis consultation was estimated by selecting the mid-point of the infant's month of birth as their date of birth (since only month and year of birth were available in the dataset). The week of the year that a bronchiolitis consultation took place was identified using the consultation date, to examine the seasonality of bronchiolitis consultations and determine whether the burden of contacts followed the winter peak pattern typical of RSV infection.[39] Any records of prescriptions associated with a bronchiolitis consultation were also identified (using the consultation date and consultation ID fields). The therapeutic codes were looked up on the GPRD medical codes browser disc which was supplied with the dataset.

Primary care consultation patterns during the first year of life among the whole birth cohort were examined, specifically comparing infants with no bronchiolitis consultations with those with a bronchiolitis consultation. To examine the most common symptoms and diagnoses among infants presenting in primary care, consultations relating to immunisations, administrative tasks, routine health checks and health promotion activities were excluded. The remaining diagnostic and symptomatic medical codes were then grouped into similar diagnosis and symptom groups, to examine the most common diagnoses among infants presenting to their GP in the first year of life (Table 6).

The relative risk (RR) and associated 95% confidence intervals of any respiratory consultation in the first year of life were calculated, among infants with a history of bronchiolitis compared with those without. The most common diagnostic and symptomatic codes recorded during any GP consultations that occurred in the 2 weeks before and 2 weeks after a bronchiolitis consultation were also identified to provide an insight into clinical events surrounding a bronchiolitis diagnosis.

Table 6: Coded diagnosis groups for consultations in the first year of life.

Diagnosis groups of consultations in 1st year of life	GPRD Medcode	GPRD Medcode description
Respiratory tract infections and cough or wheezing symptoms	1019	Acute bronchiolitis
	2637	Upper respiratory infection
	76	Upper respiratory infection NOS
	92	Cough
	1273	C/O cough
	6294	Acute upper respiratory tract infection
	2581	Chest infection NOS
	68	Chest infection
	267	Otitis Media NOS
	292	Chesty cough
	1246	Acute coryza
	2891	Wheezing symptom
	173	Wheezing
	4836	Nocturnal cough/wheeze
	78	Asthma
	7818	Neonatal snuffles
	1142	Croup
Eczema, dermatitis, rashes and other skin conditions	4703	C/O rash
	610	Infantile eczema
	1741	Atopic dermatitis/eczema
	230	Eczema NOS
	1685	Nappy rash
	5041	Non-specific viral rash
	304	Thrush
Diarrhoea, vomiting and other gastro-intestinal symptoms	176	Vomiting
	5134	Diarrhoea symptoms
	192	Diarrhoea
	139	Gastroenteritis
	8416	Constipation
	2182	Diarrhoea & vomiting symptoms
	2004	Constipation symptoms
Other infections	1864	Acute conjunctivitis
	6017	Sticky eye
	1041	Viral infection NOS
	145	Chickenpox varicella
	36525	Infectious disease: control
	196	Oral thrush

Statistical analyses

The bronchiolitis consultation rate and associated 95% confidence intervals (95% CI) were estimated for infants under the age of 1 year by dividing the total number of bronchiolitis consultations by the total cohort population. The mean and standard deviation (SD) for the age in months at which infants were diagnosed with bronchiolitis was calculated. Two-sample z tests were used to compare the proportion of primary care consultations for

different diagnoses, among infants with and without a history of bronchiolitis in their first year. All statistical analyses were carried out using STATA/MP version 10.0.

5.4 Creating birth cohorts with follow-up into early childhood

5.4.1 Developing a cohort with longer follow-up after admission to hospital with bronchiolitis, using HES

To address the third research question and objective described in chapter 2, the approach described above for creating a birth cohort using HES records was applied to earlier data to provide a sufficient number of years of follow-up. Data from all births in financial year 2003/04 were obtained and then the same data cleaning and processing as for the 2007/08 dataset was applied. I will now describe this 2003 cohort in more detail.

Study cohort

All births in NHS hospitals in England that were born and discharged during a twelve month period (from 1st April 2003 to 31st March 2004) were identified. Birth records from each individual born in this time period were then linked to any subsequent hospital admissions up to their 5st birthday in 2008/09.

Infants were then grouped according to whether they were at "high" or "low" risk of bronchiolitis, based on clinical risk factors and diagnoses identified within their birth records or from diagnoses from any other admissions during the neonatal period. Information from an in-depth review of the literature and consultation with paediatricians was used to determine which risk factors were used to classify infants as "high-risk" (see code list in Box 2).

Outcome and exposure measures

For infants in both groups, all hospital admissions during their first year of life with a primary diagnosis of acute bronchiolitis were identified (using the J21 ICD-10 codes described in the earlier cohort study). As described earlier, most bronchiolitis admissions were coded with unspecific aetiology and approximately a third were coded as being due to RSV. All bronchiolitis codes were therefore grouped into a single "RSV bronchiolitis" category, again justified on the basis that RSV is the commonest cause of bronchiolitis, but specific microbiological diagnoses are poorly coded in HES and testing for RSV is not standard clinical practice for bronchiolitis management. Bronchiolitis admissions in children over 1 year were excluded because of the uncertain nature of this clinical diagnosis in older infants.

Following an index admission for bronchiolitis, any subsequent admissions up to the age of 5 years, with a primary respiratory diagnosis, were then identified. Specifically, subsequent respiratory admissions with the following diagnoses were flagged using these ICD-10 codes:

Asthma:

J45 – asthma

J46 – status asthmaticus (acute severe asthma)

Lower respiratory tract infections (LRTI):

J12 – Viral pneumonia, not elsewhere classified

J13 – Pneumonia due to Streptococcus pneumonia

J14 – Pneumonia due to Haemophilus influenza

J15 – Bacterial pneumonia, not elsewhere classified

J16 – Pneumonia due to other infectious organisms, not elsewhere classified

J17 – Pneumonia in diseases classified elsewhere

J18 – Pneumonia, organism unspecified

J20 – acute bronchitis

J22 – unspecified acute lower respiratory infection.

Upper respiratory tract infections (URTI):

J00 - Acute nasopharyngitis

J01 – Acute sinusitis

J02 – Acute pharyngitis

J03 – Acute tonsillitis

J04 – Acute laryngitis and tacheitis

J05 – Acute obstructive laryngitis (croup) and epiglottitis

J06 – Acute upper respiratory infections of multiple and unspecified sites

J09 – Influenza due to identified avian influenza virus

J10 – Influenza due to other identified influenza virus

J11 – Influenza, virus not identified

Wheezing:

R062 – wheezing

For comparison, infants at both high and low risk of bronchiolitis who did not then have any record of a bronchiolitis admission, were also followed-up to age 5 years. In these groups any admissions for the above respiratory diagnoses from age 6 months onwards were also identified. As these infants have no bronchiolitis admission to use as an index date, subsequent respiratory admissions were identified from age 6 months onwards as this is the typical age at which the bronchiolitis admissions from the comparison group occurred.

Statistical analyses

The relative risk (RR) of a subsequent respiratory admission in the first 5 years of life was calculated, with associated 95% CI, for children with a history of prior bronchiolitis admission compared with those without. These RRs were calculated separately for children at high- and low-risk of severe RSV infection. Associated 95% CI were calculated using Poisson approximation. The attributable risk (AR) and population AR in these groups were also calculated to examine the difference in subsequent respiratory admission rates between children with and without a history of bronchiolitis admission in infancy. Data were analysed using the SAS 9.2 software package (SAS Institute, Cary, North Carolina, USA).

5.4.2 Developing a cohort with longer follow-up after a bronchiolitis consultation in infancy, using the GPRD

As described earlier in this chapter (see 5.3), GP records from the GPRD from infants born in 2003 were selected to develop a cohort. The year 2003 was chosen specifically to provide a cohort with a sufficient number of years of follow-up to address the third research question and objective of this thesis (outlined in chapter 2.0). So selecting data from the year 2003 meant I was able to examine consultations for asthma and wheezing in the first 7 years of life, comparing rates between infants with and without prior history of a bronchiolitis episode in the first year of life. The findings from this study are presented in chapter 8.2.2.

Outcome and exposure measures

Following consultation with GPs and a consultant paediatrician, a diagnosis code list was developed to identify consultations where a diagnosis of asthma or wheezing was made. The GPRD medical codes used to identify subsequent consultations for asthma and wheezing are listed within Table 5. Consultations were included if a child was aged between 1 and 7 years when the diagnosis was made. In the 2003/04 HES cohort with follow-up to age 5 years described earlier, subsequent admissions for LRTI and URTI were examined. It was not deemed feasible to explore subsequent GP consultations for LRTI and URTI because of the difficulties associated with trying to define a consultation for RTI. There are hundreds of possible READ codes that can be used to denote a RTI consultation, including many vague, symptomatic codes which are not specific enough to reliably attribute to being lower or upper RTIs. Although both this GPRD cohort and the 2003/04 HES cohort relate to children born at a similar time, there was no way of idenitfying whether there was any overlap between the cohorts to determine if any children were included in both.

Statistical analyses

The relative risks (RR) of a subsequent asthma or wheezing consultations in the first 7 years of life were calculated, with associated 95% CI, for children with a history of prior bronchiolitis consultation compared with those without. The attributable risk (AR) and population AR in these groups were also calculated to examine the difference in subsequent asthma and wheezing consultation rates between children with and without a history of bronchiolitis consultation in infancy and the reduction in subsequent respiratory consultations that would be expected in a population unexposed to bronchiolitis. All statistical analyses were carried out using STATA/MP version 10.0.

5.5 Summary of chapter

In this chapter the methods I have used to develop birth cohort studies using routine health records have been described in detail, using both hospital and GP records. These contain valuable clinical information that can be utilised to create large, population-based studies for epidemiological research. Longitudinal cohorts can be developed by linking records belonging to the same individual over time, providing information on longer-term health outcomes among children.

I have tested the feasibility of linking HES birth records to subsequent admission records to develop a longitudinal cohort with long-term follow-up. HES birth records contain valuable clinical information on prematurity, birth weight, complications and comorbidities, but the completeness of recording of this birth information is highly variable between different hospitals. Where key birth information such as birth weight and gestational age are missing, it may be preferable to select data only from hospitals with high levels of completeness of recording. The analysis presented here suggests that a cohort developed in this way is representative of births across England. The later part of this chapter described how GP records were obtained and can also be linked over time to develop a cohort of individuals with detailed clinical information about the primary care they have received, including diagnoses, prescriptions and vaccinations. Finally I have explained how two cohorts of infants born in 2003, one from the GPRD and one from HES, were developed to examine subsequent respiratory health in early childhood following an episode of bronchiolitis.

In the following three chapters the results of three separate studies are described, relating to each of the three research questions and objectives presented in chapter 2.0.

6.0 Results – Study 1: Bronchiolitis hospital admissions

Chapter Overview

This chapter presents findings in relation to research question 1 and objective 1.

Research Question 1 - What are the risk factors for admission to hospital with RSV bronchiolitis and how do admission rates vary between infants with comorbidities, preterm and term born infants?

Objective 1 - To describe the clinical burden of RSV bronchiolitis among infants admitted to NHS hospitals in England, comparing admission rates between infants with and without clinical risk factors for the condition.

Using HES data a birth cohort from the year 2007/08 with 1 year of follow-up admissions data was developed, as described in detail in the earlier methods chapter. Within the cohort I examined risk factors for hospital admission with bronchiolitis in the first year of life. These findings are presented here along with additional HES data showing trends in RSV-associated hospital admissions in England, to provide contextual information on background trends.

6.1 Introduction

In the UK, bronchiolitis admission rates in infants under 1 year have previously only been estimated from small regional studies, with the most frequently cited estimate of 30.8 per 1000 infants calculated from data from just one health authority region.[31] The geographical area the study was carried out in (Shropshire) had lower than average population density and deprivation for England, limiting the generalisability of their findings. Another study used data from the late 1990s to estimate the annual incidence of hospital admissions attributable to RSV, reporting 28.3 per 1000 infants under 1 year.[30] The only recent national data reporting trends in RSV infection rates over time for England and Wales are the weekly laboratory reports published by the HPA.[39] However, RSV reporting is not mandatory so it is fairly random as to whether or not a laboratory consistently submits their RSV reports and denominator data to indicate the number of infants tested is not readily available either. So there are no recent studies reporting the RSV bronchiolitis disease burden at a population level in the UK.

Infants known to be at high risk of developing severe RSV infection include those born preterm, with chronic lung disease, congenital heart disease, low birth weight or immunodeficiency.[3;22;44;88] Although there is a sizeable base of evidence supporting this, much of what is known to date has been concluded from studies with small sample sizes which may be underpowered to detect smaller effect sizes. For certain rarer conditions such as Down's syndrome and cerebral palsy, there has generally been an insufficient amount of published data available from the UK, to determine their risk of RSV bronchiolitis. Furthermore, despite evidence from the USA reporting that the majority of infants with severe RSV infection are born at term,[44] most UK studies have so far focused on the burden of RSV-associated illness among high-risk infants born preterm.[11;103;155;156;288;289]

The aim of this study therefore, was to examine which infants are most at risk of an RSV bronchiolitis admission in England in term, preterm and other high-risk infants and to determine the age at and duration of an RSV bronchiolitis admission in England as a proxy indicator of severity of illness (using the methods described in detail in chapter 5.2). In addition, trends in RSV-associated hospital admissions over time were examined to provide background context to this study (see methods chapter 5.1).

6.2 Results

6.2.1 Trends in RSV-associated admissions

During the 14 year time period from 1997/98 to 2010/11, trends in admission rates (per 1000 live births) for diagnoses specifying RSV as the causative agent, remained quite stable (Figure 2). In 1997/98 there were 6996 RSV admissions (12 per 1000 live births), compared with 7651 admissions in 2010/11 (11 per 1000 live births). From 1997/98 to 2004/05 unspecified bronchiolitis remained relatively stable (Figure 2). There were 22 unspecified bronchiolitis admissions per 1000 live births in 1997/98 and 23 per 1000 in 2004/05. Similarly, unspecified viral pneumonia rates remained stable during this period, with 45 admissions per 1000 live births in 1997/98 and 44 per 1000 in 2004/05.

Since 2004/05, trends in unspecified bronchiolitis and unspecified viral pneumonia have been steadily rising. There was a 43.5% increase in unspecified bronchiolitis admissions from 23 per 1000 live births in 2004/05 to 33 per 1000 in 2010/11. Unspecified viral pneumonia admission rates increased by 36.3% over the same time period, from 44 to 60 per 1000 live births.

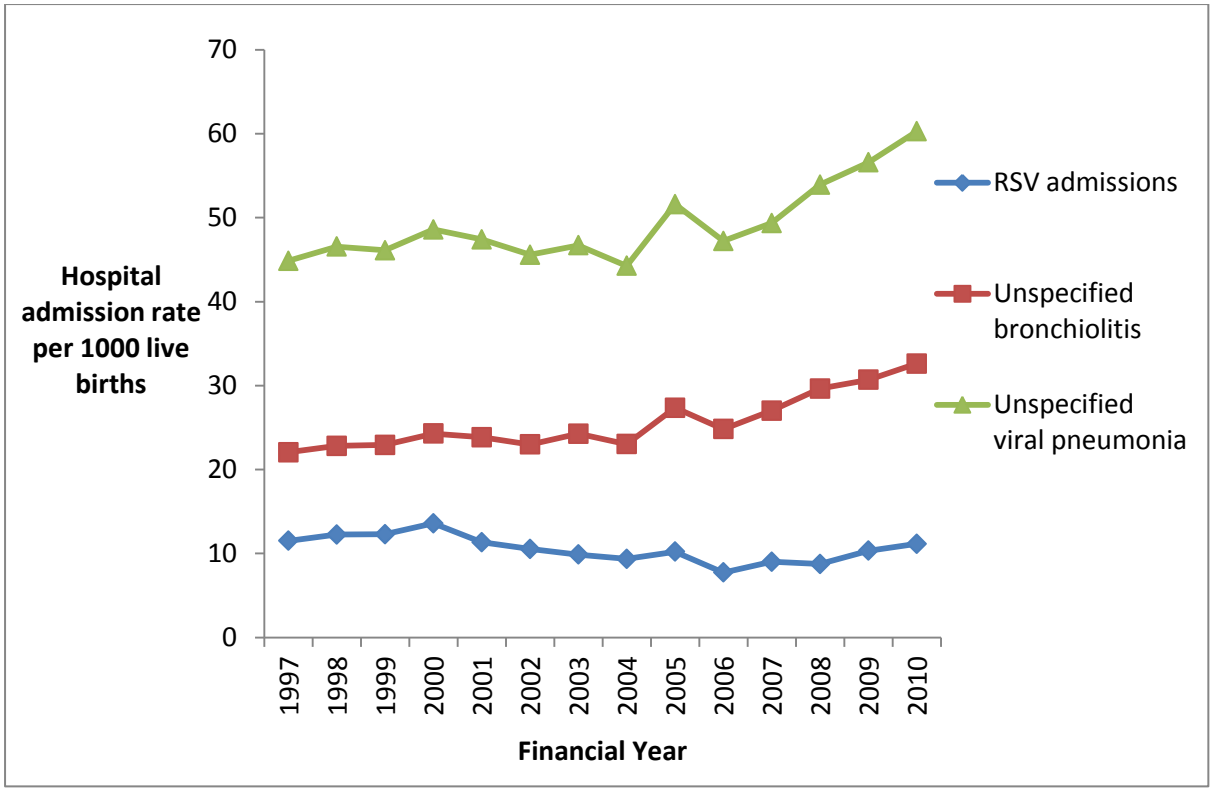


Figure 2: Trends in RSV-associated hospital admissions per 1000 live births, from 1997/98 to 2010/11.

6.2.2 Birth cohort study

Cohort characteristics

296618 infants were included in the birth cohort, from 71 NHS hospitals in England. 410 infants in the cohort died during the study year. 51% (151897/296618) of the cohort were boys, 1% (2891) were multiple births and 7·5% (22215) were born preterm before 37 weeks gestation (Table 7).

Table 7: Characteristics of birth cohort and infants admitted with bronchiolitis.

	Birth cohort N	%	Infants admitted with any bronchiolitis n (% of birth cohort)	Infants admitted with RSV bronchiolitis n (% of group)	Infants admitted with unspecified bronchiolitis n (% of group)
Number of births	296618	100.0	7189 (2.4)	2015 (100.0)	5174 (100.0)
Sex					
Male	151897	51.2	4257 (2.8)	1156 (57.4)	3101 (59.9)
Female	144659	48.8	2927 (2.0)	859 (42.6)	2068 (40.0)
Unknown	62	0.0	5 (8.1)	0 (0.0)	5 (0.1)
Place of birth					
NHS hospital	290057	97.8	6942 (2.4)	1968 (97.7)	4974 (96.1)
Home	5728	1.9	212 (3.7)	35 (1.7)	177 (3.4)
Unknown	833	0.3	35 (4.2)	12 (0.6)	23 (0.4)
Birth weight (grams)					
< 1500	5442	1.8	262 (4.8)	60 (3.0)	202 (3.9)
≥1500 < 2000	2357	0.8	159 (6.7)	57 (2.8)	102 (2.0)
≥2000 < 2500	9748	3.3	407 (4.2)	118 (5.9)	289 (5.6)
≥2500 < 3000	40632	13.7	1164 (2.9)	334 (16.6)	830 (16.0)
≥3000 < 3500	86654	29.2	2003 (2.3)	542 (26.9)	1461 (28.2)
3500+	96090	32.4	2070 (2.2)	572 (28.4)	1498 (29.0)
Unknown	55695	18.8	1124 (2.0)	332 (16.5)	792 (15.3)
Gestational age (weeks)					
<37	22215	7.5	1050 (4.7)	328 (16.3)	722 (14.0)
≥ 37 ≤ 41	212282	71.6	4637 (2.2)	1352 (67.1)	3285 (63.5)
≥ 42	10587	3.6	189 (1.8)	56 (2.8)	133 (2.6)
Unknown	51534	17.4	1313 (2.5)	279 (13.8)	1034 (20.0)
Multiple births					
	2891	1.0	94 (3.3)	30 (1.5)	64 (1.2)
Specialist neonatal care admission at birth recorded					
	18284	6.2	845 (4.6)	280 (13.9)	565 (10.9)

Bronchiolitis admissions

Among our birth cohort there were 7189 admissions to hospital with a primary diagnosis of bronchiolitis during their first year of life, 24.2 admissions per 1000 infants (95% CI 23.7 to 24.8). Only 2015 (28.0%) of these were specifically coded as being due to RSV, the remainder were unspecified. 1529 (21.3%) infants admitted with bronchiolitis, had more than one bronchiolitis admission during their first year of life. The modal age of bronchiolitis admission was 1 month (Figure 3) and the median age was 120 days (Inter-quartile range (IQR) = 61 to 209). The median length of hospital stay was 1 day (IQR = 0 to 3 days). 51.5% of the bronchiolitis admissions were reported to have been admitted via an Emergency Department and 36.4% via referral from a GP.

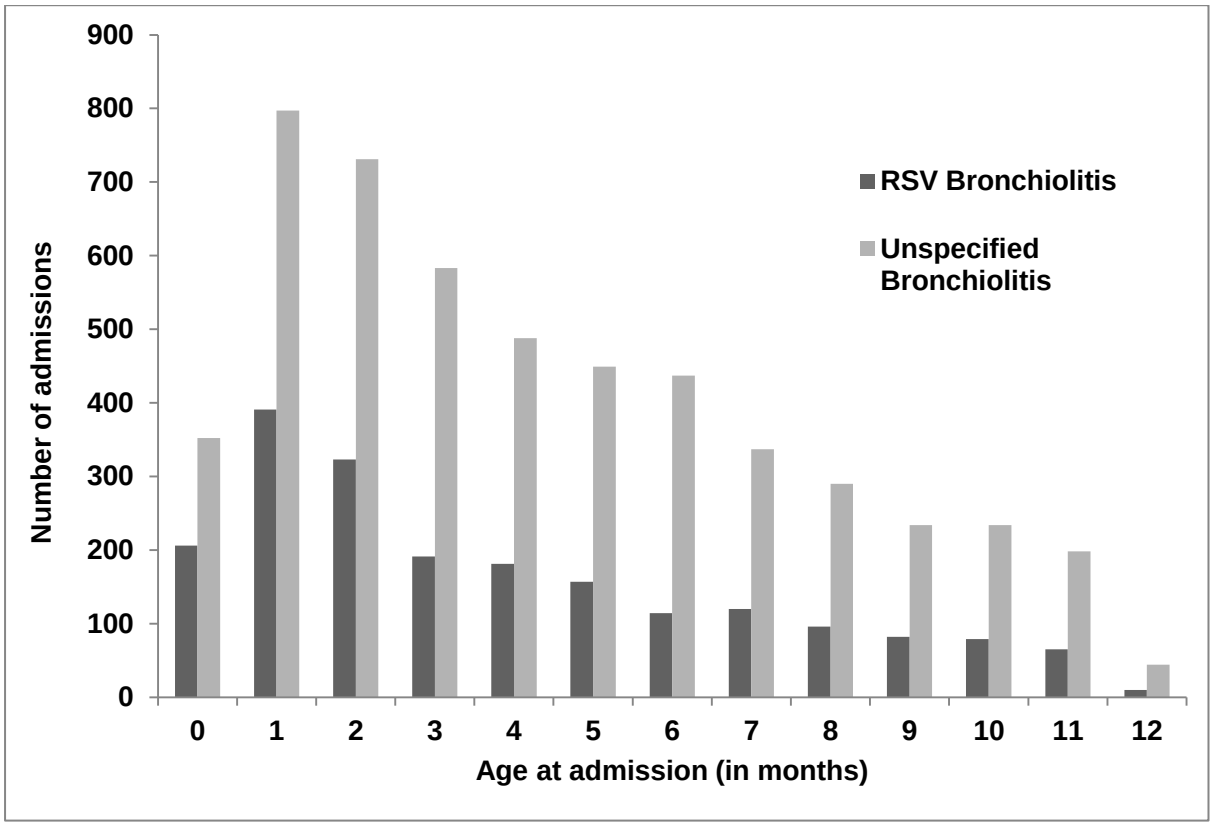


Figure 3: Age of bronchiolitis admission in months, for admissions coded as RSV or unspecified bronchiolitis.

Risk factors for bronchiolitis admission

15% (1050/7189) of infants who had a bronchiolitis admission were born preterm. Admission rates were higher among infants born preterm (47.3 per 1000 infants (95% CI 44.4-50.2)) compared with those born at term (22.4 per 1000 infants (95% CI 21.8-22.9)) (Table 8). The median length of stay for a bronchiolitis admission among infants born preterm was 1 day (IQR = 0 to 3) and among those born at term it was also 1 day (IQR = 0 to 3). The median age of the bronchiolitis admission among infants born preterm was 136 days (IQR = 71 to 221) compared with 118 days (IQR = 59 to 207) among infants born at term ($p<0.001$).

1722 (24.0%) of infants admitted with bronchiolitis had one or more known risk factors for severe RSV infection. The adjusted RR of a bronchiolitis admission was high among infants with known risk factors for severe RSV disease, including those born preterm (RR 1.9 (95% CI 1.8-2.0)) or with congenital heart conditions (RR 3.4 (95% CI 2.9-3.8)) or chronic lung disease (RR 1.6 (95% CI 1.4-1.8)), compared with healthy infants without the risk factor (Table 8). Other conditions also significantly increased an infant's risk of bronchiolitis admission, including cystic fibrosis (RR=2.5 (95% CI 1.4 to 4.4)), Down's syndrome (RR 2.5 (95% CI 1.7-3.7)), cerebral palsy (RR 2.4 (95% CI 1.5-4.0)) and other nervous system congenital abnormalities (RR=1.7 (95% CI 1.3 to 2.4)).

Table 8: The relative risk of a bronchiolitis hospital admission among infants in different risk groups.

Risk Group†	Number of bronchiolitis admissions (% of infants in risk group)	Total number of infants in risk group (% of whole birth	Median length of stay in days (IQR)	Risk of bronchiolitis admission per 1000 infants under 1 year (95%	RR† (95% CIs)	Adjusted RR (95% CI)
Born at term	6139 (2.2)	274403	1 (0 to	22.4 (21.8 to 22.9)	-	-
Premature birth	1050 (4.7)	22215 (7.5)	1 (0 to	47.3 (44.4 to 50.2)	2.11 (1.98 to 2.26)	1.89 (1.77 to 2.02)
Cystic fibrosis	11 (6.4)	171 (0.1)	2 (0 to	64.3 (32.1 to 115.1)	2.66 (1.33 to 4.76)	2.45 (1.36 to 4.43)
Congenital heart Disease	272 (12.1)	2239 (0.8)	2 (0 to	121.5 (107.5 to 136.8)	5.17 (4.56 to 5.84)	3.35 (2.92 to 3.84)
Chronic lung disease	282 (5.6)	5016 (1.7)	2 (0 to	56.2 (49.9 to 63.2)	2.37 (2.10 to 2.67)	1.61 (1.42 to 1.82)
Immunodeficiency	7 (11.7)	60 (0.0)	8 (1 to	116.7 (46.9 to 240.4)	4.82 (1.94 to 9.93)	1.69 (0.80 to 3.58)
Nervous system congenital anomalies	42 (8.6)	489 (0.2)	2 (1 to	85.9 (61.9 to 116.1)	3.56 (2.56 to 4.82)	1.73 (1.26 to 2.36)
Down's syndrome	28 (15.4)	182 (0.1)	3 (0 to	153.9 (102.2 to 222.4)	6.37 (4.23 to 9.21)	2.53 (1.72 to 3.72)
Cerebral Palsy	16 (10.7)	149 (0.1)	3 (1 to 5)	107.4 (61.4 to 174.4)	4.44 (2.54 to 7.21)	2.43 (1.48 to 3.99)

†Relative risk compared with infants without risk factor of interest.

A higher proportion of infants born in September (3.9%) or October (3.8%) were admitted with bronchiolitis than among infants born in any other month ($p<0.001$). Bronchiolitis admission rates among infants born in the North East (3.3%) and Yorkshire and the Humber (3.3%) SHAs, were higher ($p<0.001$) compared with in infants born in London (1.6%) (Figure 4).

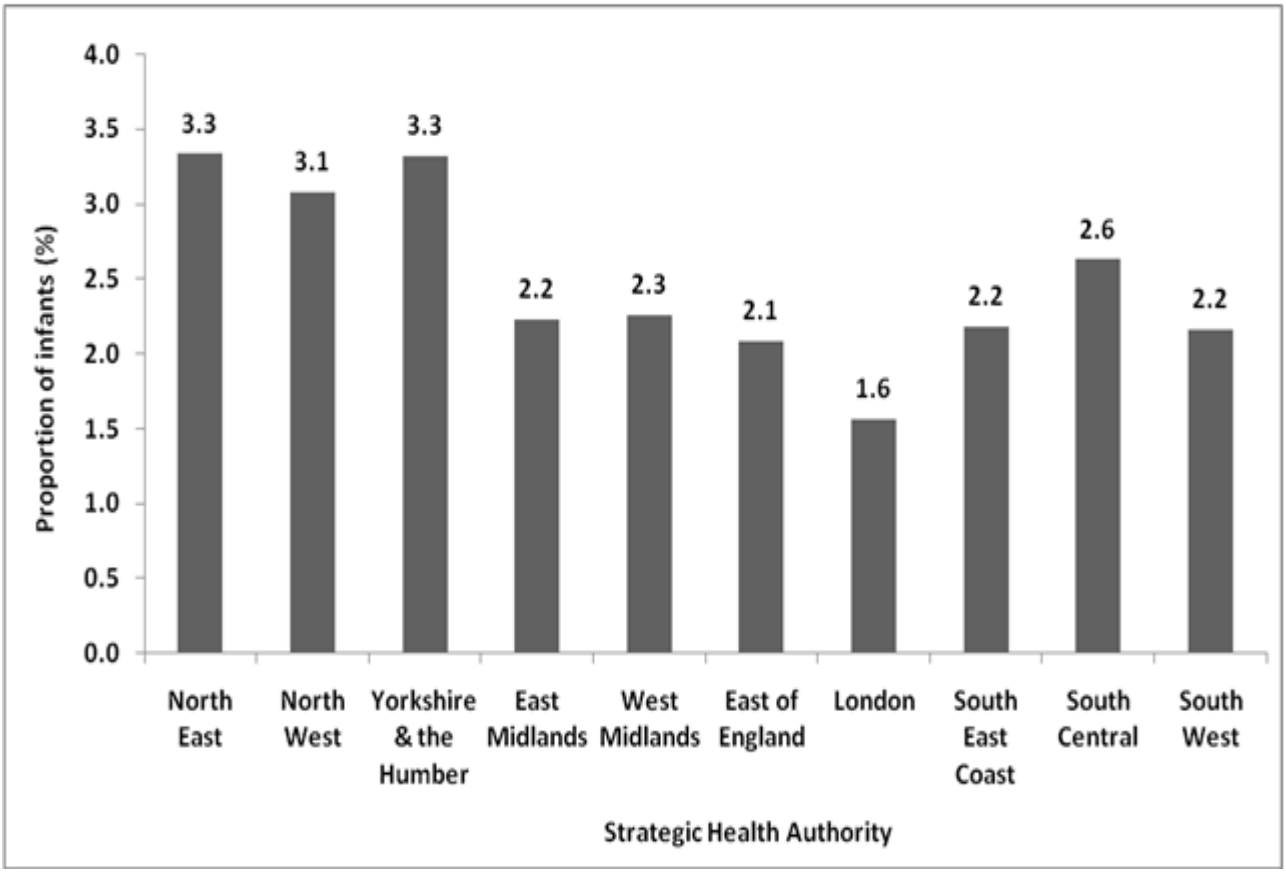


Figure 4: Incidence of bronchiolitis hospital admissions among birth cohort, by Strategic Health Authority (SHA) of admission.

6.3 Summary of chapter findings

In this chapter trends in diagnoses specifying RSV as the causative organism, unspecified viral pneumonias and unspecified bronchiolitis admission rates were estimated, over a 14 year period from 1997/98 to 2010/11. A birth cohort from the year 2007/08 with follow-up to age 1 year was created by linking HES birth and admission records and used to investigate risk factors for hospital admission with bronchiolitis. The findings from these analyses of admission data from NHS hospitals in England are summarised here:

- Infant admissions for diagnoses specifying RSV infection remained stable over the 14 year period from 1997/98 (12 per 1000 live births) to 2010/11 (11 per 1000 live births).

- From 2004/05 to 2010/11, trends in unspecified bronchiolitis and unspecified viral pneumonia admissions have steadily risen by 43.5% and 36.3% respectively.

- Among a birth cohort of 296618 infants born in 2007/08, the bronchiolitis admission rate was 24.2 admissions per 1000 infants under 1 year (95% CI 23.7 to 24.8).

- The median age at bronchiolitis admission was 120 days (IQR 61 to 209 days) and the modal age group at bronchiolitis admission was infants aged 1 month. The median length of stay for a bronchiolitis admission was 1 day (IQR 0 to 3 days).

- 51.5% of infants with bronchiolitis were admitted via the Emergency Department.

- 15% (1050/7189) of infants who had a bronchiolitis admission were born preterm, 47.3 admissions per 1000 infants (95% CI 44.4 to 50.2). The bronchiolitis admission rate among infants born at term was 22.4 per 1000 infants (95% CI 21.8 to 22.9).

- 24.0% of infants admitted with bronchiolitis had one or more known clinical risk factor.

- The relative risk of a bronchiolitis admission was high among infants with known risk factors for severe RSV disease, including those born preterm (RR 1.9 95% CI 1.8 to 2.0) or with congenital heart conditions (RR 3.4 95% CI 2.9 to 3.8) or chronic lung disease (RR 1.6 95% CI 1.4 to 1.8), compared with healthy infants without the risk factor.

- Other conditions also significantly increased an infant's risk of bronchiolitis admission, including cystic fibrosis (RR=2.5 (95% CI 1.4 to 4.4)), Down's syndrome (RR 2.5 (95% CI 1.7-3.7)) and cerebral palsy (RR 2.4 (95% CI 1.5-4.0)).

7.0 Results - Study 2: Bronchiolitis in the community setting

Chapter Overview

This chapter presents findings in relation to research question 2 and objective 2.

Research question 2 - What is the incidence of acute bronchiolitis among infants presenting to general practitioners and how is it being managed in the community?

Objective 2 - To estimate the incidence of RSV bronchiolitis among infants presenting in primary care and to describe how it is being managed by general practitioners.

For the first time, this study provides an insight into the incidence and management of bronchiolitis illness in UK primary care. Using GPRD data a birth cohort from 2003 with 1 year of follow-up was created, from 600 representative general practices across the UK, as described in the earlier methods chapter. Within this cohort, GP consultations during the first year of life were examined, focusing particularly on those where a diagnosis of bronchiolitis or related symptoms were recorded. Prescriptions given in relation to a bronchiolitis consultation were identified, to determine how GPs have been managing the condition. These findings are presented followed by a discussion of the implications of what this study tells us about the burden of bronchiolitis illness in UK primary care.

7.1 Introduction

Previous studies have focused on identifying risk factors for hospitalisation with severe disease but there are no reliable community-based estimates of the incidence of bronchiolitis and little is known about the presentation, natural history and management of the condition in the community.

It can be challenging for GPs, presented with an infant with respiratory symptoms, to differentiate their diagnosis between bronchiolitis, bacterial respiratory tract infections and wheezy bronchitis/early asthma. Evidence from systematic reviews suggests no available treatments shorten the natural course or provide clinically relevant improvements in bronchiolitis symptoms.[33;61;62]

SIGN guidelines are the main source of information for bronchiolitis management in the UK but lack guidance applicable to general practice.[36] The guidelines recommends further research is needed to examine the prevalence of bronchiolitis in primary care and current

treatment practice in the community and risk profiling to identify those who may have future respiratory symptoms.[36] Assessing the burden of bronchiolitis in UK primary care could improve the evidence base for management of this condition by GPs and improve our understanding of the epidemiology of RSV disease which will be particularly important should an effective vaccine be developed.

Early findings from The Houston Family Study published in 1986, a birth cohort of 125 infants with community follow-up for the first year of life, estimated the rate of RSV-confirmed lower respiratory tract disease to be 21.6 per 100 infants under 1 year.[46] Large prospective birth cohorts are needed to determine the current clinical burden of bronchiolitis in the community, but these are expensive and not practical. With widespread use of electronic patient records in general practices across the UK[47], it is now possible to estimate the frequency of acute bronchiolitis consultations in the community.

The main aim of this study was to identify how frequently acute bronchiolitis is diagnosed by GPs. The methods used are described in the earlier chapter 5.3.

7.2 Results

7.2.1 Cohort characteristics

The birth cohort included 80,976 infants born in the year 2003, from 600 general practices across the UK. 51% (41457/80976) of the cohort were males (39519 were females).

7.2.2 Bronchiolitis consultations

Among the birth cohort, 3543 infants were diagnosed with bronchiolitis during a general practice consultation in their first year of life, with a total of 4707 bronchiolitis consultations among the cohort, 58.1 per 1000 infants under 1 year (95% CI 56.5-59.8). The mean age of infants diagnosed with bronchiolitis in a GP consultation (Figure 5) was 5.5 months (SD = 3.2). There was a peak in bronchiolitis consultations during the winter months, particularly late December to early January (Figure 6), corresponding with the typical peak in laboratory reports of RSV infection sent to the HPA.

Using the broader bronchiolitis definition 16744 “bronchiolitis” consultations were identified among the birth cohort, giving a consultation rate of 206.7 per 1000 infants under 1 year (95% CI 204.0 – 209.6).

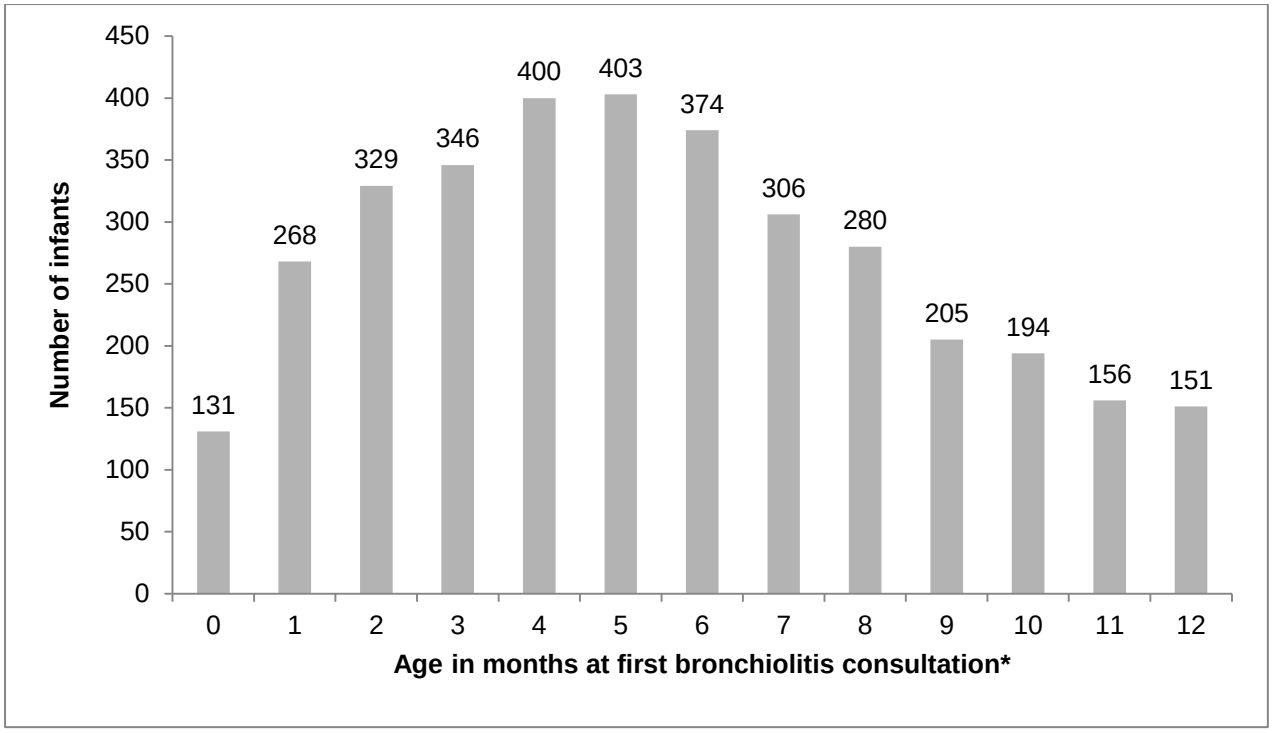


Figure 5: Age in months at first diagnosis of bronchiolitis in a GP consultation.

* Dataset contains only month and year of birth (no date) so age in months is approximate, based on the assumption that individuals are born in the middle of the month.

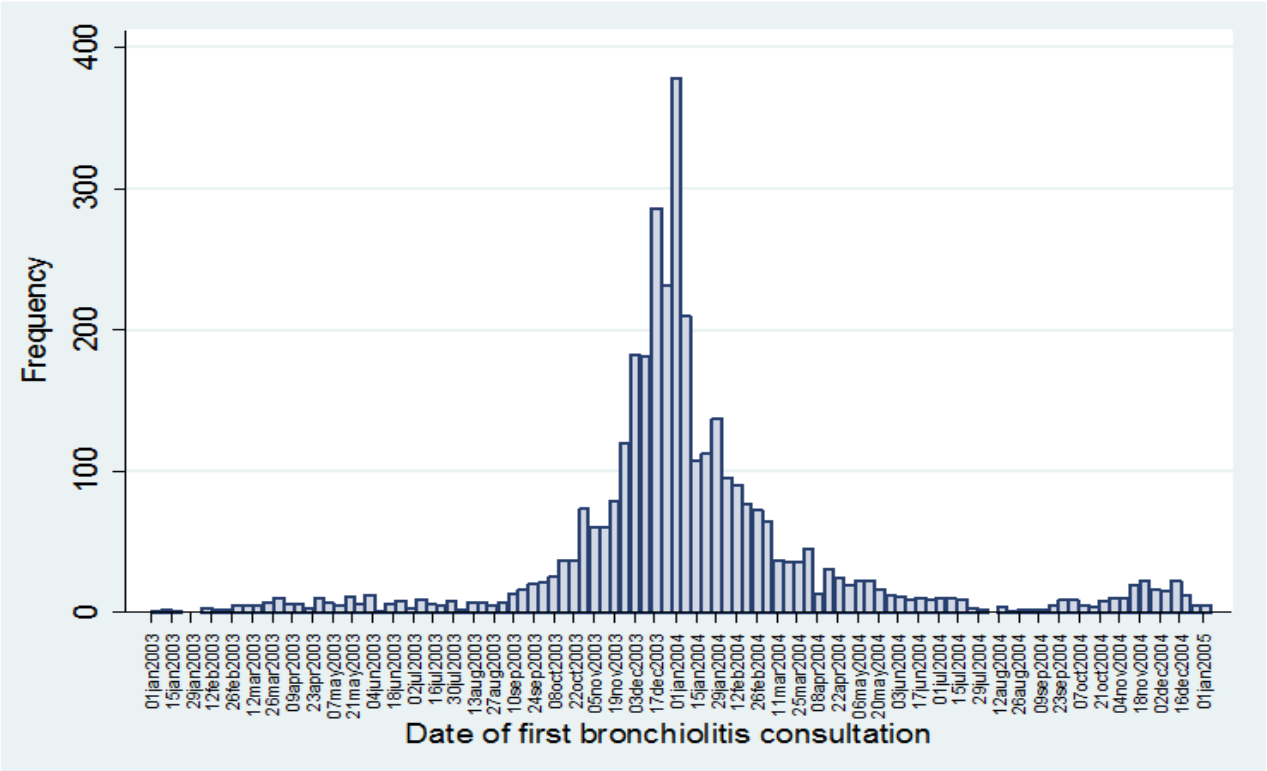


Figure 6: Seasonality of acute bronchiolitis consultations in general practices across the UK.

7.2.3 Consultations in the first year of life

Among infants with a bronchiolitis consultation, 176 (5.0%) had had no prior consultations with their general practitioner for any reason .The majority of those diagnosed with bronchiolitis, 2943/3543 (83.1%), had previously visited their GP for at least one immunisation. Among infants with no history of bronchiolitis, the most common diagnoses and symptoms recorded during primary care consultations in their first year of life were respiratory conditions (47.6%). In comparison, among infants who *did* consult with bronchiolitis during their first year, respiratory symptoms made up significantly more (60.1%) of all their primary care consultations (Table 9). The relative risk of having a respiratory consultation in the first year of life was 1.26 (CI 1.23 to 1.29) among infants who consulted with bronchiolitis compared with those who did not.

Table 9: The most common diagnoses recorded during consultations among infants in the cohort, comparing those with and without a bronchiolitis consultation during their first year of life.

Diagnosis groups of consultations during the first year of life	Infants with no bronchiolitis consultations **(n=77433)** Frequency (%)	Infants with a bronchiolitis consultation **(n=3543)** Frequency (%)	Comparison of proportions (z test) P value
Respiratory tract infections and cough or wheezing symptoms	64791 (47.6)	6940 (60.1)	<0.0001
Eczema, dermatitis, rashes and other skin conditions	27647 (20.3)	1577 (9.7)	<0.0001
Diarrhoea, vomiting and other gastro-intestinal symptoms	21469 (15.8)	1333 (8.2)	<0.0001
Other infections	22077 (16.2)	1689 (10.4)	<0.0001

7.2.4 Consultations before and after bronchiolitis

Infants who consulted with bronchiolitis had an average of 1.1 consultations during the 2 weeks before their bronchiolitis diagnosis and an average of 1.0 consultation during the following 2 weeks. During any GP consultation in the 2 weeks *before* a bronchiolitis consultation, the top 5 diagnostic or symptomatic codes recorded were upper respiratory tract infection (9.3%, n=329), cough (8.4%, n=297), chest infection (2.6%, n=93), acute conjunctivitis (1.3%, n=45) and wheezing symptoms (0.9%, n=33). The most common diagnostic or symptomatic codes recorded during GP consultations in the 2 weeks *following* a bronchiolitis consultation were acute bronchiolitis (9.5%, n=336), cough (2.9%, n=102), upper respiratory tract infection (2.0%, n=71), acute conjunctivitis (0.8%, n=27) and chest infection (0.7%, n=26).

7.2.5 Prescriptions

In total, there were 1684 prescriptions associated with a bronchiolitis consultation (35.8% of all bronchiolitis consultations resulted in a prescription). Among these, 464 were for antibiotics (27.6%), 452 for beta agonists (26.8%) and 140 for antipyretics or analgesics (8.3%) and 12 for inhaled corticosteroids (0.7%). The mean age of infants receiving any prescription associated with a bronchiolitis consultation was 6.6 months (SD=2.9).

7.3 Summary of chapter findings

In this chapter, findings from the first study to estimate the incidence of acute bronchiolitis in UK primary care are presented. A cohort was created using medical records from the GPRD database and used to examine the natural history and management of bronchiolitis in the community setting. The findings reported in this study have important implications for GP training and potentially for passive and future active RSV immunisation policy in the UK (which will be discussed in chapter 10). The main findings from there analyses are of primary care data are summarised here:

- From a cohort of 80,976 infants born in the UK in 2003, 3543 were diagnosed with bronchiolitis during a general practice consultation in their first year of life.

- In total there were 4707 bronchiolitis consultations among the cohort, 58.1 per 1000 infants under 1 year (95% CI 56.5 to 59.8), with a peak during winter months.

- The mean age at bronchiolitis consultation was 5.5 months (SD = 3.2).

- A broader bronchiolitis definition which included non-specific symptomatic codes, identified 16744 potential bronchiolitis consultations, giving a rate of 206.7 per 1000 infants under 1 year (95% CI 204.0 to 209.6).

- The relative risk of having a respiratory consultation in the first year of life was 1.26 (95% CI 1.23 to 1.29) among infants who consulted with bronchiolitis compared with those who did not.

- Infants who consulted with bronchiolitis had an average of 1.1 consultations during the 2 weeks before their bronchiolitis diagnosis and an average of 1.0 consultation during the following 2 weeks.

- 35.8% of bronchiolitis consultations resulted in a prescription, of which 464 (27.6%) were for antibiotics, 452 (26.8%) for beta agonists and 140 (8.3%) for antipyretics or analgesics.

8.0 Results – Study 3: Impact of bronchiolitis on respiratory health in early childhood

Chapter Overview

This chapter presents findings in relation to research question 3 and objective 3.

Research Question 3 - What is the long-term impact of RSV bronchiolitis on subsequent respiratory health in early childhood?

Objective 3 - To determine the longer-term effects of RSV bronchiolitis on subsequent respiratory health in early childhood.

Using the methods described earlier (chapter 5), two cohort studies were developed, one from the primary care setting using GPRD data and one from the secondary care setting using HES data. Both studies included individual medical records from birth to age 5 years (for HES) or 7 years (for GPRD), providing sufficient follow-up data to examine respiratory outcomes in early childhood, following an episode of bronchiolitis in infancy.

8.1 Introduction

Evidence suggests RSV infection in infancy may be associated with subsequent, longer-term respiratory morbidity in childhood and even into early adulthood. In particular, research has shown a link between RSV infection in infancy and recurrent wheezing and asthma in childhood.[14;138;140-146] However, there remains some uncertainty about the physiology and immunological processes that may drive this association, with some research suggesting it is the disease phenotype that is of more importance than the causative agent, in predicting future asthma and wheeze.[160] It is important to improve our understanding of this association given the long-lasting impact that a chronic respiratory illness such as asthma can have, on the quality of life of that individual and their family, as well as the associated healthcare costs. However, long-term follow-up of infants following RSV bronchiolitis is often precluded by practical and financial issues and much of the existing evidence has come from studies with a small number of participants. Longitudinal linkage of routinely collected healthcare records presents an opportunity to examine any link between RSV bronchiolitis and longer-term respiratory morbidity, among a large, nationally representative cohort of children.

The aim of this study was to determine the longer-term effects of severe RSV bronchiolitis on subsequent respiratory health, specifically on admissions and consultations for wheezing and asthma, in early childhood. The methods used in this study are described in chapter 5.

8.2 Results

8.2.1 HES cohort study with follow-up to age 5 years

Description of birth cohort

211772 individuals born in financial year 2003/04 were identified for inclusion in the birth cohort. Among whom, 7.6% (n = 16084) were born preterm. 13% (n = 28503) of infants in the cohort were identified as having one or more risk factor for severe RSV infection and were categorised as "high-risk". The remaining 87% of the cohort were categorised as being at "low-risk" of bronchiolitis admission as they were previously healthy.

Bronchiolitis admissions

In total there were 5947 admissions to hospital with a primary diagnosis of bronchiolitis among the cohort, 28.1 per 1000 infants under 1 year (95% CI 27.4 to 28.8). 5.1% (1451/28503) of the high-risk infants were admitted to hospital with a diagnosis of bronchiolitis, compared with 2.5% (4496/183269) among the group of infants at low-risk ($p<0.05$). Bronchiolitis admissions among infants in the high-risk group had a median length of stay of 3 days (IQR = 0 to 6 days), compared with a median of just 1 day in the low-risk group (IQR = 0 to 3 days) ($p<0.05$). Infants admitted with bronchiolitis in the high-risk group were more likely to be admitted to an ICU during their stay (32% compared with 4% in the low-risk group). Among the high-risk infants with a bronchiolitis admission, 28% had at least 1 further bronchiolitis admission, compared with 15% among the low-risk group ($p<0.05$). The median age group at bronchiolitis admission was infants aged 3 to 6 months, for both the high- and low-risk groups.

Subsequent respiratory morbidity

A higher proportion of children with a history of bronchiolitis admission in infancy had subsequent admissions for asthma in the first five years of life (12.0%), compared with children with no prior bronchiolitis admissions (1.7%). Hospital admissions for wheezing in the first 5 years of life were also more common among those who had been admitted to hospital with bronchiolitis in infancy (8.5%) compared with those who had not (1.3%). Among infants with a history of bronchiolitis admission, 14.5% were admitted to hospital with LRTI and 23.0% with URTI before the age of 5 years. A smaller proportion of those children with no prior bronchiolitis admission were admitted to hospital with LRTI (2.7%) or URTI (7.7%).

The relative, attributable and population attributable risks of hospital admission with asthma, wheezing, LRTI or URTI in the first five years of life, among infants who had a bronchiolitis admission in infancy compared with those who did not, are presented in Table 10.

Table 10: Risk of subsequent respiratory hospital admissions in first five years of life, following a bronchiolitis admission in infancy.

	Relative Risk (RR) (95% CI)	**Attributable Risk (AR) (95% CI)**	**Population AR**
Asthma	**6.96 (6.45 – 7.51)**	**0.10 (0.09 – 0.11)**	**0.14**
Wheezing	**6.52 (5.95 – 7.15)**	**0.07 (0.06 – 0.08)**	**0.13**
URTI	**3.00 (2.85 – 3.15)**	**0.15 (0.14 – 0.16)**	**0.05**
LRTI	**5.43 (5.07 – 5.80)**	**0.12 (0.11 – 0.13)**	**0.11**

Subsequent respiratory morbidity in high- and low-risk groups

Among the high-risk group who had been admitted with bronchiolitis, 15% were subsequently admitted with a primary diagnosis of asthma, before the age of 5 years, compared with 11% among the low-risk group with a history of bronchiolitis (Figure 7). In comparison, among high- and low- risk infants who had no bronchiolitis admissions during infancy, only 2% were admitted between the age of 6 months and 5 years with a primary diagnosis of asthma. Subsequent admissions for wheezing in the first 5 years of life occurred among 7% of the low-risk and 13% of the high-risk group, who had a history of bronchiolitis admission. In comparison, only 1% of the low-risk and 2% of the high-risk infants who had no previous bronchiolitis admission were admitted with a diagnosis of wheezing before the age of 5 (Figure 7). Following a bronchiolitis admission, subsequent admissions for LRTIs and URTIs were common among both groups but highest among the high-risk infants (Figure 7).

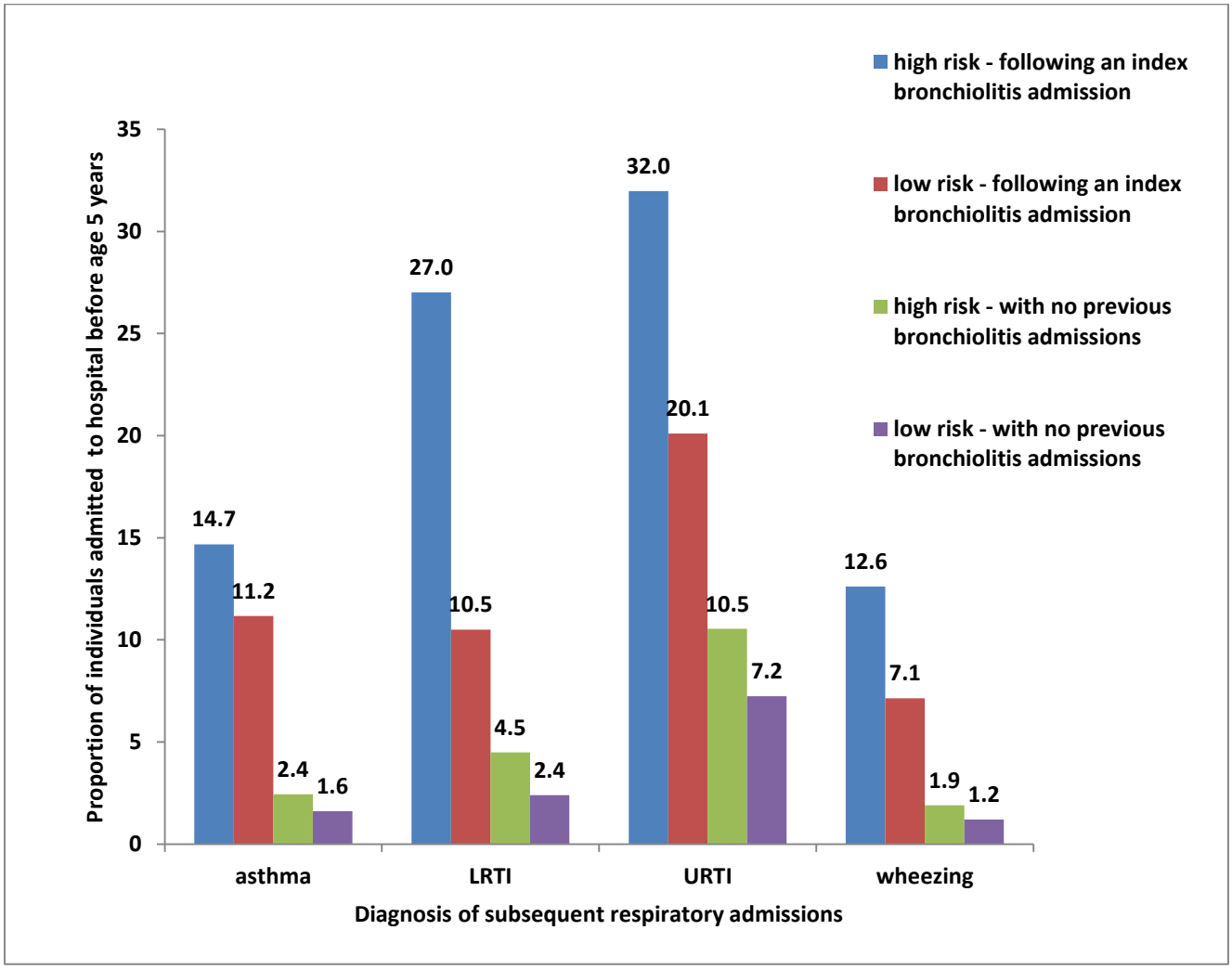


Figure 7: Subsequent respiratory admissions in early childhood (up to age 5 years) among infants in high- and low- risk groups, comparing those with and without a history of bronchiolitis admission during childhood.

8.2.2 GPRD cohort study with follow-up to age 7 years

Study cohort

The birth cohort consisted of 80,976 infants born in the year 2003 and registered across 600 general practices in the UK. 51% (41457/80976) of the cohort were males (39519 were females). Complete primary care records up to the age of 7 years were available for 54540 children in the cohort (67.4%), as others transferred out of the cohort upon registering at a different GP practice.

Bronchiolitis consultations

As described in the earlier results section (7.2), 3543 infants had at least one consultation in the first year of life with a diagnosis of acute bronchiolitis. There were a total of 4707 bronchiolitis consultations among the cohort, 58.1 per 1000 infants under 1 year (95% CI 56.5-59.8).

Subsequent respiratory consultations up to age 7 years

Among children with a history of bronchiolitis consultation in infancy, 25.4% had at least one consultation with a diagnosis of asthma by the age of 7 years. In comparison, 11.8% of children with no prior bronchiolitis consultations had an asthma consultation by age 7 years.

A higher proportion of children who consulted with bronchiolitis in infancy had at least one consultation with a diagnosis of wheezing by the age of 7 years (57.6%), compared with children with no history of bronchiolitis, among whom 13.9% had a wheezing consultation by age 7 years.

The relative, attributable and population attributable risks of subsequent GP consultations for asthma or wheezing in the first 7 years of life, comparing infants with a history of bronchiolitis consultation in infancy to those without, are presented in table 11. The findings show that children with a history of bronchiolitis are over twice as likely to consult their GP with asthma and 4 times more likely to consult with wheezing, compared with children with no prior bronchiolitis consultations.

Table 11: Risk of subsequent GP consultations for asthma or wheeze in the first 7 years of life, comparing children with a history of bronchiolitis consultation to those without.

	Relative Risk (RR) (95% CI)	Attributable Risk (AR) (95% CI)	Population AR
Asthma	**2.15 (2.01 – 2.30)**	**0.14 (0.12 – 0.15)**	**0.05**
Wheezing	**4.14 (3.95 – 4.35)**	**0.44 (0.41 – 0.46)**	**0.12**

8.3 Summary of chapter findings

In this chapter the findings of two cohort studies are described, examining the longer-term effect of RSV bronchiolitis on subsequent respiratory admissions and GP consultations in early childhood. The findings from these analyses are summarised here:

- In a cohort of 211772 children born in English NHS hospitals in 2003/04, 13% had one or more risk factors for severe RSV infection and were categorised as "high-risk". The bronchiolitis admission rate among the cohort was 28.1 per 1000 infants under 1 year. 5.1% of high-risk and 2.5% of low-risk infants had a bronchiolitis hospital admission.

- The relative risk of subsequent admissions for asthma in the first five years of life was 6.96 (95% CI 6.45 to 7.51) and for wheezing was 6.52 (95% CI 5.95 to 7.15), among children with a history of bronchiolitis admission compared with children without.

- 14.5% of children with a history of bronchiolitis admission were admitted to hospital with LRTI and 23.0% with URTI before the age of 5 years, compared with 2.7% and 7.7% respectively among children with no prior bronchiolitis admission.

- 15% of high-risk children with a history of bronchiolitis admission were subsequently admitted for asthma by age 5 years, compared with 11% in the low-risk group with a history of bronchiolitis. Among high- and low- risk children with no prior bronchiolitis admissions, only 2% had an asthma admission between age 6 months and 5 years.

- Subsequent admissions for wheezing in the first 5 years of life occurred among 7% of low-risk and 13% of high-risk children with a history of bronchiolitis admission and in only 1% of low-risk and 2% of high-risk infants with no prior bronchiolitis admissions.

- From a cohort of 80,976 infants born in the UK in 2003, 3543 (4.4%) were diagnosed with bronchiolitis during a general practice consultation in their first year of life.

- Among children with a history of bronchiolitis consultation in infancy, 25.4% had at least one consultation with a diagnosis of asthma by age 7 years, compared with 11.8% of children with no prior bronchiolitis consultations (RR = 2.15 95% CI 2.01 to 2.30).

- 57.6% of children who consulted with bronchiolitis in infancy had at least one consultation with a diagnosis of wheezing by the age of 7 years, compared with 13.9% among children with no history of bronchiolitis (RR = 4.14 95% CI 3.95 to 4.35).

9.0 Preliminary developments and progress in linking neonatal care and subsequent hospital admission records

Chapter overview

In this chapter, I describe the progress made in linking individual neonatal care records from the NNRD to subsequent hospital admissions records in the HES database. This work forms one component of the "Medicines for Neonates" programme funded by the NIHR under its Programme Grants for Applied Research funding scheme. Over the three year PhD period considerable headway has been made, despite the lengthy and bureaucratic processes that were required to ascertain all the necessary research governance and ethical approvals. Here I outline the progress that has been made to date, describe linkage methodology and summarise some potential uses of this linked longitudinal dataset.

9.1 Background

The NNRD was created to enhance the evidence base for neonatal care, (as described in the earlier chapter 4.4), harnessing the unique position of the universal care provided by the NHS, to integrate clinical and research processes and reduce the complexity and costs of conducting research. One of the objectives of developing the NNRD was to link neonatal care records to other NHS data such as hospital admission and general practice records, using unique personal identifiers, to provide long-term follow-up to study subsequent health outcomes beyond discharge. This rich source of data can facilitate applied research in neonatal medicine, as well as audit, monitoring and evaluation of interventions and service delivery. It would also be useful for clinical trials in newborns where long-term follow-up can be logistically difficult and costly.

The development of neonatal networks with coordinated care services and shared management has facilitated data collection and establishment of the NNRD.[290;291] Through engaging with and involvement of NHS staff within these networks, high completeness in data recording has been achieved and continual feedback to networks ensures the quality of inputted data has continued to rise. Since data are entered and actively used by clinicians at the point of care, data capture is likely to be more complete and accurate. These records contain clinically relevant information captured in real-time, from a large geographically defined population that is nationally generalisable. Linked neonatal and subsequent

hospital admission records could be used to establish if the contribution of previously identified variables on neonatal outcomes can be replicated, as well as to assess other important but hitherto unaddressed factors such as deprivation and ethnicity that may be major confounders in assessing long-term outcomes.[279]

9.2 Research governance and applying for ethical approval

As described earlier (see section 4.5), as part of the overall Medicines for Neonates programme of work (which this PhD research is part of) we applied to the NIGB for approval to link the NNRD managed by the NDAU to HES data at an individual level, using identifiers including NHS number, date of birth, sex, postcode and hospital number. National Research Ethics approval was also obtained in 2010 for the NDAU to form the NNRD, including the linkage to other NHS records.[279]

Although provisional NIGB approval was received in December 2010, final approval was not granted until the third year of my PhD in January 2012. This was due to the request for various clarifications and amendments to some aspects of the application, such as providing further information to parents about the potential uses of their baby's data and the inclusion of a mechanism to permit them to opt out of their baby's data being stored. Upon receipt of this final approval I was able to submit an application to the NHS Information Centre in February 2012, to receive identifiable HES data at the NDAU for the purposes of linkage to the NNRD. The HES data arrived at the end of August 2012 and I have subsequently spent considerable time uploading, cleaning and formatting these records in preparation for linkage.

The delays in receiving these necessary approvals and obtaining the identifiable HES data have consequently limited the amount of progress that could be made with the linkage stage of this work, given the 3 year time constraints of this PhD (from January 2010 to December 2012). However, I will now describe the methodology that will be used going forward with this linkage.

9.3 Proposed linkage methodology

Record linkage is the process of bringing together records pertaining to the same individual from disparate data sources.[225;241] This is becoming increasingly important in an era when clinical records are now routinely captured in electronic format and maintained in large databases, frequently on a national scale. The techniques and methodology to link individual records across different datasets are now being developed and implemented, as this is a cost-effective and efficient way to increase the amount of information available for research.[292]

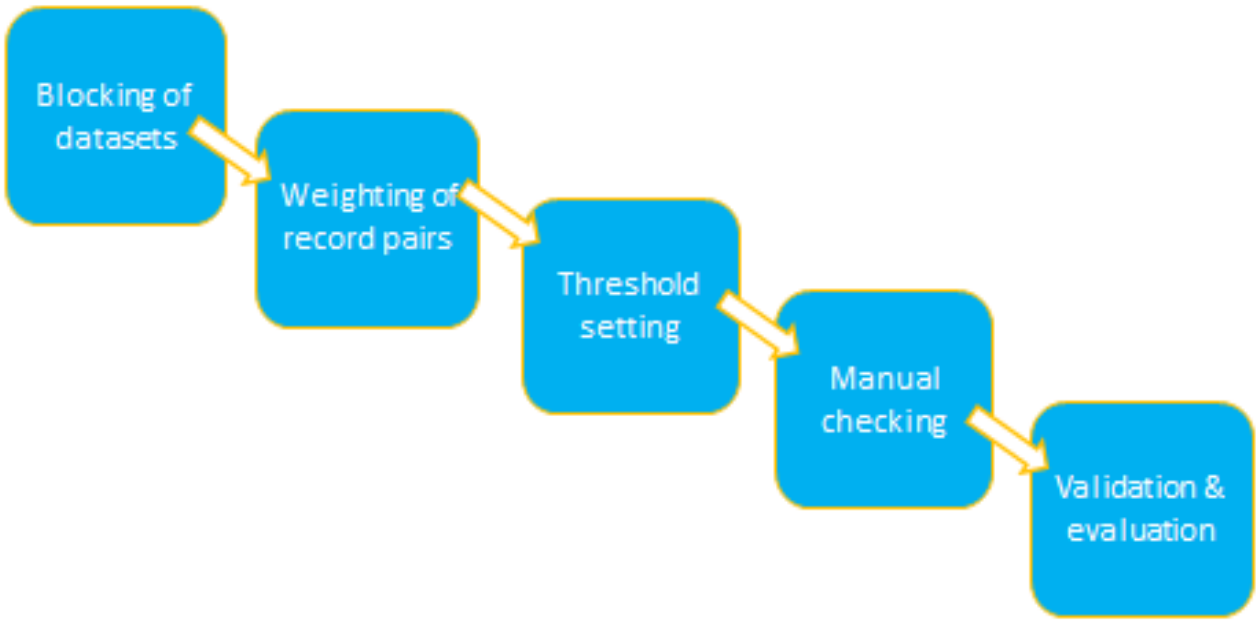
Records may be linked using either **deterministic** or **probabilistic** methods. Deterministic or exact record linkage requires a unique variable (or combination of variables) to be available in both files being linked. Probabilistic record linkage is suitable when no unique identifier is available, data are 'noisy' (i.e. contain random errors), or key identifier variables have missing values.[225;235] Probabilistic procedures calculate and assign weights or probability based criteria on the basis of the degree of matching of the identifier variables.[292-294] This allows for missing data or data entry errors by utilising the cumulative accuracy of a number of variables.[295] The likelihood that two records match is then related to the sum of these two weights, so this method requires a subjective decision to be made to accept as true matches all comparisons of record pairs with combined weights above a chosen threshold and reject all those below.[241;294] The threshold for these weights is chosen to satisfy criteria for specificity and sensitivity, which varies depending on whether the linkage is one-to-many or many-to-many.[292] The weights are based on the probability that the linking variable agrees among matches (which reflects the reliability and therefore error rate) and the probability that the linking variable agrees among non-matches (which reflects its discriminatory power).[296] A high error rate increases the number of false matches while low discriminatory power increases the number of false links.[296]

Deterministic methods result in the low false positive links but at the expense of missing matches, depending on the error rate of the linking variables.[293;294;297] A priori knowledge of the quality of linking variables is needed to help determine which level of disagreement between matching variables should be tolerated.[294] Which approach to use depends on the purpose of the record linkage. If the aim is be able to make inferences about specific individuals then a high positive predictive value is needed, so a deterministic approach may be favoured or a probabilistic method with careful clerical review of matches close to the selected threshold of acceptability.[293]

The basic theoretical and practical steps involved in probabilistic record linkage are summarised in Box 3. The first stage involves formatting and blocking of datasets. Linking variables need to be formatted consistently and ordered. The data then need to be broken down into manageable blocks for linking, for example blocking by NHS number. The next stage is matching and weighting of records. Records within each particular block in one dataset are matched to every record in the corresponding block in the other dataset.[295] More than one round of blocking may be used, for example using NHS number and date of birth. Weights are then assigned to linking variables within a pair of linked records, based on the contribution of each identifier, so that agreement on NHS number for example would carry more weight than agreement on hospital number or sex.[298] The threshold is then set, which typically has an upper and lower limit based on the maximum permissible number of

mismatching variables and the minimum number of matched variables required for a pair of records to be a possible match.[295] Records with total weights above the upper threshold are then automatically accepted as matches and those below the lower threshold are rejected as non-matches.[295] Record pairs falling between the thresholds in the so called “grey area” require manual checking to decide if they are matches or not.

Furthermore, efficient methods of validating probabilistic record linkage have been widely reported.231



Box 3: Probabilistic linkage process.

Methodological considerations

In Australia routine linkage of health record databases is particularly well established in some regions, including the Western Australian Data Linkage System.[299;300] Reporting guidelines have been developed to evaluate the methodological quality of studies using linked data, ensuring that researchers can consistently report and accurately appraise and

interpret findings derived from such studies.[301] This is important since incomplete data linkage can result in systematic bias in reported clinical outcomes.[302]

As discussed earlier (see sections 3.3 & 3.4), possible methodological limitations of linking healthcare records include the degree to which data can be anonymised without losing vital information and the successful matching of records is dependent on the quality of unique identifier variables.[189] With neonates for example, this can be particularly challenging as NHS number may not always be recorded or maternal details may be used instead. However, provided linkage methods are tailored to the datasets being used they can correctly identify a high proportion of matched records even with limited quality or availability of identifiers.[211] Furthermore, efficient methods of validating probabilistic record linkage have been widely reported.[241] The key to making linked resources useful and accessible is ensuring confidentiality and secure data storage.[189;212]

Adjustment for patient case-mix will be essential to allow for variations in co-morbidity and patient profile when comparing resource utilisation, survival and other outcomes. Other statistical issues also need to be carefully considered, such as dealing with multiple measures per patient, multiple levels of data clustering (by neonatal unit/network/hospital), transfers between neonatal unit, missing data and imputation, adjustment for confounding factors and minimising the time between data collection and its analysis to allow near real-time monitoring of outcomes.

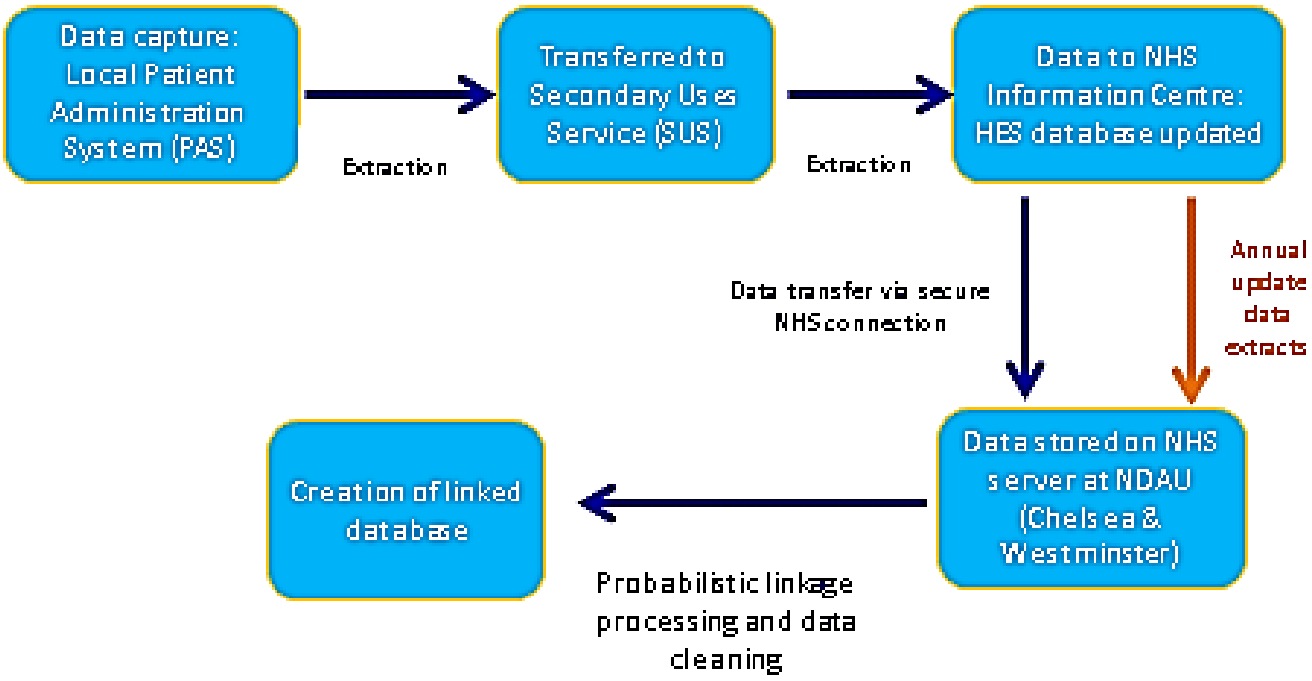
9.4 Progress to date and on-going work

Delays in receiving the necessary ethical approvals and obtaining the identifiable HES data have consequently limited the amount of progress that could be made with the linkage stage of this work, given the 3 year time constraints of this PhD (from January 2010 to December 2012). However, I will now describe the considerable progress that has been made.

HES inpatient, outpatient and Accident and Emergency (A&E) data from financial years 2006 to 2011 were obtained. The HES data capture and transfer pathway is illustrated in Box 4 to show how these data arrive at the NDAU for linkage to the NNRD. The data were uploaded, merged and saved onto the secure NHS server at the NDAU in the Chelsea & Westminster NHS hospital. The HES data fields were then cleaned and formatted, for example numeric admission and discharge date fields were converted in dates. These HES data include the following personal identifier variables which are also present in the NNRD database and will therefore be used to link records across the two databases:

- NHS number
- Date of birth
- Post code
- Sex
- Hospital number

Neonatal records are typically available in the NNRD within 3 to 6 months of collection at the point of care. There is typically a longer delay in the availability of national HES data but the NDAU will obtain annual updates of HES data when it becomes available at the NHS Information Centre.



Box 4: HES data capture and transfer pathway.

Management and analysis of the data are being conducted using SAS software (SAS Institute, Cary, N.C., USA). Initially development and testing of linkage methods will be piloted using a small subset of records from babies admitted to Chelsea and Westminster neonatal unit only, from the years 2007 to 2010 so that sufficient data from subsequent hospital admissions will be available in the HES database.

The following steps will then be taken to develop a linked database:

- Examine how many records in the two databases could be linked by straight forward deterministic methods using only patient NHS number.

- Develop algorithms to assign weights to variables for probabilistic linkage.

- Test the linkage of electronic neonatal records to HES hospital admission records using various thresholds depending on the desired specificity and sensitivity and decide on the optimum cut-off to use.

- Compare agreement between HES and NNRD neonatal data using fields common to both databases such as gestational age and birth weight (using standard Bland & Altman methodology for continuous measures and Kappa coefficients for categorical variables).

- Examine secondary care activity, post-discharge resource utilisation and case-mix adjusted clinical outcomes among the babies with linked records.

- Logistic regression modelling will be used to examine mortality after discharge from a neonatal unit (using the HES linked ONS mortality data to identify deaths occurring both in and outside of hospitals).

- Validation tested by comparing wherever possible to outcomes reported in other population-based datasets,[303] such as data from ONS birth registrations or data from large birth cohort studies such as EPICure.

Best practice will be followed, ensuring personal identifiers are removed from the dataset and a new unique child identifier key will be derived so that data that will be subsequently provided to researchers will not be identifiable.[304]

An example of a pilot study which could provide a useful demonstration of the utility of these linked records, could be examining the risk of subsequent hospital admission for RSV bronchiolitis in the first year of life after discharge from a neonatal unit. Indeed the JCVI in the UK have recommended that more detailed information on the specific risk of RSV infection by gestational age is needed. Neonatal records provide more detailed and complete information on gestational age (by weeks and days), so the NNRD data linked to HES would add the longitudinal admission records to make such a study feasible on a national scale.

9.5 Summary of chapter

In this chapter I have described the substantial progress that has been made towards linking individual neonatal care records from the NNRD to subsequent HES hospital admission records, to provide longitudinal follow-up of neonates post discharge. This project is part of a continuing body of research within the NIHR-funded "Medicines for Neonates" research programme. Having attained ethical approval to carry out this linkage and obtained the required identifiable datasets, piloting of the probabilistic methodology to link individual records is now progressing. Establishing the feasibility and methodology for linking neonatal records to extract clearly defined information post discharge, is essential to realising the potential of the NHS in research spanning the life course, providing the possibility of cradle to grave research studies. Once linkage of the NNRD to HES is completed and validated the next aim for this on-going work will be to link neonatal care records to other NHS data such as general practice records.

10.0 Discussion

Chapter overview

This chapter brings together the findings from the three results chapters, relating to each of the three objectives of this thesis. The chapter begins by restating the thesis aim and research questions with an overall summary of the main findings from this thesis. This is followed by an individual discussion of each of the three studies. I then describe the overall strengths and weaknesses of the methods used in this thesis. There is then a discussion of how the thesis findings relate to what the existing literature tells us and what the implications of this research are, for clinicians and policy makers.

10.1 Overall summary of thesis findings

The overall aim of this PhD thesis was ***to improve estimates of the wider clinical burden of RSV bronchiolitis among infants in the UK,*** by examining the full clinical spectrum of bronchiolitis illness across different healthcare settings. Three research questions were derived to tackle this and the key findings relating to each of these are summarised below and in Figure 8.

Research question 1: What are the risk factors for admission to hospital with RSV bronchiolitis and how do admission rates vary between infants with comorbidities, preterm and term born infants?

- Hospital admissions for unspecified bronchiolitis have been increasing since 2004/05.
- The estimated bronchiolitis admission rate in English NHS hospitals is 24.2 admissions per 1000 infants under 1 year (95% CI 23.7 to 24.8) with a median length of stay of 1 day (IQR 0 to 3 days) and at a median age of 120 days (IQR 61 to 209 days).
- 15% of infants admitted with bronchiolitis are born preterm (47.3 per 1000 infants) and 24% have at least one known clinical risk factor for severe RSV infection.
- Risk of bronchiolitis admission is significantly higher in infants with known clinical risk factors for RSV infection or with cystic fibrosis, Down's syndrome or cerebral palsy.

Research question 2: What is the incidence of acute bronchiolitis among infants presenting to general practitioners and how is it being managed in the community?

- The bronchiolitis consultation rate in UK general practice is estimated to be 58.1 per 1000 infants under 1 year (95% CI 56.5 to 59.8) at a mean age of 5.5 months (SD=3.2).

- 36% of bronchiolitis consultations result in a prescription, of which 28% are for antibiotics, 27% for beta agonists and 8% for antipyretics or analgesics.
- Using a broader bronchiolitis case definition which includes less specific symptomatic codes, the estimated consultation rate may be as high as 206.7 per 1000 infants.

Research question 3: What is the long-term impact of RSV bronchiolitis on subsequent respiratory health in early childhood?

- An episode of bronchiolitis in infancy is a predictor of subsequent hospital admissions and general practice consultations for asthma and wheezing in early childhood.
- Subsequent hospital admissions for asthma and wheeze are particularly likely following an episode of bronchiolitis among high-risk infants such as those born preterm.

Figure 8: Pyramid summarising estimates of RSV bronchiolitis across healthcare settings, derived from findings in this thesis.

10.2 Discussion of study 1: Bronchiolitis hospital admissions

This section discusses the findings presented in chapter 6.0 which relate to the first research question and first objective of this thesis.

10.2.1 Main findings

Infant hospital admissions for diagnoses specifying RSV as the causative organism have remained stable during the 14 year period from 1997/98 to 2010/11. Unspecified bronchiolitis and unspecified viral pneumonia admissions were relatively stable between 1997/98 and 2004/05. Unspecified bronchiolitis admissions have since increased 43.5% to 33 per 1000 live births in 2010/11, with unspecified viral pneumonia admissions increasing 36.3% over the same 6 year period, to 60 per 1000 live births.

Most (85%) infants admitted to hospital with bronchiolitis in England are born at term, with no known predisposing risk factors for severe RSV infection. Risk of admission is higher in known risk groups. Babies born with Down's syndrome, cerebral palsy and cystic fibrosis are also at higher risk of hospital admission with RSV bronchiolitis. The early age of bronchiolitis admissions has important implications for the potential impact and timing of future active and passive immunisations.

10.2.2 Strengths and weaknesses

This is among the first studies reporting the population burden of bronchiolitis across NHS hospitals in England and reporting trends in RSV-associated admissions over the last decade. The main strength of the cohort is that use of HES birth records provides a large, nationally representative sample of births across England during this time, when compared with ONS live births data.[166] It was assumed that infants with unknown gestational age and birth weight were not born preterm, as infants in this unknown group had similarly short length of stay at birth to infants born at term. Despite incomplete recording of gestational age in some birth records, 7.5% of the cohort were born at < 37 weeks gestation which is consistent with ONS data reporting that around 7.7% of live births annually are delivered preterm.[305] This suggests under-reporting of gestational age in HES birth records appears to be mostly among term babies. Similarly, rates of congenital heart disease, chronic lung disease and other risk factors for severe RSV disease compare well with published UK population data.[31;306-311]

There are a number of important limitations to this study. The case definitions for RSV-associated illness, bronchiolitis and co-morbidities are dependent on the accuracy of clinical coding and recording in diagnosis fields in HES records, although improvements in the quality of HES coding have been demonstrated in recent years.[240;312] I combined RSV and

unspecified bronchiolitis, presenting data on *all* bronchiolitis admissions. Only 28% of the bronchiolitis admissions were coded as being due to RSV (ICD-10 code J210), the remainder having unspecific bronchiolitis codes ('ICD-10 codes J218 and J219). No linked microbiological data to confirm the presence of RSV were available. A previous study in the UK has estimated that around 75% of unspecified bronchiolitis admissions were RSV related, but this was estimated using a regression modelling approach so was also not based on microbiologically confirmed RSV cases.[30] Globally it is estimated that only between 4% and 28% of children admitted with LRTI are tested for RSV.[21]

A limitation of the cohort study design is attrition, as it was not possible to identify bronchiolitis admissions occurring outside NHS hospitals in England, or to identify individuals who migrated elsewhere, though I estimate this number should be low with only one year of follow-up. Migration data from the ONS suggests the number of infants migrating out of England may be small, for example approximately 30,000 children under 15 years migrated out of England and Wales in 2008.[313] Using linked ONS mortality data it was possible to identify the small number of individuals in the cohort who died during the study period. The number of infant deaths directly attributable to RSV infection is reported to be very low.[191]

There is potentially some inherent selection bias in the study, through the inclusion of only hospitals with good completeness of recording of birth record fields, but sensitivity analyses showed few differences between included and excluded sites in terms of supply factors (such as number of maternity beds and number of deliveries). No data on other possible confounders were available, particularly environmental exposures such as passive smoking, residential overcrowding, air pollution and nursery/day care attendance. No data were available to identify any individuals who had received palivizumab prophylaxis against RSV infection. However, there are very strict criteria for its use in the UK due to its significant cost, which means that very few babies in the cohort are likely to have received it.

10.2.3 Findings in relation to other studies

No comparable national studies were found to compare to these estimates of bronchiolitis hospital admission rates in the UK but one previous study from Shropshire reported a bronchiolitis hospital admission rate of 30.8 per 1000 infants in this region of England,[31] compared with my estimate of 24.2 admissions per 1000 infants. Another study estimated the annual incidence of hospital admissions attributable to RSV in the late 1990s reporting a rate of 28.3 per 1000 infants under 1 year, but the authors used linear regression modelling to predict the proportion of admissions with unspecified aetiology that were attributable to different organisms.[30] My estimate may be slightly lower than both of these because I applied a narrow case definition including only infants admitted with a primary diagnosis of

bronchiolitis (ICD-10 ‘J21’codes). This is likely to underestimate the total burden of RSV illness in infancy as I have excluded pneumonias due to RSV and non-specific, symptomatic clinical codes from the bronchiolitis case definition, though some of these admissions may in reality have been related to RSV infection. Another population-based study in the UK reported that 1.1% of the Avon Longitudinal Study of Parents and Children (ALSPAC) birth cohort, were admitted to hospital with confirmed RSV bronchiolitis in their first year of life.[147]

A study from the USA has reported a bronchiolitis admission rate of 31.2 per 1000 infants under 1 year,[10] but a more recent study in Texas reported higher bronchiolitis admission rates of 5.5% among children under 2 years.[45] These higher incidence estimates are likely to be due to the wider inclusion of children aged 12 to 24 months in their study, coupled with variation in the clinical definition of bronchiolitis used in the US compared with that used in the UK.[13] Zhou *et al* examined hospital discharge data from 13 states representing approximately 40% of the US population and calculated an RSV hospitalisation rate in infants under 1 year of 2350 per 100,000 person-years (95% CI 2220 to 2520 per 100,000 person-years).[314] Similarly Holman and colleagues estimated a rate of 2700 RSV hospitalisations per 100,000 person-years among infants in the USA.[315] Both of these rates are more comparable to my findings particularly since infants over 1 year of age were excluded in these studies.

One previous study attempted to estimate the proportion of unspecified admissions that may be attributable to RSV,[30] but this is the first recent study to my knowledge to examine trends in RSV-associated admissions over a 14 year period, giving an insight into how coding of these admissions may have changed over time too. Admissions coded as specifically being due to RSV remained stable over the time period, but from 2004 the unspecified diagnoses increased substantially, the majority of which may well in reality be attributable to RSV. This increase may be explained in part by the general trend in rising childhood admissions since 2004, which coincides with the changes in provision of out-of-hours primary care services at this time.[6] However, the rise in child admission rates is not on the same scale as that shown in Figure 2. So although rising short stay admissions among children may in part have contributed to this increase it does not entirely explain the increases in unspecified viral pneumonias and bronchiolitis.

Diagnostic trends over time may also have impacted on the observed admission rates, particularly in the earlier years studied when bronchiolitis may commonly have been recorded as wheezy bronchitis or asthma in older infants. Further examination of trends in asthma and wheezing admissions, among both children under 1 year and those aged over 1

year is now warranted. This would help to identify any diagnostic transfer from bronchiolitis to asthma admissions, particularly in older infants.

Improvements in the completeness of diagnostic recording in hospital admission records may also explain the trend in increasing unspecified bronchiolitis admission rates. It is likely therefore that the increase has been driven by a combination of factors which may include a real increase in RSV disease incidence; rises in the incidence of other viral LRTIs (particularly influenza); improvements in completeness of diagnostic recording, changes in healthcare-seeking behaviour and a lower threshold for admitting children to hospital.

21% of infants admitted with bronchiolitis, had more than one bronchiolitis admission during their first year of life. Previous infection with RSV only confers partial immunity, so it is possible for individuals to be re-infected with the same or different RSV strains over the course of one season. Re-admissions for bronchiolitis among older infants may in reality be incorrectly diagnosed episodes of post-bronchiolitic wheezing. The modal age of bronchiolitis admission was 1 month and the median was 4 months, confirming findings reported elsewhere which show increased risk of RSV admission with younger age.[36;98;316-319] In recent years a trend in decreasing age of patients with bronchiolitis has been emerging across Europe.[14;320;321] It is unclear why RSV may be affecting babies at a younger age than has previously been reported. It could be associated with changes to the virus itself or maternal immunological priming, but further investigation into what may be driving this trend is needed. The median age of bronchiolitis admission was 4 months which is consistent with other studies where RSV hospital admissions are typically reported during the first 3 to 6 months of life.[22;322] Although, there is some debate regarding the most common age of RSV admissions.[317] A population-based cohort in Denmark found the median age of admission for RSV infections was 6 months.[323] Similarly, a recent prognostic birth cohort study reported that the median age at time of RSV LRTI was 6 months.[324] Emergency short stay hospital admissions in this age group have been increasing.[6] In 2006 the unplanned admission rate among children less than 1 year in England was 181 per 1000 child years.[6] Our findings suggest that approximately 1 in 6 of these unplanned admissions was for bronchiolitis.

In this study the majority of infants admitted with bronchiolitis were not in high-risk groups, with only 24% (1722/7189) having one or more recognised clinical risk factors for severe RSV infection. In a smaller US study, 34% of RSV-infected patients (189/564) had at least one high-risk condition for RSV infection.[44] It is difficult to determine the effect of reverse causality on bronchiolitis admissions among infants with conditions known to increase risk of RSV infection. A clinician's knowledge of an underlying co-morbid condition such as Down's syndrome may increase the likelihood of a child being admitted to hospital. Preterm infants

accounted for only 15% of bronchiolitis admissions in our study. A small study of five paediatric intensive care units (PICUs) in London also found that the majority of infants admitted were born at term, so current preventive interventions such as palivizumab have minimal impact on PICU workload.[325] Infants with chronic lung disease or congenital heart disease each accounted for around 4% of bronchiolitis admissions. This is consistent with findings reported elsewhere, suggesting that the majority of infants with severe RSV infection do not have any underlying risk factors.[44] A cohort study in the Netherlands has also demonstrated that Down's syndrome is a novel independent risk factor for severe RSV LRTI.[326] 6.4% of infants in my study who had cystic fibrosis were admitted with bronchiolitis, slightly lower than the 8.8% reported in an Australian study.[327] Low rates of bronchiolitis admissions in some high-risk groups may also be driven by palivizumab use. However, no national prescribing data for palivizumab is available for us to estimate the number of admissions its use may have prevented. The median age of bronchiolitis admission was only 18 days later in preterm infants (136 days) compared with those born at term (118 days), so onset of RSV disease appears to occur at a similar corrected age. Median LOS for bronchiolitis admissions was calculated as a proxy marker of severity of illness and was found to be similarly short (1 day) among both term and preterm born infants.

Geographical variation in bronchiolitis admission rates may be a proxy for other risk factors for hospital admission with severe RSV infection, such as maternal smoking or deprivation.[99;328] Variation in diagnostic coding practices across England could also have contributed to these regional trends in bronchiolitis admissions, although audits of discharge coding have shown improvements in accuracy throughout the UK.[329-331] I think this finding is most likely driven by variation in clinical practice, with different thresholds for admission in different regions as there are no national evidence based guidelines giving criteria for admission. Indeed, widespread variation in hospital management of bronchiolitis has previously been reported elsewhere.[13]

10.2.4 Implications and future research

This study has highlighted that the burden of RSV bronchiolitis hospital admissions among infants in England predominantly affects those born at term, without any risk factors for severe RSV infection and the age at admission appears to be significantly lower now than previously reported. Since the majority of infants hospitalised with bronchiolitis are below 6 months of age, this will limit the impact of any future vaccine which is unlikely to be capable of eliciting immunity in such young infants. This has important implications for future vaccination strategies, suggesting that the cost-benefit of both active and passive vaccines needs detailed evaluation. I found that the median length of stay for bronchiolitis admissions was only 1 day. Improved management of symptoms in the community, through both

primary care and A & E services may further reduce the need for admission. Additional studies examining the clinical burden of bronchiolitis in the community are needed and could confirm the clinical and health service burden to be greater still.[36;111;112] Improved management of RSV is achievable through the use of improved, evidence-based guidelines.[1] The regional variation in bronchiolitis admission rates I have presented suggests admission criteria vary widely and there is a need for clearer guidance about which infants to admit to hospital in the UK.[328]

The findings reported in this study suggest that in addition to the recognised risk factors for RSV bronchiolitis hospital admission, namely congenital heart disease, chronic lung disease and preterm birth, several other clinical conditions may also increase risk. The factors I have identified are cerebral palsy, cystic fibrosis and Down's syndrome. I suggest further investigation is warranted, into whether infants with these conditions would also benefit from palivizumab immunoprophylaxis.[332]

10.2.5 Conclusion

Bronchiolitis is an important cause of hospital admissions among infants in the UK. Several clinical subgroups are at increased risk of hospital admission, but the majority of bronchiolitis admissions in England are among infants born at term, with no known clinical risk factors for severe RSV infection. This confirms findings previously only reported in the USA showing that the majority of infants admitted to hospital with RSV infection are born at term. The early age at which infants are typically admitted with bronchiolitis is an important finding since this will have implications for the timing and potential effectiveness of any future active vaccine which pharmaceutical companies are trying to develop. This is the first study from a large, population-based cohort in England, which has provided evidence that babies born with Down's syndrome, cystic fibrosis and cerebral palsy are also at higher risk of hospital admission with RSV bronchiolitis. More research is needed to understand why this is and whether these babies might also benefit from palivizumab prophylaxis.

10.3 Discussion of study 2: Bronchiolitis in the community setting

This section discusses the findings presented in chapter 7.0 which relate to the second research question and second objective of this thesis.

10.3.1 Main findings

The bronchiolitis GP consultation rate was 58.1 per 1000 infants under 1 year, among our large birth cohort, at a mean age of 5.5 months. Using a broader bronchiolitis case definition the estimated bronchiolitis consultation rate could be as high as 206.7 per 1000 infants under 1 year. There was a peak in bronchiolitis consultations during December and January, corresponding with the typical seasonal peak in RSV infections. In the 2 weeks before a bronchiolitis consultation, infants commonly present to their GP with consultations for upper RTIs, coughs, chest infections and wheezing symptoms. The relative risk of having a respiratory consultation in the first year of life was 1.26 (CI 1.23 to 1.29) among infants who consulted with bronchiolitis compared with those who did not. 9.5% of infants had a subsequent bronchiolitis consultation in the 2 weeks following an initial diagnosis. Antibiotics and beta agonists have been widely prescribed by GPs to infants presenting with bronchiolitis in primary care.

10.3.2 Strengths and weaknesses

The main strength of this study was the large number of infants (80,976) included in the birth cohort, with detailed information on their general practice consultations during their first year of life. Use of records from 600 general practices in the GPRD ensures these findings reliably reflect the burden of bronchiolitis across primary care and are generalisable to the whole UK population, although I did not have access to data to determine whether GPRD practices differ in any respect from other general practices. However, the general validity of using the GPRD for epidemiological research has been widely reported[192;333] and coded diagnoses and prescriptions in the database should be highly reliable since the patient records are completed at the point of clinical contact.

This is the first study that I am aware of to estimate the burden of bronchiolitis illness in primary care, facilitated by the use of anonymised patient records from the GPRD. Bronchiolitis is a clinical diagnosis and although we did not have microbiological data to confirm the presence of RSV, the timing of bronchiolitis consultations corresponds very closely with the characteristic seasonal peak in laboratory reports of RSV infections collected by the HPA.[36;39]

A possible limitation of this study was the completeness of coding of acute bronchiolitis consultations in the GPRD. It is likely that I have underestimated the incidence of

bronchiolitis in general practices because GPs could also have recorded the diagnosis with a non-specific symptomatic code or in the free text sections which I did not have access to. An earlier study examined use of the GPRD for respiratory epidemiology research and found it to be valid when comparing consultation rates to the Fourth National Morbidity Survey in General Practice (MSGP4).[334] The authors suggest "acute bronchiolitis and bronchitis" rates in the GPRD were an order of magnitude lower than expected when compared with the MSGP4 estimates, unless combined with non-specific codes for "chest infections".[334] Thus it is likely the more conservative definition used in this study underestimates the clinical burden of bronchiolitis consultations in primary care since it identified only consultations with a specific clinical code indicating a bronchiolitis diagnosis. I tested the impact of expanding the conservative definition of acute bronchiolitis to include consultations in the first year of life with other diagnosis codes which are also likely to represent infants with bronchiolitis, such as those with "chest infection", "chesty cough", "wheezing" and "asthma". Using this wider bronchiolitis definition I identified 16744 "bronchiolitis" consultations among the birth cohort during the first year of life, giving a much higher estimated bronchiolitis consultation rate of 206.7 per 1000 infants under 1 year (95% CI 204.0 to 209.6), suggesting as many as 1 in 5 infants require primary care for an episode of bronchiolitis.

I have previously examined hospital admission rates for RSV bronchiolitis among term and preterm infants in England (data reported in chapter 6.0). Among this cohort of infants with severe bronchiolitis requiring hospital admission, 51.5% were reported to have been admitted via an Emergency Department and 36.4% via referral from a GP. It is difficult to determine the typical care pathway for an infant with bronchiolitis as it is dependent on the individual course of illness. The recording of referral information is variable between general practices and I was unable to determine the accuracy and completeness of this recording within the GPRD so could not include referrals in this analysis. It may be that only the most severely ill infants are directly referred to hospital following a GP consultation, while others may consult their GP but are then only admitted following presentation at an Emergency Department when their condition has deteriorated. It remains unknown whether parents take more severely unwell infants directly to an Emergency Department or whether infants with bronchiolitis that is similar in severity are treated differently in hospital settings. More research is needed to clarify how many infants who are admitted to hospital via an Emergency Department have previously attended primary care and how many are taken directly to an Emergency Department as the first point of contact.

Access to more detailed information about some other potentially confounding variables would have benefitted this study, such as prematurity which is known to be a significant risk

factor for RSV bronchiolitis[89;289] and also deprivation, as more deprived children have higher GP consultation rates [335] and higher risk of RSV infection.[99] I was unable to identify any infants in the cohort who had received palivizumab prophylaxis against RSV infection, since this is not typically prescribed in general practice and is only recommended for use in a few extremely high-risk infants.

10.3.3 Findings in relation to other studies

Existing knowledge of the burden of bronchiolitis has primarily come from hospital-based studies, as an estimated 3% of infants are hospitalised with bronchiolitis each year.[31;34] Until now limited information regarding the more minor end of the RSV disease spectrum and acute bronchiolitis consultation patterns in the community has been available. Some of the first serological RSV studies examined the clinical and patient characteristics of infants with RSV infection in hospital and outpatient settings[336] and an early prospective birth cohort study with longitudinal follow-up reported an RSV infection rate of 68.8 per 100 infants aged under 1 year, among whom the rate of symptomatic lower respiratory tract disease was 22.4 per 100 infants.[46] To my knowledge, this is the first study to estimate the incidence of bronchiolitis in the primary care setting. This was a key recommendation for future research outlined in the national guidance for management of bronchiolitis published in 2006.[36]

Age is an important factor in determining the potential impact of any future active vaccination against RSV infection. Data from the US have shown the highest proportion of RSV related hospital admissions occurs among the 3 to 6 month age group.[337] In the HES cohort described earlier (chapter 6.0) using data from hospitals in England, I found the median age of bronchiolitis admission to be at 120 days (inter-quartile range 61 to 209 days), with a positively skewed age distribution showing a peak in admissions at the age of 1 month. In this primary care study there was an approximately normal age distribution, with a mean age of 5.5 months for infants with bronchiolitis presenting to GPs. It is unclear why the age distribution of infants hospitalised with bronchiolitis appears to be skewed towards younger infants than those presenting in primary care. This could be a reflection of disease severity, with younger infants who may be immunologically more susceptible to infection more likely to be admitted to hospital. Publishing of the National Institute for Health and Clinical Excellence (NICE) clinical guideline for the management of children with febrile illness, may also have impacted on management of bronchiolitis in both primary and secondary care.[5]

National guidance for management of bronchiolitis was published later than this 2003 cohort study. No available treatments shorten the natural course or provide clinically relevant symptomatic improvements from this viral illness, [33;338] including inhaled and oral corticosteroids, beta agonists and antibiotics.[61;62;64;76;339-342] This study shows that despite no

evidence for their use, GPs have prescribed antibiotics and beta agonists to infants diagnosed with bronchiolitis, risking adverse effects including disruption of gut flora and the development of antimicrobial resistance.[343] Although not from primary care, studies from Canada and across Europe also provide evidence of variation in RSV bronchiolitis treatment between different hospitals.[344;345]

Following the creation of new bronchiolitis guidelines in Switzerland, an intensive implementation programme tailored to the specific needs of hospital and primary care based paediatricians and generalists, was shown to be highly effective and resulted in significant reductions in the reported prescription of bronchodilators, inhaled steroids and antibiotics for the management of bronchiolitis.[33] A particularly successful component of this quality improvement programme was an information leaflet for parents which highlighted the lack of evidence to support the use of any drugs for the management of viral, self-limiting illnesses such as bronchiolitis.[33;346] Houben et al carried out a prospective birth cohort study in the Netherlands in which they showed that a simple clinical prediction rule can be used to identify healthy newborns at risk of RSV LRTI, allowing primary care paediatricians to target preventive and monitoring strategies towards high-risk infants, such as warning parents about the risks associated with passive smoking.[99;324]

10.3.4 Implications and future research

Studies from elsewhere in Europe have shown that improved management of RSV is achievable through the use of improved, evidence-based guidelines[33] and simple clinical predictions rules can be an effective tool for identifying infants at high risk of RSV infection in primary care.[324] Clearer guidance about which infants to admit to hospital are required and a firmer evidence base from the community setting is needed to develop up-to-date recommendations for bronchiolitis management in primary care. Increasing emergency hospital admission rates among young children in England are mostly for short stay minor illness episodes such as bronchiolitis,[6] which may be potentially avoidable if better managed by primary care in the community.

A GP consultation rate of 58.1 per 1000 infants under 1 year would result in approximately 46480 bronchiolitis GP consultations across the UK, given the current live birth rate of around 800,000 per year. Although as previously discussed, the reliance on clinical coding means this is likely to be an underestimate of the total burden of illness in the community. Using the less conservative and perhaps more realistic bronchiolitis definition, I estimated a consultation rate of 206.7 per 1000 infants under 1 year, which would equate to approximately 165,360 GP consultations across the UK each year. Analyses of data from Emergency Department services are needed, which could confirm the clinical and health

service burden of RSV bronchiolitis to be greater still.[36;111;112] Diagnoses in A&E data are coded using their own specific system and not ICD-10 codes, so linkage of A&E data to HES inpatient records would be necessary to investigate this further.

Further work is needed to determine the typical care pathway of infants admitted with bronchiolitis. This could be facilitated through the use of linked patient medical records from general practices, Emergency Departments, hospitals and laboratories. Population-based linkage between hospital admission records and laboratory data in Australia has highlighted how linkage to laboratory records can improve the accuracy of estimating the burden of RSV disease compared with relying on clinical diagnosis coding alone.[347]

These findings have important implications for future vaccination strategies and identifying a target population for emerging active immunisations against RSV. Previous studies have attempted to determine the total cost of RSV associated illness, in order to estimate the potential impact and cost-effectiveness of passive and active immunisations against RSV infection.[111;112] However, the total burden of disease can only be estimated by examining the full, wider spectrum of illness and not only costs of hospitalisation but also costs of caring for infants with milder RSV-associated illness provided in the community. Prospective studies are also needed, to determine the long-term impact of RSV infection in infancy on subsequent respiratory health.

10.3.5 Conclusion

This is the first study to quantify the significant burden of acute bronchiolitis among infants presenting in general practices across the UK. Over 4% of infants are diagnosed with bronchiolitis during a GP consultation in the first year of life, typically during the seasonal winter RSV peak and around 10% of these re-consult with bronchiolitis during the following 2 weeks. Using a broader case definition which includes symptomatic diagnosis codes, it is estimated that as many as 21% of infants may have an episode of bronchiolitis requiring a GP consultation. The findings suggest antibiotics and beta agonists have been widely prescribed by GPs to infants presenting with bronchiolitis, despite no evidence base for their use. Estimating this burden of illness in the community setting, at the population level, has important policy implications for GP training and potentially for future passive and active RSV immunisation policy.

10.4 Discussion of study 3: Impact of bronchiolitis on respiratory health in early childhood

10.4.1 Main findings

Two cohort studies, one using data from primary care and one from secondary care settings, were developed to examine longer-term associations between RSV bronchiolitis in infancy and subsequent respiratory morbidity in early childhood.

HES cohort study with follow-up to age 5 years

The HES cohort study of 211772 children with follow-up to age 5 years showed that the relative risk of hospital admission with asthma or wheezing was 6.96 and 6.52 times greater respectively, among children with a history of bronchiolitis admission compared with those without. 12.0% of children admitted with bronchiolitis were subsequently admitted with asthma by age 5 years and 8.5% for wheezing illness. In comparison, among children with no prior bronchiolitis admissions 1.7% had subsequent asthma admissions and 1.3% had subsequent wheeze admissions by age 5 years. Children who had been admitted with bronchiolitis during infancy were also more likely than those who had not, to be admitted to hospital for URTI or LRTI. High-risk children with a history of bronchiolitis were most likely to have subsequent respiratory admissions, followed by low-risk children with a history of bronchiolitis, then high- and low-risk children with no previous bronchiolitis admissions.

GPRD cohort study with follow-up to age 7 years

Findings from the primary care cohort of 80,976 children showed a similar association between history of bronchiolitis and subsequent respiratory consultations in early childhood. 25.4% of children who had a bronchiolitis consultation in infancy had at least one subsequent asthma consultation by the age of 7 years and 57.6% had at least 1 consultation for wheezing illness. In comparison, among children with no prior bronchiolitis consultations 11.8% had subsequent asthma consultations and 13.9% had subsequent consultations for wheeze by age 7 years. So children with a history of bronchiolitis consultation in infancy were 2 times more likely than those with no prior bronchiolitis consultations to consult their GP for asthma (RR = 2.15) and 4 times more likely to consult for wheezing (RR = 4.14) by the age of 7 years.

10.4.2 Strengths and weaknesses

Strengths

These are the first large-scale, population-based cohort studies to report the longer-term impact of RSV bronchiolitis on subsequent respiratory consultations and hospital admissions in early childhood, using individual-level clinical data from children in the UK.

The main strength of both cohorts was the large number of children they included, from across England (for the HES cohort) or the whole of the UK (for the GPRD cohort). Existing evidence of the impact of RSV infection on subsequent risk of asthma and other respiratory morbidity has predominantly come from small, prospective cohort studies in the USA and Sweden. The only UK data on school age respiratory outcomes following RSV bronchiolitis are from small regional studies following up high-risk infants born preterm.[156;289] These are the first population scale studies from the UK to provide estimates of wheezing and asthma incidence following RSV infection.

Deriving large nationally representative cohorts by using routine clinical records in this way ensures that the study findings are likely to be highly generalisable to the wider population of children in the UK. As discussed in relation to the two earlier studies (see sections 10.2.2 and 10.3.2), another key strength of using HES and GPRD data is that their general validity for use in health services and epidemiological research has already been widely reported.[192;312;331;333] Coding of diagnoses in the GPRD should be highly reliable since the patient records are completed at the point of clinical contact and it has been shown that the accuracy of diagnoses recorded in HES has greatly improved.[331]

Weaknesses

The main weakness of this study was that neither of the databases used to derive these cohorts contained information on environmental risk factors for severe RSV infection, such as household smokers, nursery/day care attendance, number of siblings or breast feeding duration, which are all potential confounders that could not be controlled for.

A possible weakness of these retrospective cohort studies is attrition. It is not possible to determine exactly how many of the children in the HES or GPRD birth cohorts may have migrated away from the area in which they were born but as described earlier, outgoing migration rates are low among children in the UK.[313] In the GPRD cohort it was possible to identify individuals who had transferred out of the cohort because they had left the GP practice and registered elsewhere, but there may also have been “ghost patients” who changed GP practice without de-registering.[348] It was possible though in both the GPRD and HES cohorts, to identify children who died during the study period, reducing possible

biases from attrition. There is potentially some inherent selection bias in the HES cohort, since only children born in hospitals with good completeness of recording of birth record fields were included. However, the sensitivity analyses described earlier (see section 5.2) revealed this approach for creating a birth cohort should be generalisable.[257]

As the GPRD cohort study described earlier showed (chapter 7.2), using diagnostic codes for bronchiolitis alone may underestimate the total incidence, since many GPs record non-specific symptomatic codes or record diagnoses in the free text sections. A limitation of using HES and GPRD data is that it is not possible to check the validity of diagnostic coding for this particular study, since it is not feasible to identify and contact the large number of children in the cohort to clarify the diagnoses they received. A further limitation of the GPRD cohort was the lack of clinical information available about the children's birth status. Since no information on gestational age at birth was recorded in their GP records it was not possible to identify children born prematurely, who would be at higher risk of RSV bronchiolitis.

It was not possible in either the HES or GPRD cohorts to identify infants who had received palivizumab immunotherapy to prevent RSV infection, since HES contains no information on prescriptions and infants are unlikely to receive it in a primary care setting. However, I obtained national aggregate data from the Prescribing Cost Analysis (PCA) database from the NHS Information Centre, which showed that only a very small number of infants receive palivizumab annually so this is unlikely to be a major confounder in our study (Appendix 1, Chapter 11). The number of palivizumab prescriptions in this dataset is likely to be an underestimate because it does not include all in-hospital prescribing, but it is the only national estimate of palivizumab prescribing that the NHS Information Centre were able to provide. Given the stringent JCVI criteria for palivizumab prescribing,[349] the total number of infants receiving the immunotherapy is still likely to be very low and so should have minimal impact on these cohort studies.

Another limitation of both cohort studies is the lack of microbiological data to determine what proportion of the bronchiolitis consultations and admissions were due to RSV infection. However, it remains a subject of debate as to whether the etiologic agent causing bronchiolitis is clinically relevant if the course of disease caused by RSV is similar to that of other viruses.[14;320] Testing for RSV infection is not currently recommended as routine clinical practice in the UK, so both GPs and paediatricians are rarely aware of the organism causing bronchiolitis when caring for a symptomatic infant. The findings presented from these cohorts show that history of a bronchiolitis episode in infancy, regardless of the etiologic agent causing it, predicts increased risk of asthma and wheezing in early childhood. It was

not possible given the constraints of the clinical data available, to determine the impact of reverse causality on these findings, since it may be that those with pre-existing chest disease are more prone to RSV infection and to the development of asthma, rather than RSV infection itself causing the subsequent respiratory illness.

10.4.3 Findings in relation to other studies

Bronchiolitis incidence rates in this study were comparable with the HES cohort study described earlier (see section 6.2) as well as with existing estimates of RSV associated admissions from smaller regional studies in the UK and USA.[20;30;31;44] Evidence of an association between RSV bronchiolitis and subsequent respiratory complications in childhood, including recurrent wheezing and asthma, has mostly come from small prospective studies in developed countries, since long-term follow-up of children is not often feasible and expensive to do on a large scale.[10;14;140-145;164;350] A HTA study published in 2008 highlighted the major uncertainties surrounding our understanding of the effects of RSV infection on long term respiratory health among children in the UK.[111]

A study from the USA assessed respiratory outcomes and healthcare utilisation 12 months after RSV diagnosis, among commercially insured late-preterm infants.[351] They reported infantile asthma rates were six to nine times higher among infants with a history of RSV/unspecified bronchiolitis compared with those without. This is very similar to the rates reported in my cohort studies which had longer follow-up, with asthma hospital admissions seven times higher (up to age 5 years), whilst asthma GP consultations were just two times higher (up to age 7 years), among children with a history of bronchiolitis compared with those without. Their study also showed significantly higher Emergency Department and outpatient visits in children with a history of RSV LRTI compared with those without.[351] Findings from the GPRD cohort enhance this by going beyond existing evidence to show that as well as children with a history of bronchiolitis hospital admission, non-hospitalised children with bronchiolitis are also more likely to have adverse respiratory outcomes during early childhood.

This study also confirms on a larger, population scale, the findings reported by Henderson *et al* who examined data from 150 infants admitted to hospital in the UK with RSV confirmed bronchiolitis, from the ALSPAC cohort.[147] They reported the prevalence of wheezing was 28.1% in the RSV group compared with 13.1% in the control group, at age 30-42 months. Interestingly their study also found RSV bronchiolitis was associated with prevalence of asthma but not with development of atopy, by age 7 years.[147]

Subsequent respiratory illness following RSV bronchiolitis not only has an impact on an individual child in terms of their increased suffering, but also results in a significant economic burden due to the associated healthcare costs as well as wider costs from parents requiring time off work.

Palmer *et al* followed up a group of children in the USA for a year following RSV/unspecified bronchiolitis and found that they had higher healthcare costs associated with subsequent respiratory episodes, when compared with infants without a history of RSV/unspecified bronchiolitis.[352] These healthcare costs were greatest among late-preterm infants compared with full term infants. Findings from the HES cohort study confirm this, since low-risk infants had lower subsequent respiratory admission rates than high-risk infants which included those born preterm.

The subsequent costs of health care utilisation following RSV hospitalisation have been estimated in studies from the USA, UK and Canada,[103;289;353;354] but these studies have focused on following-up infants born preterm or with chronic lung disease, despite the majority of infants with bronchiolitis being born at term (see chapter 6.2). Many previous studies on this topic have been limited by an inability to generalise their findings to term infants.[127;355]

10.4.4 Implications and future research

This is the first study to identify, using cohort data from large, nationally representative populations, that subsequent healthcare utilisation in early childhood is greater across both primary and secondary healthcare settings, among not only high risk infants with CLD or preterm birth, but generally among *all* children with a history of bronchiolitis. Not only are infants with RSV bronchiolitis which required hospital admission at increased risk of adverse respiratory outcomes, but infants with perhaps milder bronchiolitis presenting to their GP are also at higher risk of further GP consultations for asthma and wheezing. This information highlights the importance of developing new preventive and therapeutic agents to tackle both the acute and longer-term effects of RSV infection in childhood. These findings also have huge implications for the NHS in terms of healthcare utilisation costs among children with a history of bronchiolitis, as this study provides the first estimate of the subsequent respiratory burden in primary as well as secondary care settings.

This is a topic of growing importance for GPs and respiratory consultants across the UK, because the burden of illness from acute bronchiolitis and subsequent respiratory morbidity is likely to rise, as preterm birth rates have increased to around 8% of births in the UK and survival of these babies is also increasing.[88;170;356;357] Early life factors are important to consider not only in childhood but throughout the life course as their impact on the respiratory system continues through adulthood.[141;358;359]

I have only been able to demonstrate the epidemiological relationship between RSV bronchiolitis and asthma and wheeze in early childhood, what remains unclear is whether there is any causal pathway underlying this association.[160] The precise role of genetic susceptibility to respiratory conditions versus interactions with the environment remains

unknown. It may be that RSV LRTI disrupts bronchiolar smooth muscle in infancy and therefore has a direct causal association with subsequent asthma, or it may simply be that affected infants already have pre-existing genetic susceptibility to sensitised airways and RSV bronchiolitis is just indicative of these susceptible, higher-risk individuals.[360-363] Recent studies from Denmark suggest severe RSV infection and asthma may share a common environmental exposure and/or genetic predisposition.[361;362] The increasing prevalence of asthma makes it even more important that this association with RSV in infancy is better understood.[364]

Rhinovirus is an important cause of wheeze in children aged 3 to 6 years.[365] There is evidence to suggest that bronchiolitis due to rhinovirus may be more severe than that due to RSV.[37;363] Indeed a study from a hospital in Finland found children with a history of bronchiolitis not caused by RSV were substantially more likely than those who had RSV bronchiolitis, to develop recurrent wheezing during a 3 year follow-up period.[366] However, another study has shown that both RSV and rhinovirus interact with atopy in infancy to promote the development of asthma.[359] Conversely, other recent evidence suggests the disease phenotype may be more important for future wheeze than the causative agent.[160] So further immunological research is needed to truly delineate the association between RSV infection and bronchiolitis illness in infancy and subsequent respiratory morbidity in childhood.

10.4.5 Conclusion

This is the first UK study to report on respiratory diagnoses in early childhood following an episode of bronchiolitis, among both high- and low-risk children, using data from nationally representative, population-based cohorts. Evidence from these cohorts of children in both primary and secondary care show that children who have an episode of bronchiolitis in infancy are at higher risk of subsequent asthma and wheezing episodes in early childhood, compared with children with no recorded history of bronchiolitis. Children who have previously been admitted to hospital with bronchiolitis are also at increased risk of subsequent hospital admissions for both upper and lower RTIs, in the first 5 years of life. Among children with a history of bronchiolitis admission, the increased risk of subsequent respiratory admissions was greatest in those classified as “high-risk” such as children born preterm. These findings have important implications for clinicians as an increased awareness of this association among GPs, paediatricians and respiratory specialists is required, to identify high-risk individuals and optimise management of affected children. When determining the potential impact of emerging treatments and vaccines against RSV infection, it will be necessary to consider how these might impact on longer-term respiratory health.

10.5 Methodological strengths and weaknesses

Earlier in this discussion chapter I have separately described the strengths and weaknesses of the methods used in each of the individual studies that constitute this thesis. I will now consider and summarise the general methodological issues that relate to the whole body of work that is my thesis.

Strengths

The primary strength of this thesis was the novel methodology developed to utilise routinely collected electronic databases of clinical records, to develop large, population-based, representative cohort studies using individual level data from children in the UK. Unlike previous epidemiological studies of RSV infection among high-risk infants, this research focused on the wider population burden of bronchiolitis illness and yielded useful, clinically relevant findings that are generalisable to infants across the UK and consistent with those reported outside of the UK.

Linkage between routinely collected hospital birth and other records offers the potential to study health outcomes by developing population level cohorts from birth with longitudinal cradle-to-grave follow-up. The strength of the retrospective cohort study design is that several different outcomes can be assessed and the sequence of events from exposure to outcome can be studied. In the HES cohort detailed clinical information from birth records such as gestational age, birth weight and comorbidities, provides crucial data to examine how early life factors might predict future health outcomes. The methods I have developed to construct cohorts using routine hospital data therefore have the potential to answer a plethora of different research questions that might otherwise not be feasible to study in child populations.

Another strength of utilising clinical data from the HES and GPRD databases was that previous studies have already tested their robustness and validity for use in health services and epidemiological research.[192;312;331;333] Unlike the artificial settings established in many experimental study designs, using routinely collected clinical data increases the likelihood of findings relating and being transferable to, real clinical practice.

Weaknesses

A limitation of HES data is that they pertain to NHS hospitals in England only. The demography and health of populations living elsewhere in the UK does differ from England but the findings reported are likely to be broadly generalisable to the wider UK population. As previously discussed, another limitation of the databases used in this thesis was the lack of microbiological data available, to confirm whether RSV was the causative agent in bronchiolitis cases. RSV bronchiolitis and unspecified bronchiolitis admissions were collated

into a single “RSV bronchiolitis” diagnosis group which was deemed justified since RSV is the commonest cause of bronchiolitis. Specific microbiological diagnoses are poorly coded in HES and testing for RSV is not standard clinical practice for bronchiolitis management. It was not possible in any of the studies described to identify individuals who had received passive immunotherapy against RSV infection, but I estimate that only a very small number of infants qualify for its use (as described earlier in section 10.4).

A weakness of not using prospectively collected data is that it makes it difficult to discount alternative explanations for reported findings, since data on potentially confounding variables may not be available. For this reason it was not possible in either of the cohorts examining the longer-term impact of RSV bronchiolitis, to conclusively determine from these findings alone, whether RSV infection in infancy has any concrete causal role in the onset of childhood asthma or wheeze. Data on many other possible confounders of this association were not available in these databases, particularly environmental exposures such as passive smoking, air pollution and nursery/day care attendance. This information would be necessary to narrow the field of doubt surrounding any proposed possible causal mechanism.

Chance, bias and confounding

When interpreting the results of any epidemiological study it is important to consider alternative explanations for any observed effect of an exposure on the outcome of interest. In particular, one must consider the role of chance (or random error), bias and confounding in generating the observed results. These are major methodological problems that can affect our interpretation of observational research studies, leading to type I (rejecting a null hypothesis that is true) and type II (not rejecting a null hypothesis that is false) errors, which result in associations that do not truly exist being found or associations that are present being missed.

When conducting a large number of statistical analyses, for example when investigating associations between exposures at birth and risk of bronchiolitis admission, around 1 in 20 of these associations will have p values below the adopted standard threshold of statistical significance of 0.05, simply by chance. It is therefore advisable to be cautious when interpreting findings from such large numbers of analyses.

Biases are systematic errors in the collection or analysis of data, or the study design itself, which violate the internal validity of a study.[367] The large cohorts studied in this thesis are more likely to capture the natural variation that exists in the populations from which they were taken and are therefore less likely than much smaller, regional studies to be biased by lack of comparability in their exposure and outcome information to the wider population.

There may be some inherent selection bias in the hospitals and GP practices which were selected for inclusion in the cohorts. However, the sensitivity analyses described earlier showed no significant variation in terms of key hospital and maternity factors, between hospitals which were included and excluded from the HES study birth cohorts. The 600 GP practices from which the GPRD cohort was derived, are selected as being up-to-standard for research and are geographically widespread, covering an estimated 8% of the UK population.[258] The retrospectively derived cohort studies used in this thesis were affected by losses to follow-up but such attrition bias is difficult to minimise and also affects prospective cohort studies.

In both the HES and GPRD cohort studies described, there was likely to be some missingness in the outcome measures (bronchiolitis consultations/admissions), due to the reliance on clinical diagnosis coding which may miss some cases coded with less specific symptomatic codes. This is likely to have resulted in underestimates of the burden of bronchiolitis in primary and secondary care. Missing data on exposures was also a problem in the studies I carried out but where possible I used alternative, national data sources to provide validation. For example, despite missing gestational age information in the HES birth cohorts, the overall preterm birth rate was very similar to national birth registration records collected by the ONS, justifying the decision to consider babies with missing gestational age to have been born at term. In addition, the approach described to minimise the impact of missing HES birth record information and the sensitivity analyses presented, provide sufficient justification that selecting a cohort of infants born in hospitals with higher completeness of recording of gestational age at birth is a valid and generalisable approach.

I considered potentially confounding variables in each study and tried to take these into account either in the study design or by controlling for them at the analysis stage (wherever this was possible given the constraints of routinely collected data). For example, when calculating the relative risk of bronchiolitis hospital admission for infants with different clinical risk factors, I adjusted for any other clinical risk factors which might be confounders. Residual confounding due to measurement error in confounders, or poorly controlled or unmeasured factors, may result in biased estimates of the effect of an exposure on the outcome of interest.[368] It was not possible to adjust for all possible confounders, particularly in the GPRD cohort where information on gestational age at birth and other neonatal exposures were not available. However, wherever data on a confounder was available, I attempted to adjust for it in my analyses to minimise the chances of identifying spurious associations between exposures and outcomes.

10.6 Implications for policy and practice

Implications for health care policy and the development of clinical guidance

The improved estimates of the clinical burden of RSV bronchiolitis which this thesis has provided are essential, because such quantitative evidence is required to form the basis of health care policy, informs the development of clinical guidelines and is useful for predicting and planning health service use. Both health professionals and policy-makers require up-to-date information on the country-specific impact of RSV infection and the extent to which it is concentrated in particular groups, in order to improve care and calculate the costs to health care providers.[325]

The importance of this burden of RSV illness is growing because the pool of susceptible individuals is increasing as the UK birth rate rises and particularly because preterm birth rates have increased to around 8% of births, plus survival of preterm babies is also increasing.[88] Furthermore, the burden is not limited to the acute episode of illness itself but is also associated with an increased risk of subsequent primary and secondary care consultations/admissions for asthma and wheezing in early childhood. So these findings have huge implications for the NHS in terms of healthcare utilisation costs among children with a history of bronchiolitis.

Despite caveats relating to the database used and reliance on clinical coding, this thesis has provided for the first time to my knowledge, an estimate of the bronchiolitis GP consultation rate in the UK, which can be used to inform future guidance for the optimal management of bronchiolitis. SIGN guidelines highlight this as a priority area for research since there were previously no estimates of bronchiolitis incidence in primary care, which I have now attempted to address.

This is the first large, population-based study in the UK which has provided evidence that babies born with Down's syndrome, cystic fibrosis and cerebral palsy are at higher risk of hospital admission with RSV bronchiolitis. More research is therefore needed to understand why this is and whether these babies might also benefit from palivizumab prophylaxis. If palivizumab is found to be efficacious and cost-effective in these subgroups of high-risk infants, then JCVI recommendations for palivizumab use may need to be revised. Indeed current recommendations for the use of palivizumab from the American Academy of Pediatrics (AAP) suggest it should be considered for use in infants with cystic fibrosis, since evidence suggests their course of illness may be prolonged with impaired lung function for several months following a LRTI.[369;370] Neither the AAP in the USA nor the JCVI in the UK specifically recommend the use of palivizumab for infants with Down's syndrome, but the

AAP do suggest it should be used in infants with compromised ability to handle secretions due to neuromuscular conditions.[127;349;369] A randomised trial of palivizumab would be required in order to evaluate whether it may reduce hospital admissions for RSV LRTI among infants with these specific conditions, but the cost implications of this are likely to be prohibitive.[122;127]

This thesis has provided valuable information which can be used to estimate the potential risk-benefit ratio of any emerging vaccines or other passive immunisations against RSV infection, at both the individual and societal level, which could then determine the target recipients.[371;372] The total burden of RSV bronchiolitis can only be estimated by examining the full, wider spectrum of illness across health care settings, hence this was previously underestimated by research focused solely on hospitalisation costs among only the highest risk groups. This thesis has highlighted the need to consider that the majority of infants affected are not born preterm and there are significant costs associated with caring for infants with bronchiolitis in the community. Furthermore, my findings have implications for the timing of any novel prophylaxis as the analyses confirmed that the majority of infants hospitalised with bronchiolitis are below 6 months of age, limiting the impact of any future vaccine since it would be very difficult to elicit immunity in such young infants. Developing an effective vaccine is challenging because immunogenicity needs to be effective during the most vulnerable period in the first 6 months of life but also balanced with the need for adequate attenuation.[127]

Implications for clinical management and service provision

Although the findings related to GP prescribing for bronchiolitis occurred before the SIGN guidance was published in 2006, prior to this it was still widely known that no treatment shortens the duration of bronchiolitis illness and antibiotics have never been recommended for use with this typically self-limiting viral illness. So it is still worrying that antibiotics and beta agonists have been prescribed by GPs to infants presenting with bronchiolitis. I believe this warrants further investigation of GP prescribing in relation to bronchiolitis consultations, particularly since the SIGN guidelines were made available. Not only would these prescriptions be ineffective in improving the infants condition, but the use of antibiotics carries the risk of adverse effects including disruption of gut flora, wastes healthcare resources, contributes to the development of antimicrobial resistance and may increase future demand from parents presenting children for similar illnesses.[192;343;373-375] Evidence based guidelines for bronchiolitis should be revised to include evidence from more recent studies and awareness of the recommended best practice may need reinforcement through GP training. The development of a simple clinical prediction rule would be a useful next step, as this could be an effective tool to help GPs identify infants at high-risk of RSV

infection and to help determine which of them require hospital admission.[113] Admission criteria have not been well established to date.[36;376]

The short length of stay for many bronchiolitis hospital admissions implies some of these may be potentially avoidable, through improved symptom management and parental reassurance from clinicians in primary care and EDs. The annual number of bronchiolitis consultations in UK general practices was estimated to be between 46480 and 165,360, depending on the diagnosis codes included in the bronchiolitis case definition. This has important implications for planning and commissioning health services, since no estimates of primary care utilisation relating to acute bronchiolitis and subsequent respiratory consultations have previously been published. Likewise, the finding that most infants admitted to hospital with bronchiolitis are previously healthy and born at term, is useful for planning health service provision. Bolton *et al* have recently highlighted how important it is for clinicians specialising in respiratory medicine to take into account early life factors.[358] They suggest that given the increasing evidence of the impact of early life factors on late respiratory health (as shown in this thesis too), awareness of this needs to be promoted among respiratory specialists and greater emphasis placed on this in their training.[358]

10.7 Remaining research questions and future work

The findings presented in each of the results chapters of this thesis provide improved estimates of the total clinical burden of RSV bronchiolitis and enhance our epidemiological understanding of this illness, among infants in the UK. However, as outlined earlier in this chapter in the separate discussion sections for each of the studies making up this thesis (see sections 10.2, 10.3 and 10.4), this research has also identified key areas where insufficient evidence currently exists and further research is now required. Here I summarise these emerging themes to guide the direction and prioritising of future studies in this field.

10.7.1 Improving our understanding of the typical care pathway for infants with RSV bronchiolitis

Further observational studies are needed to examine the clinical burden of bronchiolitis in the community. Investigation of data from Emergency Departments is particularly warranted, to help elicit what the typical care pathway is for infants with bronchiolitis symptoms. This could be facilitated by linking individual medical records from general practices, Emergency Departments, hospitals and laboratories, to provide a more comprehensive picture of the full

clinical spectrum of illness and evidence of the typical patterns of bronchiolitis presentation to healthcare services.

Such a database of linked records could then be used to evaluate the impact of any new guidance for the management of infants with bronchiolitis or to monitor the effect of any interventions implemented to reduce potentially avoidable, short stay bronchiolitis admissions. For example, simple clinical prediction rules have been shown to be a highly effective tool for identify infants at high-risk of RSV infection in primary care. [324] A similar risk calculator tool needs to be developed and piloted in the UK, which may be used to help clinicians identify individuals at risk of severe RSV infection. An electronic risk prediction tool could be useful across healthcare settings, from helping neonatologists in NICUs to identify infants requiring passive immunotherapy, to guiding management and referrals of infants presenting with bronchiolitis in general practice.[377;378] There has been some recent debate about the utility of testing the etiologic diagnosis of bronchiolitis, with evidence suggesting that it may help to predict prognosis and reduce further unnecessary diagnostic tests.[320;379] Prospective studies in the UK are needed to determine the value of routine testing for RSV among infants with LRTI.

Although it may not be feasible to conduct a prospective cohort study on such a large scale as the retrospective cohort I developed using the GPRD (consisting of over 80,000 infants from 600 GP practices across the UK), a prospective study on a smaller scale could be used to validate the findings presented in chapter 7.0. Such a study could be used to deduce how GPs are typically recording cases of bronchiolitis and whether the inclusion of symptomatic codes to identify cases of bronchiolitis provides a truer reflection of the burden of bronchiolitis illness, than relying on specific bronchiolitis codes alone. Access to the free text in patient records could provide further insight into whether reliance on clinical diagnosis coding significantly underestimates bronchiolitis incidence.

Other sources of data from primary care settings could be used to validate the findings I have presented using the GPRD. For example, The Health Improvement Network (THIN) is a database of electronic anonymised medical records from general practices in the UK using the Vision software system.[380] THIN data could be used to carry out a similar study estimating the frequency of bronchiolitis consultations in primary care.

Further investigation is needed to determine whether the significant geographical differences in bronchiolitis admission rates reflect true variation in bronchiolitis incidence or are related to specific health services factors such as hospital admission policies; differences in access to specialty services; or due to variations between tertiary centres versus district hospitals.

10.7.2 Prevention and treatment of RSV bronchiolitis

The findings presented in this thesis have highlighted the urgent need to develop novel preventive and therapeutic agents to target both the acute and longer-term effects of RSV infection in childhood. This ought to be a high priority for research since the burden of bronchiolitis illness is greater than previously estimated, has an impact across healthcare settings, effects a wide spectrum of infants and is associated with longer-term respiratory morbidity in childhood.

Further research is also needed to determine whether palivizumab immunoprophylaxis might be clinically and cost-effective if given to other high-risk infants such as those with cerebral palsy, cystic fibrosis or Down's syndrome. To test this robustly, a multicentre, randomised, double-blind placebo controlled trial would be required, but this is unlikely to be feasible because of the considerable costs of such a study.

Up-to-date economic evaluations are also needed to estimate the total costs of NHS healthcare utilisation associated with RSV bronchiolitis, incorporating not only costs of hospitalisation but also costs of caring for infants in the community and subsequent costs related to longer-term respiratory problems.

10.7.3 Improve understanding of the longer-term impact of RSV infection on airways

Further studies are needed to determine the role of genetic versus environmental factors on susceptibility to RSV infection in infancy. The immune response to RSV infection remains poorly understood, so further microbiological studies are needed if a successful vaccination candidate is to be identified. We need to determine whether there is a definite causal association between RSV infection and subsequent asthma or wheeze, or whether the disease phenotype plays a larger role in this than the etiologic agent.

10.7.4 The future of using routine data for child health research

With limited funding, restricted resources and improved recognition of the necessity for life-long follow-up for infant research, it is imperative that the research community, funders and the public maximise the value of information that can be gleaned from existing routine data sources and avoid duplication of effort. This was highlighted by the public inquiry report into the Bristol Royal Infirmary paediatric cardiac unit, which recommended that HES should be supported as a national resource and used to monitor healthcare outcomes.[381;382]

Previously HES had been commonly used to examine disease time trends and standards of care at an aggregate level. For the first time, I have demonstrated how HES data can be used to develop birth cohorts for epidemiological research. Recording of birth information needs to improve in some NHS hospitals to provide robust baseline population estimates of

birth outcomes. Encouragingly, the data I have presented from recent years indicate completeness of recording of HES birth records are improving. Individual follow-up from birth is feasible using HES but the limitations and processing issues I have described are important methodological considerations that need to be taken into account.

Data capture is likely to be more complete and most accurate in records that are entered by clinicians at the point of care. In the UK for example, Neonatal Networks submit neonatal specialist care records to the NNRD, managed by the NDAU (based at Imperial College London in collaboration with Chelsea & Westminster NHS Foundation Trust).[271] These records capture real-time operational data which are actively used in the processes of clinical care. Collaboration between NHS organisations, academic institutions, the commercial sector and parents has enabled the successful development of this NNRD. A White Paper from the UK coalition government published in 2010 called for an information revolution, using data to drive up the quality of healthcare, improve patient choice and provide public accountability.[279;383] Specialist neonatal care has already been leading the way regarding this challenge, since the establishment of the NNRD which provides comprehensive data collection on a national scale.[279] The NHS Newborn Infant Physical Examination Programme (NIPE) and the Newborn Hearing and Bloodspot Screening programmes provide additional sources of neonatal healthcare data in the UK.[384-387] However, there are huge inconsistencies between some of these datasets in the types of data captured and how they are defined.

The NHS Information Centre has been commissioned to develop a maternity record. If this is to be truly useful it is essential that current uncertainties regarding the retention of identifiers, access for secondary use, clarity regarding data definitions and validation are addressed. An integrated approach will be required to ensure that such data are captured once, serve multiple needs and can be responsive to changing requirements.[279]

A logical next step must also be to develop robust linkage between birth records and more detailed clinical datasets. For example, linkage between hospital and general practice records (belonging to both the mother and baby) would add substantial utility to these data, providing insight into the effects of exposures during pregnancy on subsequent infant health outcomes. Rather than existing ad hoc approaches to linkage, the aim should be routine database linkage on a national scale.

10.8 Overall thesis conclusion

The clinical burden of RSV bronchiolitis across healthcare settings in the UK is greater than previously estimated. This thesis examines the wider clinical spectrum of bronchiolitis illness across primary and secondary healthcare settings, using routine data from electronic health records to develop longitudinal cohorts with follow-up from birth through early childhood.

This thesis provides, to my knowledge, the first estimate of bronchiolitis incidence in UK primary care. Between 4% and 21% of infants have a bronchiolitis GP consultation in their first year. National hospital admission data have been utilised to improve estimates of the population burden of severe bronchiolitis illness and examine trends in hospital admissions. 2.4% of the national birth cohort are admitted to hospital with bronchiolitis in the first year of life, the majority of whom are born at term, with none of the known clinical risk factors for severe RSV infection. An episode of bronchiolitis in infancy is a predictor of subsequent hospital admissions and general practice consultations for asthma and wheezing in early childhood.

The findings presented in this thesis provide evidence for the need to investigate the cost effectiveness of passive immunotherapy in other high-risk groups, such as infants with cerebral palsy, cystic fibrosis and Down's syndrome. Further research is also needed to quantify the clinical and healthcare burden of RSV bronchiolitis among infants presenting in Emergency Departments in the UK. Linkage between healthcare databases such as the NNRD and HES, will provide a valuable source of data to quantify risk of RSV bronchiolitis admission in specific groups such as by gestational age, as recommended by the JCVI.

The improved estimates of the clinical burden of RSV bronchiolitis among infants in the UK presented in this thesis have important policy implications for clinical training and potentially for passive and future active RSV immunisation policy. This thesis highlights the need to prioritise the development of new approaches for the prevention and treatment of RSV infection, given that the clinical and healthcare burden of bronchiolitis in the UK is greater than previous studies have been powered to estimate.

11.0 Appendices

Appendix 1: Palivizumab prescribing data from the Prescribing Cost Analysis (PCA) database from the NHS Information Centre

Annual number of Palivizumab prescriptions dispensed across England, recorded in the PCA system.

Year	Items (000s)	NIC £ (000s)
2003	0.059	50.768
2004	0.054	43.720
2005	0.110	65.106
2006	0.155	90.984
2007	0.197	131.593
2008	0.152	99.651
2009	0.123	77.023

PCA Data

Prescription information is taken from the Prescribing Cost Analysis (PCA) system, supplied by the Prescription Services Division of the NHS Business Services Authority (BSA), and is based on a full analysis of all prescriptions dispensed in the community i.e. by community pharmacists and appliance contractors, dispensing doctors, and prescriptions submitted by prescribing doctors for items personally administered in England. Also included are prescriptions written in Wales, Scotland, Northern Ireland and the Isle of Man but dispensed in England. The data do not cover drugs dispensed in hospitals, including mental health trusts, or private prescriptions. Prescribers are GPs, hospital doctors, dentists and non-medical prescribers such as nurses and pharmacists.

Prescription Items

Prescriptions are written on a prescription form. Each single item written on the form is counted as a prescription item.

Net Ingredient Cost (NIC)

NIC is the basic cost of a drug. It does not take account of discounts, dispensing costs, fees or prescription charges income.

Quantity

The quantity of a drug dispensed is measured in units depending on the formulation of the product. Quantities are not added together across preparations because of different strengths and formulations.

Units

All the data are measured in units of a thousand, with the exception of net ingredient cost per quantity which is measured in pounds sterling (£).

Appendix 2: Published methodology paper

Journal of Public Health Advance Access published September 11, 2012

Journal of Public Health | pp. 1–10 | doi:10.1093/pubmed/fds077

Quality of routine hospital birth records and the feasibility of their use for creating birth cohorts

Joanna Murray[1], Sonia Saxena[1], Neena Modi[2], Azeem Majeed[1], Paul Aylin[1], Alex Bottle[1], on behalf of the Medicines for Neonates Investigator Group[†]

[1]Department of Primary Care and Public Health, Imperial College London, 3rd Floor Reynolds Building, St Dunstan's Road, London W6 8RP, UK
[2]Department of Neonatal Medicine, Imperial College London, London, UK
Address correspondence to Joanna Murray, E-mail: joanna.murray@imperial.ac.uk

ABSTRACT

Background Linkage between routinely collected hospital birth and other records offers the potential for epidemiological and public health research by developing population-level birth cohorts with cradle-to-grave follow-up. Data from births in English National Health Service hospitals are collected in the Hospital Episode Statistics (HES) database but are of uncertain quality.

Methods We examined the range and completeness of birth information recorded in HES and tested an approach for minimizing the effect of hospital-level variations by selecting hospitals with high completeness of recording (≥90%) for key fields. We discuss important methodological considerations when using routine healthcare data to develop a birth cohort.

Results The proportion of missing data in key birth record fields has been decreasing annually, such as gestational age and birth weight (from 46.2 and 43.9% in 2005/06 to 18.1 and 16.9% in 2009/10, respectively). We compared the important characteristics such as size and access to specialist neonatal care between 71 high-coding and 85 low-coding hospitals and found no significant differences, suggesting hospitals with high birth record completeness may be generalizable and representative of all hospitals.

Conclusions The completeness of recording of hospital birth information varies greatly between hospitals in England but is improving. It may be preferable and valid to construct cohorts from only hospitals with high completeness of recording.

Keywords birth cohort, hospital admission records, linkage, longitudinal data, hospital episode statistics

Introduction

Potential uses of maternity data to monitor health outcomes

Information on birth characteristics, such as gestational age and birth weight, is needed for many health services and epidemiological research studies examining short- and longer term clinical outcomes.[1] Hospital birth records are routinely collected throughout England and can provide a rich electronic source of clinical information about the health status of individuals at birth, as well as for measuring the quality of maternity care. Life-long assessment of outcomes is difficult in practice and much infant research is therefore selective, poorly generalizable and focused on short-term outcomes. Secondary use of electronic health records cannot replace nor precisely replicate the value of data collected for bespoke prospective observational studies, but the potential benefits for research and patient care are well established[2–4] and the use of these data involves comparatively minimal costs. These data can facilitate benchmarking and comparison of outcomes between different hospitals or geographical areas. Cohorts based on these data can provide representative population-level information that

[†]Members of the Medicines for Neonates Investigator Group: Neena Modi, Peter Brocklehurst, Jane Abbott, Kate Costeloe, Elizabeth Draper, Azeem Majeed, Jacquie Kemp, Deborah Ashby, Alys Young, Stavros Petrou.

Joanna Murray, Doctoral Research Student
Sonia Saxena, Clinical Senior Lecturer in Primary Care
Neena Modi, Professor of Neonatal Medicine
Azeem Majeed, Professor of Primary Care
Paul Aylin, Clinical Reader in Epidemiology and Public Health
Alex Bottle, Senior Lecturer in Medical Statistics

© The Author 2012, Published by Oxford University Press on behalf of Faculty of Public Health. All rights reserved

is highly generalizable, has power to detect small effect sizes and relates directly to real clinical practice. Such cohorts have the potential to answer an extensive range of research questions that require longitudinal follow-up, from examining long-term health outcomes after preterm birth, to identifying risk factors for childhood hospital admission with influenza or bronchiolitis.

Linkage with other health records

In health systems with universal coverage, the value of clinical databases for research can be significantly enhanced by linking different data sources, creating the potential to develop cradle-to-grave data sets for the whole population. Probabilistic or deterministic linkage methods can be used to match records for the same individual from different databases, using unique personal identifiers such as National Health Service (NHS) number. Linkage between data sets can provide validation of recording, coding and completeness of data, help to develop more robust clinical case definitions and provide information on events that happen outside healthcare settings including deaths.[5,6] Other possible benefits include adverse drug reaction monitoring,[7,8] improved healthcare-associated infection surveillance with linkage between microbiological and clinical records[9] and also identification and long-term follow-up of clinical trial participants.[10] Area-level linkage to data on deprivation, pollution and other environmental data sources could provide valuable information for public health research studies. Large-scale population-based linkage presents its own complex ethical challenges of consent, confidentiality and secure data storage. These are best addressed by robust information governance and engaging with patients and the public about how their records are to be used.[11,12]

How have administrative birth records been used so far?

Non-UK

The effectiveness of routine health record linkage in *adults* has been demonstrated in Australia and Canada,[5,11,13,14] where it has improved both data quality and utility.[5] Birth records have been successfully linked to hospital discharge data in Australia, with matching rates of 99%.[15] In several regions of the USA, data from birth certificates have been linked to hospital discharge records to examine maternal outcomes.[16,17] In Scandinavian countries, the assignment of unique personal identification numbers permits the development of population-based cohorts, facilitating a broad array of epidemiological studies, such as research in Denmark investigating the impact of place of birth and familial risk factors on risk of autism.[18]

UK

In Wales, the Secure Anonymized Information Linkage (SAIL) databank collates anonymized person-based electronic health and social care data, which are now being combined to establish a Wales-wide Electronic Cohort for Children.[19,20] This databank has already been used to identify potential clinical trial participants from primary care data.[19,21] The Scottish Health Informatics Programme (SHIP) provides linked health and social care data from birth through to death.[22] To date, SHIP has primarily been used for pharmacovigilance and diabetes epidemiology research.[22,23] The Oxford Record Linkage Study contains over 10 million records for 5 million people since 1963.[24] This has been used for longitudinal studies identifying maternal and perinatal risk factors for conditions such as inflammatory bowel disease,[25] asthma[26] and coeliac disease.[27]

Administrative birth data are captured in England in both hospital discharge records and also birth registration information collected by the Office for National Statistics (ONS). The use of information in the English Hospital Episode Statistics (HES) delivery and birth records for research has been limited. A previous feasibility study showed high rates of linkage between HES maternity and ONS birth records.[28] In addition to these national resources, the NHS Numbers for Babies (NN4B) Service, introduced in 2002, collects a small data set containing some key fields that are not recorded in birth registrations, such as gestational age.[29] The NN4B service ensures every baby is allocated a unique NHS number shortly after birth. Linkage between NN4B and HES birth records indicated that improvements in the quality and completeness of HES maternity data are needed.[29,30] HES maternity data have also been used to validate birth information recorded in the Millennium Cohort Study, a UK-wide longitudinal observational cohort of nearly 19 000 babies.[31]

We aimed to examine the range and completeness of information recorded in hospital birth records in England and to identify the methodological issues when using such routine healthcare data to develop a birth cohort. First, we describe linkage issues and the HES data set. Second, we summarize the data quality and assess the feasibility of creating a birth cohort using records from hospitals with high data completeness.

Methods

Overview of HES data

HES is a complex administrative database containing details of all admissions to NHS (public) hospitals in England

electronically since 1989.[32,33] HES contains a wide range of data including coded clinical diagnoses, procedures, geographical information on where patients are treated and demographic information such as patient age, sex and ethnicity. Further information on the HES database is presented in Appendix 1.

HES maternity records and baby tails

When a mother gives birth, her hospital admission record changes from a general inpatient admission record to a maternity record and is updated as such before it is submitted to HES.[32] HES contains two types of maternity record, the delivery and the birth record, both of which contain a 'baby tail' of an additional 19 fields (http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=925). A list of some of the maternity data collected within HES can be found in Box 1 and HES maternity records are described in more detail in Appendix 1.[32]

Box 1 Key information collected in HES maternity records

- Method of onset of labour
- Place and date of delivery
- Miscarriages and ectopic pregnancies
- Live and still births
- Deliveries and births with complications
- Multiple deliveries
- Length of gestation
- Birth weight
- Neonatal unit admissions and level of care
- Antenatal and postnatal length of stay

Linkage across time to produce a HES cohort

The processing of HES birth records described for the first time here can be used to develop cohorts of children with the potential of long-term follow-up. Using a unique personal identifier, we can link individual birth records through time to subsequent admission records to provide longitudinal information on severe episodes of illness requiring hospitalization. The unique identifier used in HES ('HESID') is derived by matching records for the same patient using a combination of NHS number and local patient identifier, plus the patient's date of birth, sex and postcode.[32]

Processing and methodological issues using HES maternity data

HES data are divided into financial years from 1 April in a given year to 31 March in the following year. We outline the key data processing and methodological issues required to use these HES birth records for research purposes, described in detail in Appendix 1.

Analysis

The NHS Information Centre publish annual quality reports examining the coverage of the data submitted by NHS hospitals to HES, which have repeatedly found that for many hospitals completeness of birth fields is often inadequate.[34,35] We examined the completeness of recording for baby tail fields over 5 years (2005/06–2009/10) and also compared the total number of births with ONS birth registrations. We compared the proportion of missing data for each baby tail field in 2005/06 to the most recent year available (2009/10), using chi-squared tests, with associated *P*-values presented in Table 1. Using a cut-off of 90% completeness of recording for key birth fields (gestational age and birth weight), we compared hospitals with highly complete data with the others in terms of hospital characteristics from available data to assess how representative of all of England a birth cohort formed from babies at high-coding hospitals would be. To test for significant differences between these hospitals with high versus low completeness of birth record fields, we used chi-squared tests to compare proportions and *t*-tests to compare mean values (associated *P*-values presented in Table 2).

Results

Coverage and completeness of recording

The completeness of recording of all birth fields has significantly improved over the 5 years between 2005/06 and 2009/10 (Table 1). The values presented differ from those published by HES online because of our extensive data cleaning processes, which resulted in the removal of some invalid birth records (such as those with an invalid date of birth). Among our 2009/10 data set, birth weight and gestational age were recorded for 83 and 82% of births, respectively.

Comparison with data from ONS birth registrations

Overall, the HES cohort captured 87% of all live births recorded by the ONS in England during the time period. There are various explanations for this difference. An estimated 97% of live births in England occur in NHS hospitals. Those occurring in private hospitals or at home may not be recorded in HES. In addition, because HES contains discharge records, some births which occurred within the

Table 1. Completeness of recording of baby tail fields in HES birth records (2005/06–2009/10)

Baby tail fields in HES birth records (field name)	*% Missing or unknown*					*P-value*
	2005/06	*2006/07*	*2007/08*	*2008/09*	*2009/10*	
Anaesthetic given during labour or delivery (delpren)	41.9	41.8	44.8	29.6	16.5	<0.001
Anaesthetic given post-labour or delivery (delposn)	48.1	46.0	49.6	34.8	21.6	<0.001
Antenatal days of stay (antedur) (derived from other HES fields)	0.2	0.2	0.2	0.1	0.1	<0.001
Baby's age in days (neodur) (derived from other HES fields)	0.1	0.0	0.0	0.0	0.0	<0.001
Birth order (birorder)	33.9	36.9	39.4	24.9	13.7	<0.001
Birth weight (birweit)	43.9	47.1	50.1	31.3	16.9	<0.001
Delivery place change reason (delchang)	45.2	45.8	47.4	34.6	21.7	<0.001
Delivery method (delmeth)	35.1	35.8	44.3	30.6	14.9	<0.001
Delivery place (actual) (delplace)	44.0	46.8	57.0	41.3	17.9	<0.001
Delivery place (intended) (delinten)	41.3	42.7	43.6	30.2	14.9	<0.001
First antenatal assessment date (anasdate)	41.7	44.3	44.6	34.7	20.4	<0.001
Gestation in weeks at first antenatal assessment (anagest)	54.5	63.9	55.2	45.6	28.3	<0.001
Length of gestation (gestat)	46.2	54.2	48.0	34.6	18.1	<0.001
Birth status (birstat)	43.9	47.0	48.0	32.9	16.2	<0.001
Labour/delivery onset method (delonset)	36.2	37.7	41.1	25.5	11.5	<0.001
Mother's age at delivery (matage)	42.4	43.3	43.0	34.5	30.5	<0.001
Neonatel level of care (neocare)	16.1	16.0	17.1	18.4	12.4	<0.001
Number of babies (numbaby)	31.8	33.3	36.1	23.6	11.9	<0.001
Resuscitation method (biresus)	44.2	45.3	48.0	34.2	21.1	<0.001
Status of person conducting delivery (delstat)	38.9	42.6	48.4	33.9	19.1	<0.001
Total number of births[a]	554 521	566 749	575 493	589 684	603 786	—

[a]Total number of births in a cohort after the removal of duplicate episodes and completion of data cleaning processes as described in Appendix 1.

cohort year but were discharged from hospital after this period would not be captured in our data set. Most of the births that are not captured within our HES cohort may have been excluded as a result of the quality criteria we applied to remove records with invalid information reported in key fields, or due to the incomplete coverage issues described earlier.

Approach for minimizing the effect of variable data completion between hospitals

In our 2007/08 data set, 50% of records had no birth weight recorded, 48% were missing gestational age, 43% had no maternal age recorded and 17% had no indication of whether the baby received any specialist neonatal care, with considerable variation in the completeness of information recorded in birth records between NHS providers. Some hospitals with many births had recorded gestational age and birth weight in <10% of birth records, whilst other similarly large hospitals recorded this information for >90% of births. There appears to be no correlation between the number of births occurring in a hospital and the completeness of recording of these prematurity indicators. Therefore, depending on the study purpose and the exposure and outcome measures of interest, we suggest selecting birth records only from hospitals with high completeness of recording. We tested this by creating a 2007/08 birth cohort where we selected only birth records from hospitals where ≥90% of their birth records contained complete recording of key variables, birth weight and gestational age. The resulting cohort included 296 618 babies born at 71 hospitals across England. Table 2 shows a comparison of characteristics of included ($n = 71$) and excluded ($n = 85$) hospitals. The mean number of births, maternity beds and access to neonatal intensive care were mostly similar among hospitals with high and low completeness of recording of birth record information. The mean maternal age, the proportion of babies of non-white British ethnicity and the proportion of babies in the most deprived Carstairs quintile were similar among the two groups of hospitals, but both groups had high variability between hospitals in terms of data completeness for these fields (Table 2).

Table 2. Comparison of maternity characteristics between hospitals with high and low completeness of recording in their birth admission records[a], from financial year 2007/08

Hospital maternity factors	*Hospitals with* high *completeness (n = 71)*	*Hospitals with* low *completeness (n = 85)*	P-*value*
Mean number of births in year (SD)	3957 (2011)	3465 (1997)	0.1287
Mean number of maternity beds available (SD)	55.1 (30.3)	55.3 (26.4)	0.9649
Mean number of maternity beds reported to be occupied (SD)	35.4 (21.4)	35.2 (18.0)	0.9495
Number with access to onsite neonatal intensive care (%)	52 (73)	68 (80)	0.3022
Mean number of available beds in neonatal intensive care unit (SD)	10.6 (11.7)	11.4 (10.9)	0.6595
Mean maternal age (% birth records missing the data)	28.9 (18.4)	29.0 (70.1)	0.3964
Proportion of births per hospital in most deprived Carstairs Deprivation score quintile (% birth records missing the data)	0.472 (69.4)	0.435 (56.2)	0.6406
Proportion of births per hospital of non-white British ethnicity (% birth records missing the data)	0.527 (7.1)	0.564 (5.1)	0.6438

SD, standard deviation.

[a]Low completeness of recording was defined as hospitals where <90% of their birth admission records contained complete recording of birth weight and gestational age.

Discussion

Main findings of this study

The completeness of recording of information in hospital birth records is highly variable between hospitals but is generally improving. The proportion of missing or unknown data in key birth record fields has decreased over the 5-year period, such as gestational age (from 46.2% in 2005/06 to 18.1% in 2009/10) and birth weight (from 43.9% in 2005/06 to 16.9% in 2009/10). We compared characteristics of hospitals with high ($n = 71$) and low ($n = 85$) completeness of recording of key birth fields and found no significant differences between them in terms of important hospital factors such as size or access to specialist neonatal care. This suggests a birth cohort derived from these 71 hospitals would be representative of the whole country.

What is already known on this subject

Detailed clinical information is routinely captured in hospital birth and other records but their value for use in public health, epidemiological and health services research has yet to be fully exploited. The Healthcare Commission's review of maternity services found that even among larger, well-respected maternity units, information technology was poor and data collection systems inadequate.[36] Coverage of hospital deliveries was estimated to be around 73% between 2001/02 and 2005/06, but just 14% for home deliveries.[32] As a result of these quality concerns, additional rigorous validation was carried out by the NHS Information Centre on HES maternity data, including intensive cleaning of systematic coding errors.[32] Some of the quality and coverage issues specifically affecting HES maternity data are listed in Box 2.

Box 2 Quality and coverage issues affecting HES maternity data

- Stand-alone maternity systems in around 20 hospitals are not linked to their patient administration system, from which HES data are obtained (via the Secondary Uses Service).
- Some hospitals return data on the number of birth episodes or delivery episodes but not both.
- Transfer of maternity information between systems leaves scope for data errors and shortfalls.
- Stillbirths are not reliably recorded in every hospital and are not allocated an NHS number.[28,31,34]
- Lack of *a priori* definitions for data variables resulting in inconsistencies in data entry.
- Use of aggregate or categorized fields rather than the raw data.

What this study adds

Individual birth records linked to other healthcare records can provide a valuable source of longitudinal, population level data for developing cohorts from birth. We describe

some of the methodological and processing issues that need to be considered when using birth records in this context. Completeness of birth information recording is highly variable between English NHS hospitals but has improved over the 5-year period from 2005/06 to 2009/10. Where key birth information such as gestational age and birth weight are missing, it may be necessary and valid to select data only from hospitals with high completeness of recording.

The future of routine data collection for research

With limited funding, restricted resources and improved recognition of the necessity for life-long follow-up for infant research, it is imperative that the research community, funders and the public maximize the value of information that can be gleaned from existing routine data sources and avoid duplication of effort. This was highlighted by the public inquiry report into the Bristol Royal Infirmary paediatric cardiac unit, which recommended that HES should be supported as a national resource to monitor healthcare outcomes.[37] HES has been commonly used to examine disease time trends and standards of care at an aggregate level. Recording of birth information needs to improve in some NHS hospitals to provide robust baseline population estimates of birth outcomes. Encouragingly, the data we have presented from recent years indicates that completeness of recording of HES birth records is improving. Individual follow-up from birth is feasible using HES but the limitations and processing issues we have described are important methodological considerations to be taken into account.

Data capture is likely to be more complete and most accurate in records that are entered by clinicians at the point of care. In the UK for example, Neonatal Networks submit neonatal specialist care records to a National Neonatal Research Database managed by the Neonatal Data Analysis Unit (based at Imperial College London in collaboration with the Chelsea and Westminster NHS Foundation Trust). These records capture real-time operational data that are actively used in the processes of clinical care. The NHS Newborn Infant Physical Examination Programme and Newborn Hearing and Bloodspot Screening programmes provide additional sources of neonatal healthcare data in the UK. However, there are huge inconsistencies between data sets in the types of data captured and how they are defined. The NHS Information Centre has been commissioned to develop a maternity record. If this is to be truly useful it is essential that current uncertainties regarding the retention of identifiers, access for secondary use, clarity regarding data definitions and validation are addressed. A logical next step must also be to develop robust linkage between birth records and more detailed clinical data sets. For example, linkage between hospital and general practice records (belonging to both the mother and baby) would add substantial utility to the data, providing insight into the effects of exposures during pregnancy on subsequent infant health outcomes. Rather than existing *ad hoc* approaches to linkage, our aim should be routine database linkage on a national scale.

Limitations of this study

We were unable to determine how accurately the identifier we used to deterministically link records across time is allocated to unique individuals, although we know that an improved, complex algorithm is used to derive 'HESID', based on a combination of other identifiers as described earlier. Probabilistic approaches can facilitate linkage with records from other databases with differing availability and quality of unique identifiers, by developing algorithms to assign individual unique identifiers based on a combination of variables such as date of birth and postcode. Provided methods are tailored to the data sets used and the context of how data are collected is considered, linkage methods can correctly identify a high proportion of matched records even with limited quality or availability of identifiers.[12,38]

Our comparison of maternity characteristics between hospitals with high and low completeness of recording was limited to specific fields for which we had access to information from all hospitals, such as their mean number of births, maternity beds and specialist neonatal care facilities. However, we are confident that our comparison of these key hospital factors between hospitals with high and low completeness of recording of birth information provides sufficient evidence that these do not significantly differ. We also compared the mean maternal age, the proportion of babies of non-white British ethnicity and the proportion of babies in the most deprived Carstairs quintile. These appear similar between the two groups, though data incompleteness means we cannot be certain of this.

Conclusion

Detailed clinical information including prematurity, birth weight, complications and comorbidities is routinely captured in hospital birth and other records. Despite some complex data-processing requirements, HES and other birth records can be used for population-scale epidemiological research. However, the completeness of birth information recorded in English hospitals is variable. Where key birth information such as gestational age and birth weight are missing, it may be preferable to select data only from hospitals with high levels of completeness of recording; our

analysis suggests that the results may be generalizable to all hospitals. HES data can be used to develop longitudinal cohorts by linking individual birth records to subsequent hospital admission and other healthcare, microbiological or treatment records for many vital purposes.

Details of contributors and guarantor

J.M., A.B. and S.S. conceived the study. J.M. performed all analyses and wrote the first draft. All authors contributed in the revision of the manuscript. J.M. is the guarantor.

Ethical approval

We have permission from the NIGB under Section 251 of the NHS Act 2006 (formerly Section 60 approval from the Patient Information Advisory Group) to hold confidential data. We have ethical approval to use them for research and measuring quality of delivery of health care from the South East Ethics Research Committee.

Funding

J.M. is funded through a National Institute for Health Research (NIHR) Programme Grant for Applied Research (RP-PG-0707-10010) held by N.M.; J.M., A.M. and N.M. declare financial support from this award for the submitted work. Support for data analysis was received from the Neonatal Data Analysis Unit led by N.M. Unrestricted support for the Neonatal Data Analysis Unit from Abbott International and Danone UK is acknowledged. A.B. and P.A. are part of The Dr Foster Unit at Imperial, which is principally funded via a research grant by Dr Foster Intelligence, an independent healthcare information company and joint venture with the Information Centre of the NHS. The Unit is affiliated with the NIHR Imperial Centre for Patient Safety and Service Quality at Imperial College Healthcare NHS Trust. The Department of Primary Care & Public Health at Imperial College is grateful for support from the NIHR Collaboration for Leadership in Applied Health Research & Care (CLAHRC) for North West London (a partnership between Chelsea and Westminster NHS Foundation Trust and Imperial College London), the NIHR Imperial Biomedical Research Centre (a partnership between Imperial College Healthcare NHS Trust and Imperial College London) and the NIHR Imperial Centre for Patient Safety and Service Quality. This publication presents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research scheme (RP-PG-0707-10010). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. The funders had no role in the study design, collection, analysis and interpretation of data, in the writing of the report or in the decision to submit the paper for publication.

Conflict of interest

All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author). The views expressed in this publication are those of the authors and not necessarily those of the NHS, NIHR or Department of Health.

References

1 Hockley C, Quigley MA, Hughes G *et al.* Linking millennium cohort data to birth registration and hospital episode records. *Paediatr Perinat Epidemiol* 2008;**22**:99–109.

2 The Academy of Medical Sciences. Personal data for public good: using health information in medical research. London, UK, 2006. http://www.acmedsci.ac.uk/p99puid62.html. Accessed June 2011.

3 UK Clinical Research Collaboration and the Wellcome Trust. Frontiers meeting on the use of electronic patient records for research and health benefit. London, UK, 2007. http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtd038686.pdf. Accessed June 2011.

4 UKCRC Research and Development Advisory Group to Connecting for Health. Report of the research simulations, 2007. http://www.ukcrc.org/publications/reports/. Accessed June 2011.

5 Field K, Kosmider S, Johns J *et al.* Linking data from hospital and cancer registry databases: should this be standard practice? *Intern Med J* 2009;**40**:566–73.

6 Murray GD, Lawrence AE, Boyd J. Linkage of Hospital Episode Statistics (HES) data to Office for National Statistics (ONS) mortality records, 2000. http://www.bristol-inquiry.org.uk/Documents/Hes%20ons%20linkage%20report.pdf. Accessed June 2011.

7 Wu TY, Jen MH, Bottle A *et al.* Ten-year trends in hospital admissions for adverse drug reactions in England 1999–2009. *J R Soc Med* 2010;**103**:239–50.

8 Meropol SB, Chan KA, Chen Z *et al.* Adverse events associated with prolonged antibiotic use. *Pharmacoepidemiol Drug Saf* 2008;**17**:523–32.

9 Garcia AL, Aylin P, Tian J *et al.* Data linkage between existing healthcare databases to support hospital epidemiology. *J Hosp Infect* 2011;**79**:231–5.

10 Staa TP, Goldacre B, Gulliford M *et al.* Pragmatic randomised trials using routine electronic health records: putting them to the test. *BMJ* 2012;**344**:e55.

11 Chamberlayne R, Green B, Barer ML *et al.* Creating a population-based linked health database: a new resource for health services research. *Can J Public Health* 1998;**89**:270–3.

12 Souhami R. Governance of research that uses identifiable personal data. *BMJ* 2006;**333**:315–6.

13 Roos LL, Brownell M, Lix L *et al.* From health research to social research: privacy, methods, approaches. *Soc Sci Med* 2008;**66**:117–29.

14 Taylor LK, Travis S, Pym M *et al.* How useful are hospital morbidity data for monitoring conditions occurring in the perinatal period? *Aust N Z J Obstet Gynaecol* 2005;**45**:36–41.

15 Ford JB, Roberts CL, Taylor LK. Characteristics of unmatched maternal and baby records in linked birth records and hospital discharge data. *Paediatr Perinat Epidemiol* 2006;**20**:329–37.

16 Bradford H, C+írdenas V, Camacho-Carr K *et al.* Accuracy of birth certificate and hospital discharge data: a certified nurse-midwife and physician comparison. *Matern Child Health J* 2007;**11**:540–8.

17 Lydon-Rochelle MT, Holt VL, Nelson JC *et al.* Accuracy of reporting maternal in-hospital diagnoses and intrapartum procedures in Washington State linked birth records. *Paediatr Perinat Epidemiol* 2005;**19**:460–71.

18 Lauritsen MB, Pedersen CB, Mortensen PB. Effects of familial risk factors and place of birth on the risk of autism: a nationwide register-based study. *J Child Psychol Psychiatry* 2005;**46**:963–71.

19 Lyons R, Jones K, John G *et al.* The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak* 2009;**9**:3.

20 Welsh Electronic Cohort for Children (WECC), 2011. http://www.swan.ac.uk/ils/research/chiral/methodologies/healthinformatics/wecc/. http://www swan ac uk/ils/research/chiral/methodologies/healthinformatics/wecc/. Accessed June 2011.

21 McGregor J, Brooks C, Chalasani P *et al.* The Health Informatics Trial Enhancement Project (HITE): using routinely collected primary care data to identify potential participants for a depression trial. *Trials* 2010;**11**:39.

22 Scottish Health Informatics Programme, 2011. http://www.scot-ship.ac.uk. Accessed June 2011.

23 Govan L, Wu O, Briggs A *et al.* Inpatient costs for people with type 1 and type 2 diabetes in Scotland: a study from the Scottish Diabetes Research Network Epidemiology Group. *Diabetologia* 2011;**54**:2000–8.

24 Gill L, Goldacre M, Simmons H *et al.* Computerised linking of medical records: methodological guidelines. *J Epidemiol Community Health* 1993;**47**:316–9.

25 Roberts SE, Wotton CJ, Williams JG *et al.* Perinatal and early life risk factors for inflammatory bowel disease. *World J Gastroenterol* 2011;**17**:743–9.

26 Davidson R, Roberts SE, Wotton CJ *et al.* Influence of maternal and perinatal factors on subsequent hospitalisation for asthma in children: evidence from the Oxford record linkage study. *BMC Pulm Med* 2010;**10**:14.

27 Roberts SE, Williams JG, Meddings D *et al.* Perinatal risk factors and coeliac disease in children and young adults: a record linkage study. *Aliment Pharmacol Ther* 2009;**29**:222–31.

28 Abrahms C, Davy K. Linking HES maternity records with ONS birth records. *Health Stat Q* 2002;(**Spring Edition**):22–30.

29 Dattani N, Datta-Nemdharry P, Macfarlane A. Linking maternity data for England, 2005–06: methods and data quality. *Health Stat Q* 2011;**49**:53–79.doi:10.1057/hsq.2011.3.

30 Hilder L, Moser K, Dattani N *et al.* Pilot linkage of NHS numbers for babies data with birth registrations. *Health Stat Q* 2007;**33**:25–33.

31 Dezateux C, Hockley C, Johnson J *et al.* Millennium cohort study: birth registration and maternity. Hospital Episode Statistics Linkage. Centre for Longitudinal Studies Report, 2006. Published by the Centre for Longitudinal Studies, Bedford Group for Lifecourse and Statistical Studies, Institute of Education, University of London. In collaboration with the National Perinatal Epidemiology Unit and the Institute for Child Health.

32 The NHS Information Centre for Health and Social Care HESonline Hospital Episode Statistics. www.hesonline.nhs.uk, 2010.

33 Aylin P, Williams S, Bottle A *et al.* Counting hospital activity: spells or episodes? *BMJ* 2004;**329**:1207.

34 The NHS Information Centre for Health and Social Care HESonline Hospital Episode Statistics. NHS Maternity Statistics, 2010. http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=1475 (June 2011, date last accessed).

35 The NHS Information Centre for Health and Social Care. NHS Maternity Statistics, England: 2007–2008, 2009. http://www.ic.nhs.uk/statistics-and-data-collections/hospital-care/maternity/nhs-maternity-statistics-england:-2007-08. Accessed June 2011.

36 Care Quality Commission Periodic Review. Maternity data quality indicator specifications, 2010. http://archive.cqc.org.uk/_db/_documents/Maternity_data_quality_indicator_specifications.pdf.

37 Kennedy I, Howard R, Jarman B *et al.* Learning from Bristol: the report of the public inquiry into children's heart surgery at the Bristol Royal Infirmary 1984–1995. Command Paper: CM 5207, 2001. Published and presented to Parliament by the Secretary of State for Health by Command of Her Majesty.

38 MacLeod MC, Bray CA, Kendrick SW *et al.* Enhancing the power of record linkage involving low quality personal identifiers: use of the best link principle and cause of death prior likelihoods. *Comput Biomed Res* 1998;**31**:257–70.

39 The NHS Information Centre for Health and Social Care. HES Online Data Dictionary: inpatients, 2009. http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=571 (June 2011, date last accessed).

40 The Healthcare Commission. Towards better births: a review of maternity services in England, 2008. http://www.cqc.org.uk/_db/_documents/Towards_better_births_200807221338.pdf.

41 The NHS Information Centre for Health and Social Care. Hospital episode statistics: inpatient cleaning rules, 2009. http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=376.

Appendix 1

Processing and methodological issues using HES maternity data

Overview of HES data

The basic unit recorded in HES is the finished consultant episode (the period of time during which a patient is under the care of one consultant). A 'spell' or admission is defined

as the continuous period of time spent as a patient within one hospital,[33] from admission to discharge or transfer to another provider and may therefore include more than one consultant episode.[33]

HES maternity records and baby tails

The delivery record is for the mother and contains the same information as a general record plus an additional baby tail with information about the delivery. The birth record is for the baby and also contains general record information plus the baby tail, which contains identical information to the corresponding baby tail in the mother's delivery record. Diagnoses and procedures recorded in the birth record refer to the baby and, conversely, diagnoses and procedures in the delivery record refer to the mother. For multiple births, separate tails for each baby will appear in the mother's delivery record, but each birth record will contain only the individual baby's relevant tail. HES data are collected from births in NHS and non-NHS hospitals and at home, although information from births outside NHS hospitals is often incomplete.

Identifying birth episodes

The first step in developing a birth cohort study using HES was identifying all individual birth episodes within a given year. To identify birth episodes, we used the 'admimeth' variable which contains a code recording how the patient was admitted to hospital.[39] This field was used to select all records with an admission method coded as 82 (other: babies born in healthcare provider) or 83 (other: babies born outside the healthcare provider, except when born at home as intended).

Removing duplicate birth episodes

The next step was identifying any duplicate birth episodes (a typical but limited problem with administrative data) using the unique personal identifier ('HESID'). We looked specifically at the first episode within the given spell (i.e. the initial birth episode itself) using the 'epiorder' field which contains the number of the episode within the current spell.[39] In our data set from financial year 2007/08, we identified 5600 individuals with more than one birth episode recorded, of which 5545 had two and 55 had more than two birth episodes recorded. We excluded individuals with more than two birth episodes from the cohort because it would be almost impossible to decipher which birth record ought to be retained when comparing more than two duplicates. Many records from individuals with two birth episodes were identical matches (2750 individuals), having two exact duplicate birth episodes with identical information recorded in all fields. Where this was the case, one of the identical records was deleted.

Among the remaining individuals with two birth episodes, the records did not contain matching information. It is very difficult to determine which of two birth records with the same unique identifier is likely to be the true record of that individual's birth. In previous studies using HES and similar databases, the common approach has been to retain only the first observation recorded. We observed that the basic demographic information for each pair of birth episodes tended to be common to both and for most it was only the diagnostic fields that differed. Consequently, records for individuals with duplicate birth episodes were compared using recording of *diagnostic information* and only the birth episode with the most diagnostic information recorded (number of non-empty diagnosis fields) was retained in the cohort. It was not possible to determine whether this approach is better than randomly retaining one birth episode. However, the record with the most diagnostic information might be the more accurate of the two because it is common for hospitals to resubmit data after they have carried out all their diagnosis coding more thoroughly (their first submissions may contain only very basic diagnostic information). Therefore, records can appear as duplicates if hospitals fail to report them as re-submissions to the secondary uses service (the NHS data repository). Consequently, for an individual with two birth episodes, the one with more diagnostic information may be the more accurate record. Handling duplicate records is a common, challenging problem when using routine data such as HES and requires careful consideration of the study context.

Summarizing information from additional episodes in a birth spell

Babies can have more than one episode of care within their birth admission, for example if a baby receives specialist care from a different consultant, is transferred between hospitals or is admitted to a neonatal unit. These additional episodes occur within the same birth spell but, where the initial birth event would have an 'epiorder' value of 1, subsequent episodes in the birth spell have an 'epiorder' value >1. To facilitate one-to-many linkage to subsequent hospital admission records, it is easiest to develop a data set consisting of one observation (or row) per individual. Seven thousand four hundred and ten individuals in our 2007/08 birth cohort had more than one episode in their birth spell. To simplify the data set to consist of only one birth record per patient, we summarized key information such as any

premature or congenital anomaly diagnoses using flags. These were then linked back into the original birth episode and subsequent episodes in the birth spell were dropped.

Cleaning variables

A range of exclusion criteria were developed to clean key variable fields and examine the quality of coding. The Care Quality Commission (CQC) conducted a review exploring quality indicator specifications, used to assess the quality of HES maternity data from 2009/10.[36,40] We combined the criteria identified within the CQC review[36] and HES inpatient cleaning rules[41] and applied these to the HES birth fields to ensure suspicious data and invalid records were removed. In addition, we identified any stillbirths that were recorded (using the birth status and discharge method fields) and removed these from the final cohort of live births.

Identical baby tail information recorded for each baby can be found in their mother's delivery record. Maternity systems can record up to nine birth tails for each delivery, allowing information from multiple births to appear in the mother's delivery record. In some instances, we found a baby's information was not recorded in the first field of a given variable. For example, if a baby was the second twin, their gestational age at birth ('gestat') may have appeared in the second field ('gestat_2') with the first field blank ('gestat_1') because in the mother's delivery record this would contain the first twin's gestation. To simplify analyses, we condensed individual birth records so that only one field for each variable existed. So for the above example, we transferred the relevant information for the gestation variable from 'gestat_2' into 'gestat_1' and then removed all additional fields for that variable (i.e. 'gestat_2' to 'gestat_9').

Contextual issues to consider

As well as the completeness and quality of recording, attrition is also important to consider because although an individual may be born in a English NHS hospital and therefore included in a HES birth cohort, migration, deaths and admissions outside of English NHS hospitals are difficult to identify without individual patient consent and follow-up.

References

Reference List

(1) Beck S, Wojdyla D, Say L, Betran AP, Merialdi M, Requejo JH et al. The worldwide incidence of preterm birth: a systematic review of maternal mortality and morbidity. *Bull World Health Organ* 2010; 88(1):31-38.

(2) United Nations. The Millenium Development Goals Report (2009) http://www.un.org/millenniumgoals/pdf/MDG_Report_2009_ENG.pdf. Accessed October 2012.

(3) Hall CB. Respiratory syncytial virus in young children. *The Lancet* 2010; 375(9725):1500-1502.

(4) Rudan I, Boschi-Pinto C, Biloglav Z, Mulholland K, Campbell H. Epidemiology and etiology of childhood pneumonia. *Bull World Health Organ* 2008; 86(5):408-416.

(5) National Institute for Health and Clinical Excellence. NICE Clinical Guideline 47: Feverish illness in children: assessment and initial management in children younger than 5 years. (2007) Accessed October 2012. http://www.nice.org.uk/nicemedia/pdf/CG47NICEGuideline.pdf.

(6) Saxena S, Bottle A, Gilbert R, Sharland M. Increasing short-stay unplanned hospital admissions among children in England; time trends analysis '97-'06. *PLoS One* 2009; 4(10):e7484.

(7) Wise PH. The transformation of child health in the United States. *Health Aff (Millwood)* 2004; 23(5):9-25.

(8) Christopher A and Crowley S. The burden of respiratory disease in childhood. Factsheet http://www.laia.ac.uk/factsheets/20032.pdf. (2003) Lung and Asthma Information Agency. Accessed October 2012.

(9) Kusel MM, de KN, Holt PG, Landau LI, Sly PD. Occurrence and management of acute respiratory illnesses in early childhood. *J Paediatr Child Health* 2007; 43(3):139-146.

(10) Shay DK, Holman RC, Newman RD, Liu LL, Stout JW, Anderson LJ. Bronchiolitis-associated hospitalizations among US children, 1980-1996. *JAMA* 1999; 282(15):1440-1446.

(11) Broughton S, Bhat R, Roberts A, Zuckerman M, Rafferty G, Greenough A. Diminished lung function, RSV infection, and respiratory morbidity in prematurely born infants. *Arch Dis Child* 2006; 91(1):26-30.

(12) Shay DK, Holman RC, Roosevelt GE, Clarke MJ, Anderson LJ. Bronchiolitis-associated mortality and estimates of respiratory syncytial virus-associated deaths among US children, 1979-1997. *J Infect Dis* 2001; 183(1):16-22.

(13) Paes BA, Mitchell I, Banerji A, Lanctot KL, Langley JM. A decade of respiratory syncytial virus epidemiology and prophylaxis: translating evidence into everyday clinical practice. *Can Respir J* 2011; 18(2):e10-e19.

(14) Simoes EA, Carbonell-Estrany X. Impact of severe disease caused by respiratory syncytial virus in children living in developed countries. *Pediatr Infect Dis J* 2003; 22(2 Suppl):S13-S18.

(15) Aherne W, Bird T, Court SD, Gardner PS, McQuillin J. Pathological changes in virus infections of the lower respiratory tract in children. *J Clin Pathol* 1970; 23(1):7-18.

(16) van Drunen Littel-van den Hurk, Watkiss ER. Pathogenesis of respiratory syncytial virus. *Curr Opin Virol* 2012; 2(3):300-305.

(17) Stick S. Pediatric origins of adult lung disease. 1. The contribution of airway development to paediatric and adult lung disease. *Thorax* 2000; 55(7):587-594.

(18) Walsh EE, McConnochie KM, Long CE, Hall CB. Severity of respiratory syncytial virus infection is related to virus strain. *J Infect Dis* 1997; 175(4):814-820.

(19) Panozzo CA, Fowlkes AL, Anderson LJ. Variation in timing of respiratory syncytial virus outbreaks: lessons from national surveillance. *Pediatr Infect Dis J* 2007; 26(11 Suppl):S41-S45.

(20) Hall CB. Respiratory Syncytial Virus and Parainfluenza Virus. *New England Journal of Medicine* 2001; 344(25):1917-1928.

(21) 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. *The Lancet* 2010; 375(9725):1545-1555.

(22) Smyth RL, Openshaw PJ. Bronchiolitis. *Lancet* 2006; 368(9532):312-322.

(23) Oshansky CM, Zhang W, Moore E, Tripp RA. The host response and molecular pathogenesis associated with respiratory syncytial virus infection. *Future Microbiol* 2009; 4(3):279-297.

(24) Walsh EE, Peterson DR, Falsey AR. Risk factors for severe respiratory syncytial virus infection in elderly persons. *J Infect Dis* 2004; 189(2):233-238.

(25) Welliver RC, Sr. The immune response to respiratory syncytial virus infection: friend or foe? *Clin Rev Allergy Immunol* 2008; 34(2):163-173.

(26) U.S.Department of Health and Human Services NIoHNIoAaID. The Jordan Report: Accelerated Development of Vaccine 2012. (2012) Accessed October 2012.

(27) Howard TS, Hoffman LH, Stang PE, Simoes EA. Respiratory syncytial virus pneumonia in the hospital setting: length of stay, charges, and mortality. *J Pediatr* 2000; 137(2):227-232.

(28) Leader S, Kohlhase K. Recent trends in severe respiratory syncytial virus (RSV) among US infants, 1997 to 2000. *J Pediatr* 2003; 143(5 Suppl):S127-S132.

(29) Health Protection Agency. Hospital Admissions for Respiratory Infections Data (2008) (www.hpa.org.uk). Accessed October 2012.

(30) Muller-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.

(31) 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):1065-1069.

(32) Fleming DM, Pannell RS, Cross KW. Mortality in children from influenza and respiratory syncytial virus. *J Epidemiol Community Health* 2005; 59(7):586-590.

(33) Barben J, Kuehni CE, Trachsel D, Hammer J. Management of acute bronchiolitis: can evidence based guidelines alter clinical practice? *Thorax* 2008; 63:1103-1109.

(34) Bush A, Thomson AH. Acute bronchiolitis. *BMJ* 2007; 335(7628):1037-1041.

(35) Koehoorn M, Karr CJ, Demers PA, Lencar C, Tamburic L, Brauer M. Descriptive Epidemiological Features of Bronchiolitis in a Population-Based Cohort. *Pediatrics* 2008; 122(6):1196-1203.

(36) Scottish Intercollegiate Guidelines Network (SIGN). Bronchiolitis in children - A national clinical guideline. 2006. http://www.sign.ac.uk/pdf/sign91.pdf Report No 91, SIGN. Accessed October 2012.

(37) Papadopoulos NG, Moustaki M, Tsolia M, Bossios A, Astra E, Prezerakou A et al. Association of Rhinovirus Infection with Increased Disease Severity in Acute Bronchiolitis. *Am J Respir Crit Care Med* 2002; 165(9):1285-1289.

(38) Health Protection Agency Centre for Infections. Laboratory reports of RSV received by CfI from NHS and Health Protection Agency microbiology laboratories, by date of specimen, 2010/11 and recent years. http://www.hpa.org.uk/web/HPAwebFile/HPAweb_C/1194947413920. (2011) Accessed October 2012.

(39) Health Protection Agency Centre for Infections. Weekly Laboratory Reports of RSV Infection. (2012) http://www.hpa.org.uk/web/HPAweb&HPAwebStandard/HPAweb_C/1195733841831. Accessed October 2012.

(40) Halfhide C, Smyth RL. Innate immune response and bronchiolitis and preschool recurrent wheeze. *Paediatr Respir Rev* 2008; 9(4):251-262.

(41) Thorburn K, Harigopal S, Reddy V, Taylor N, van Saene HK. High incidence of pulmonary bacterial co-infection in children with severe respiratory syncytial virus (RSV) bronchiolitis. *Thorax* 2006; 61(7):611-615.

(42) Semple MG, Cowell A, Dove W, Greensill J, McNamara PS, Halfhide C et al. Dual infection of infants by human metapneumovirus and human respiratory syncytial virus is strongly associated with severe bronchiolitis. *J Infect Dis* 2005; 191(3):382-386.

(43) Midulla F, Scagnolari C, Bonci E, Pierangeli A, Antonelli G, De AD et al. Respiratory syncytial virus, human bocavirus and rhinovirus bronchiolitis in infants. *Arch Dis Child* 2010; 95(1):35-41.

(44) Hall CB, Weinberg GA, Iwane MK, Blumkin AK, Edwards KM, Staat MA et al. The Burden of Respiratory Syncytial Virus Infection in Young Children. *The New England Journal of Medicine* 2009; 360(6):588-598.

(45) Garcia 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-e1460.

(46) 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):543-546.

(47) Majeed A. Sources, uses, strengths and limitations of data collected in primary care in England. *Health Stat Q* 2004;(21):5-14.

(48) National Institute for Health and Clinical Excellence. NICE Clinical Guideline 69: Respiratory tract infections - antibiotic prescribing. Prescribing of antibiotics for self-limiting respiratory tract infections in adults and children in primary care. (2008) http://www.nice.org.uk/nicemedia/pdf/CG69FullGuideline.pdf. Accessed October 2012.

(49) Viswanathan M, King VJ, Bordley C, Honeycutt AA, Wittenborn J, Jackman AM et al. Management of bronchiolitis in infants and children. *Evid Rep Technol Assess (Summ)* 2003;(69):1-5.

(50) Fitzgerald DA, Kilham HA. Bronchiolitis: assessment and evidence-based management. *Med J Aust* 2004; 180(8):399-404.

(51) Rakshi K, Couriel JM. Management of acute bronchiolitis. *Arch Dis Child* 1994; 71(5):463-469.

(52) Lakhanpaul M, MacFaul R, Werneke U, Armon K, Hemingway P, Stephenson T. An evidence-based guideline for children presenting with acute breathing difficulty. *Emerg Med J* 2009; 26(12):850-853.

(53) Madge P, Paton JY, McColl JH, Mackie PL. Prospective controlled study of four infection-control procedures to prevent nosocomial infection with respiratory syncytial virus. *Lancet* 1992; 340(8827):1079-1083.

(54) Woo PC, Chiu SS, Seto WH, Peiris M. Cost-effectiveness of rapid diagnosis of viral respiratory tract infections in pediatric patients. *J Clin Microbiol* 1997; 35(6):1579-1581.

(55) Mackie PL, McCormick EM, Williams C. Evaluation of Binax NOW RSV as an acute point-of-care screening test in a paediatric accident and emergency unit. *Commun Dis Public Health* 2004; 7(4):328-330.

(56) Mackie PL, Joannidis PA, Beattie J. Evaluation of an acute point-of-care system screening for respiratory syncytial virus infection. *J Hosp Infect* 2001; 48(1):66-71.

(57) Collins CL, Pollard AJ. Respiratory syncytial virus infections in children and adults. *J Infect* 2002; 45(1):10-17.

(58) Schauer U, Ihorst G, Rohwedder A, Petersen G, Berner R, Frank HD et al. Evaluation of respiratory syncytial virus detection by rapid antigen tests in childhood. *Klin Padiatr* 2007; 219(4):212-216.

(59) Reina J, Gonzalez GM, Ruiz de GE, Padilla E, Ballesteros F, Mari M et al. Prospective evaluation of a dot-blot enzyme immunoassay (Directigen RSV) for the antigenic detection of respiratory syncytial virus from nasopharyngeal aspirates of paediatric patients. *Clin Microbiol Infect* 2004; 10(11):967-971.

(60) Ventre K, Randolph A. Ribavirin for respiratory syncytial virus infection of the lower respiratory tract in infants and young children. *Cochrane Database Syst Rev* 2004;(4):CD000181.

(61) Spurling GK, Fonseka K, Doust J, Del MC. Antibiotics for bronchiolitis in children. *Cochrane Database Syst Rev* 2007;(1):CD005189.

(62) Patel H, Platt R, Lozano JM, Wang EE. Glucocorticoids for acute viral bronchiolitis in infants and young children. *Cochrane Database Syst Rev* 2004;(3):CD004878.

(63) Ventre K, Randolph AG. Ribavirin for respiratory syncytial virus infection of the lower respiratory tract in infants and young children. *Cochrane Database Syst Rev* 2007;(1):CD000181.

(64) Spurling GK, Doust J, Del Mar CB, Eriksson L. Antibiotics for bronchiolitis in children. *Cochrane Database Syst Rev* 2011;(6):CD005189.

(65) Fuller H, Del MC. Immunoglobulin treatment for respiratory syncytial virus infection. *Cochrane Database Syst Rev* 2006;(4):CD004883.

(66) Rodriguez WJ, Arrobio J, Fink R, Kim HW, Milburn C. Prospective follow-up and pulmonary functions from a placebo-controlled randomized trial of ribavirin therapy in respiratory syncytial virus bronchiolitis. Ribavirin Study Group. *Arch Pediatr Adolesc Med* 1999; 153(5):469-474.

(67) Everard ML, Swarbrick A, Rigby AS, Milner AD. The effect of ribavirin to treat previously healthy infants admitted with acute bronchiolitis on acute and chronic respiratory morbidity. *Respir Med* 2001; 95(4):275-280.

(68) Edell D, Khoshoo V, Ross G, Salter K. Early ribavarin treatment of bronchiolitis: effect on long-term respiratory morbidity. *Chest* 2002; 122(3):935-939.

(69) King VJ, Viswanathan M, Bordley WC, Jackman AM, Sutton SF, Lohr KN et al. Pharmacologic treatment of bronchiolitis in infants and children: a systematic review. *Arch Pediatr Adolesc Med* 2004; 158(2):127-137.

(70) Goh A, Chay OM, Foo AL, Ong EK. Efficacy of bronchodilators in the treatment of bronchiolitis. *Singapore Med J* 1997; 38(8):326-328.

(71) Patel H, Gouin S, Platt RW. Randomized, double-blind, placebo-controlled trial of oral albuterol in infants with mild-to-moderate acute viral bronchiolitis. *J Pediatr* 2003; 142(5):509-514.

(72) Cade A, Brownlee KG, Conway SP, Haigh D, Short A, Brown J et al. Randomised placebo controlled trial of nebulised corticosteroids in acute respiratory syncytial viral bronchiolitis. *Arch Dis Child* 2000; 82(2):126-130.

(73) Wong JY, Moon S, Beardsmore C, O'Callaghan C, Simpson H. No objective benefit from steroids inhaled via a spacer in infants recovering from bronchiolitis. *Eur Respir J* 2000; 15(2):388-394.

(74) Wainwright C, Altamirano L, Cheney M, Cheney J, Barber S, Price D et al. A multicenter, randomized, double-blind, controlled trial of nebulized epinephrine in infants with acute bronchiolitis. *N Engl J Med* 2003; 349(1):27-35.

(75) Hartling L, Bialy LM, Vandermeer B, Tjosvold L, Johnson DW, Plint AC et al. Epinephrine for bronchiolitis. *Cochrane Database Syst Rev* 2011;(6):CD003123.

(76) Hartling L, Fernandes RM, Bialy L, Milne A, Johnson D, Plint A et al. Steroids and bronchodilators for acute bronchiolitis in the first two years of life: systematic review and meta-analysis. *BMJ* 2011; 342:d1714.

(77) Zhang L, Mendoza-Sassi RA, Wainwright C, Klassen TP. Nebulized hypertonic saline solution for acute bronchiolitis in infants. *Cochrane Database Syst Rev* 2008;(4):CD006458.

(78) Perrotta C, Ortiz Z, Roque M. Chest physiotherapy for acute bronchiolitis in paediatric patients between 0 and 24 months old. *Cochrane Database Syst Rev* 2007;(1):CD004873.

(79) Perrotta C, Ortiz Z, Roque M. Chest physiotherapy for acute bronchiolitis in paediatric patients between 0 and 24 months old. *Cochrane Database Syst Rev* 2005;(2):CD004873.

(80) Rochat I, Leis P, Bouchardy M, Oberli C, Sourial H, Friedli-Burri M et al. Chest physiotherapy using passive expiratory techniques does not reduce bronchiolitis severity: a randomised controlled trial. *Eur J Pediatr* 2012; 171(3):457-462.

(81) Eriksson M, Bennet R, Rotzen-Ostlund M, von SM, Wirgart BZ. Population-based rates of severe respiratory syncytial virus infection in children with and without risk factors, and outcome in a tertiary care setting. *Acta Paediatr* 2002; 91(5):593-598.

(82) Greenough A, Cox S, Alexander J, Lenney W, Turnbull F, Burgess S et al. Health care utilisation of infants with chronic lung disease, related to hospitalisation for RSV infection. *Arch Dis Child* 2001; 85(6):463-468.

(83) Pedersen O, Herskind AM, Kamper J, Nielsen JP, Kristensen K. Rehospitalization for respiratory syncytial virus infection in infants with extremely low gestational age or birthweight in Denmark. *Acta Paediatr* 2003; 92(2):240-242.

(84) Fjaerli HO, Farstad T, Bratlid D. Hospitalisations for respiratory syncytial virus bronchiolitis in Akershus, Norway, 1993-2000: a population-based retrospective study. *BMC Pediatr* 2004; 4(1):25.

(85) Alexander PM, Eastaugh L, Royle J, Daley AJ, Shekerdemian LS, Penny DJ. Respiratory syncytial virus immunoprophylaxis in high-risk infants with heart disease. *J Paediatr Child Health* 2012; 48(5):395-401.

(86) Kristensen K, Stensballe LG, Bjerre J, Roth D, Fisker N, Kongstad T et al. Risk factors for respiratory syncytial virus hospitalisation in children with heart disease. *Arch Dis Child* 2009; 94(10):785-789.

(87) Weisman LE. Populations at risk for developing respiratory syncytial virus and risk factors for respiratory syncytial virus severity: infants with predisposing conditions. *Pediatr Infect Dis J* 2003; 22(2 Suppl):S33-S37.

(88) Office for National Statistics. News Release: 1 in 13 live births in England and Wales are born preterm. 24-5-2007. http://www.statistics.gov.uk/pdfdir/preterm0507.pdf Accessed October 2012.

(89) Horn SD, Smout RJ. Effect of prematurity on respiratory syncytial virus hospital resource use and outcomes. *J Pediatr* 2003; 143(5 Suppl):S133-S141.

(90) Moser K and Hilder L. Assessing quality of NHS Numbers for Babies data and providing gestational age statistics. *Health Statistics Quarterly* 2008; Spring 2008.

(91) Moser K SKMLDA. Birthweight and gestational age by ethnic group,England and Wales 2005: introducing new data on births. *Health Statistics Quarterly* 2008; Autumn 2008.

(92) Gouyon JB, Roze JC, Guillermet-Fromentin C, Glorieux I, Adamon L, DI MM et al. Hospitalizations for respiratory syncytial virus bronchiolitis in preterm infants at <33 weeks gestation without bronchopulmonary dysplasia: the CASTOR study. *Epidemiol Infect* 2012;1-11.

(93) Simon A, Ammann RA, Wilkesmann A, Eis-Hubinger AM, Schildgen O, Weimann E et al. Respiratory syncytial virus infection in 406 hospitalized premature infants: results from a prospective German multicentre database. *Eur J Pediatr* 2007; 166(12):1273-1283.

(94) Manzoni P, Paes B, Resch B, Carbonell-Estrany X, Bont L. High risk for RSV bronchiolitis in late preterms and selected infants affected by rare disorders: a dilemma of specific prevention. *Early Hum Dev* 2012; 88 Suppl 2:S34-S41.

(95) Larroque B, Breart G, Kaminski M, Dehan M, Andre M, Burguet A et al. Survival of very preterm infants: Epipage, a population based cohort study. *Arch Dis Child Fetal Neonatal Ed* 2004; 89(2):F139-F144.

(96) Spencer N, Logan S, Scholey S, Gentle S. Deprivation and bronchiolitis. *Arch Dis Child* 1996; 74(1):50-52.

(97) Colosia AD, Masaquel A, Hall CB, Barrett AM, Mahadevia PJ, Yogev R. Residential crowding and severe respiratory syncytial virus disease among infants and young children: a systematic literature review. *BMC Infect Dis* 2012; 12:95.

(98) Carbonell-Estrany X A V I, Quero J O S, The IRIS Study Group. Hospitalization rates for respiratory syncytial virus infection in premature infants born during two consecutive seasons. *The Pediatric Infectious Disease Journal* 2001; 20(9).

(99) Semple MG, Taylor-Robinson DC, Lane S, Smyth RL. Household tobacco smoke and admission weight predict severe bronchiolitis in infants independent of deprivation: prospective cohort study. *PLoS One* 2011; 6(7):e22425.

(100) Strachan DP, Cook DG. Health effects of passive smoking. 1. Parental smoking and lower respiratory illness in infancy and early childhood. *Thorax* 1997; 52(10):905-914.

(101) Bradley JP, Bacharier LB, Bonfiglio J, Schechtman KB, Strunk R, Storch G et al. Severity of respiratory syncytial virus bronchiolitis is affected by cigarette smoke exposure and atopy. *Pediatrics* 2005; 115(1):e7-14.

(102) DiFranza JR, Masaquel A, Barrett AM, Colosia AD, Mahadevia PJ. Systematic literature review assessing tobacco smoke exposure as a risk factor for serious respiratory syncytial virus disease among infants and young children. *BMC Pediatr* 2012; 12:81.

(103) Broughton S, Roberts A, Fox G, Pollina E, Zuckerman M, Chaudhry S et al. Prospective study of healthcare utilisation and respiratory morbidity due to RSV infection in prematurely born infants. *Thorax* 2005; 60(12):1039-1044.

(104) Nielsen HE, Siersma V, Andersen S, Gahrn-Hansen B, Mordhorst CH, Norgaard-Pedersen B et al. Respiratory syncytial virus infection--risk factors for hospital admission: a case-control study. *Acta Paediatr* 2003; 92(11):1314-1321.

(105) Karr CJ, Rudra CB, Miller KA, Gould TR, Larson T, Sathyanarayana S et al. Infant exposure to fine particulate matter and traffic and risk of hospitalization for RSV bronchiolitis in a region with lower ambient air pollution. *Environ Res* 2009; 109(3):321-327.

(106) Carbonell-Estrany X, Figueras-Aloy J, Law BJ. Identifying risk factors for severe respiratory syncytial virus among infants born after 33 through 35 completed weeks of gestation: different methodologies yield consistent findings. *Pediatr Infect Dis J* 2004; 23(11 Suppl):S193-S201.

(107) Bulkow LR, Singleton RJ, Karron RA, Harrison LH. Risk factors for severe respiratory syncytial virus infection among Alaska native children. *Pediatrics* 2002; 109(2):210-216.

(108) Simoes EA, King SJ, Lehr MV, Groothuis JR. Preterm twins and triplets. A high-risk group for severe respiratory syncytial virus infection. *Am J Dis Child* 1993; 147(3):303-306.

(109) Moore HC, de KN, Holt P, Richmond PC, Lehmann D. Hospitalisation for bronchiolitis in infants is more common after elective caesarean delivery. *Arch Dis Child* 2012; 97(5):410-414.

(110) Law BJ, Carbonell-Estrany X, Simoes EA. An update on respiratory syncytial virus epidemiology: a developed country perspective. *Respir Med* 2002; 96 Suppl B:S1-S7.

(111) 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-86.

(112) Wang D, Bayliss S, Meads C. Palivizumab for immunoprophylaxis of respiratory syncitial 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):1-124.

(113) Marlais M, Evans J, Abrahamson E. Clinical predictors of admission in infants with acute bronchiolitis. *Arch Dis Child* 2011; 96(7):648-652.

(114) Schroeder AR, Marmor AK, Pantell RH, Newman TB. Impact of pulse oximetry and oxygen therapy on length of stay in bronchiolitis hospitalizations. *Arch Pediatr Adolesc Med* 2004; 158(6):527-530.

(115) Mansbach JM, Clark S, Christopher NC, LoVecchio F, Kunz S, Acholonu U et al. Prospective multicenter study of bronchiolitis: predicting safe discharges from the emergency department. *Pediatrics* 2008; 121(4):680-688.

(116) Paes BA. Current strategies in the prevention of respiratory syncytial virus disease. *Paediatr Respir Rev* 2003; 4(1):21-27.

(117) Macartney KK, Gorelick MH, Manning ML, Hodinka RL, Bell LM. Nosocomial respiratory syncytial virus infections: the cost-effectiveness and cost-benefit of infection control. *Pediatrics* 2000; 106(3):520-526.

(118) Thwaites R, Piercy J. Nosocomial respiratory syncytial virus infection in neonatal units in the United Kingdom. *Acta Paediatr Suppl* 2004; 93(444):23-25.

(119) Strutton DR, Stang PE. Prophylaxis against respiratory syncytial virus (RSV), varicella, and pneumococcal infections: Economic-based decision-making. *J Pediatr* 2003; 143(5 Suppl):S157-S162.

(120) Simoes EA, Groothuis JR. Respiratory syncytial virus prophylaxis--the story so far. *Respir Med* 2002; 96 Suppl B:S15-S24.

(121) Embleton ND, Harkensee C, Mckean MC. Palivizumab for preterm infants. Is it worth it? *Arch Dis Child Fetal Neonatal Ed* 2005; 90(4):F286-F289.

(122) The IMpact-RSV Study Group. Palivizumab, a humanized respiratory syncytial virus monoclonal antibody, reduces hospitalization from respiratory syncytial virus infection in high-risk infants. *Pediatrics* 1998; 102(3 Pt 1):531-537.

(123) Feltes TF, Cabalka AK, Meissner HC, Piazza FM, Carlin DA, Top J et al. Palivizumab prophylaxis reduces hospitalization due to respiratory syncytial virus in young children with hemodynamically significant congenital heart disease. *The Journal of Pediatrics* 2003; 143(4):532-540.

(124) Joint Committee on Vaccination and Immunisation - Respiratory Syncitial Virus (RSV) Subgroup. Minutes of the RSV subgroup on 8th June 2010. (2010) http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@ab/documents/digitalasset/dh_118086.pdf. Accessed October 2012.

(125) Meissner HC, Bocchini JA, Jr., Brady MT, Hall CB, Kimberlin DW, Pickering LK. The role of immunoprophylaxis in the reduction of disease attributable to respiratory syncytial virus. *Pediatrics* 2009; 124(6):1676-1679.

(126) NIHR Health Technology Assessment Programme. Palivizumab for immunoprophylaxis of respiratory syncitial virus (RSV) bronchiolitis in high risk infants, additional subgroup analysis.(2011) http://www.hta.ac.uk/project/2056.asp. Accessed October 2012.

(127) Shadman KA, Wald ER. A review of palivizumab and emerging therapies for respiratory syncytial virus. *Expert Opin Biol Ther* 2011; 11(11):1455-1467.

(128) Carbonell-Estrany X, Simoes EA, Dagan R, Hall CB, Harris B, Hultquist M et al. Motavizumab for prophylaxis of respiratory syncytial virus in high-risk children: a noninferiority trial. *Pediatrics* 2010; 125(1):e35-e51.

(129) US Food and Drug Administration. Summary Minutes of the Antiviral Drugs Advisory Committee (2010) http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/AntiviralDrugsAdvisoryCommittee/UCM241510.pdf. Accessed October 2012.

(130) Handforth J, Friedland JS, Sharland M. Inhaled corticosteroids after respiratory syncytial virus infection. *BMJ* 2009; 338(mar31_2):b164.

(131) Astrazeneca MedImmune. MedImmune discontinues development of motavizumab for RSV prophylaxis indication. (2012) http://pressroom.medimmune.com/press-releases/2010/12/21/medimmune-discontinues-development-of-motavizumab-for-rsv-prophylaxis-indication/. Gaithersburg, MD, MedImmune/Social media press releases. Accessed October 2012.

(132) Habibi MS, Openshaw PJ. Benefit and harm from immunity to respiratory syncytial virus: implications for treatment. *Curr Opin Infect Dis* 2012; 25(6):687-694.

(133) Openshaw PJ. Potential therapeutic implications of new insights into respiratory syncytial virus disease. *Respir Res* 2002; 3 Suppl 1:S15-S20.

(134) Habibi MS, Patel S, Openshaw P. Hot topics in the prevention of respiratory syncytial virus disease. *Expert Rev Vaccines* 2011; 10(3):291-293.

(135) 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):435-448.

(136) 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):405-421.

(137) Meijboom MJ, Rozenbaum MH, Benedictus A, Luytjes W, Kneyber MC, Wilschut JC et al. Cost-effectiveness of potential infant vaccination against respiratory syncytial virus infection in The Netherlands. *Vaccine* 2012; 30(31):4691-4700.

(138) Leung AK, Kellner JD, Davies HD. Respiratory syncytial virus bronchiolitis. *J Natl Med Assoc* 2005; 97(12):1708-1713.

(139) Sigurs N. A cohort of children hospitalised with acute RSV bronchiolitis: impact on later respiratory disease. *Paediatr Respir Rev* 2002; 3(3):177-183.

(140) Sigurs N.E.L.E., Bjarnason R.A.G.N., Sigurbergsson F.R.I.D., Kjellman B.E.N.G. Respiratory Syncytial Virus Bronchiolitis in Infancy Is an Important Risk Factor for Asthma and Allergy at Age 7. *Am J Respir Crit Care Med* 2000; 161(5):1501-1507.

(141) Sigurs N, Aljassim F, Kjellman B, Robinson PD, Sigurbergsson F, Bjarnason R et al. Asthma and allergy patterns over 18 years after severe RSV bronchiolitis in the first year of life. *Thorax* 2010; 65(12):1045-1052.

(142) Simoes EA. RSV disease in the pediatric population: epidemiology, seasonal variability, and long-term outcomes. *Manag Care* 2008; 17(11 Suppl 12):3-6, discussion.

(143) Henderson J, Hilliard TN, Sherriff A, Stalker D, Shammari NA, Thomas HM et al. Hospitalization for RSV bronchiolitis before 12-ámonths of age and subsequent asthma, atopy and wheeze: A longitudinal birth cohort study. *Pediatric Allergy and Immunology* 2005; 16(5):386-392.

(144) Stein RT, Sherrill D, Morgan WJ, Holberg CJ, Halonen M, Taussig LM et al. Respiratory syncytial virus in early life and risk of wheeze and allergy by age 13 years. *The Lancet* 1999; 354(9178):541-545.

(145) Noble V, Murray M, Webb MS, Alexander J, Swarbrick AS, Milner AD. Respiratory status and allergy nine to 10 years after acute bronchiolitis. *Arch Dis Child* 1997; 76(4):315-319.

(146) Mailaparambil B, Grychtol R, Heinzmann A. Respiratory syncytial virus bronchiolitis and asthma - insights from recent studies and implications for therapy. *Inflamm Allergy Drug Targets* 2009; 8(3):202-207.

(147) Henderson J, Hilliard TN, Sherriff A, Stalker D, Al SN, Thomas HM. Hospitalization for RSV bronchiolitis before 12 months of age and subsequent asthma, atopy and wheeze: a longitudinal birth cohort study. *Pediatr Allergy Immunol* 2005; 16(5):386-392.

(148) Stein RT. Long-term airway morbidity following viral LRTI in early infancy: recurrent wheezing or asthma? *Paediatr Respir Rev* 2009; 10 Suppl 1:29-31.

(149) Stein RT, Martinez FD. Respiratory syncytial virus and asthma: still no final answer. *Thorax* 2010; 65(12):1033-1034.

(150) Mikalsen IB, Halvorsen T, Eide GE, Oymar K. Severe bronchiolitis in infancy: Can asthma in adolescence be predicted? *Pediatr Pulmonol* 2012.

(151) Bacharier LB, Cohen R, Schweiger T, Yin-Declue H, Christie C, Zheng J et al. Determinants of asthma after severe respiratory syncytial virus bronchiolitis. *J Allergy Clin Immunol* 2012; 130(1):91-100.

(152) Escobar GJ, Ragins A, Li SX, Prager L, Masaquel AS, Kipnis P. Recurrent wheezing in the third year of life among children born at 32 weeks' gestation or later: relationship to laboratory-confirmed, medically attended infection with respiratory syncytial virus during the first year of life. *Arch Pediatr Adolesc Med* 2010; 164(10):915-922.

(153) Ruotsalainen M, Piippo-Savolainen E, Hyvarinen MK, Korppi M. Respiratory morbidity in adulthood after respiratory syncytial virus hospitalization in infancy. *Pediatr Infect Dis J* 2010; 29(9):872-874.

(154) You D, Becnel D, Wang K, Ripple M, Daly M, Cormier SA. Exposure of neonates to respiratory syncytial virus is critical in determining subsequent airway response in adults. *Respir Res* 2006; 7:107.

(155) Greenough A, Broughton S. Chronic manifestations of respiratory syncytial virus infection in premature infants. *Pediatr Infect Dis J* 2005; 24(11 Suppl):S184-7, discussion.

(156) Greenough A, Alexander J, Boit P, Boorman J, Burgess S, Burke A et al. School age outcome of hospitalisation with respiratory syncytial virus infection of prematurely born infants. *Thorax* 2009; 64(6):490-495.

(157) Bont L, Steijn M, Van Aalderen WM, Kimpen JL. Impact of wheezing after respiratory syncytial virus infection on health-related quality of life. *Pediatr Infect Dis J* 2004; 23(5):414-417.

(158) Korppi M, Piippo-Savolainen E, Korhonen K, Remes S. Respiratory morbidity 20 years after RSV infection in infancy. *Pediatr Pulmonol* 2004; 38(2):155-160.

(159) Drysdale SB, Milner AD, Greenough A. Respiratory syncytial virus infection and chronic respiratory morbidity - is there a functional or genetic predisposition? *Acta Paediatr* 2012; 101(11):1114-1120.

(160) Everard ML. The relationship between respiratory syncytial virus infections and the development of wheezing and asthma in children. *Curr Opin Allergy Clin Immunol* 2006; 6(1):56-61.

(161) Kim CK, Kim SW, Kim YK, Kang H, Yu J, Yoo Y et al. Bronchoalveolar lavage eosinophil cationic protein and interleukin-8 levels in acute asthma and acute bronchiolitis. *Clin Exp Allergy* 2005; 35(5):591-597.

(162) Blanken MO, Koffijberg H, Nibbelke EE, Rovers MM, Bont L. Prospective validation of a prognostic model for respiratory syncytial virus bronchiolitis in late preterm infants: a multicenter birth cohort study. *PLoS One* 2013; 8(3):e59161.

(163) Blanken MO, Rovers MM, Molenaar JM, Winkler-Seinstra PL, Meijer A, Kimpen JL et al. Respiratory syncytial virus and recurrent wheeze in healthy preterm infants. *N Engl J Med* 2013; 368(19):1791-1799.

(164) Simoes EA, Carbonell-Estrany X, Rieger CH, Mitchell I, Fredrick L, Groothuis JR. The effect of respiratory syncytial virus on subsequent recurrent wheezing in atopic and nonatopic children. *J Allergy Clin Immunol* 2010; 126(2):256-262.

(165) Piedimonte G. Respiratory syncytial virus and asthma: speed-dating or long-term relationship? *Curr Opin Pediatr* 2013; 25(3):344-349.

(166) Office for National Statistics. Birth statistics: Births and patterns of family building England and Wales (FM1). 15-12-2008. http://www.statistics.gov.uk/downloads/theme_population/FM1_36/FM1-No36.pdf Accessed October 2012.

(167) Hertzman C, Siddiqi A, Hertzman E, Irwin LG, Vaghri Z, Houweling TA et al. Bucking the inequality gradient through early child development. *BMJ* 2010; 340:c468.

(168) Commission on the Social Determinants of Health. Closing the gap in a generation: health equity through action on the social determinants of health. World Health Organisation [2008

(169) O'Lonergan T, Milgrom H. Ethical considerations in research involving children. *Current Allergy and Asthma Reports* 2005; 5(6):451-458.

(170) Modi N. Survival after extremely preterm birth. *BMJ* 2008; 336(7655):1199-1200.

(171) Smyth RL. Research with children. Paediatric practice needs better evidence--gained in collaboration with parents and children. *BMJ* 2001; 322(7299):1377-1378.

(172) Smyth RL, Weindling AM. Research in children: ethical and scientific aspects. *Lancet* 1999; 354 Suppl 2:SII21-SII24.

(173) Waters KL, Wiebe N, Cramer K, Hartling L, Klassen TP. Treatment in the pediatric emergency department is evidence based: a retrospective analysis. *BMC Pediatr* 2006; 6:26.

(174) Smyth RL. Making a difference: the clinical research programme for children. *Arch Dis Child* 2007; 92(10):835-837.

(175) Rudolf MC, Lyth N, Bundle A, Rowland G, Kelly A, Bosson S et al. A search for the evidence supporting community paediatric practice. *Arch Dis Child* 1999; 80(3):257-261.

(176) Modi N, Clark H, Wolfe I, Costello A, Budge H. A healthy nation: strengthening child health research in the UK. *Lancet* 2012.

(177) Medicines for Children Research Network (MCRN). Annual Report 2007/2008. MCRN (2010) Available from: http://ctuprod.liv.ac.uk/mcrnweb/images/stories/1b_mcrnccannualreport07-08.pdf Accessed October 2012.

(178) Modi N. Turning the Tide: increasing and strengthening child health research. *Archives of Disease in Childhood* 2011; 96(10):988.

(179) Bray RJ. Propofol infusion syndrome in children. *Paediatr Anaesth* 1998; 8(6):491-499.

(180) Choonara I. Clinical trials of medicines in children. *BMJ* 2000; 321(7269):1093-1094.

(181) Conroy S, Choonara I, Impicciatore P, Mohn A, Arnell H, Rane A et al. Survey of unlicensed and off label drug use in paediatric wards in European countries. European Network for Drug Investigation in Children. *BMJ* 2000; 320(7227):79-82.

(182) Choonara I, Bauchner H. Publishing the evidence for children's medicines. *Arch Dis Child* 2008; 93(10):815.

(183) Choonara I. Improving children's medicines. *Arch Dis Child* 2006; 91(7):550-551.

(184) Campbell H, Surry SA, Royle EM. A review of randomised controlled trials published in Archives of Disease in Childhood from 1982-96. *Arch Dis Child* 1998; 79(2):192-197.

(185) Napoleone E. Children and ADRs (Adverse Drug Reactions). *Ital J Pediatr* 2010; 36(1):4.

(186) Bodeau-Livinec F, Marlow N, Ancel PY, Kurinczuk JJ, Costeloe K, Kaminski M. Impact of intensive care practices on short-term and long-term outcomes for extremely preterm infants: comparison between the British Isles and France. *Pediatrics* 2008; 122(5):e1014-e1021.

(187) Fewtrell MS, Kennedy K, Singhal A, Martin RM, Ness A, Hadders-Algra M et al. How much loss to follow-up is acceptable in long-term randomised trials and prospective studies? *Archives of Disease in Childhood* 2008; 93(6):458-461.

(188) Ellenberg JH. Selection bias in observational and experimental studies. *Statist Med* 1994; 13(5-7):557-567.

(189) Souhami R. Governance of research that uses identifiable personal data. *BMJ* 2006; 333(7563):315-316.

(190) Sharland M, Kendall H, Yeates D, Randall A, Hughes G, Glasziou P et al. Antibiotic prescribing in general practice and hospital admissions for peritonsillar abscess, mastoiditis, and rheumatic fever in children: time trend analysis. *BMJ* 2005; 331(7512):328-329.

(191) Ladhani S, Pebody RG, Ramsay ME, Lamagni TL, Johnson AP, Sharland M. Continuing impact of infectious diseases on childhood deaths in England and Wales, 2003-2005. *Pediatr Infect Dis J* 2010; 29(4):310-313.

(192) Thompson PL, Gilbert RE, Long PF, Saxena S, Sharland M, Wong IC. Effect of antibiotics for otitis media on mastoiditis in children: a retrospective cohort study using the United kingdom general practice research database. *Pediatrics* 2009; 123(2):424-430.

(193) Murray ML, de Vries CS, Wong IC. A drug utilisation study of antidepressants in children and adolescents using the General Practice Research Database. *Arch Dis Child* 2004; 89(12):1098-1102.

(194) Koshy E, Murray J, Bottle A, Aylin P, Sharland M, Majeed A et al. Significantly increasing hospital admissions for acute throat infections among children in England: is this related to tonsillectomy rates? *Arch Dis Child* 2012.

(195) Mayor S. Oxford hospital suspends paediatric cardiac surgery after four deaths. *BMJ* 2010; 340(mar04_2):c1303.

(196) Davies C, Collins R. Balancing potential risks and benefits of using confidential data. *BMJ* 2006; 333(7563):349-351.

(197) Knowles RL, Bull C, Wren C, Dezateux C. Ethics, governance and consent in the UK: implications for research into the longer-term outcomes of congenital heart defects. *Arch Dis Child* 2009.

(198) Metcalfe C, Martin RM, Noble S, Lane JA, Hamdy FC, Neal DE et al. Low risk research using routinely collected identifiable health information without informed consent: encounters with the Patient Information Advisory Group. *J Med Ethics* 2008; 34(1):37-40.

(199) Cassell J, Young A. Why we should not seek individual informed consent for participation in health services research. *J Med Ethics* 2002; 28(5):313-317.

(200) Kalra D, Gertz R, Singleton P, Inskip HM. Confidentiality of personal health information used for research. *BMJ* 2006; 333(7560):196-198.

(201) Hansson MG. Need for a wider view of autonomy in epidemiological research. *BMJ* 2010; 340:c2335.

(202) McKinney PA, Jones S, Parslow R, Davey N, Darowski M, Chaudhry B et al. A feasibility study of signed consent for the collection of patient identifiable information for a national paediatric clinical audit database. *BMJ* 2005; 330(7496):877-879.

(203) Iversen A, Liddell K, Fear N, Hotopf M, Wessely S. Consent, confidentiality, and the Data Protection Act. *BMJ* 2006; 332(7534):165-169.

(204) Tu JV, Willison DJ, Silver FL, Fang J, Richards JA, Laupacis A et al. Impracticability of informed consent in the Registry of the Canadian Stroke Network. *N Engl J Med* 2004; 350(14):1414-1421.

(205) Hockley C, Quigley MA, Hughes G, Calderwood L, Joshi H, Davidson LL. Linking Millennium Cohort data to birth registration and hospital episode records. *Paediatr Perinat Epidemiol* 2008; 22(1):99-109.

(206) The Academy of Medical Sciences. Personal data for public good: using health information in medical research. (2006) http://www.acmedsci.ac.uk/p99puid62.html. London, UK. Accessed October 2012.

(207) UK Clinical Research Collaboration and the Wellcome Trust. Frontiers meeting on the use of electronic patient records for research and health benefit. (2007) http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtd038686.pdf. London, UK. Accessed October 2012.

(208) UKCRC Research and Development Advisory Group to Connecting for Health. Report of the research simulations (2007) http://www.ukcrc.org/publications/reports/. Accessed October 2012.

(209) Vergnano S, Menson E, Kennea N, Embleton N, Russell AB, Watts T et al. Neonatal infections in England: the NeonIN surveillance network. *Arch Dis Child Fetal Neonatal Ed* 2011; 96(1):F9-F14.

(210) Field K, Kosmider S, Johns J, Farrugia H, Hastie I, Croxford M et al. Linking Data from Hospital and Cancer Registry Databases: Should this be Standard Practice? *Intern Med J* 2009.

(211) MacLeod MC, Bray CA, Kendrick SW, Cobbe SM. Enhancing the power of record linkage involving low quality personal identifiers: use of the best link principle and cause of death prior likelihoods. *Comput Biomed Res* 1998; 31(4):257-270.

(212) Chamberlayne R, Green B, Barer ML, Hertzman C, Lawrence WJ, Sheps SB. Creating a population-based linked health database: a new resource for health services research. *Can J Public Health* 1998; 89(4):270-273.

(213) Roos LL, Brownell M, Lix L, Roos NP, Walld R, MacWilliam L. From health research to social research: privacy, methods, approaches. *Soc Sci Med* 2008; 66(1):117-129.

(214) Taylor LK, TRAVIS S, PYM M, OLIVE E, HENDERSON-SMART DJ. How useful are hospital morbidity data for monitoring conditions occurring in the perinatal period? *Australian and New Zealand Journal of Obstetrics and Gynaecology* 2005; 45(1):36-41.

(215) Ford JB, Roberts CL, Taylor LK. Characteristics of unmatched maternal and baby records in linked birth records and hospital discharge data. *Paediatric and Perinatal Epidemiology* 2006; 20(4):329-337.

(216) Bradford H, C+írdenas V, Camacho-Carr K, Lydon-Rochelle M. Accuracy of Birth Certificate and Hospital Discharge Data: A Certified Nurse-Midwife and Physician Comparison. *Maternal and Child Health Journal* 2007; 11(6):540-548.

(217) Lydon-Rochelle MT, Holt VL, Nelson JC, C+írdenas V, Gardella C, Easterling TR et al. Accuracy of reporting maternal in-hospital diagnoses and intrapartum procedures in Washington State linked birth records. *Paediatric and Perinatal Epidemiology* 2005; 19(6):460-471.

(218) Lauritsen MB, Pedersen CB, Mortensen PB. Effects of familial risk factors and place of birth on the risk of autism: a nationwide register-based study. *J Child Psychol Psychiatry* 2005; 46(9):963-971.

(219) Linnet KM, Wisborg K, Secher NJ, Thomsen PH, Obel C, Dalsgaard S et al. Coffee consumption during pregnancy and the risk of hyperkinetic disorder and ADHD: a prospective cohort study. *Acta Paediatr* 2009; 98(1):173-179.

(220) Lyons R, Jones K, John G, Brooks C, Verplancke JP, Ford D et al. The SAIL databank: linking multiple health and social care datasets. *BMC Medical Informatics and Decision Making* 2009; 9(1):3.

(221) Welsh Electonic Cohort for Childern (WECC) 2011 http://www.swan.ac.uk/ils/research/chiral/methodologies/healthinformatics/wecc/. Accessed October 2012.

(222) McGregor J, Brooks C, Chalasani P, Chukwuma J, Hutchings H, Lyons RA et al. The Health Informatics Trial Enhancement Project (HITE): Using routinely collected primary care data to identify potential participants for a depression trial. *Trials* 2010; 11:39.

(223) Scottish Health Informatics Programme. http://www.scot-ship.ac.uk (2011) Accessed October 2012.

(224) Govan L, Wu O, Briggs A, Colhoun HM, McKnight JA, Morris AD et al. Inpatient costs for people with type 1 and type 2 diabetes in Scotland: a study from the Scottish Diabetes Research Network Epidemiology Group. *Diabetologia* 2011; 54(8):2000-2008.

(225) Gill L, Goldacre M, Simmons H, Bettley G, Griffith M. Computerised linking of medical records: methodological guidelines. *J Epidemiol Community Health* 1993; 47(4):316-319.

(226) Roberts SE, Wotton CJ, Williams JG, Griffith M, Goldacre MJ. Perinatal and early life risk factors for inflammatory bowel disease. *World J Gastroenterol* 2011; 17(6):743-749.

(227) Davidson R, Roberts SE, Wotton CJ, Goldacre MJ. Influence of maternal and perinatal factors on subsequent hospitalisation for asthma in children: evidence from the Oxford record linkage study. *BMC Pulm Med* 2010; 10:14.

(228) Roberts SE, Williams JG, Meddings D, Davidson R, Goldacre MJ. Perinatal risk factors and coeliac disease in children and young adults: a record linkage study. *Aliment Pharmacol Ther* 2009; 29(2):222-231.

(229) Abrahms C, Davy K. Linking HES maternity records with ONS birth records. *Health Statistics Quarterly* 2002; Spring Edition:22-30.

(230) Dattani N, Datta-Nemdharry P, Macfarlane A. Linking maternity data for England, 2005-06: methods and data quality. *Health Statistics Quarterly* 2011; 49(Spring):53-79 doi:10.1057/hsq.2011.3.

(231) Hilder L, Moser K, Dattani N, Macfarlane A. Pilot linkage of NHS Numbers for Babies data with birth registrations. *Health Statistics Quarterly* 2007; Spring(33):25-33.

(232) Dezateux C, Hockley C, Johnson J, Joshi H, Quigley M, Rosenburg R. Millenium Cohort study: Birth registration and maternity Hospital Episode Statistics Linkage. Centre for Longitudinal Studies Report (2006) Accessed October 2012.

(233) UKCRC Research and Development Advisory Group to Connecting for Health. Report of the research simulations (2007) Accessed October 2012.

(234) Black N, Payne M. Directory of clinical databases: improving and promoting their use. *Qual Saf Health Care* 2003; 12(5):348-352.

(235) Newcombe HB, KENNEDY JM, AXFORD SJ, JAMES AP. Automatic linkage of vital records. *Science* 1959; 130:954-959.

(236) The Scottish Medical Record Linkage System.(2010) http://www.isdscotland.org/Products-and-Services/eDRIS/SHIP/ Accessed October 2012.

(237) Roberts CL, Algert CS, Ford JB. Methods for dealing with discrepant records in linked population health datasets: a cross-sectional study. *BMC Health Serv Res* 2007; 7:12.

(238) Campbell SE, Campbell MK, Grimshaw JM, Walker AE. A systematic review of discharge coding accuracy. *Journal of Public Health* 2001; 23(3):205-211.

(239) The Audit Commission. Information and data quality in the NHS: Key messages from three years of independent review. (2004) http://www.audit-commission.gov.uk/SiteCollectionDocuments/AuditCommissionReports/NationalStudies/20040330dataquality.pdf. Accessed October 2012.

(240) Holt PJ, Poloniecki JD, Thompson MM. Multicentre study of the quality of a large administrative data set and implications for comparing death rates. *Br J Surg* 2012; 99(1):58-65.

(241) Tromp M, Ravelli AC, Meray N, Reitsma JB, Bonsel GJ. An efficient validation method of probabilistic record linkage including readmissions and twins. *Methods Inf Med* 2008; 47(4):356-363.

(242) The NHS Information Centre for Health and Social Care HESonline Hospital Episode Statistics. (2010) www.hesonline.nhs.uk. Accessed October 2012.

(243) Aylin P, Williams S, Bottle A, Jarman B. Counting hospital activity: spells or episodes? *BMJ* 2004; 329(7476):1207.

(244) World Health Organisation (WHO). International Classification of Disease (ICD) Version 10. (2010) http://apps.who.int/classifications/icd10/browse/2010/en Accessed October 2012.

(245) The NHS Information Centre for Health and Social Care. HES Online Data Dictionary: Inpatients. (2009) http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=571 Accessed October 2012.

(246) Campbell SE, Campbell MK, Grimshaw JM, Walker AE. A systematic review of discharge coding accuracy. *J Public Health Med* 2001; 23(3):205-211.

(247) Williams JG, Mann RY. Hospital episode statistics: time for clinicians to get involved? *Clin Med* 2002; 2(1):34-37.

(248) Bottle A, Hansell A, Aylin P. Hospital episode statistics: time for clinicians to get involved? *Clin Med* 2002; 2(5):483-484.

(249) Bottle A, Aylin P. Application of AHRQ patient safety indicators to English hospital data. *Qual Saf Health Care* 2009; 18(4):303-308.

(250) Aylin P, Yunus A, Bottle A, Majeed A, Bell D. Weekend mortality for emergency admissions. A large, multicentre study. *Qual Saf Health Care* 2010; 19(3):213-217.

(251) Aylin P, Bottle A, Majeed A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ* 2007; 334(7602):1044.

(252) Hansell A, Bottle A, Shurlock L, Aylin P. Accessing and using hospital activity data. *J Public Health Med* 2001; 23(1):51-56.

(253) Koshy E, Murray J, Bottle A, Sharland M, Saxena S. Impact of the seven-valent pneumococcal conjugate vaccination (PCV7) programme on childhood hospital admissions for bacterial pneumonia and empyema in England: national time-trends study, 1997-2008. *Thorax* 2010; 65(9):770-774.

(254) Jen MH, Saxena S, Bottle A, Pollok R, Holmes A, Aylin P. Assessment of administrative data for evaluating the shifting acquisition of Clostridium difficile infection in England. *J Hosp Infect* 2012; 80(3):229-237.

(255) Farrar S, Yi D, Sutton M, Chalkley M, Sussex J, Scott A. Has payment by results affected the way that English hospitals provide care? Difference-in-differences analysis. *BMJ* 2009; 339:b3047.

(256) The NHS Information Centre for Health and Social Care HESonline Hospital Episode Statistic. NHS Maternity Statistics (2010) http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=1475. Accessed October 2012.

(257) Murray J, Saxena S, Modi N, Majeed A, Aylin P, Bottle A. Quality of routine hospital birth records and the feasibility of their use for creating birth cohorts. *J Public Health (Oxf)* 2012.

(258) The General Practice Research Database. Using GPRD for primary care research. http://www.gprd.com/academia/primarycare.asp. (2011) Accessed October 2012.

(259) The Clinical Practice Research Datalink. http://www.cprd.com/home/. Accessed October 2012.

(260) Dave S, Petersen I. Creating medical and drug code lists to identify cases in primary care databases. *Pharmacoepidemiol Drug Saf* 2009; 18(8):704-707.

(261) Chisholm J. The Read clinical classification. *BMJ* 1990; 300(6732):1092.

(262) Wong IC, Murray ML. The potential of UK clinical databases in enhancing paediatric medication research. *Br J Clin Pharmacol* 2005; 59(6):750-755.

(263) Ackers R, Besag FM, Hughes E, Squier W, Murray ML, Wong IC. Mortality rates and causes of death in children with epilepsy prescribed antiepileptic drugs: a retrospective cohort study using the UK General Practice Research Database. *Drug Saf* 2011; 34(5):403-413.

(264) Schneider-Lindner V, Quach C, Hanley JA, Suissa S. Antibacterial drugs and the risk of community-associated methicillin-resistant Staphylococcus aureus in children. *Arch Pediatr Adolesc Med* 2011; 165(12):1107-1114.

(265) Stowe J, Andrews N, Bryan P, Seabroke S, Miller E. Risk of convulsions in children after monovalent H1N1 (2009) and trivalent influenza vaccines: a database study. *Vaccine* 2011; 29(51):9467-9472.

(266) Smeeth L, Cook C, Fombonne E, Heavey L, Rodrigues LC, Smith PG et al. MMR vaccination and pervasive developmental disorders: a case-control study. *Lancet* 2004; 364(9438):963-969.

(267) Kaye JA, del MM-M, Jick H. Mumps, measles, and rubella vaccine and the incidence of autism recorded by general practitioners: a time trend analysis. *BMJ* 2001; 322(7284):460-463.

(268) Report of the Department for Health Neonatal Intensive Care Services Review Group. Neonatal Intensive Care Review - Strategy for Improvement. 2003. Department of Health. Accessed October 2012.

(269) Department of Health. Neonatal intensive care review: strategy for improvement. www nao org uk/publications/nao_reports/07-08/0708101 pdf [2003

(270) National Audit Office. Caring for vulnerable babies: the reorganisation of neonatal services in England. (2007) www.nao.org.uk/publications/nao_reports/07-08/0708101.pdf. Accessed October 2012.

(271) The Neonatal Data Analysis Unit. NDAU Annual Reports http://www1.imperial.ac.uk/departmentofmedicine/divisions/infectiousdiseases/paediatrics/neonatalmedicine/ndau/. Accessed October 2012.

(272) The British Association of Perinatal Medicine. http://www.bapm.org/. Accessed October 2012.

(273) The Royal College of Paediatrics and Child Health. The National Neonatal Audit Programme http://www.rcpch.ac.uk/nnap. Accessed October 2012.

(274) Clevermed. Badgernet Neonatal http://clevermed.com/Neonatal.aspx. Accessed October 2012.

(275) The Neonatal Data Analysis Unit. NDAU 2011 report http://www1.imperial.ac.uk/departmentofmedicine/divisions/infectiousdiseases/paediatrics/neonatalmedicine/ndau/reports_1/. Accessed October 2012.

(276) The Neonatal Data Analysis Unit. 2012. http://www1.imperial.ac.uk/departmentofmedicine/divisions/infectiousdiseases/paediatrics/neonatalmedicine/ndau/ Accessed October 2012.

(277) National Information Governance Board for Health and Social Care. Role of the NIGB http://www.nigb.nhs.uk/nigb. (2012) Accessed October 2012.

(278) National Information Governance Board for Health and Social Care. Section 251 of the NHS Act 2006 http://www.nigb.nhs.uk/s251. (2012) Accessed October 2012.

(279) Spencer A, Modi N. National neonatal data to support specialist care and improve infant outcomes. *Archives of Disease in Childhood - Fetal and Neonatal Edition* 2012.

(280) Care Quality Commission Periodic Review. Maternity data quality indicator specifications. 14-6-2010. http://archive.cqc.org.uk/_db/_documents/Maternity_data_quality_indicator_specifications.pdf. Accessed October 2012.

(281) The Healthcare Commission. Towards better births: A review of maternity services in England. http://www.cqc.org.uk/_db/_documents/Towards_better_births_200807221338.pdf. (2008) Accessed October 2012.

(282) The NHS Information Centre for Health and Social Care. Hospital Episode Statistics: Inpatient cleaning rules http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=376. (2009) Accessed October 2012.

(283) The NHS Information Centre for Health and Social Care. NHS Maternity Statistics, England: 2007-2008. http://www.ic.nhs.uk/statistics-and-data-collections/hospital-care/maternity/nhs-maternity-statistics-england:-2007-08 . (2009) Accessed October 2012.

(284) Agency for Healthcare Research and Quality (AHRQ). Clinical Classifications Software for ICD-10 Data. 2003. http://www.ahrq.gov/data/hcup/icd10usrgd.htm Accessed June 2011.

(285) National Statistics. News Release: 1 in 13 live births in England and Wales are born preterm. 24-5-2007. http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-50818 Accessed October 2012.

(286) Mason WM. Multilevel methods of statistical analysis. 2001. International Encyclopedia of the Social and Behaviorual Sciences (Edited by N J Smelser & P B Baltes), Department of Sociology and Statistics, University of California. Online working paper series. Accessed October 2012.

(287) Curcin V, Bottle A, Molokhia M, Millett C, Majeed A. Towards a scientific workflow methodology for primary care database studies. *Stat Methods Med Res* 2010; 19(4):378-393.

(288) Greenough A. Does low birth weight confer a lifelong respiratory disadvantage? *Am J Respir Crit Care Med* 2009; 180(2):107-108.

(289) Greenough A, Alexander J, Burgess S, Bytham J, Chetcuti PA, Hagan J et al. Health care utilisation of prematurely born, preschool children related to hospitalisation for RSV infection. *Arch Dis Child* 2004; 89(7):673-678.

(290) Gale C, Santhakumaran S, Nagarajan S, Statnikov Y, Modi N. Impact of managed clinical networks on NHS specialist neonatal services in England: population based study. *BMJ* 2012; 344:e2105.

(291) Marlow N, Bryan Gill A. Establishing neonatal networks: the reality. *Archives of Disease in Childhood - Fetal and Neonatal Edition* 2007; 92(2):F137-F142.

(292) Goldstein H, Harron K, Wade A. The analysis of record-linked data using multiple imputation with data value priors. *Statist Med* 2012;n/a.

(293) Clark DE. Practical introduction to record linkage for injury research. *Inj Prev* 2004; 10(3):186-191.

(294) Clark DE, Hahn DR. Comparison of probabilistic and deterministic record linkage in the development of a statewide trauma registry. *Proc Annu Symp Comput Appl Med Care* 1995;397-401.

(295) Potz N, Powell D, Pebody R, Lamagni T, Bridger D, Duckworth G. Development of a method to link infection and mortality data. Health Protection Agency Conference Poster Presentation [2006 Available from: URL:http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1200660056163 Accessed October 2012.

(296) Tromp M, Reitsma JB, Ravelli AC, Meray N, Bonsel GJ. Record linkage: making the most out of errors in linking variables. *AMIA Annu Symp Proc* 2006;779-783.

(297) Tromp M, Ravelli AC, Bonsel GJ, Hasman A, Reitsma JB. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *J Clin Epidemiol* 2011; 64(5):565-572.

(298) Harron K, Wade A, Muller-Pebody B, Goldstein H, Gilbert R. Opening the black box of record linkage. *Journal of Epidemiology and Community Health* 2012.

(299) Brook EL, Rosman DL, Holman CD. Public good through data linkage: measuring research outputs from the Western Australian Data Linkage System. *Aust N Z J Public Health* 2008; 32(1):19-23.

(300) Holman CD, Bass AJ, Rosman DL, Smith MB, Semmens JB, Glasson EJ et al. A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system. *Aust Health Rev* 2008; 32(4):766-777.

(301) Bohensky MA, Jolley D, Sundararajan V, Evans S, Ibrahim J, Brand C. Development and validation of reporting guidelines for studies involving data linkage. *Aust N Z J Public Health* 2011; 35(5):486-489.

(302) Bohensky MA, Jolley D, Sundararajan V, Evans S, Pilcher DV, Scott I et al. Data linkage: a powerful research tool with potential problems. *BMC Health Serv Res* 2010; 10:346.

(303) Lain SJ, Hadfield RM, Raynes-Greenow CH, Ford JB, Mealing NM, Algert CS et al. Quality of data in perinatal population health databases: a systematic review. *Med Care* 2012; 50(4):e7-20.

(304) Kelman CW, Bass AJ, Holman CD. Research use of linked health data--a best practice protocol. *Aust N Z J Public Health* 2002; 26(3):251-255.

(305) Office for National Statistics. Gestation-specific infant mortality in England and Wales (2010) http://www.ons.gov.uk/ons/dcp171778_282579.pdf Accessed October 2012.

(306) Wren C, Richmond S, Donaldson L. Temporal variability in birth prevalence of cardiovascular malformations. *Heart* 2000; 83(4):414-419.

(307) Manktelow BN, Draper ES, Annamalai S, Field D. Factors affecting the incidence of chronic lung disease of prematurity in 1987, 1992, and 1997. *Archives of Disease in Childhood - Fetal and Neonatal Edition* 2001; 85(1):F33-F35.

(308) Wren C, O'Sullivan JJ. Survival with congenital heart disease and need for follow up in adult life. *Heart* 2001; 85(4):438-443.

(309) Dodge JA, Morison S, Lewis PA, Coles EC, Geddes D, Russell G et al. Incidence, population, and survival of cystic fibrosis in the UK, 1968-95. UK Cystic Fibrosis Survey Management Committee. *Arch Dis Child* 1997; 77(6):493-496.

(310) Odding E, Roebroeck ME, Stam HJ. The epidemiology of cerebral palsy: incidence, impairments and risk factors. *Disabil Rehabil* 2006; 28(4):183-191.

(311) Morris JK, Alberman E. Trends in Down's syndrome live births and antenatal diagnoses in England and Wales from 1989 to 2008: analysis of data from the National Down Syndrome Cytogenetic Register. *BMJ* 2009; 339:b3794.

(312) Audit Commission. An Evaluation of Three Years of the PbR Data Assurance Framework http://www.audit-commission.gov.uk/sitecollectiondocuments/downloads/20110120AnEvaluationOfThreeYearsOfThePbRDataAssuranceFramework.pdf. (2011) Accessed October 2012.

(313) Office for National Statistics. Long-term international migration statistics: 1991 to date. (2011) http://www.ons.gov.uk/ons/rel/migration1/long-term-international-migration/2011/index.html Accessed October 2012.

(314) Zhou H, Thompson WW, Viboud CG, Ringholz CM, Cheng PY, Steiner C et al. Hospitalizations associated with influenza and respiratory syncytial virus in the United States, 1993-2008. *Clin Infect Dis* 2012; 54(10):1427-1436.

(315) Holman RC, Curns AT, Cheek JE, Bresee JS, Singleton RJ, Carver K et al. Respiratory syncytial virus hospitalizations among American Indian and Alaska Native infants and the general United States infant population. *Pediatrics* 2004; 114(4):e437-e444.

(316) Figueras-Aloy J, Carbonell-Estrany X, Quero J, IRIS Study Group. Case-Control Study of the Risk Factors Linked to Respiratory Syncytial Virus Infection Requiring Hospitalization in Premature Infants Born at a Gestational Age of 33-35 Weeks in Spain. *The Pediatric Infectious Disease Journal* 2004; 23(9).

(317) Bont L, Houben ML. Commentary: why are young healthy term infants protected against respiratory syncytial virus bronchiolitis? *Pediatr Infect Dis J* 2011; 30(9):785-786.

(318) Chan PW, Lok FY, Khatijah SB. Risk factors for hypoxemia and respiratory failure in respiratory syncytial virus bronchiolitis. *Southeast Asian J Trop Med Public Health* 2002; 33(4):806-810.

(319) Boyce TG, Mellen BG, Mitchel EF, Jr., Wright PF, Griffin MR. Rates of hospitalization for respiratory syncytial virus infection among children in medicaid. *J Pediatr* 2000; 137(6):865-870.

(320) Hervas D, Reina J, Yanez A, del Valle JM, Figuerola J, Hervas JA. Epidemiology of hospitalization for acute bronchiolitis in children: differences between RSV and non-RSV bronchiolitis. *Eur J Clin Microbiol Infect Dis* 2012; 31(8):1975-1981.

(321) Tsolia MN, Kafetzis D, Danelatou K, Astral H, Kallergi K, Spyridis P et al. Epidemiology of respiratory syncytial virus bronchiolitis in hospitalized infants in Greece. *Eur J Epidemiol* 2003; 18(1):55-61.

(322) Tregoning JS, Schwarze J. Respiratory viral infections in infants: causes, clinical symptoms, virology, and immunology. *Clin Microbiol Rev* 2010; 23(1):74-98.

(323) Stensballe LG, Hjuler T, Andersen A, Kaltoft M, Ravn H, Aaby P et al. Hospitalization for respiratory syncytial virus infection and invasive pneumococcal disease in Danish children aged <2 years: a population-based cohort study. *Clin Infect Dis* 2008; 46(8):1165-1171.

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

(325) Crowcroft NS, Zambon M, Harrison TG, Mok Q, Heath P, Miller E. Respiratory syncytial virus infection in infants admitted to paediatric intensive care units in London, and in their families. *Eur J Pediatr* 2008; 167(4):395-399.

(326) Bloemers BL, van Furth AM, Weijerman ME, Gemke RJ, Broers CJ, van den Ende K et al. Down syndrome: a novel risk factor for respiratory syncytial virus bronchiolitis--a prospective birth-cohort study. *Pediatrics* 2007; 120(4):e1076-e1081.

(327) Armstrong D, Grimwood K, Carlin JB, Carzino R, Hull J, Olinsky A et al. Severe viral respiratory infections in infants with cystic fibrosis. *Pediatr Pulmonol* 1998; 26(6):371-379.

(328) The Child and Maternal Health Observatory. NHS Atlas of Variation in Healthcare for Children and Young People. Reducing unwarranted variation to increase value and improve quality. http://www.rightcare.nhs.uk/index.php/atlas/children-and-young-adults/. 58-59. 2012. Accessed October 2012.

(329) Audit Commission, Audit. Improving coding, costing and commissioning - Annual report on the PbR data assurance programme 2010/11. 2011. http://archive.audit-commission.gov.uk/auditcommission/nationalstudies/health/pbr/pbr2011/Pages/pbr2011.aspx.html. Accessed October 2012.

(330) Audit Commission. Improving data quality in the NHS: 2010 - Annual report on the PbR assurance programme. 2010. http://archive.audit-commission.gov.uk/auditcommission/nationalstudies/health/pbr/pbr2010/pages/default.aspx.html. Accessed October 2012.

(331) Burns EM, Rigby E, Mamidanna R, Bottle A, Aylin P, Ziprin P et al. Systematic review of discharge coding accuracy. *J Public Health (Oxf)* 2012; 34(1):138-148.

(332) Wilkesmann A, Ammann RA, Schildgen O, Eis-Hubinger AM, Muller A, Seidenberg J et al. Hospitalized children with respiratory syncytial virus infection and neuromuscular impairment face an increased risk of a complicated course. *Pediatr Infect Dis J* 2007; 26(6):485-491.

(333) Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. *Br J Gen Pract* 2010; 60(572):e128-e136.

(334) Hansell A, Hollowell J, Nichols T, McNiece R, Strachan D. Use of the General Practice Research Database (GPRD) for respiratory epidemiology: a comparison with the 4th Morbidity Survey in General Practice (MSGP4). *Thorax* 1999; 54(5):413-419.

(335) Saxena S, Majeed A, Jones M. Socioeconomic differences in childhood consultation rates in general practice in England and Wales: prospective cohort study. *BMJ* 1999; 318(7184):642-646.

(336) Hall CB, Walsh EE, Schnabel KC, Long CE, McConnochie KM, Hildreth SW et al. Occurrence of groups A and B of respiratory syncytial virus over 15 years: associated epidemiologic and clinical characteristics in hospitalized and ambulatory children. *J Infect Dis* 1990; 162(6):1283-1290.

(337) Fryzek JP, Martone WJ, Groothuis JR. Trends in chronologic age and infant respiratory syncytial virus hospitalization: an 8-year cohort study. *Adv Ther* 2011; 28(3):195-201.

(338) Lenney W, Boner AL, Bont L, Bush A, Carlsen KH, Eber E et al. Medicines used in respiratory diseases only seen in children. *Eur Respir J* 2009; 34(3):531-551.

(339) Hartling L, Wiebe N, Russell K, Patel H, Klassen TP. Epinephrine for bronchiolitis. *Cochrane Database Syst Rev* 2004;(1):CD003123.

(340) Gadomski AM, Bhasale AL. Bronchodilators for bronchiolitis. *Cochrane Database Syst Rev* 2006;(3):CD001266.

(341) Davison C, Ventre KM, Luchetti M, Randolph AG. Efficacy of interventions for bronchiolitis in critically ill infants: a systematic review and meta-analysis. *Pediatr Crit Care Med* 2004; 5(5):482-489.

(342) Kneyber MC, Kimpen JL. Antibiotics in RSV bronchiolitis: still no evidence of effect. *Eur Respir J* 2007; 29(6):1285.

(343) Keith T, Saxena S, Murray J, Sharland M. Risk-benefit analysis of restricting antimicrobial prescribing in children: what do we really know? *Curr Opin Infect Dis* 2010; 23(3):242-248.

(344) Kimpen JL, Schaad UB. Treatment of respiratory syncytial virus bronchiolitis: 1995 poll of members of the European Society for Paediatric Infectious Diseases. *Pediatr Infect Dis J* 1997; 16(5):479-481.

(345) Wang EE, Law BJ, Boucher FD, Stephens D, Robinson JL, Dobson S et al. Pediatric Investigators Collaborative Network on Infections in Canada (PICNIC) study of admission and management variation in patients hospitalized with respiratory syncytial viral lower respiratory tract infection. *J Pediatr* 1996; 129(3):390-395.

(346) Swiss Association of Paediatric Pulmonology (SAPP/SGPP). Parent information leaflet on bronchiolitis. 2003. http://www.sgpp-schweiz.ch/go2/de/publikationen. Accessed October 2012.

(347) Moore HC, de KN, 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* 2011.

(348) Bremner SA, Carey IM, DeWilde S, Richards N, Maier WC, Hilton SR et al. Early-life exposure to antibacterials and the subsequent development of hayfever in childhood in the UK: case-control studies using the General Practice Research Database and the Doctors' Independent Network. *Clin Exp Allergy* 2003; 33(11):1518-1525.

(349) Joint Committee on Vaccination and Immunisation - Respiratory Syncitial Virus (RSV) Subgroup. Joint Committee on Vaccination and Immunisation Statement on immunisation for Respiratory Syncytial Virus http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@ab/documents/digitalasset/dh_120395.pdf. 11-10-2010. Accessed October 2012.

(350) Langley JM, LeBlanc JC, Smith B, Wang EE. Increasing incidence of hospitalization for bronchiolitis among Canadian children, 1980-2000. *J Infect Dis* 2003; 188(11):1764-1767.

(351) Palmer L, Hall CB, Katkin JP, Shi N, Masaquel AS, McLaurin KK et al. Respiratory outcomes, utilization and costs 12 months following a respiratory syncytial virus diagnosis among commercially insured late-preterm infants. *Curr Med Res Opin* 2011; 27(2):403-412.

(352) Palmer L, Hall CB, Katkin JP, Shi N, Masaquel AS, McLaurin KK et al. Healthcare costs within a year of respiratory syncytial virus among Medicaid infants. *Pediatr Pulmonol* 2010; 45(8):772-781.

(353) Sampalis JS. Morbidity and mortality after RSV-associated hospitalizations among premature Canadian infants. *J Pediatr* 2003; 143(5 Suppl):S150-S156.

(354) Romero JR, Stewart DL, Buysman EK, Fernandes AW, Jafri HS, Mahadevia PJ. Serious early childhood wheezing after respiratory syncytial virus lower respiratory tract illness in preterm infants. *Clin Ther* 2010; 32(14):2422-2432.

(355) Simoes EA, Groothuis JR, Carbonell-Estrany X, Rieger CH, Mitchell I, Fredrick LM et al. Palivizumab prophylaxis, respiratory syncytial virus, and subsequent recurrent wheezing. *J Pediatr* 2007; 151(1):34-42, 42.

(356) Moore T, Hennessy EM, Myles J, Johnson SJ, Draper ES, Costeloe KL et al. Neurological and developmental outcome in extremely preterm children born in England in 1995 and 2006: the EPICure studies. *BMJ* 2012; 345:e7961.

(357) Costeloe KL, Hennessy EM, Haider S, Stacey F, Marlow N, Draper ES. Short term outcomes after extreme preterm birth in England: comparison of two birth cohorts in 1995 and 2006 (the EPICure studies). *BMJ* 2012; 345:e7976.

(358) Bolton CE, Bush A, Hurst JR, Kotecha S, McGarvey L, Stocks J et al. Are early life factors considered when managing respiratory disease? A British Thoracic Society survey of current practice. *Thorax* 2012.

(359) Kusel MM, de Klerk NH, Kebadze T, Vohma V, Holt PG, Johnston SL et al. Early-life respiratory viral infections, atopic sensitization, and risk of subsequent development of persistent asthma. *J Allergy Clin Immunol* 2007; 119(5):1105-1110.

(360) Palmer L, Hall CB, Katkin JP, Shi N, Masaquel AS, McLaurin KK et al. Respiratory outcomes, utilization and costs 12 months following a respiratory syncytial virus diagnosis among commercially insured late-preterm infants. *Curr Med Res Opin* 2010; 27(2):403-412.

(361) Thomsen SF, van der Sluis S, Stensballe LG, Posthuma D, Skytthe A, Kyvik KO et al. Exploring the association between severe respiratory syncytial virus infection and asthma: a registry-based twin study. *Am J Respir Crit Care Med* 2009; 179(12):1091-1097.

(362) Stensballe LG, Simonsen JB, Thomsen SF, Larsen AM, Lysdal SH, Aaby P et al. The causal direction in the association between respiratory syncytial virus hospitalization and asthma. *J Allergy Clin Immunol* 2009; 123(1):131-137.

(363) Walton RP, Johnston SL. Role of respiratory viral infections in the development of atopic conditions. *Curr Opin Allergy Clin Immunol* 2008; 8(2):150-153.

(364) Akinbami LJ, Schoendorf KC. Trends in childhood asthma: prevalence, health care utilization, and mortality. *Pediatrics* 2002; 110(2 Pt 1):315-322.

(365) Lemanske RF, Jr., Jackson DJ, Gangnon RE, Evans MD, Li Z, Shult PA et al. Rhinovirus illnesses during infancy predict subsequent childhood wheezing. *J Allergy Clin Immunol* 2005; 116(3):571-577.

(366) Valkonen H, Waris M, Ruohola A, Ruuskanen O, Heikkinen T. Recurrent wheezing after respiratory syncytial virus or non-respiratory syncytial virus bronchiolitis in infancy: a 3-year follow-up. *Allergy* 2009; 64(9):1359-1365.

(367) Rothman K.J. GSLTL. Precision and statistics in epidemiologic studies. In: Rothman K. J., editor; Greenland S., editor; Lash T. L., editor. Modern Epidemiology. 3rd Edition ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008. 148-167.

(368) Fewell Z, Davey SG, Sterne JA. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am J Epidemiol* 2007; 166(6):646-655.

(369) From the American Academy of Pediatrics: Policy statements--Modified recommendations for use of palivizumab for prevention of respiratory syncytial virus infections. *Pediatrics* 2009; 124(6):1694-1701.

(370) Hiatt PW, Grace SC, Kozinetz CA, Raboudi SH, Treece DG, Taber LH et al. Effects of viral lower respiratory tract infection on lung function in infants with cystic fibrosis. *Pediatrics* 1999; 103(3):619-626.

(371) Joint Committee on Vaccination and Immunisation - Respiratory Syncitial Virus (RSV) Subgroup. Minutes of the RSV subgroup on Friday 27 March 2009 http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@ab/documents/digitalasset/dh_103572.pdf. 31-7-2009. Department of Health. Accessed October 2012.

(372) Sharland M, Bedford-Russell A. Preventing respiratory syncitial virus bronchiolitis. *BMJ* 2001; 322(7278):62-63.

(373) Thompson PL, Spyridis N, Sharland M, Gilbert RE, Saxena S, Long PF et al. Changes in clinical indications for community antibiotic prescribing for children in the UK from 1996 to 2006: will the new NICE prescribing guidance on upper respiratory tract infections just be ignored? *Arch Dis Child* 2009; 94(5):337-340.

(374) Little P, Gould C, Williamson I, Warner G, Gantley M, Kinmonth AL. Reattendance and complications in a randomised trial of prescribing strategies for sore throat: the medicalising effect of prescribing antibiotics. *BMJ* 1997; 315(7104):350-352.

(375) Francis NA, Butler CC, Hood K, Simpson S, Wood F, Nuttall J. Effect of using an interactive booklet about childhood respiratory tract infections in primary care consultations on reconsulting and antibiotic prescribing: a cluster randomised controlled trial. *BMJ* 2009; 339:b2885.

(376) Lind I, Gill JH, Calabretta N, Polizzoto M. Clinical inquiries. What are hospital admission criteria for infants with bronchiolitis? *J Fam Pract* 2006; 55(1):67-69.

(377) Dixon J, Bardsley M. Predictive risk modelling using routine data: underexploited potential to benefit patients. *J Health Serv Res Policy* 2012; 17(3):131-132.

(378) Paes B, Cole M, Latchman A, Pinelli J. Predictive value of the respiratory syncytial virus risk-scoring tool in the term infant in Canada. *Curr Med Res Opin* 2009; 25(9):2191-2196.

(379) Harris JA, Huskins WC, Langley JM, Siegel JD. Health care epidemiology perspective on the October 2006 recommendations of the Subcommittee on Diagnosis and Management of Bronchiolitis. *Pediatrics* 2007; 120(4):890-892.

(380) The Health Improvement Network (2012) http://csdmruk.cegedim.com/ Accessed October 2012.

(381) Kennedy I, Howard R, Jarman B, Maclean M. Learning from Bristol: the report of the public inquiry into children's heart surgery at the Bristol Royal Infirmary 1984 - 1995 Command Paper: CM 5207 www.bristol-inquiry.org.uk. (2001) Accessed October 2012.

(382) Aylin P, Bottle A, Jarman B, Elliott P. Paediatric cardiac surgical mortality in England after Bristol: descriptive analysis of hospital episode statistics 1991-2002. *BMJ* 2004; 329(7470):825.

(383) Department of Health. Liberating the NHS: An Information Revolution. http://www.ico.gov.uk/about_us/consultations/~/media/documents/consultation_responses/Liberating_the_NHS_An_Information_Revolution_Consultation.ashx. (2010) Accessed October 2012.

(384) Black N. Maximising research opportunities of new NHS information systems. *BMJ* 2008; 336(7636):106-107.

(385) NHS Screening Programmes. The NHS Newborn and Infant Physical Examination Programme (NIPE) http://newbornphysical.screening.nhs.uk/. (2012) Accessed October 2012.

(386) NHS Screening Programmes. The NHS Newborn Bloodspot Screening Programme http://newbornbloodspot.screening.nhs.uk/. (2012) Accessed October 2012.

(387) NHS Screening Programmes. The NHS Newborn Hearing Screening Programme http://hearing.screening.nhs.uk/index.php. (2012) Accessed October 2012.