

Pathways from socioeconomic deprivation to bronchiolitis and subsequent childhood asthma

Kate Marie Lewis Ba MSc

A thesis submitted for the degree of Doctor of Philosophy

University College London (UCL)

Great Ormond Street Institute of Child Health

March 2021

DECLARATION

I, Kate Marie Lewis, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Date:

Signature:

ABSTRACT

Introduction: Bronchiolitis and childhood asthma are major causes of morbidity among children in the UK, yet there are no preventative or curative measures for most children that develop these conditions. A better understanding of the longitudinal pathways to these conditions is warranted to design effective prevention policies. Using a social determinants of health framework, I explored the pathways between socioeconomic position, bronchiolitis and childhood asthma.

Methods: I used national birth cohorts created from linked administrative datasets in my thesis. I used harmonic Poisson regression models to examine associations between socioeconomic deprivation and the seasonality of bronchiolitis admissions in England. I modelled typical trajectories of asthma/wheeze among children in Scotland using latent class growth analysis. Using causal inference methods, I estimated: the socioeconomic disparities in the risk of bronchiolitis admissions that would remain if maternal smoking during pregnancy were eliminated; and the socioeconomic disparities in the risk of chronic trajectories of asthma that would remain if bronchiolitis admissions were eliminated.

Results: The peak timing of bronchiolitis admissions varied marginally across England, with earlier peaks in areas with higher population densities. After accounting for seasonal patterns, the North of England had disproportionately higher rates of admissions and, nationwide, disparities followed a socioeconomic gradient. I estimated that eliminating maternal smoking would reduce 20% of socioeconomic disparities in the risk of bronchiolitis admission. I identified four asthma/wheeze trajectories in children: no/infrequent, early-transient, early-persistent and intermediate-onset. Eliminating bronchiolitis admissions could reduce up to 18% of the disparities in the risk of chronic asthma by age ten.

Conclusions: Intervening early on the most socioeconomically deprived populations should be central to policies aiming to reduce the incidence of bronchiolitis admissions and asthma. The contribution of other socioeconomically patterned risk factors, including pollution and housing conditions, should be investigated in future work.

IMPACT STATEMENT

The work in this thesis addresses two key child public health issues in the UK. Bronchiolitis, the leading cause of hospital admission in infancy, and asthma, the most common chronic respiratory condition in children. Other epidemiological studies have addressed these conditions (and the potential link between the two), but none with the systematic methods presented here. I focus on socioeconomic deprivation, adding to the evidence base showing a socioeconomic gradient in the incidence of bronchiolitis and paediatric asthma in the UK. Further, my work begins to untangle some of the pathways through which inequities in these conditions develop.

Drawing from the conceptual framework developed from the World Health Organisation's Commission of the Social Determinants of Health, I create a structure to study the complex layers of risk factors that influence the development of bronchiolitis and subsequent asthma. My work includes two studies that look at the mechanisms of the seasonality of bronchiolitis admissions using epidemiological methods. The seasonal pattern of bronchiolitis has long been recognised by paediatricians, but there is no research to date that has studied and quantified this effect in a UK setting. I then use causal inference methods to quantify the contribution of specific risk factors in the development of bronchiolitis and asthma. These findings can be used to enhance public health strategies towards the alleviation of inequity in bronchiolitis and asthma.

I provide an overview of causal mediation estimands recently proposed in the literature, critically assessing the potential application of each to health inequity research using observational data. As part of this, I apply a novel method of evidence synthesis to design causal diagrams to inform analysis. I show the complexities involved in the process of finding unbiased estimates of mediation effects, but also provide some workable solutions to this. This work provides a framework through which other researchers can apply these methods to health inequity research in the future. I have made the code for the statistical analyses in my published studies freely and publicly available on GitHub. This is important for transparency and reproducibility of research, and I will continue to do this for further papers published from my thesis. I have published three open-access journal articles directly based on work presented in this thesis and have presented my findings at several conferences.

I list specific outputs produced over the course of my PhD in Appendix 0, p.195.

TABLE OF CONTENTS

Declaration	2
Abstract	3
Impact Statement	4
Table of contents	5
List of tables.....	11
List of boxes.....	12
List of figures	13
Abbreviations	15
Acknowledgements	17
Chapter 1. Background.....	18
1.1. Introduction.....	18
1.1.1. Bronchiolitis.....	18
1.1.2. Asthma.....	19
1.1.3. Why focus on these two conditions?	20
1.2. Health inequities and socioeconomic position	22
1.2.1. Key definitions.....	22
1.2.2. Measuring SEP	22
1.3. Conceptual framework	24
1.3.1. Literature search strategy.....	25
1.4. Structural determinants	27
1.4.1. SEP.....	27
1.4.2. Ethnicity and migration status (not included in Figure 1.1).....	30
1.4.3. Geographical location	32
1.4.4. Built environment	33
1.4.5. Neighbourhood	33
1.5. Intermediary determinants	34
1.5.1. Maternal smoking during pregnancy.....	34
1.5.2. Maternal asthma	36
1.5.3. Outdoor air pollution.....	36
1.5.4. Maternal age at delivery	37
1.5.5. Congenital heart disease, chronic lung disease and other early-life conditions	38
1.5.6. Birth weight and gestational age	39
1.5.7. Delivery method	40
1.5.8. Number of older siblings.....	41
1.5.9. Breastfeeding.....	42
1.5.10. Indoor living conditions.....	43

1.6. The Health system	44
1.7. Biological factors.....	46
1.7.1. Age	46
1.7.2. Season of birth.....	47
1.7.3. Sex	47
1.7.4. Genetic risk of bronchiolitis and asthma	47
1.8. Public health policy and practice	48
1.8.1. RSV infection and bronchiolitis.....	49
1.8.2. Asthma.....	51
1.8.3. Looking forward	52
1.9. Chapter summary	53
Chapter 2. Rationale, aims and thesis structure.....	54
2.1. Literature summary	54
2.1.1. Key findings from Chapter 1.....	54
2.1.2. Priorities driving my work	55
2.2. Thesis rationale	56
2.3. Thesis aims and objectives	56
2.4. Thesis structure	57
Chapter 3. Introduction to thesis datasets.....	59
3.1. Administrative data	59
3.2. English Hospital Episode Statistics linked to mortality records	60
3.2.1. Overview	60
3.2.2. Using HES APC in my PhD.....	62
3.3. Scottish administrative health databases	64
3.3.1. Overview	64
3.3.2. Using Scottish Administrative Health Databases in my thesis	67
3.4. National data opt-out programme: consequences for maternity statistics in England .	69
3.4.1. Background.....	69
3.4.2. Application to NHS maternity statistics	71
3.4.3. Are opt outs inevitable?.....	74
3.5. Chapter summary	75
Chapter 4. Geospatial and seasonal variation of bronchiolitis in England.....	76
4.1. Background.....	76
4.2. Methods.....	77

4.2.1. Data sources and study cohort.....	77
4.2.2. Study outcome	78
4.2.3. Covariates.....	79
4.2.4. Statistical methods	80
4.3. Results.....	83
4.3.1. Descriptive statistics.....	83
4.3.2. Objective 4A: Seasonality of bronchiolitis admissions	87
4.3.3. Objective 4B: Multivariable regression models	90
4.3.4. Sensitivity analysis	92
4.4. Discussion	92
4.4.1. Strengths and weaknesses	92
4.4.2. Research findings: comparison with other studies.....	93
4.4.3. Implications of study findings	94
4.5. Chapter summary	95
Chapter 5. Is socioeconomic position associated with bronchiolitis seasonality?	96
5.1. Background.....	96
5.2. Methods.....	97
5.2.1. Data sources and study cohort.....	97
5.2.2. Study outcome	97
5.2.3. Covariates.....	98
5.2.4. Statistical methods	99
5.3. Results.....	102
5.3.1. Objective 5A: the proportion of admissions attributable to bronchiolitis	102
5.3.2. Cohort description	103
5.3.3. Objective 5B: effect modification by socioeconomic position	106
5.3.4. Sensitivity analysis: excluding the health domain from IMD.....	107
5.4. Discussion	108
5.4.1. Strengths and weaknesses	108
5.4.2. Research findings: comparison with other studies.....	109
5.4.3. Implications of study findings	110
5.5. Chapter summary	111
Chapter 6. Socioeconomic position and bronchiolitis—mediation through maternal smoking.....	112
6.1. Background.....	112
6.2. Causal inference framework	114
6.2.1. Basic principles	114
6.2.2. Key assumptions.....	115

6.3.	Causal diagrams.....	117
6.4.	Causal mediation analysis	123
6.4.1.	Controlled, natural and interventional (in)direct effects.....	123
6.4.2.	Multiple mediators	127
6.5.	Study cohort and variable selection	128
6.5.1.	Study DAG	128
6.5.2.	Data sources and study cohort.....	129
6.5.3.	Study variables	130
6.6.	Statistical methods.....	134
6.6.1.	Admission rates	134
6.6.2.	Disparities in bronchiolitis admissions	134
6.7.	Results.....	139
6.7.1.	Missing data.....	139
6.7.2.	Cohort description (complete cases)	141
6.7.3.	Admission rates (all infants)	141
6.7.4.	Mediation analyses	143
6.8.	Discussion	144
6.8.1.	Strengths and weaknesses	144
6.8.2.	Research findings and implications (Objective 6D).....	148
6.9.	Chapter summary	148
Chapter 7. Trajectories of asthma/wheeze symptoms and their association with early-life risk factors		150
7.1.	Background.....	150
7.2.	Study cohort and variable selection	152
7.2.1.	Study DAG for mediation analyses.....	152
7.2.2.	Data sources and study cohort.....	155
7.2.3.	Study variables	156
7.3.	Statistical methods.....	160
7.3.1.	Overview of the analytical steps to address the 3 objectives	160
7.3.2.	Typical trajectories of asthma/wheeze (objective 7A)	161
7.3.3.	Risk factors by asthma/wheeze trajectory group (objective 7B).....	162
7.3.4.	Mediation analysis (objective 7C).....	164
7.4.	Results.....	167
7.4.1.	Missing data.....	167
7.4.2.	Cohort characteristics	169
7.4.3.	Growth modelling	170
7.4.1.	Association between asthma/wheeze trajectories and risk factors	173

7.4.2. Mediation analysis.....	176
7.5. Discussion	176
7.5.1. Strengths and weaknesses	177
7.5.2. Comparison to other studies	179
7.6. Chapter summary	181
Chapter 8. Discussion	183
8.1. Critical assessment of my PhD	183
8.1.1. Restatement of thesis rationale and aim	183
8.1.2. Key findings	185
8.1.3. Strengths and limitations	187
8.2. Implications of my research findings	189
8.2.1. Continuing high and unequal burden of disease.....	189
8.2.2. Clinical practice and policy	189
8.2.3. Future research.....	192
8.3. The impact of the SARS-CoV-2 pandemic	193
8.4. Concluding remarks	194
Appendix 0: research outputs.....	195
Appendix 0.1. Publications during PhD period	195
Appendix 0.2. Conferences and presentations.....	196
Appendix 0.3. Secondment	196
Appendix - Chapter 1	197
Appendix 1.1. EJE publication: Mother’s education and offspring asthma risk in 10 European cohort studies	197
Appendix - Chapter 3	206
Appendix 3.1. Linking Mother and Infant Data in HES APC	206
Appendix 3.2. Data linkage feedback form	209
Appendix 3.3. Feedback request sent to the Farr Institute Public Panel.....	212
Appendix 3.4. IJPDS publication: National data opt out programme	213
Appendix - Chapter 4	219
Appendix 4.1. Thorax publication: Geospatial and seasonal variation of bronchiolitis in England	219
Appendix 4.2. Stata code for Chapter 4 harmonic Poisson regression analyses	227
Appendix 4.3. Harmonic Poisson regression models	231
Appendix 4.4: Sensitivity analyses	233
Appendix - Chapter 5	235
Appendix 5.1. JECH publication: Is socioeconomic position associated with bronchiolitis seasonality?	235

Appendix 5.2. Stata code for Chapter 5 harmonic regression analyses.....	244
Appendix 5.3. Harmonic Poisson regression models	247
Appendix - Chapter 6	249
Appendix 6.1. ESC-DAG output.....	249
Appendix 6.2. Lancet abstract: Could socioeconomic inequity in bronchiolitis admissions be reduced through intervening on gestational age?	257
Appendix 6.3. Stata code for causal mediation analyses.....	258
Appendix - chapter 7.....	264
Appendix 7.1. Defining chronic conditions.....	264
Appendix 7.2. Example Mplus code	266
Appendix 7.3. Stata code for mediation analysis.....	270
Appendix 7.4. Risk factor distribution by asthma/wheeze trajectories.....	277
Bibliography	279

LIST OF TABLES

Table 1.1. Literature search terms for two overarching research questions	26
Table 1.2. Rates of hospital admissions for bronchiolitis (birth to 1 year) and asthma (1 year to 5 years), by mothers smoking status at first antenatal booking (from Lawder at al.) ¹³⁸	35
Table 3.1. Description of datasets comprising the English birth cohort used in my thesis ^{271,286}	62
Table 3.2. Number of singleton birth records, proportion unlinked to delivery records in original and updated linkage and available delivery records, by financial year.....	64
Table 4.1. Cohort characteristics and the number and rate (per 1000 infant-years) of bronchiolitis hospital admissions before 1 year of age among cohort members	85
Table 4.2. Derived average annual seasonal estimates following harmonic Poisson regression (Table A4.1, Appendix 4.3), by region	87
Table 4.3. Multivariable linear regression analyses assessing the effect of population density and IMD score on IRR compared to the national average and peak timing of CCG epidemic	90
Table 5.1. IMD 2015 domain weightings ⁶⁵	99
Table 5.2. Postnatal hospital admissions in children <1 year old, any condition and bronchiolitis-related, by year, month and IMD quintile: England, 2012-2016	102
Table 5.3. Distribution of births and bronchiolitis admissions (per 1000 infant-years) in the cohort.....	105
Table 5.4. Average annual seasonal estimates derived from harmonic Poisson regression models (see Appendix 4.3), by IMD quintile	107
Table 6.1. Search strategy for literature used in ESC-DAG.....	118
Table 6.2. Final directed edge index	120
Table 6.3. Scottish datasets used to create the study cohort for this study	130
Table 6.4. Cohort characteristics, missing data and admission rates	140

Table 6.5. Final cohort characteristics, by socioeconomic deprivation group, and ≥ 1 hospital admission for bronchiolitis in infancy	142
Table 6.6. Potential risk of ≥ 1 admission to hospital for bronchiolitis during infancy, estimated counterfactual disparity in admissions and corresponding disparity reduction explained by maternal smoking, by g-method and SIMD group	144
Table 7.1. Scottish datasets used to create the study cohort in this study.....	155
Table 7.2. Drugs that are used to treat asthma, defined using the BNFC ⁴³²	157
Table 7.3. Characteristics of final study cohort (pre-imputation), number of children with missing data and estimated adjusted odds ratios (aORs) for missingness	168
Table 7.4 Observed completeness and frequency of asthma/wheeze symptoms, by age of child	169
Table 7.5. Model fit indices for LCGA and GMM on asthma/wheeze symptoms (selected model highlighted).....	171
Table 7.6. Univariable RRRs (95% CI) of belonging to asthma/wheeze groups compared to the never or infrequent asthma/wheeze group, by risk factor	174
Table 7.7. Potential risk of chronic asthma* by age 9, estimated counterfactual disparity in risk and corresponding disparity reduction involving bronchiolitis admissions, by SIMD group*176	

LIST OF BOXES

Box 4.1. Models' specifications and derived parameters ³⁴⁰	82
Box 5.1. Model specifications	101
Box 6.1. Could socioeconomic inequity in bronchiolitis admissions be reduced through intervening on gestational age?*	126
Box 6.2. Steps for estimating CDM($m=0$) using IPW estimation of MSMs with a categorical exposure and binary mediator	137
Box 6.3. Steps for estimating CDM($m=0$) using g-estimation of structural nested models with a categorical exposure and binary mediator.....	138

LIST OF FIGURES

Figure 1.1. Framework for pathways to socioeconomic inequities in the risk of bronchiolitis and paediatric asthma.....	25
Figure 1.2. Prevalence of wheezing in the last 12 months (dashed lines) and doctor diagnosis of asthma ever (solid line) among children <16 years, by measure of SEP: Scotland, 2019	28
Figure 3.1. Scottish health datasets and relevant variables used to create birth cohort.....	68
Figure 3.2. Rates of patients opting out, by CCG: England (with inlay map of London CCGs), as at December 2018. Data from NHS Digital. ⁴⁰	70
Figure 3.3. Observed and simulated rates of births with very low birth weight (<1500g, 2016/17) and deliveries by caesarean section (2016): by CCG.....	73
Figure 4.1. Flow diagram showing derivation of final study cohort and bronchiolitis hospital admissions (linked via unique HESID)	84
Figure 4.2. Observed weekly rates (per 1000 infant-years) of hospital admissions for bronchiolitis in England, from January 2012 to December 2016.....	85
Figure 4.3. Observed weekly rates of hospital admissions for bronchiolitis, by Region: England, 2012-2016 averaged.....	86
Figure 4.4. Predicted weekly rates of hospital admissions for bronchiolitis, by Region: England, 2012-2016 averaged.....	88
Figure 4.5. Predicted IRRs of bronchiolitis hospital admissions compared with the national average, by CCG: England (with London inset).....	89
Figure 4.6. Predicted peak timing (weeks) of bronchiolitis admission rates, by CCG: England (with London inset).....	89
Figure 4.7. Observed and predicted variation in CCG-level rates of bronchiolitis hospital admissions compared with the national average, by IMD score and log population density (including London-based CCGs)	91
Figure 4.8. Observed and predicted peak timing of CCG-level bronchiolitis hospital admissions, by IMD score and log population density (including London-based CCGs).....	91

Figure 5.1. Flow diagram to show study participant and hospital admission selection	104
Figure 5.2. Observed weekly rates of hospital admission for bronchiolitis, by IMD group: England 2012-2016.....	106
Figure 5.3. Predicted weekly rates of hospital admission for bronchiolitis, by IMD: England 2012-2016.....	108
Figure 6.1. Overall integrated-DAG created through the ESC-DAG protocol*	122
Figure 6.2. Assumed causal DAG*	129
Figure 6.3. Map of Scotland, by postal area; key showing 5 grouping used in this study ..	133
Figure 6.4. Flow diagram showing derivation of final study cohorts for mediation analysis and calculations of admission rates	139
Figure 7.1. Assumed causal DAG for the mediator-outcome (bronchiolitis admission-chronic asthma) association	154
Figure 7.2. Analysis steps in this Chapter, split by software	160
Figure 7.3. Flow diagram showing derivation of final study cohort.....	167
Figure 7.4. Average trajectory group probabilities*	172
Figure 7.5. Probability and 95% CI of symptoms at each time point for latent trajectory groups*	172
Figure 7.6. Forest plots showing univariable RRR estimates with 95% CI for membership of asthma/wheeze groups compared to the never or infrequent asthma/wheeze group, by risk factor.....	175
Figure 8.1. Framework for pathways to socioeconomic inequities in the risk of bronchiolitis and paediatric asthma.....	184

ABBREVIATIONS

A&E	Accident and emergency
aOR	Adjusted odds ratio
BIC	Bayesian information criterion
BLRT	Bootstrapped likelihood ratio test
BNFC	British National Formulary for Children
CCG	Clinical Commissioning Group
CDE	Controlled direct effect
CDM	Counterfactual disparity measure
CHI	Community Health Index
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
CSDH	Commission on Social Determinants of Health
DAG	Directed acyclic graph
eDRIS	Electronic Data Research and Innovation Service
ESC-DAG	Evidence synthesis for constructing directed acyclic graphs
GP	General practitioner
GMM	Growth mixture modelling
GWAS	Genome-wide association study
HES APC	Hospital episode statistics admitted patient care
HR	Hazard ratio
IMD	Index of multiple deprivation
IRR	Incidence rate ratio
ISAAC	The International Study of Asthma and Allergies in Childhood
IPW	Inverse probability weighting
LCGA	Latent class growth analysis
LSOA	Lower super output area
MHCLG	Ministry of Housing, Communities and Local Government
MICE	Multiple imputation by chained equations
MSM	Marginal structural model
NDE	Natural direct effect
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIE	Natural indirect effect
NO₂	Nitrogen dioxide

NRS	National Records of Scotland
NS-SEC	National Statistics Socio-Economic Classification
ONS	Office for National Statistics
OR	Odds ratio
PHE	Public Health England
PHS	Public Health Scotland
PICU	Paediatric intensive care unit
PIS	Prescription Information System
PM_{2.5/10}	Particulate matter (less than 2.5/10 micrometres in diameter)
RIA	Randomised interventional analogues
RRR	Relative risk ratio
RSV	Respiratory syncytial virus
SBR	Scottish Birth Records
SGA	Small for gestational age
SEP	Socioeconomic position
SIGN	Scottish Intercollegiate Guidelines Network
SIMD	Scottish index of multiple deprivation
SMR	Scottish Morbidity Record
TCE	Total causal effect
UK	United Kingdom
WHO	World Health Organisation

ACKNOWLEDGEMENTS

I would like to thank my brilliant supervisors, Dr Pia Hardelid and Professor Bianca De Stavola, from whom I have learnt and continue to learn so much. Ever supportive, knowledgeable and encouraging, they are the ultimate role models for any early career researcher. I feel very lucky to have been supervised by them and for being part of the wider Child Health Informatics Group whilst doing my PhD.

A big thank you to my fellow PhD students and (pre-pandemic) office colleagues who have provided much needed laughter, advice and encouragement during my PhD. To my friends, particularly Eloisa and Jess, and my family, for being there for me through the ups and downs of the last few years. Fellow DTP student, Tyrone, for the walks that kept my spirits up during the pandemic. And, not forgetting, my closest companion, Ziggo.

I cannot thank my partner, Tom, enough for the unwavering support he has given me, particularly in the last few months of writing up my thesis.

My project would not have been possible without the funding from a Medical Research Council UK DTP studentship, nor the data provided by patients and collected by the National Health Service as part of their care and support. I am also very grateful to Professor Steve Cunningham for his expert guidance on bronchiolitis and asthma and colleagues, Dr Ania Zylbersztejn and Maximiliane Verfürden, for their hard work constructing the English and Scottish birth cohorts, respectively.

Chapter 1. Background

Chapter overview

The overarching aim of Chapter 1 is to present an overview of the current literature and public health policies relevant to this thesis, focusing on the relationships between socioeconomic deprivation, bronchiolitis and the subsequent development of asthma. Firstly, I will present an overview of the epidemiology of bronchiolitis and paediatric asthma in the United Kingdom, and introduce how the two conditions may be linked. Where possible I will focus specifically on England and Scotland, the nations from which my study data originates. Secondly, I will define health inequity and socioeconomic position—key terms in my work—and explain the conceptual framework that underpins this research. Thirdly, I will present an overview of pre- and post-natal risk factors for bronchiolitis and asthma. Lastly, I will outline current policy and practice relevant to the prevention and treatment of bronchiolitis and childhood asthma in England and Scotland, with a particular focus on efforts to lessen socioeconomic inequities in these conditions.

1.1. Introduction

1.1.1. Bronchiolitis

Bronchiolitis is an acute lower respiratory tract illness caused by viral infections that affects young children, particularly infants.¹ Diagnosis of bronchiolitis is based on respiratory signs and symptoms rather than a diagnostic test or universal scoring system.^{2,3} This infection typically begins with symptoms similar to the common cold, but progresses to a persistent cough, fast and shallow breathing, chest recession and crackles and/or wheezing within a few days.^{1,3} Mild cases of bronchiolitis can be managed at home, but a significant number of infants require supportive hospital care to aid feeding and respiration.^{4,5} In infants admitted to hospital in England with bronchiolitis, approximately 80% of cases are caused by respiratory syncytial virus (RSV),⁶ however, other respiratory viruses including human metapneumovirus (HMPV), rhinovirus, adenovirus, influenza and parainfluenza are also known to cause bronchiolitis.¹ These viruses are mainly transmitted through direct and indirect contact (via contaminated intermediate objects) with an infected person and by airborne droplets, which are spread through coughing, sneezing and talking.⁷ There is early evidence to suggest that

RSV may also be spread by aerosolised (airborne) particles, which are smaller than droplets and have a longer range of transmission.⁸

Whilst mortality due to this condition is uncommon in the United Kingdom (UK) (in 2019 there were 7 deaths in England and 2 deaths in Scotland among young children with bronchiolitis recorded as the underlying cause),^{9,10} its associated morbidity is substantial. In 2018/19, there were 48,744 hospital admissions in England with a primary diagnosis of bronchiolitis among infants (children <1 year old), making it the most common reason for hospital admission in the first year of life.¹¹ Qualitative research from Canada and France highlights the emotional strain, disruption and anxiety bronchiolitis places on the families of infants hospitalised with this infection.¹²⁻¹⁴ The rate of hospital admissions for this condition have been rising substantially over time,⁴ leading to warnings about needing to “stem the rising tide” of these admissions.¹⁵ RSV-attributable bronchiolitis in children less than 2 years old is also estimated to contribute an more than 82,000 general practitioner visits annually in the UK.¹⁶ Reflecting the seasonal pattern of RSV, cases of bronchiolitis peak in early winter in the UK, contributing to pressures on the National Health Service (NHS) during a particularly stretched time of year.^{1,4}

1.1.2. Asthma

Asthma, a respiratory disease involving airway inflammation and hyper-responsiveness, is one of the most common chronic conditions among children in the UK.¹⁷ In 2019, the estimated prevalence of asthma (defined by self-/parent-reported doctor’s diagnosis of asthma and wheeze in the last year) among under 15’s in the UK was 9.6% (95% confidence interval (CI) 6.5 to 13.7).^{18,19} This compares to a Europe-wide average of 5.7% (95% CI 3.9 to 8.1) and a global average of 4.7% (95% CI 3.3 to 6.7). Symptoms of asthma can vary over time and with intensity, but commonly include wheezing, chest tightness, breathlessness and coughing.^{20,21} Exacerbations of symptoms, commonly known as an asthma attack, can lead to variable airflow obstruction and inflammation within the lungs, which may be fatal if left untreated.¹⁷ It is increasingly recognised that asthma is not a singular disease but an umbrella term for a heterogeneous group of conditions.^{22,23} Relatedly, there is no single gold standard definition or diagnostic test for asthma.¹⁷

Different trajectories of childhood asthma, defined by the evolution over time in reported symptoms, have been distinguished based on longitudinal data.²⁴⁻²⁶ This includes early-transient asthma (beginning before age 3 and subsiding after a few years), early-persistent asthma (beginning before age 3 and continuing into later childhood) and intermediate/late-onset asthma (beginning between 3 and 8 years). In an estimated 40-60% of cases, allergic sensitisation is also a characteristic of the disease; a category of asthma sometimes defined

as atopic asthma.²⁷⁻²⁹ Early asthma symptoms may be indicative of viral or transient wheeze caused by respiratory infections and, therefore, diagnosis of asthma is not recommended in children under the age of 5.^{30,31} Where symptoms are still present at age 5, the National Institute for Health and Care Excellence (NICE) recommends that clinicians use objective tests (e.g. spirometry) to then diagnose asthma in these children.¹⁷

The direct costs for asthma across all NHS services has been estimated to be £1.1 billion, including 85,000 hospital admissions and 108,000 general practitioner (GP) consultations across all age groups per year in the UK.³² In 2018/19, there were 25,342 recorded hospital admissions with a primary diagnosis of asthma among under-18s in England.¹¹ Asthma-related deaths in children are rare, with 15 deaths in England and 1 death in Scotland among 0-19 year olds with asthma recorded as the underlying cause in 2019.^{9,10} However, these deaths are almost entirely, if not completely, avoidable. Of the 28 deaths in people less than 20 years old assessed in the 2014 National Review of Asthma Deaths, 26 (or 93%) were ascertained to be “potentially avoidable”.³³ Further, data published by the Global Health Data Exchange shows that, in 2017, the UK had the second highest rate of death from asthma amongst 15-19 year olds in a comparison with 18 other high income countries.³⁴ Childhood asthma is associated with other adverse outcomes, including an increased risk of anxiety and depression, more absentee days from school and poorer educational outcomes than peers, and diminished self-rated quality of life.^{32,35-38}

1.1.3. Why focus on these two conditions?

Despite being the cause of substantial morbidity and cost to the NHS, there are currently no preventative or curative measures for the majority of children that develop either bronchiolitis or asthma.^{30,39}

Under UK guidelines, infants at high risk of RSV-related hospital admissions, including preterm infants with chronic heart or lung conditions that are less than six months of age at the beginning of the RSV season, are recommended to receive palivizumab, a humanised monoclonal antibody, prophylactically during peak months of RSV circulation.^{40,41} Costing between £3000 and £5000 to treat each child with a full course, it is not deemed cost-effective to administer this preventative treatment to the wider infant population (who, on the majority, will not develop severe symptoms from RSV infection). There is also no vaccination currently available against RSV, the dominant pathogen causing bronchiolitis, although several candidates are currently being assessed in clinical trials.^{42,43} Given that approximately 80% of children admitted to hospital in England with bronchiolitis are reported as otherwise healthy,

this means that there is no preventative healthcare interventions for the majority of infants at risk of bronchiolitis.^{44,45}

In the last decade, previous gains in improving key outcomes for asthma globally, including reducing mortality and hospital admission rates, have reportedly stalled.^{46,47} Asthma care focuses on managing symptoms through the use of medicines that prevent and relieve symptoms, such as inhaled corticosteroids.²⁰ Poor management of symptoms, particularly around the use of preventative medicine, is referenced at the outset of the Scottish Intercollegiate Guidelines Network (SIGN) guidelines on the management of asthma as the driver of “much of this morbidity”.⁴⁸ This document highlights the failure of primary prevention methods, including allergen avoidance, modified milk formulae, weaning and nutritional supplementation, to offer clear beneficial outcomes on asthma outcomes. Breastfeeding, childhood immunisations and reduced smoking exposure are recommended for all children (and weight loss interventions for overweight and obese children), but as factors that “promote general health” rather than specifically in relation to preventing asthma. Trials of pharmacological prevention using antihistamines have been undertaken but here is little evidence to show that these stop the long-term development of asthma so far.²⁰

A better understanding of the risk factors for these conditions is therefore imperative to plan how to best intervene early and reduce the incidence and population impact of these conditions in children. Prominently, severe symptoms of RSV infection have been consistently associated with an increased risk of wheeze and asthma in later childhood across the research literature.^{49–53} For example, infants born in English NHS hospital in 2007/08 were 4.4 times more likely to have a hospital admission for asthma between ages 1 and 5 if they had one or more admissions for bronchiolitis in infancy, compared to those with no admission for bronchiolitis (adjusted hazard ratio (HR) 4.35, 95% CI 4.00 to 4.73).⁵⁴ There is a debate to whether this association represents a true causal mechanism (i.e. early infection impairs pulmonary function, thereby directly influencing the development of asthma) or it is due to other shared influences, such as environmental exposures or a familial disposition to respiratory ill health.^{49,51,55,56} Notable recent evidence from the MAKI randomised controlled trial in the Netherlands has shown the effectiveness of palivizumab prophylaxis, given to otherwise healthy preterm infants, in reducing wheeze up to age 3 but not at age 6.^{57,58} Or, to put this another way, RSV infection as an important mechanism for early, but not later, wheeze.^{57,58} These results further open up the question of alternative—and potentially modifiable—factors that may jointly influence the development of both conditions.⁵⁵

My PhD will focus specifically on the common socioeconomic determinants of bronchiolitis and asthma to elucidate the pathway to and between the two conditions, and the potential for interventions, further.

1.2. Health inequities and socioeconomic position

1.2.1. Key definitions

Health inequity and socioeconomic position (SEP) are concepts that I will refer to often throughout this thesis. By inequities in health, I mean the disparities in health outcomes (or in the opportunity to be healthy) across strata of society that are created by systematic social, political and economic conditions.^{59,60} I use the term health *inequity* as opposed to *inequality* to emphasise the avoidable and unjust nature of these differences in circumstance and health outcomes.⁶¹ Within my PhD different social strata will be characterised by SEP, defined by Galobardes and colleagues as “the social and economic factors that influence what positions individuals or groups hold within the structure of a society”.⁶² Socioeconomic deprivation will be used to refer to the proportion of the population at the most deprived end of SEP.

1.2.2. Measuring SEP

Although commonly employed in health research, there is no agreed upon definition of such a multidimensional concept such as SEP, which may encompass education, occupation, income, ownership of goods, social prestige and other factors. Therefore, there is no standard method of measuring SEP across health inequity or epidemiological studies.^{46,62} As argued by Bartley, the definition of SEP used should be guided by research hypotheses. For purposefully designed studies on health inequity, careful consideration of how to measure SEP and related concepts may be achieved. Unfortunately, as in many cases, the measurement of SEP in my PhD, is driven by the available data. However, by discussing the available indicator at the outset of my thesis, I hope to be transparent about the conclusions about SEP that I can make in my research based on my employed measure.

My measure of SEP: the Index of Multiple Deprivation

The Index of Multiple Deprivation (IMD) is a relative measure of deprivation across small areas, constructed by assigning each area a value calculated from weighted scores across seven domains of deprivation: income, employment, health, education, access to services, living environment and crime.⁶³ England and Scotland have their own versions of IMD, which differ in respect to the coverage of each small area (approximately 1500 residents in England

compared to 760 in Scotland) and the precise components of the seven domains.^{63,64} IMD is an official statistic, constructed by the Governments of England and Scotland for a variety of purposes, including: to shape policy and strategy; identify priority areas to target resources; and monitor inequalities across small areas.⁶⁵ The indices are updated every 4 to 5 years to reflect changes in the underlying data, with the same methodology and domains used.⁶⁵ There are two other notable area-level measures of material deprivation that cover the whole of the UK; the Carstairs Index and Townsend Deprivation Index.⁶⁶ These indices are both derived from components of census data including car ownership, unemployment, overcrowding, low social class (Carstairs Index only) and home ownership (Townsend Index only).

I will now outline the strengths and limitations of IMD with reference to the two goals of measuring inequity in my work.⁶⁷

Goal one: to monitor and describe the social distribution of disease

Monitoring the social determinants linked to health outcomes is an essential step to changing the course of health inequities. IMD is an apt measure for monitoring as it provides a consistent and nationally applicable measure of inequity. IMD has been used extensively in health inequity research, including in studies of bronchiolitis and asthma,^{4,68,69} which means that my results can be compared with previous and future work. IMD captures a wide range of factors that contribute to inequity, which is arguably more suited to the similarly comprehensive concept of SEP than a non-composite measure. In addition, IMD is an area-level measure, which means it can characterise some living conditions that individual or household measures cannot, such as ambient air pollution.⁷⁰ On the other hand, use of this ecological measure as a proxy for an individual level indicator can, in some circumstances, overestimate the true individual-level effect of SEP.⁶⁷ Notably, as IMD was constructed to measure specific aspects of deprivation rather than affluence, the indices are less able to discriminate between areas in the less deprived end of the distribution.⁶³ A criticism levelled at the IMD's identification of the most deprived areas is its heavy reliance on benefit-recipient data, which means it undercounts "entitled nonrecipients".⁷¹

Goal two: to identify interventions to weaken the relationship between socioeconomic deprivation and disease

There are two feasible ways to go about decreasing socioeconomic inequities in health outcomes. The first is by directly intervening on SEP and decreasing overall socioeconomic inequity across society. This is an ambitious but not impossible goal, with Heyman writing that this "major social change requires sustained public support and political will".^{72 p.369} A universal

basic income and formal (high quality) early years education and childcare are widely promoted ways to reduce inequities.^{73,74} The 2010 Marmot Review advocates the universal application of actions such as these, but at a scale and intensity that varies proportionately across the social gradient, a method coined as proportionate universalism.⁷⁵ Other measures of SEP, such as income or education, are arguably more suited to measuring this type of policy change due to their specificity. However, a composite measure of SEP may be a better indicator of the broader societal change described by Heyman.⁷² IMD in particular, though, may not be the best measure of changes within areas because it is a relative measure of deprivation and is not suited to observing absolute change in deprivation over time.⁶³

The second approach to reducing health inequities is to intervene on risk factors that sit on the pathway between SEP and disease (for example, by improving housing conditions or reducing rates of smoking). This, arguably, more pragmatic and achievable approach is what my thesis focuses on. To achieve the goal of identifying interventions to reduce inequities, we need to first quantify the extent to which risk factors contribute to health inequities. As a nationally applicable measure of SEP, IMD is a helpful way of categorising the population by levels of socioeconomic deprivation to further investigate these underlying risk factors. Indeed, IMD was designed as a tool for local policy makers and communities to identify the most deprived areas for “the effective targeting of resources”.⁶⁴ On the other hand, some of the components that are used to create IMD may include indicators of the very risk factors that we wish to investigate. For example, the living environment domain includes measures of the quality of housing and air pollution. To avoid the potential of ‘mathematical coupling’ (whereby two variables correlate if one shares part of the other), some researchers choose to remove the health domain from IMD when conducting health inequity research.^{70,76}

1.3. Conceptual framework

The pathways between SEP and bronchiolitis, and later asthma, are highly complex. To develop specific, yet policy relevant, research questions that can be answered with the available data, I will draw from the multifaceted conceptual framework developed for the Commission on Social Determinants of Health (CSDH) by the World Health Organisation (WHO).⁶⁰ A key strength of the CSDH framework lies in the differentiation between levels of causation. This framework purports that the socioeconomic and political context of a society creates socioeconomic strata by factors such as income, education, occupation, gender and/or ethnicity. In turn, these structural dimensions operate through intermediary factors, which broadly include material and psychosocial circumstance and behavioural factors, to shape health outcomes.

Figure 1.1, which I have adapted from the CSDH, displays the guiding framework of this PhD in reference to inequity in bronchiolitis and asthma outcomes. Alongside SEP, which is the structural determinant of inequity central to my thesis, I include neighbourhood, built environment, and geographical location and climate. These environmental dimensions are strongly associated with socioeconomic factors, but are also independently associated with risk factors important to respiratory disease, such as pollutants and viruses.^{20,77} In addition, although not explicitly displayed in the framework, the life course perspective is an essential component of the social determinants of health. This perspective helps to frame the temporal element to my project; namely, that exposures earlier in life contribute to respiratory disease outcomes months or years later in life. The effects of risk factors may accumulate over life and/or specific insults may happen during a critical or sensitive period, leading to long-lasting effects on later disease outcomes.^{78,79} For example, children with asthma by the age of seven have been shown to have lung function deficits at just one month of age, emphasising the early origins of at least some asthma phenotypes.^{80,81}

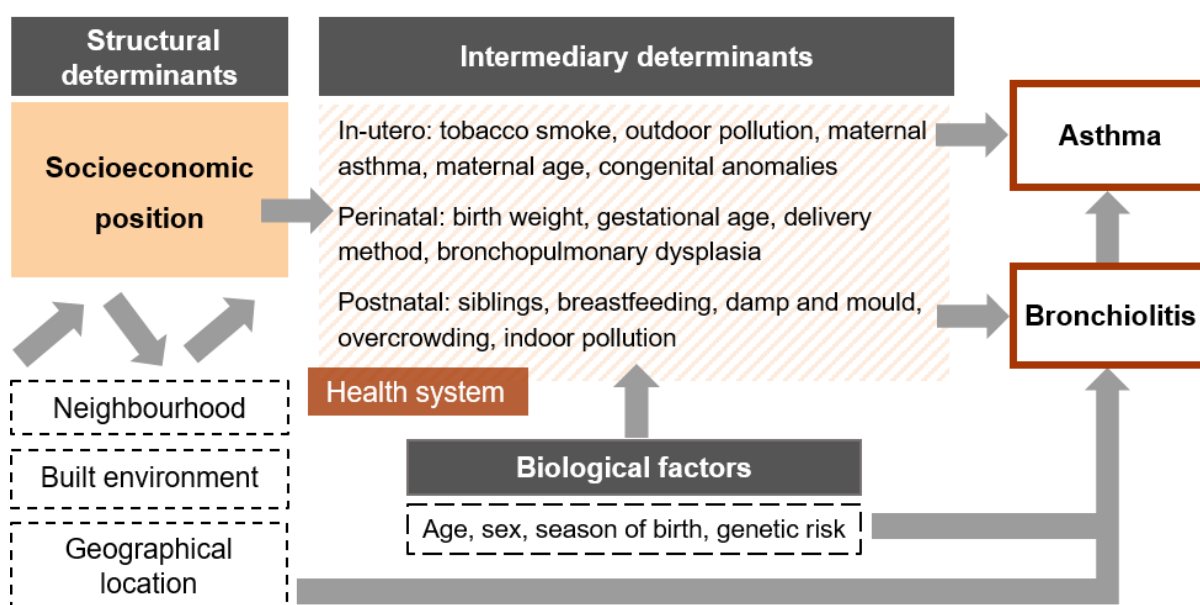


Figure 1.1. Framework for pathways to socioeconomic inequities in the risk of bronchiolitis and paediatric asthma

In the next sections, I will review the evidence for each component of my framework in turn.

1.3.1. Literature search strategy

The information presented in this narrative review of the literature was initially selected by searching electronic databases (Medline, Embase, Maternity and Infant care database and

PubMed) and national statistics with keywords relevant to each section. For example, bronchiolitis and ethnicity, or asthma and smoking. I applied a snowball method (pursuing references of references) to these papers to find the literature most relevant to describing inequities in bronchiolitis infections and asthma in the UK. To ensure that I also included the most recent research available in the field, I subsequently added to the review using the papers identified using the search criteria outlined in Table 1.1.

Table 1.1. Literature search terms for two overarching research questions

Research question	(1) How is early life socioeconomic position linked to bronchiolitis?	(2) How is early life socioeconomic position linked to asthma in childhood?
<u>Search strategy</u>		
Databases	Medline, Embase, Maternity and Infant care database (using OvidSP) and PubMed	Medline, Embase, Maternity and Infant care database (using OvidSP) and PubMed
Search terms	OvidSP: (Social determinants of health/ OR Social status/ OR Social structure/ or Socioeconomics/) AND (bronchiolitis OR human respiratory syncytial virus/). PubMed: (Socioeconomic factors [MeSH term] OR Social class [MeSH term]) AND (bronchiolitis OR human respiratory syncytial virus)	OvidSP: (Social determinants of health/ OR Social status/ OR Social structure/ or Socioeconomics/) AND asthma. PubMed: (Socioeconomic factors [MeSH term] OR Social class [MeSH term]) AND asthma
Inclusion criteria	<ul style="list-style-type: none"> • English language • Human • High-income countries • Study participants <2 years • Published ≥2018 • Outcome: rates or cases of bronchiolitis 	<ul style="list-style-type: none"> • English language • Human • High-income countries • Study participants <18 years • Published ≥2018 • Outcome: rates or cases of asthma
<u>Number of identified papers</u>		
Initial	66 studies	157 studies
Removed	11 duplicates 12 not relevant (from title/abstract) 2 not relevant (paper)	6 duplicates 63 not relevant (from title/abstract) 52 not relevant (paper)
Final	41 studies	36 studies

1.4. Structural determinants

In this section, I will present evidence on the association between structural determinants of bronchiolitis and asthma in a UK context. Whilst the primary focus of my PhD is on SEP, I will also introduce evidence on the relationship between ethnicity and immigration status—other structural determinants that interact closely with SEP to create inequities—and these health outcomes in this section. Geographical location, the built environment and neighbourhood are included here as additional factors that interact with SEP, as well as work independently, to influence the development of bronchiolitis and asthma.

1.4.1. SEP

Evidence from several studies of administrative health data show a socioeconomic gradient in the incidence of severe bronchiolitis (i.e. cases that require hospitalisation), a pattern that has persisted over many years. Using data from the Oxford Record Linkage Study, Green et al. looked at hospital admissions among children born between 1970 and 1989.⁴ After adjustment for other risk factors, they found a 2.34 greater odds of admission to hospital for bronchiolitis among infants born into the lowest fifth of SEP (measured using the Registrar General's occupational classification, based on the head of the household), compared to the highest fifth of SEP (odds ratio (OR) 2.34, 95% CI 1.65 to 3.30). Among births in England between 2003 and 2012, infants with a birth address in the highest IMD decile had a 1.94 greater incidence of hospitalisation for bronchiolitis (sex and year-adjusted incidence rate ratio (IRR) 1.94, 95% 1.73 to 2.19) and those in the 50%-89% of IMD a 1.48 greater incidence rate (sex and year-adjusted IRR 1.48, 95% CI 1.07 to 2.06), compared to infants in the lowest IMD decile.⁸² A similar pattern but at a lower risk was observed among infants born in Scotland over the same period (sex and year-adjusted IRR 1.28, 95% CI 1.16 to 1.42, highest compared to the lowest Carstairs index decile).⁸² The Carstairs index is constructed at the postcode sector level, covering a population on average more than 3 times the size of the English IMD, which likely explains the lower risk found in Scotland compared to England.⁸³

A study of all 29 Paediatric Intensive Care Units (PICUs) in England and Wales between 2004 and 2007 also found that admissions for acute respiratory failure and bronchiolitis were more likely among infants from the most deprived fifth of the community (measured using the Townsend deprivation index), compared to the least deprived fifth (OR 1.37, 95% CI 1.22 to 1.54).⁸⁴ However, this study did not find an association between deprivation and length of PICU stay, indicating that there may be no difference in disease severity. Another study, which looked at all hospital admissions for bronchiolitis in England (not just those admitted to

intensive care), did not find an association between inpatient duration of stay and IMD at the Primary Care Trust level.¹⁵ Research from the USA, which investigated additional indicators of disease severity, including receipt of mechanical ventilation, days since breathing problems began and presence of apnoea, also found no difference between bronchiolitis severity among hospitalised infants from different median household income groups (based on ZIP-codes, which covered a median of 2,751 people at the 2010 census).^{85,86} Higher income groups were more likely to be admitted to an ICU, however, pointing towards a disparity in treatment based on non-clinical factors. The authors suggest that wealthier families, who can afford the “best care” may push for ICU treatment; a factor which may be less relevant in the context of the (mostly) free at the point of use health service in the UK.⁸⁷

Associations between low SEP (i.e. more deprivation), measured by parental education, income and IMD, and a doctor diagnosis of asthma in childhood, have been noted across cross-sectional and cohort studies in the UK.^{21,29,88} Estimates of asthma and wheeze among children in 2019 derived from the cross-sectional Scottish Health Survey are displayed in Figure 1.2.⁸⁹ This chart shows a broad increase in asthma prevalence by increased deprivation, particularly where SEP is measured by the Scottish index of multiple deprivation (SIMD).

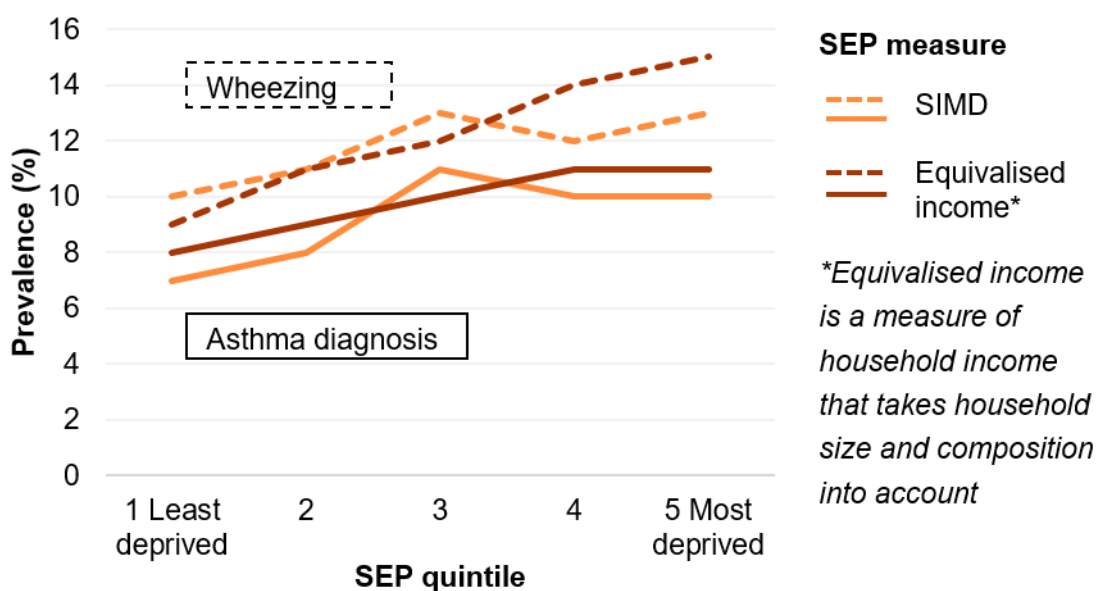


Figure 1.2. Prevalence of wheezing in the last 12 months (dashed lines) and doctor diagnosis of asthma ever (solid line) among children <16 years, by measure of SEP: Scotland, 2019

There are also inequities in the risk of emergency hospital admissions due to the condition. A meta-analysis of 3 randomised controlled trials and 33 observational studies (mostly from North America) estimated that the odds of emergency department re-attendance was 97% greater (OR 1.97, 95% CI 1.17 to 3.34) and hospital admission 25% greater (OR 1.25, 95% CI 1.07 to 1.47) in children from low compared to high socioeconomic groups (measured by a several indicators including income, insurance status and working rank).⁹⁰ On average, in England between 2002 and 2015, the rate of emergency hospital admissions for asthma was 3.34 times greater among people aged 5-44 years in the most deprived quintile of IMD compared to the least deprived quintile (IRR 3.34, 95% CI 3.30 to 3.38).⁶⁹

Inequities in the prevalence of asthma in the UK are relatively high compared to similar countries. This is shown in a paper that I wrote as part of a rotation in the first year of my PhD (the paper, published in the *European Journal of Epidemiology*, is included in Appendix 1.1).⁸⁸ This study investigated the association between maternal education and reported doctor diagnosis of asthma in children ≤ 8 years across 10 European countries. The UK data source was the Millennium Cohort Study, which is a longitudinal cohort study of approximately 19,000 young people born across the four nations of the UK between 2000-2001.⁹¹ This study found that, of the 10 included countries, the UK had the second highest relative difference in offspring asthma prevalence between the highest and lowest maternal education groups. Children born to mothers with up to lower secondary level education were 73% more likely to report that their child had a doctor's diagnosis of asthma by age 5 compared to mothers with post-secondary level education (95% CI 1.44 to 2.09), after adjustment for child sex, smoking during pregnancy, parity, maternal age and ethnicity.

Not all studies present a clear association between socioeconomic position and the prevalence of asthma.^{21,92} Differences in survey design, the age of the population sampled and the definition of SEP likely contribute to differing results; however, the measurement of asthma is likely the biggest contributing factor.⁹² As explained in Section 1.1.2, asthma is a descriptive label applied to diverse array of symptoms, and these also likely have a diverse array of aetiologies.^{24,47} It therefore follows that the influence of SEP may vary among children with different types of asthma phenotypes and, by treating asthma as one homogenous disease in research, we fail to tease out these associations. Research using the Millennium Cohort Study found a slightly stronger relationship between maternal education and early remitting childhood wheeze (relative risk ratio (RRR) of 1.48 when comparing children of mothers with no qualifications to those with a degree) than for persistent wheeze (RRR 1.31 (95% CI 1.04 to 1.65)).⁹³

The association between SEP and asthma also differs in studies that consider comorbid atopy. In 2014, Uphoff et al.⁹² conducted a systematic review of studies looking at the association between SEP and asthma and allergic diseases in children and adults. Overall, they found a positive association between lower SEP (i.e. more deprivation) and asthma (OR 1.38, 95% CI 1.37 to 1.39), but the opposite association was found when the outcome was limited to atopy (OR 0.67, 95% CI 0.62 to 0.72). Uphoff and colleagues write that this finding may simply represent differences in the reporting of allergies, but the evidence for this is unclear. Another explanation is derived from Strachan's 1989 'hygiene hypothesis', which proposed that declining family size, household improvements and higher personal cleanliness were contributing to the rise in allergic diseases, especially amongst richer families.⁹⁴ Subsequent studies have provided inconsistent evidence for this theory, with the newer "microflora hypothesis" focussing on disruption to early life microbial exposure as the pathway to allergic diseases.^{95,96}

1.4.2. Ethnicity and migration status (not included in Figure 1.1)

Ethnicity is another grouping that can represent a structural determinant of health. Ethnicity (or ethnic group) is defined as "the social group a person belongs to, and either identifies with or is identified with by others, as a result of a mix of cultural and other factors including language, diet, religion, ancestry, and physical features".⁹⁷ The term ethnic minority is used to describe people regarded by themselves or others as not being from the majority ethnic group in the population (i.e. white or white British in the UK).⁹⁸ People from ethnic minority groups in the UK tend to experience greater disparities in socioeconomic and health outcomes (although experiences differ vastly by ethnicity).^{99,100} Krishna et al. suggest that poor clinical health outcomes for allergy and asthma, in particular, are due to a multitude of factors including "generally lower socio-economic status, language barriers, cultural, social, religious aspects and stigma/taboo associated with chronic disease that may impact on engagement with clinical services and compliance to therapy."¹⁰⁰

Migration status, defined by residence in a country that is not an individual's country of birth, often overlaps with ethnic minority status, but is not the exact same entity. There are joint and separate pathways that may operate from ethnicity, migration status and socioeconomic position to inequities in health outcomes. Prominently, the relationships between inequity, ethnic group and migration status are not heterogeneous and, as written by the Migration Observatory, "classifying people according to their ethnic origin or immigrant background is tendentious and difficult".¹⁰¹ The complexities in unravelling the relationship between ethnicity and health outcomes is reflected in the respiratory health literature. One study of bronchiolitis admissions to hospital in England in 2008/09 observed no difference in the distribution of

ethnicity (split into white and “other” ethnicities).⁵⁴ However, the findings results could be biased by missingness; 15% of children had no ethnicity recorded. Moreover, the broad grouping of all ethnic minority groups will hide underlying disparities relevant to some groups. The highest rates of bronchiolitis were observed among infants from black and Hispanic ethnic groups and the lowest in Asian groups in two USA studies (hospitalisations in New York and medical records in Minnesota).^{102,103} Between 2004 and 2007, the odds of admittance to a PICU in the UK with a primary diagnosis of bronchiolitis was lower in South Asian infants compared to non-South Asian infants (OR 0.73, 95% CI 0.63 to 0.84).⁸⁴ There is no further breakdown of ethnicity beyond South Asian.

Differences in the prevalence of doctor diagnosed asthma and severity of asthma symptoms by ethnicity and immigration status have been noted in the UK. A national cohort study using Scottish administrative data showed significant differences in the rate of hospitalisations for asthma by ethnic group.¹⁰⁴ Compared to the white Scottish population, age-adjusted rates of admission for asthma were 20-50% greater among people from a South Asian ethnicity and 30–40% lower among people with a Chinese ethnicity. These associations remained after adjustment for SIMD and country of birth. A study using Born in Bradford cohort data found a 59% increased risk of doctor diagnosed asthma among Pakistani children compared to white children at age 4 in unadjusted logistic regression analysis (OR 1.59, 95% CI 1.11 to 2.28).¹⁰⁵ Parental immigration status was marginally associated with an increased odds of asthma diagnosis (OR 1.33 95% CI 0.97 to 1.81). However, with no pattern seen among other measures of asthma, including reported wheeze and asthma medication, the study authors concluded that “traditional risk factors such as gender, family history, socio-economic status and child's medical history may be stronger risk factors than ethnicity or familial migration patterns.”

In a study conducted in Leicestershire, reported rates of ever having asthma, wheezing or bronchitis were twice as high among white women compared to South Asian women (21.8% vs. 10.9%).¹⁰⁶ However, this gap closed when restricting south Asian women to those born in the UK (16.2%), highlighting the potential influence of early life exposures in the UK. This reflects findings from a linked administrative data study in Sweden, which showed a persistent decline in the odds of asthma medication with age at migration in children from a range of low or median income regions.¹⁰⁷ The pattern was noted both among children who were adopted from abroad by a Swedish family and those who emigrated with their parents, further highlighting the role of environmental factors in asthma.

1.4.3. Geographical location

Across the globe RSV exhibits a latitudinal gradient in the timing of epidemics.¹⁰⁸ In the Northern hemisphere this manifests as RSV activity beginning in July in tropical sites and spreading further North to reach high-latitude sites in January. A systematic analysis of global viral patterns found, in Europe, where there is more longitudinal than latitudinal variation, a later RSV onset in the East compared to the West (0.8 months across 20°E), $r=0.45$, $p=0.03$, but no clear association between North and South, $r=0.11$, $p=0.61$.¹⁰⁸ In temperate maritime climates such as the UK, RSV epidemics occur during the colder months.^{109,110} Climatic factors, including temperature, humidity, precipitation and vapour pressure have all been linked to RSV outbreaks.^{108,110,111} A study of the infectious disease data reported to Public Health England (PHE) observed a positive correlation between air frost ($r=0.69$), and a negative correlation between max temperature ($r=-0.60$), and number of cases of RSV infection one month later.¹¹²

Climate change has been implicated in changes to the length of RSV epidemics among young children in the UK. A study of bronchiolitis hospital admissions and laboratory isolations of RSV in England and Wales between 1990 and 2004 found that the epidemic length was associated with a 2.5 to 3.1 week decrease per 1 degrees increase in temperature over this time.¹¹³ Alongside infectious and cardiovascular diseases, asthma and allergies are described as being “at the front line of the sequelae of climate change”.¹¹⁴ Cecchi et al. describe the impact of climate change on allergies in terms of the exposome—an increase in environmental exposures, such as pollen load and moulds, that cumulatively impact on an individual from birth onwards.¹¹⁵ Air pollution, a precursor to climate change that is socially patterned in terms of exposure,¹¹⁶ is another important component of this sequence (discussed further in Section 1.5.3, below).

There is some evidence suggesting that geographical location influences the development of asthma. One study of current reported asthma across the USA and Australia, found that a 10° change in latitude from South to North (approximately 220km) was associated with a 2.0% increase in the prevalence of asthma.¹¹⁷ In mutually adjusted analyses of the USA sample, annual mean air temperature (°C) and solar radiation (kWh/m²/day) were negatively associated with asthma prevalence, $r=-0.42$ and $r=-0.10$, respectively. A 2016 Cochrane review of clinical trials found a “clinically and statistically significant protective effect of vitamin D against severe exacerbation of asthma”,^{118 p.21} further suggesting that sunlight plays a role in the disease. The relative size of the UK means there is unlikely to be substantial geographical variation in the prevalence of asthma within the country due to latitude or climate;

however, changing climatic features across the year, including temperature fluctuations and humidity have been associated with asthma symptoms and mortality across the country.^{119,120}

1.4.4. Built environment

I use the term built environment to refer to the complex of buildings, streets, greenspace and infrastructure that make up the physical spaces that humans live within.¹²¹ These physical spaces can contribute positively to health and wellbeing by influencing lifestyle factors such as exercise and social interactions.^{122,123} On the other hand, the built environment can facilitate exposure to harmful agents, such as viruses, outdoor pollution and toxins. These negative exposures are particularly important in the context of bronchiolitis and asthma. There is some evidence from non-UK countries that concentrated populations (typically found in high density built environments) affects the transmission of RSV. In Pitzer et al.'s study of the spatiotemporal dynamics of RSV in the USA, population density at the state level was positively correlated with the reproduction number of RSV ($r = 0.77$).¹¹¹ In another study, at the ZIP code-level in Connecticut, population density was correlated with earlier RSV epidemic peak and higher incidence rates.¹²⁴ The built environment is inherently linked to levels of outdoor air pollution with road transport sources producing two of the main pollutants associated with respiratory morbidity, nitrogen dioxide (NO₂) and particulate matter less than 2.5/10 micrometres (PM_{2.5/10}) (discussed further in Section 1.5.3).^{24,125} Facets of the indoor environment, particularly damp and mould, are also relevant to poor respiratory health (see Section 1.5.10).

1.4.5. Neighbourhood

SEP and neighbourhood are intrinsically linked, and in some instances are inseparable, as neighbourhood level indicators are sometimes used to define SEP, as in this study. There is evidence to suggest that measurements at the neighbourhood level are less volatile than other measures such as income.¹²⁶ Of 13 studies in Uphoff et al.'s systematic review reporting on area deprivation or average household income within a country, 12 showed an association between higher deprivation and asthma or wheeze and 1 found 'no significant result'.⁹² There are also risk factors for ill-health that exist at the neighbourhood level, such as physical deterioration and crime, that individual or household measures cannot characterise.¹²⁷⁻¹²⁹ Unequal access to resources may further enhance inequities (discussed further in the health system Section, 1.6).¹²¹ Moving between neighbourhoods may also have an impact on respiratory health in children. Using a cohort of 2,619 children from California, USA, Cantu et al. investigated the impact of mobility between poor and non-poor neighbourhoods.¹³⁰ After a period of at least 4 years, 6.2% of the cohort moved into and 11.5% moved out of a poor

neighbourhood. Membership of the former group was associated with higher rates of childhood asthma than children who remained in a poor area at all times of the study (19.5% vs. 16.7%, 13.2% overall prevalence). The authors theorise that moving may be a sign of family instability or stress; however, the models did not account for possible mediators such as passive smoking that may help explain this association.

1.5. Intermediary determinants

For the most part, structural determinants of health do not act directly on health outcomes but, rather, work through intermediary determinants (or risk factors) to determine health. In statistical terms, these intermediary determinants are known as mediators, a concept which will be discussed further in Chapter 6. In this section, I outline how each of the intermediary determinants listed in Figure 1.1 is associated with socioeconomic deprivation and inequities in bronchiolitis/asthma. For illustrative purposes, these are grouped by timing of exposure (in-utero, perinatal and postnatal), although some may exert their influence at more than one time (e.g. smoke exposure in-utero and passive smoking in the early years). Notably, this list is not exhaustive and other risk factors have been identified, with varying levels of association noted.

1.5.1. Maternal smoking during pregnancy

Rates of smoking in the population follows a clear social gradient, and this pattern is also observed among pregnant women.^{131,132} The Scottish Public Health Observatory reports that in 2018/19, 26.6% of women living in the most deprived fifth of SIMD reported smoking during pregnancy compared to just 3.3% in the least deprived fifth.¹³³ In England in 2018/19, 12.8% of women living in the most deprived clinical commission group (CCG), as measured by IMD, reported that they smoked at the time of delivery compared to 6.1% in the least deprived CCG.^{134,135}

Two Northern American studies have also looked at the association between maternal smoking during pregnancy and medical visits for bronchiolitis. An observational cohort study, based on term non low-birth weight infants enrolled in the Tennessee Medicaid Program from 1995 to 2003, found that maternal smoking was associated with a 14% increased risk of a bronchiolitis diagnosis (adjusted HR 1.14, 95% CI 1.10 to 1.18) and a 28% increased odds of hospitalisation for the condition (adjusted OR 1.28, 95% CI 1.20 to 1.36).¹³⁶ A Canadian study, conducted on administrative records of all infants born the geographical area known as the Georgia Air Basin (British Columbia) between 1999 and 2002, found a stronger association between maternal smoking and hospitalisation for bronchiolitis (adjusted HR 1.47, 95% CI

1.27 to 1.69).¹³⁷ A study of Scottish administrative data presented the association between maternal smoking in pregnancy and subsequent offspring respiratory admissions. Hazard ratios by smoking status on bronchiolitis admissions within the first year of life and asthma admissions between the 1st and 5th year are displayed in Table 1.2. Women who reported being a current smoker during pregnancy had the highest risk of subsequent offspring bronchiolitis (unadjusted HR 1.79, 95% CI 1.73 to 1.84) and asthma (unadjusted HR 1.51, 95% CI 1.43 to 1.59) when compared to women reporting never smoking.

Table 1.2. Rates of hospital admissions for bronchiolitis (birth to 1 year) and asthma (1 year to 5 years), by mothers smoking status at first antenatal booking (from Lawder et al.)¹³⁸

	Infants at risk	Infants with ≥1 admission	% admitted	Crude HR (95% CI)	Adjusted HR* (95% CI)
Bronchiolitis	697,003	20,399	2.9		
<i>Never smoked</i>	391,861	9,499	2.4	Reference	Reference
<i>Current smoker</i>	159,819	6,861	4.3	1.79 (1.73–1.84)	1.43 (1.38–1.48)
<i>Former smoker</i>	59,326	1,534	2.6	1.07 (1.01–1.13)	1.03 (0.97–1.08)
<i>Unknown</i>	85,997	2,505	2.9	1.21 (1.16–1.27)	1.15 (1.09–1.21)
Asthma	679,588	6,943	1.0		
<i>Never smoked</i>	382,327	3,436	0.9	Reference	Reference
<i>Current smoker</i>	156,123	2,164	1.4	1.51 (1.43–1.59)	1.29 (1.22–1.37)
<i>Former smoker</i>	58,025	608	1.0	1.16 (1.07–1.27)	1.10 (1.00–1.20)
<i>Unknown</i>	83,113	735	0.9	1.04 (0.96–1.12)	1.01 (0.92–1.10)

*Adjusted for: maternal age; infant gender, SIMD; mother's socioeconomic status (measured by occupation); father's socioeconomic status (measured by occupation); parity; mode of delivery, infant feeding at 6–8 weeks and country of birth.

Associational studies have also presented an association with passive smoking exposure after birth and bronchiolitis/asthma. In many cases, the size of the association is larger than that of smoke exposure in-utero. For example, a meta-analysis pooling results of 7 studies identified a 2.51 increased odds of bronchiolitis where infants were exposed to passive smoke by any household member (adjusted OR 2.51, 95% CI 1.96–3.21).¹³⁹ However, there is no clear indication of how exposure to smoke exposure in-utero was accounted for in these studies. In another systematic review, contradictory findings were noted among studies that looked at the association between passive smoking in early years and non-atopic asthma.¹⁴⁰ Worsened bronchiolitis outcomes due to passive smoke exposure have also been noted. A cohort study of children admitted to a children's hospital in Liverpool for bronchiolitis found that infants living with a tobacco smoker had increased odds of needing oxygen supplementation and mechanical ventilation than infants without a household tobacco smoker.¹³⁰

1.5.2. Maternal asthma

As discussed above in Section 1.1.2 in reference to the paediatric population, the prevalence of asthma in adults also exhibits a socioeconomic gradient. In 2018, 21% of Scottish women in the most deprived quintile of SIMD reported a doctor's diagnosis of asthma compared to 15% in the least deprived SIMD quintile.¹⁴¹ Maternal asthma, in turn, is with an increased odds of offspring bronchiolitis and asthma. A study of otherwise healthy infants who visited a health care facility for bronchiolitis found a risk 39% greater (HR 1.39, 95% CI 1.30 to 1.48) in infants of mothers with asthma compared to mothers without asthma (identified using clinical records), after adjustment for a multitude of risk factors.¹³⁶ Analysis of data from the Oxford record linkage study showed a strong association between maternal asthma and offspring hospital admission for asthma after 2 years of age (OR 3.07, 95% CI 2.61 to 3.60 in multivariable analyses).¹⁴²

It has been suggested that perinatal factors, including birthweight and gestational age, may sit on the pathway between maternal asthma and child respiratory ill-health.¹⁴³ However, the association between maternal and offspring asthma is complex and likely due to a mixture of factors including familial disposition to respiratory ill health, direct effects of maternal asthma on their offspring and shared environmental factors. In Strina et al.'s systematic review, familial history of asthma or atopy was the most consistently associated risk factor with paediatric non-atopic asthma, with the authors referencing the influence of genetics and/or shared environment.¹⁴⁰ However, other studies that examine both maternal and paternal symptoms suggest that there is also an effect attributable to the mother only. For example, Schoos et al. examined the distinct association between mother and offspring asthma and atopy using 685 parent-child trios from the Copenhagen Prospective Studies on Asthma in Childhood 2010 cohort.¹⁴⁴ They found a greater (but overlapping) association between mother and child asthma/persistent wheeze symptoms between 0 and 6 years, than with father and child symptoms (adjusted OR for mothers 2.11, 95% CI 1.46 to 3.05 compared to OR 1.55 for fathers, 95% CI 1.03 to 2.33). The authors write that "this suggests that maternal non-genetic factors seem to confer an added disease risk to the child, particularly in early life".^{144 p.915}

1.5.3. Outdoor air pollution

The relationship between socioeconomic position and ambient air pollution levels is complex, varying by urban-rural status and across countries, but the evidence broadly points to inequities in exposure to air quality across England and Scotland.^{116,128,145,146} A 2006 report for the Department of Environment, Food and Rural Affairs summarised that there are higher relative concentrations of NO₂ and PM₁₀ in more deprived areas (measured by IMD/SIMD

deciles) in England and Scotland.¹¹⁶ In England, sulphur dioxide (SO₂, produced during fossil fuel combustion at industrial plants) levels were also higher in areas with greater deprivation. These patterns are driven by urban areas (where more people live in deprived areas and there are far bigger populations generally) and, notably, there is wide variation within deprivation deciles.

A link between short-term spikes in air pollution and asthma symptom exacerbations has been well established in the research literature;¹⁴⁷ however, ascertaining a causal role in the chronic exposure to ambient air pollution and the development of respiratory conditions has been more difficult.¹⁴⁸ Research in this area is particularly limited by difficulties separating the effects of specific pollutants and in assigning the correct exposure level to individuals.¹⁴⁹ Nonetheless, there is a large body of evidence showing an association between concentrations of NO₂ and PM_{2.5} (measured prenatally and in early-life) and the onset of asthma in children.^{24,150–155} As highlighted in a nationwide case-control study conducted in Denmark, this is even the case where PM_{2.5} levels are below the proposed European annual limit.¹⁵²

The evidence of a link specifically with bronchiolitis is less strong, particularly for chronic exposure to air pollution. One systematic review looked at the effect of several outdoor air pollutants on the risk of hospitalisation for bronchiolitis in children under the age of 2.¹⁵⁶ The findings were inconsistent, hampered by variation between studies in measurement of study variables and the lack of cohort study designs. However, there is evidence of an association between air pollution and other early-life acute lower respiratory tract infections, particularly pneumonia. As part of the ESCAPE project, combined results of 10 European birth cohorts presented a 76% increased odds of doctor-diagnosed pneumonia by age 3 per 10-µg/m³ increase in PM₁₀ (adjusted OR 1.76, 95% CI 1.00 to 3.09) and 30% increased odds of pneumonia per 10 µg/m³ increase in NO₂ (adjusted OR 1.30, 95% CI 1.02 to 1.65).¹²⁵ A meta-analysis of studies, the majority of which were in high-income countries, produced an estimate of 12% increased risk (95% CI 1.03 to 1.30) in acute lower respiratory tract infection occurrence per 10 µg/m³ increase in annual average PM_{2.5}.¹⁵⁷

1.5.4. Maternal age at delivery

Maternal age increases with higher levels of socioeconomic position. Of live births in England and Wales in 2019, 32% of mothers in the lowest category of the National Statistics Socio-economic classification (a measure of SEP based on occupations) were less than 25 years at the time of birth compare to just 3% in the highest category.¹⁵⁸ In contrast, 16% of mothers in the lowest category were 35 and older at the delivery of their child compared to 37% in the highest category. A similar pattern was observed in Scotland in 2019/20, where the most

common maternal age at first birth was 24 years among mothers from the most deprived SIMD quintile compared to 30 years in mothers from the least deprived quintile.¹⁵⁹

Risk of RSV-related hospital admission before the age of 3 was associated with earlier maternal age at delivery in a Scottish birth cohort.⁴⁴ Children of mothers aged less than 20 at birth had a 43% increased adjusted risk compared to mothers between 30 and 39 years old (adjusted HR 1.43, 95% CI 1.26 to 1.32). Higher maternal age during pregnancy was associated with lower odds of offspring asthma without atopy at 7/8 years of age (OR 0.95, 95 CI% 0.93 to 0.97) in the Avon Longitudinal Study of Parent and Children.²⁹ Given the strong correlation between SEP and maternal age, it is thought that SEP-related risk factors, unmeasured in these studies, explain the majority of the noted associations.^{44,137} Although evidence for a causal association has not been shown,²⁹ Koehoorn et al. theorise that a potential biological mechanism could be “related to maternal nutrition, health status, or stress that alters the development of the respiratory system of the fetus”.^{137 p.1201}

1.5.5. Congenital heart disease, chronic lung disease and other early-life conditions

There are a number of early-onset chronic health conditions, which are associated with a greater risk of respiratory ill health, particularly RSV-related hospitalisation. The foremost examples of these are congenital heart disease and chronic lung disease, which are two of the conditions specifically listed (alongside other criteria) in recommendations for infants to receive prophylaxis against RSV infection.^{40,41} Congenital heart disease encompasses a range of defects, which develop prenatally and affect the structure of the heart.¹⁶⁰ Congenital heart diseases are the most prevalent type of birth defects in the UK, affecting 60.9 per 10,000 births in England in 2018 according to the National Congenital Anomaly and Rare Disease Registration.¹⁶¹ Chronic lung disease includes some congenital conditions (i.e. those that develop prenatally), but primarily consists of bronchopulmonary dysplasia—chronic lung dysfunction related to extreme premature birth.¹⁶² Other notable early life conditions include Down’s syndrome, cystic fibrosis and neuromuscular disease.^{163,164}

Socioeconomic patterns of these conditions differ by the specific condition and underlying cause^{165,166}. Notably, conditions caused by chromosomal anomalies (including Down’s syndrome) are more prevalent in higher socioeconomic positions, which is primarily explained by the increased prevalence of these conditions at older maternal ages.¹⁶⁶ Non-chromosomal anomalies, including congenital heart disease, tends to be concentrated in more deprived populations.^{165,166} It is thought that socially patterned risk factors, including maternal nutrition, alcohol, tobacco and chemical exposure likely contribute to this, although there are still many

unknowns.¹⁶⁷ Parental decision making about terminations following prenatal identification is another noted factor in the prevalence of congenital anomalies.^{165,168} A study of the East Midlands and South Yorkshire congenital anomaly register found similar rates of anomalies detected prenatally across IMD deciles (after adjusting for maternal age). However, lower proportions of terminations after antenatal diagnosis meant that the most deprived areas had a 61% higher rate of live births (IRR 1.61, 1.21 to 2.15) associated with a congenital anomaly.¹⁶⁸

The risk of RSV-related hospital admission in children less than 3 years old was 3.4 times higher among children with a chronic condition (including chronic heart, lung and neurological conditions) born in Scotland between 2009 and 2015 (unadjusted HR 3.40, 95% CI 3.11 to 3.72).⁴⁴ In England in 2008/07, infants with a congenital heart disease comprised 12.1% of those admitted to hospital with bronchiolitis in the first year of life, although only comprising <1% of the infant population.⁴⁵ After adjustment for other risk factors, infants with a congenital heart disease had a 3.35 increased risk of bronchiolitis hospitalisation compared to infants without this condition (95% CI 2.92 to 3.84). Once admitted, children with congenital anomalies are also at an increased risk of complications, including secondary infections, and mortality.¹⁶⁹ In a longitudinal study covering the Province of Manitoba in Canada, congenital anomalies of the respiratory and circulatory system were associated with a 43% (HR 1.43, 95% CI 1.27 to 1.61) and 16% (HR 1.16, 95% CI 1.03 to 1.30) increased risk of an asthma-associated health care visit by age 6, respectively.¹⁷⁰

1.5.6. Birth weight and gestational age

Gestational age and birthweight are strongly related to one another, with gestational age likely on the pathway to birthweight.¹⁷¹ Children born into lower SEPs are more likely to be born before full term (< 37 weeks) and at a lower birthweight. Millennium Cohort Study data presents a social gradient in the proportion of children born with low birth weight.¹⁷² 10.1% of children born to mothers with no secondary school qualifications had low birthweight (< 2500g) compared to 4.5% of those born to mothers with an undergraduate degree or higher. A national cohort study of NHS records in England and Wales found that 6.7% of singleton children in the most deprived 10th (measured by IMD) were born prematurely compared to 4.5% in the least deprived 10th.¹⁷³

Hardelid et al.⁴⁴ explored risk factors for RSV-confirmed hospital admissions in the first 3 years of life using a birth cohort created using linked Scottish administrative data. The association between RSV admission increased with earlier gestational age; at the lowest end of gestational age, children born at less than 34 weeks had an adjusted hazard ratio of 2.54 (95%

CI 2.10 to 3.06) compared to children born at term (37-40 weeks). Alongside gestational age, fetal growth is also implicated in birthweight, and babies that have fetal growth restrictions may have different health problems than those that are light due to prematurity.¹⁷⁴ Small for gestational age (SGA)—commonly defined as birthweight below the 10th percentile for sex-specific gestational age—is commonly used in studies as a proxy for fetal growth. In their study of Scottish administrative data, Hardelid et al.⁴⁴ observed an increased risk of RSV-related hospital admissions in the first 3 years of life among children born SGA after adjustment for gestational age and other factors (HR 1.14, 95% CI 1.03 to 1.25).

A meta-analysis of 13 cohort studies estimated that the risk of asthma in childhood was 16% higher in those born with a low birthweight (pooled RR 1.16, 95% CI 1.13 to 1.20).¹⁷⁵ Birthweight below 2500g was associated with increased risk of asthma and wheezing disorders compared to children weighing between 2500g and 4000g at birth in the Born in Bradford study.¹⁷⁶ A large cohort study of children from Denmark, Sweden and Finland estimated a 17% increased risk of hospitalisation for asthma between 3 and 18 years of age for every 1000g decrease in birthweight (RR 1.17, 95% CI 1.15 to 1.18) and 5% increased risk for each reduced week of gestational age (RR 1.05, 95% CI 1.02 to 1.09).¹⁷⁷ On the other hand, two studies investigating the association of prematurity with non-atopic asthma found no relationship.¹⁴⁰ There is mixed evidence on the role of SGA in asthma development. The combined Denmark, Sweden and Finland cohort study mentioned above found an association between SGA and asthma hospitalisation risk in children born at term only.¹⁷⁷

1.5.7. Delivery method

Caesarean section is a life-saving intervention when specific complications arise during pregnancy and childbirth, but has also been associated with a range of risks to the mother and child.¹⁷⁸ Compared to many other intermediary factors discussed in this section, the evidence on socioeconomic differences in delivery method is less clear. Crude results suggest higher rates of caesareans in the least deprived mothers;^{179,180} however, this relationship changes when more nuanced analyses are used. A large study of administrative hospital data in England found no association in rates of caesarean sections by IMD after adjustment for clinical indication (including breech presentation and previous caesarean section).¹⁸⁰ A study of the 2010 National Maternity Survey in England indicated a slightly raised odds of caesarean section in increasingly deprived groups (measured by IMD) when adjusted for ethnicity, maternal age and parity (OR 1.04, 95% CI 0.99 to 1.10).¹⁷⁹ Analyses of the Millennium Cohort Study found an association between caesarean section and increased deprivation that differed depending on the parity status of the mother and whether SEP was defined by education or occupation.¹⁸¹ For first-time mothers, only (lower) household occupation was associated with

an increased risk, whereas for multiparous women only (lower level of) education was associated with an increased risk of a planned caesarean birth.

A cohort study from Western Australia found that, compared to spontaneous vaginal delivery, children born by elective caesarean had an 11% increased risk of admission to hospital for bronchiolitis in the first year of life, after adjustment for other risk factors, which again points to the influence of underlying clinical risk (IRR 1.11, 95% CI 1.01 to 1.23).¹⁸² A meta-analysis of 26 studies estimated that birth by caesarean section increased the odds of an asthma diagnosis between ages 0 and 18 by more than 20% compared to spontaneous vaginal delivery (elective caesarean, OR 1.21, 95% CI 1.17 to 1.25; emergency caesarean OR 1.23, 95% CI 1.19 to 1.26).¹⁸³ Differences in the intestinal flora between babies born by caesarean and vaginal birth, has led to the hypothesis that immune system development lies on the pathway between delivery method and respiratory ill-health.¹⁸⁴ However, two sibling control studies, a study design that enables researchers to look at the association independent of environment and genetic factors, in Sweden have found no evidence to support a causal association between caesarean section and asthma.^{185,186} These indicate that caesarean section is actually an indicator of other shared familial or intrauterine factors which increase the risk of asthma.¹⁸⁷

1.5.8. Number of older siblings

On average in the UK, the number of children in a family increases with socioeconomic deprivation. Berrington and colleagues investigated fertility histories in the UK using retrospectively reported information from two panel studies, the British General Household Survey and the UK Household Longitudinal Study.¹⁸⁸ Among all women born between 1960 and 1969 (including those who have no children), lower levels of completed formal education were associated with a lower average completed family size. This ranged from a mean of 2.35 children among women who left school before or at 16 with no formal qualifications to 1.68 among women with a degree-level education or higher. This pattern is strongly related to the later age at entry into motherhood for women with higher levels of formal education. The presence of, particularly older, siblings increases a young child's exposure to infection. Hardelid et al.¹⁸⁹ estimated that, in a Scottish setting, RSV hospital admissions in children under 3 years old could be reduced by up to 34% by removing the risk posed by older siblings (population attributable fraction 34.0%, 95% CI 31.0 to 36.9). A study of RSV hospitalisations among children under 2 years in Denmark found that this association persisted after adjustment for a number of other risk factors including childcare.¹⁹⁰

The relationship between the number of siblings and asthma is less clear. As with many studies of asthma, the heterogeneous definition of asthma used by different studies likely affects the comparability of results. More than one older sibling was associated with early transient wheeze in a birth cohort study from the Netherlands (OR 1.31, 95% CI 1.07 to 1.60).¹⁹¹ However the outcome of transient wheeze, which peaked at ages 1 and 2, may be indicative of respiratory infection rather than asthma. In Strina et al.'s systematic review of the literature, six of seven studies presented no association between family size or birth order and non-atopic asthma/wheeze,¹⁴⁰ The final study, which reportedly found a protective effect of siblings, was from rural Ecuador, and the results may not be applicable to a UK-setting. Using International Study of Asthma and Allergies in Childhood (ISAAC) data on 13-14 year olds from 16 affluent counties, different trends were observed based on the outcome studied. Each additional older sibling was associated with an 8% greater odds of reporting severe wheeze in the last 12 months (OR 1.08, 95% CI 1.04 to 1.12), whereas the opposite effect was shown for asthma ever (OR per additional sibling 0.96, 0.94 to 0.99).¹⁹²

1.5.9. Breastfeeding

To provide the ideal nutrition for the healthy growth and development of infants, the WHO recommends that babies receive breast milk exclusively for the first 6 months of life.¹⁹³ The latest national estimate of rates of breastfeeding, from 2010, placed the UK as having one of the lowest breastfeeding rates in the world. The (now discontinued) Infant Feeding Survey reported that 34% of women were still breastfeeding their child at 6 months in 2010.¹³¹ This figure followed a social gradient, from 31% of women in the most deprived IMD quintile to 40% in the least deprived IMD quintile. There is evidence suggesting that this pattern very much persists a decade later. Public Health Scotland (PHS) reports that, in 2019/20, 43.9% of women were breastfeeding at their 6 to 8 week child health review, an estimate that ranged from 29.6% for women living in the most deprived Scottish IMD quintile to 63.8% in the least deprived quintile.¹⁹⁴

Breastfeeding has broadly shown to be associated with lower risk of early respiratory infections and, to a lesser extent, asthma, although there are differences based on the definition of breastfeeding used. There was an increased risk of hospitalisation for bronchiolitis among infants in the Georgia air basin region of Canada, where mothers had not initiated breastfeeding in the hospital (unadjusted HR 1.86, 95% CI 1.61 to 2.16).¹³⁷ Pooled results from 3 hospital-based studies rated high quality in a systematic review showed a 2.24 increased odds of hospitalisation for pneumonia and bronchiolitis in under 5s that had not been breastfed at all (OR 2.24, 95% CI 1.56 to 3.20).¹⁹⁵ In a Dutch population-based cohort study, breastfeeding at 6 months or longer was associated with a reduced risk of offspring

bronchitis or pneumonia up to age 4 years, compared to no breastfeeding (OR 0.63, 95% CI 0.46 to 0.87).¹⁹⁶ Similar but slightly lower effect estimates were produced for breastfeeding less than 3 months (OR 0.75, 95% CI 0.56 to 1.02) and 3-6 months (OR 0.72, 95% CI 0.48 to 1.06), compared to no breastfeeding.

In a meta-analysis of wheeze trajectory-specific risk factors in children, any breastfeeding was associated with a lower odds of early transient wheeze symptoms, compared to never/infrequent wheeze (pooled OR 0.86, 95% CI 0.77 to 0.95).¹⁹⁷ This association was not apparent in children with other trajectories of wheeze. Seven studies included in Strina et al.'s systematic review looked at the relationship between breastfeeding and non-atopic asthma.¹⁴⁰ Of these, five found a reduced risk from breastfeeding (three looked at no breastfeeding compared to any amount of breastfeeding, one at <3 vs. ≥3 months and one at <20 vs. ≥20 weeks) and two did not report an association.

1.5.10. Indoor living conditions

Components of the indoor environment that may increase a child's susceptibility to respiratory illness include dampness and mould, residential overcrowding and indoor air pollution (produced by cigarette smoke, heating or cooking).^{198,199} These housing conditions are more common among poorer households in the UK. The English Housing Survey, a national cross-sectional survey commissioned by the Ministry of Housing, Communities and Local Government (MHCLG), collects data about housing conditions at private addresses in England.²⁰⁰ This survey estimated that, in 2018, 7.2% of households in the lowest quintile of income did not have a reasonable degree of thermal comfort (related to insulation and heating efficiency) compared to 3.7% of households in the highest quintile.²⁰¹ In addition, 5.1% of the lowest income quintile households were categorised as having "any damp" compared to 1.3% of households in the highest income quintile. It has been proposed that damp indoor air conditions prolong virus survival, which may lead to a higher likelihood of RSV infection.^{202,203} Persistent indoor dampness provides conditions in which microbial pollution, particularly mould, can grow.²⁰² The government's bedroom standard indicator (used in the English Housing Survey) defines household overcrowding by those with fewer bedrooms than the number needed based on age, sex and marital status of the occupants.²⁰⁴ According to the latest survey results (combined over the last 3 years to enable reliable estimates), 5.1% of households in the lowest socioeconomic group were overcrowded compared to 1.6% in the highest group (defined by occupations).²⁰⁵ Moreover, overcrowding in England was at the highest recorded figure since records began (in 1995/96) for both the social rented (8.7%) and the private rented sector (6.7%).²⁰⁴

In their 2009 report on dampness and mould, the WHO wrote that “there is sufficient evidence of an association between indoor dampness-related factors and a wide range of respiratory health effects...including asthma development, asthma exacerbation, current asthma, respiratory infections”.²⁰² p.27 A meta-analysis published in 2010 found “moderate but statistically significant” associations between residential dampness and mould and respiratory tract infections and bronchitis.¹⁹⁹ In pooled analyses, Fisk et al. estimated that children living in residences with dampness and mould had a 48% increased odds of respiratory tract infection (unadjusted OR 1.48, 95% CI 1.34 to 1.62). The proportion of the risk in respiratory infection attributable to dampness and mould in this group was estimated to be between 8.8-19.4% after adjustment for age, sex, smoking exposure and SEP. A singular study from Wellington, New Zealand using an unmatched case–control design estimates that if all housing were free from damp and mould, then 19% of these types of hospital admission for any acute respiratory infection in under 2 year olds would be prevented.²⁰⁶ In a systematic review published in 2012, visible mould was associated with a 49% increased odds of asthma (OR 1.49, 95% CI 1.28 to 1.72) and 68% increased odds of wheeze (OR 1.68, 95% CI 1.48 to 1.90) in children aged between 0 and 15.²⁰⁷

A systematic review of 20 studies on residential crowding and laboratory-confirmed RSV hospitalisations among children younger than 5 years found a consistent increased odds of disease across different settings.²⁰⁸ It is thought that crowding can facilitate the spread of RSV through close contact conditions. However, the authors also offer caution when interpreting results from these observational studies, pointing out the difficulty in separating the effect of residential crowding from other socioeconomic-related risk factors such as indoor pollution. Another study used the Growing Up in New Zealand longitudinal child cohort study to look at a range of home environment risk factors in the risk of hospitalisation for acute respiratory infection during the first five years of life.¹⁹⁸ In unadjusted analysis, associations were found between household crowding (2+ versus <2 on the crowding index, HR 1.63, 95% CI 1.42 to 1.88), as well as no heating in the house, rented tenancy (as opposed to owner), cigarette smoking in the house, dampness of the house, and heavy condensation mould or mildew in the room where the child sleeps at night. In models adjusted for a range of child, maternal and environmental factors, there were notable associations between hospitalisation for acute respiratory infection and the use of a flued or unflued gas heater in the child’s bedroom (flued HR 1.69, 95% CI 1.21 to 2.36; unflued 1.68, 95% CI 1.12 to 2.53) or gas heater as the sole type of household heating (HR 1.64, 95% CI 1.29 to 2.09). See section 1.5.1 for information about passive smoking.

1.6. The Health system

In my framework, the health system is placed just outside the other intermediary determinants of health to highlight the unique role it plays in determining the equitability of health outcomes. Access within the health system not only directly impacts on health outcomes, but it can also address differences in the exposure of other intermediary determinants and vulnerability to their effects.⁶⁰ Understanding social determinants of health within the context of the UK health system is particularly important in my thesis, as it is through this system that the two health outcomes of interest are measured.

The national health system in the UK, the NHS, is designed so that there is equal and free access for all UK residents.²⁰⁹ In fact, economic analysis of English hospital care has shown that spending is highest amongst people living in the most deprived areas of the country.²¹⁰ However, rather than evidencing equity, increased hospital visits may be indicative of health seeking behaviour of those in a lower SEP. In general, poorer patients tend to present symptoms later to health care professionals, are less likely to avail of preventative NHS care and report worse patient experience than their richer counterparts.²¹¹ The inverse care law, coined by Tudor Hart in 1971 to refer to the phenomena of good medical care varying inversely with the needs of the population it was serving is,²¹² it seems, (as put by Michael Marmot) “as relevant as ever”.²¹³

There is evidence suggesting that differences in access and uptake of preventative health care access can affect the respiratory health of very young children in the UK. Hardelid et al. found an association between delayed vaccinations at 6 months (a proxy measure for poor access to preventative health services) and risk of a RSV-related hospital admission before the age of 3 in Scotland (adjusted HR 1.14, 95% CI 1.03 to 1.25).⁴⁴ The impact of preventative health care is particularly apparent in asthma outcomes. Asthma is considered an ambulatory care sensitive condition, which is a term used to describe conditions with symptoms that ought to be preventable by interventions in primary care.²¹⁴ Using Hospital Episode Statistics from 2002 to 2015, Gupta et al. found a social gradient to hospital admissions with a primary diagnosis of asthma in England, indicating a similar inverse gradient to preventative care uptake.⁶⁹ People aged 5-44 years in the most deprived quintile of IMD were over three times more likely to have an emergency hospital admission for asthma compared with the least deprived quintile (IRR 3.34, 95% CI 3.30 to 3.38).

There is also evidence of disparities in service provision for bronchiolitis and asthma once healthcare is sought. These differences are perhaps most clearly illustrated by variation in healthcare use and treatment by geographical location. Two studies have found significant variation in admission rates for bronchiolitis across hospitals in England, and concluded that healthcare provider factors, particularly admission thresholds, are likely the biggest

contributors to the variation in admission rates noted.^{4,215} A survey of 1,009 GPs in England in 2016 found that 19% prescribed oral corticosteroids for bronchiolitis despite NICE guidelines stating that this is not a recommended treatment.²¹⁵ Other studies have shown that the availability of primary care, particular out-of-hours GP services, in England is associated with visits to the emergency department.^{216,217} Using cross-sectional data from English practices in 2010-2011, Fleetwood et al. showed that a 10% increase in access (measured by a composite score) was associated with a 32.1% reduction in emergency admissions for asthma across all patient ages.²¹⁷ In another study, between April 2011 and March 2012, GP surgeries with the lowest levels of access (ranked using patients' reported ability to schedule appointments from national GP Patient Survey data) had a risk of an unplanned hospitalisation for asthma in patients under 15 years more than twice than those with the highest levels of access (RR 2.22, 95% CI 2.08 to 2.38).²¹⁶

1.7. Biological factors

There are several individual-level risk factors that increase the risk of both bronchiolitis and asthma but are not necessarily affected by SEP or other structural determinants of health. These non-socially determined risk factors are presented in the conceptual model under the broad (and loosely fitting) heading of biological factors (Figure 1.1).

1.7.1. Age

Young age is the factor most strongly predictive of the risk of hospital admission for bronchiolitis. Bronchiolitis hospital admissions peak between 1 and 4 months of age and two thirds of all admissions for bronchiolitis are in children less than 6 months old.^{6,45,218-220} In comparison, there are very few bronchiolitis admissions after the first year of life. Between mid-2007 and mid-2012 in England, only 7.4% admissions with a primary diagnosis of bronchiolitis were amongst children 1 year and older.⁶ There is also a strong age-related pattern to the risk to asthma. Although, as mentioned in Section 1.1.2, it is advised that asthma is not diagnosed in young children, clinical symptoms of wheeze are frequently seen in children in the first few years of life.^{20,24} Additionally, differences in the timing of symptomology before the age of five have been shown to be associated with different risk factors for asthma.²⁴ The progression of asthma is also not uniform across children. The European White Lung book estimates that of all children with asthma: about one third have persistent asthma through to adulthood; about one fifth have remission in their early teens, but see a re-appearance of asthma in adulthood; and in the remaining children (over two fifths of those with asthma), the disease remits in early teens and does not reappear.²⁰ Although less common, asthma can

also emerge for the first time in adulthood. Occupations that expose workers to airborne agents such as chemicals and detergents have been associated with adult-onset asthma.²⁰

1.7.2. Season of birth

In England, infants born in September or October make up a higher proportion of admissions for bronchiolitis compared to children born in other months of the year.⁴⁵ These birth months are approximately 2 months before peak viral activity, and it is suggested that a natural trough in antibody availability (alongside general immaturity of the lungs) in the first 2-4 months of life may leave these infants more susceptible to the chronic effects on the airways.⁵³ In addition, as the NHS recommends seeking help sooner for infants under 3 months who have a temperature, this likely leads to more admission for this age group in the first place.²²¹ An association between timing of birth in relation to winter virus peak and asthma development has also been noted.⁵³ Wu et al. found, in a large cohort from Tennessee, USA, that children born ~4 months before the winter RSV peak had a 29% increased risk of developing asthma at 4 to 5.5 years (defined using prescriptions and hospital diagnoses data) compared to those who were 1 at viral peak (adjusted OR 1.29, 95% CI 1.19 to 1.40).

1.7.3. Sex

Clear sex differences exist in the prevalence of respiratory illness in children. There are prenatal sex differences in lung maturation growth, with females showing earlier foetal breathing, increased surfactant production and lower air resistance than males before birth.^{154,222} Accordingly, males have higher rates of bronchiolitis, and asthma/wheeze in the first 10 to 15 years of life, following which females begin to have an increased risk of asthma compared to males, possibly due to pubertal changes.^{25,93} The risk of hospitalisation for bronchiolitis amongst infants was approximately 50% higher for boys compared to girls in England and Scotland between 2000 and 2012.⁸² At age 10.7, 13.9% of boys were reported to have a doctor diagnosis of asthma in the last year compared to 10.1% of girls in the Avon Longitudinal Study of Parent and Children.²⁵ Males continued to have excess asthma until 16.5 year of age when the prevalence in girls exceeded that of boys (12.6% versus 11.9%).

1.7.4. Genetic risk of bronchiolitis and asthma

Genetic background has been proposed as a contributing factor to explaining why only a proportion of infants, most with no underlying conditions, develop bronchiolitis when the majority of the population are infected with RSV by age 2.^{223,224} However, research findings on genetic risk factors of bronchiolitis is limited and no evidence has so far been produced from Genome-wide Association Studies (GWAS)—described as the preferred method for

gene discovery.²²³ Based on the results of twin heritability studies, it is through that asthma is a more heritable condition than bronchiolitis, with an estimated genetic contribution of about 50% compared to less than 20%.²²⁴ In addition, GWAS have been used to identify several genes that are associated with asthma.²²⁵ However, only a small proportion of asthma heritability has so far been explained by these genetic variants.²²⁵

Moreover, genetic factors cannot explain the sudden rise in the prevalence of asthma at the end of the 20th Century.²⁴ Interactions between genetics and environmental risk factors, alongside epigenetic exposures, have been proposed to explain differences in disease susceptibility and severity. Epigenetics describes modification of gene expression that can then be passed down to offspring, and offers advancement in understanding asthma susceptibility, although, as warned by Meyers et al. “caution must be exercised where the direction of causality is unknown, as epigenetic marks could be either a cause or consequence of disease”.²²⁵ As with non-genetic studies of asthma, research has been hindered by the use of one simple measure of asthma, rather than considering the range of phenotypes under this umbrella condition.^{225 p.4}

A prominent theory in the bronchiolitis-asthma causality debate is that there is a shared genetic predisposition to both conditions.²²⁶ A Danish population-based twin study demonstrated that genetic determinants of severe RSV infection overlap with asthma.²²⁷ Other studies have looked at asthma and atopy within families, and found that family history of asthma or atopic disease is associated with childhood acute lower respiratory tract infections, further indicating a genetic component across these diseases.¹⁴⁰ Candidate gene association studies have previously identified some immune response genes implicated in both bronchiolitis and asthma, although only a handful of common polymorphisms have been identified so far from GWASs.^{228–230}

1.8. Public health policy and practice

In this section, I will introduce current and planned English and Scottish public health policy and practice related to the prevention of bronchiolitis and asthma, with particular reference to attempts to lessen inequity in the rates of these two conditions. Notably, although England and Scotland both reside under the same sovereign state (the UK), health and social services fall under one of many devolved matters.²³¹ Since the Scotland Act 1998,²³² the NHS, funding, health regulation and public health, amongst other aspects of health care in Scotland, are the responsibility of the Scottish Parliament. PHE has been responsible for improving the population’s health and addressing inequalities in England since its inception in 2013.²³³ The

Scottish equivalent, PHS, was formed in April 2020, taking over the function of several pre-existing public health bodies.²³⁴

I write about the implications of the SARS-CoV-2 pandemic in my discussion chapter.

1.8.1. RSV infection and bronchiolitis

Infection control

Public guidance on RSV infection by PHE was published in 2008.²³⁵ This document gives brief information about transmission, at risk groups, symptoms and diagnosis. It has one preventative statement, 'transmission can be prevented through standard infection control practices such as hand washing'. A 2011 Cochrane review concluded that simple and low cost interventions such as handwashing were effective at reducing the spread of respiratory viruses, especially around younger children.²³⁶ However, there is limited research on the population impact of such campaigns.²³⁷ An evaluation of the national CleanYourHands campaign—a Department of Health funded campaign to improve hand hygiene in hospitals in England and Wales—showed a reduction in some infection rates, although a causal association cannot be assumed from the study design.²³⁸ Health Protection Scotland, the public health domain previously concerned with health protection in Scotland (before PHS was created), reports on seasonal respiratory pathogens without specifically referencing prevention.²³⁹

Palivizumab prophylaxis

PHE guidance on RSV infection mentions the lack of treatment available for RSV infection, although palivizumab is cited as a preventative therapy that is available for infants at high risk of infection.²³⁵ The risk factors that infer eligibility for palivizumab (chronic lung conditions or congenital heart disease, with preterm birth) are more common in low SEP groups,^{168,172} therefore we may expect this preventative treatment to reduce inequity in hospital admission rates to some degree. However, as only about 15% of infants admitted for bronchiolitis have one or more known predisposing risk factors, and a minority of these are eligible for palivizumab, this is unlikely to have a big impact on overall inequity.⁴⁵ Notably, there is no national record of access to palivizumab in England, making it difficult to assess the potential impact of this drug. The limited research points to a large variation between hospital trusts, and limited uptake of this preventative medication overall. Using a linked administrative dataset, Zylbersztejn and Hardelid estimated that less than 20% of eligible infants were

actually prescribed palivizumab across 43 acute hospital trusts in England between 2012 and 2016.²⁴⁰

Vaccination

RSV vaccines are cited as the most promising avenue for the prevention of bronchiolitis,¹⁵ with research and development for this vaccine is listed as a strategic priority for the WHO.²⁴¹ It is hoped that maternal immunisation against RSV will protect offspring during the first few months of life, although results from Novovax's phase 3 randomised controlled trial (PREPARE) in early 2019 did not meet this primary endpoint.^{242,243} Promisingly, GlaxoSmithKline begun a phase 3 study of another RSV maternal candidate vaccine in November 2020.⁴² Once a RSV vaccine is available, there is still the barrier of immunisation uptake, which has typically been lower among pregnant women compared to other populations. For example, since a national outbreak of pertussis (whooping cough) that led to 14 deaths in babies under 3 month olds in 2012, immunisation for pertussis has been offered to all pregnant women by the NHS.²⁴⁴ However, according to statistics published by Public Health England, around 70% of women took up this vaccine in 2018/19. The influenza vaccine, which has been recommended in this population since 2010, had an uptake of only 45% among pregnant women in England during the 2018-19 flu season.²⁴⁵ This is compared to an average coverage of 90.9% for the 9 routine childhood vaccinations in England over the same period.²⁴⁶

Childhood immunisation statistics point to further differences by level of deprivation. Figures published by the Scottish Public Health Observatory, for example, show that there is a high uptake for all vaccines routinely recommended for children by 24 months of age, but that rates follow a social gradient.²⁴⁷ At year-end 2019, children aged 24 months in the least deprived SIMD quintile had an average uptake of 95.8% across the four recommended vaccines compared to an average of 91.9% in the most deprived quintile. This suggests that whilst a RSV vaccination will significantly reduce overall bronchiolitis rates, it is unlikely to decrease inequity (and may even increase it) if current methods to encourage uptake are used.

Awareness of RSV and bronchiolitis is important when implementing preventive interventions, particularly given the reported decline in childhood vaccination uptake in recent years.²⁴⁶ A study of pregnant women at 4 hospital sites in Southern England found that, whilst only 14% reported never hearing of bronchiolitis, 71% reported never hearing of RSV.²⁴⁸ 36% of midwives, also surveyed in this study, had also never heard of RSV. Women who were more familiar with RSV said that they were slightly more likely to accept RSV vaccination if routinely recommended compared with women with no understanding of RSV. 'More than a cold' is a

campaign run by the biopharmaceutical company, AbbVie, in conjunction with several children's charities to raise awareness of bronchiolitis, especially amongst parents.²⁴⁹ The impact of this campaign is unclear.

1.8.2. Asthma

NHS England states that childhood asthma is the most common chronic condition amongst young people in the UK, and the number one reason for paediatric emergency hospital admissions in England.²⁵⁰ In 2011, the Department of Health set out an outcomes strategy for chronic obstructive pulmonary disease (COPD) and asthma in England.²⁵¹ This strategy states that "the causes of asthma are not well understood, so prevention of asthma is not currently possible" and the focus of recommendations are on accurate diagnosis, treatment and symptom control. The National review of asthma deaths, published by the Royal College of Physicians in 2014, similarly focused on poor control of asthma symptoms, inappropriate prescribing and patient monitoring.³³ The English Quality and Outcomes Framework, a voluntary incentive program for GP practices, includes several indicators intended to improve the care of patients with asthma.²⁵² Specific quality indicators include maintaining an asthma register, updating patients' symptoms and co-morbidities and offering support/treatment to current smokers. Whilst socioeconomic deprivation is associated with severe asthma symptoms and, therefore, clinical factors are likely an important focus in reducing inequity in outcomes of people with asthma, the focus of this PhD is on preventing asthma in the first place.⁶⁹

All our health is a framework of evidence that was published in 2015 by PHE, "to guide healthcare professionals in preventing illness, protecting health and promoting wellbeing".²⁵³ One section of this report is related specifically to respiratory disease. This document refers to inequality in avoidable mortality due to respiratory disease and the cost these conditions have on the health system and wider economy. In 2015, the website *Respiratory Futures* was set up by NHS England in collaboration with the British Thoracic Society to provide professionals with local area-level resources related to respiratory health.²⁵⁴ Smoking is cited as a preventative priority for almost all sustainability and transformation partnerships in England; a factor also frequently mentioned in the Department of Health's 2011 COPD and asthma outcomes strategy.²⁵¹ Other priorities highlighted in *Respiratory Futures* include integrated services, early diagnosis and reducing emergency admissions.²⁵⁴ Leicester, Leicestershire and Rutland Sustainability and Transformation Partnership are specifically acknowledged for their work in tackling the wider determinants of the health through improved transport, air quality and housing.

Good Places Better Health for Scotland's Children was part of the Scottish Government's 2008 strategy on health and environment.²⁵⁵ This strategy, which focussed on the role of creating health environments, considered asthma as one of four major health challenges facing children in Scotland. As part of this strategy a health evidence assessment was conducted on asthma in 2012, which reviewed the risk factors for asthma and potential actions to reduce the scale of the problem.²⁵⁶ Socioeconomic factors are cited in this assessment as an influence in the prevalence of asthma, with inadequate housing and smoking mentioned specifically in relation to deprivation. Two related actions emerged from expert workshops: "educational initiatives on smoking should acknowledge socio-economic, stress and social class contexts"; and "address socio-economic related smoking pressures". The outcome of this work is unclear, and no clear continuation of the strategy documented.

1.8.3. Looking forward

The NHS (England) Long Term Plan was first published on 7 January 2019.²⁵⁷ This report outlines the plan for the NHS in England over the next decade. It references respiratory conditions as an area where the NHS will be intensifying its focus, citing that lung conditions cost wider society £9.9 billion annually.²⁵⁸ Section 3.83 of the NHS Long Term Plan refers specifically to social deprivation and respiratory disease:

"Incidence and mortality rates for those with respiratory disease are higher in disadvantaged groups and areas of social deprivation, where there is often higher smoking incidence, exposure to higher levels of air pollution, poor housing conditions and exposure to occupational hazards."

The NHS Long Term Plan cites specific areas of change needed to reduce incidence of and improve outcomes for respiratory disease, including reductions in hospital stays due to these conditions, tackling risk factors that contribute to higher rates in disadvantaged groups, improved detection and diagnosis, and pulmonary rehabilitation. In addition, the 'best start in life' is one of the ten priorities listed in the PHE strategy 2020 to 2025.²⁵⁹ This report also notes that social disadvantage in health is persistent and in some cases, the gap is widening. Two particularly pertinent actions set out to prevent health inequalities are specialist smoking cessation services for pregnant women and commitment to lower air pollution.

NHS Health Scotland published a five-year Strategic Framework in 2017 entitled *A Fairer Healthier Scotland 2017-22*.²³⁴ This framework has a strong focus on prevention and reducing health inequalities across the board. Relevant particularly to bronchiolitis and asthma is the commitment to reduce rates of smoking, increase the proportion of infants with a healthy

birthweight and improve suitable housing options. Putting this strategy into action, the *NHS Health Scotland delivery plan for 2019-20* includes specific short term goals to improve Scottish health.²⁶⁰ These commitments include: progressing the Scottish Burden of Disease Study to identify risk factors for health inequality; supporting the implementation of the tobacco control action plan; and providing expert input into the development of strategies that aim to improve health and reduce inequalities for children, young people and families.

1.9. Chapter summary

To conceptualise the potential pathways to inequity in bronchiolitis and asthma rates in the UK, this chapter presents findings from the literature on the association between SEP, bronchiolitis and asthma using a social determinants of health framework. Through this framework, individual socioeconomic position is viewed as a structural determinant of health that, along with other structural factors including geographical location and the built environment, influences exposure to intermediary determinants of health. The intermediary risk factors discussed are focussed on the pre- and post-natal period and include tobacco smoke and air pollution exposure, premature birth, chronic conditions and housing conditions. Together with non-modifiable factors, such as age, sex and genetic risk, intermediary determinants ultimately modify the likelihood that a child will develop bronchiolitis and/or asthma. Current policy around bronchiolitis and asthma in England and Scotland is focussed, for the most part, on treating the symptoms of the conditions although future strategies place emphasis on preventative risk factors (particularly of reducing tobacco smoke exposure) and reducing socioeconomic inequity.

Chapter 2. Rationale, aims and thesis structure

Chapter overview

In this chapter, I summarise the key findings and future priorities that have emerged from the literature presented in Chapter 1. These form the justification for my work. I present the rationale for my PhD project, followed by my specific aims and objectives. I will finish by outlining the structure of this thesis.

2.1. Literature summary

2.1.1. Key findings from Chapter 1

- Bronchiolitis is the most common cause of hospital admission among infants and asthma is estimated to affect approximately 10% of children by their mid-teens in England and Scotland.^{18,261}
- In line with the seasonal patterns of RSV, which is the dominant pathogen that causes bronchiolitis infections, the burden of bronchiolitis on NHS hospitals is substantial in winter months.^{4,6}
- Asthma is estimated to cost the NHS around £1.1 billion per annum, including 85,000 hospital admissions and 108,000 GP consultations each year.³²
- It has been established that hospitalisation with RSV-related bronchiolitis in infancy is associated with an increased risk of developing wheeze and asthma in childhood, but there is ongoing debate about the nature of this association.^{49–53}
- There is a socioeconomic gradient in the incidence rate of hospitalisation for bronchiolitis and the prevalence of asthma in England and Scotland.^{4,82,93}
- Alongside socioeconomic factors, wider structural factors implicated in the prevalence of these respiratory conditions include geographical location and the associated climate, as well as the built environment, particularly in relation to air pollution.^{24,108,262}
- A multitude of socially patterned prenatal and perinatal risk factors have been shown to have an association with the development of bronchiolitis and asthma, including tobacco smoke exposure, congenital anomalies and prematurity.^{44,138,263}
- Other non-socially patterned risk factors for bronchiolitis and asthma include child's age, sex and, to a lesser extent, genetic variation.^{6,24,25,82,225,226}

- The health system in the UK is designated free and equal for residents and there are national clinical guidelines in place for the treatment of bronchiolitis and asthma;^{3,17,209} however, there are disparities in the uptake of care, particularly preventative treatments, and in service provision across the UK.^{44,69,211}
- Preventative treatment of bronchiolitis is primarily focussed on the development of a RSV vaccine, but this programme is some years away from implementation and it is not currently clear whether this will tackle inequity in bronchiolitis incidence.¹⁵
- In the past, asthma policy has been concentrated on treatment through improvements in symptom management and more streamlined and effective clinical pathways; however, more recent policy documents appear to be introducing a new focus on reducing inequalities and wider preventative action.^{251,257}
- The NHS Long Term Plan for England cites specific areas of change needed for managing and preventing respiratory disease, including reductions in hospital stays due to respiratory conditions and tackling risk factors that contribute to higher rates in disadvantaged groups.²⁵⁷
- NHS Scotland is similarly concerned with inequalities in health outcomes, particularly childhood asthma, and is taking active steps to understand risk factors for disease burden.²³⁴

2.1.2. Priorities driving my work

- Without a good understanding of the longitudinal pathways to bronchiolitis and asthma, it is difficult to design effective policies for prevention.
- There is a wealth of research identifying factors that are associated with an increased risk of bronchiolitis and asthma; however, a major limitation of the presented evidence lies with how risk factors and their interrelated pathways are considered in adjusted models.
- Methodological constraints have further hampered efforts to get unbiased causal estimates of disease risk.
- It is well established that early-RSV infection (and therefore bronchiolitis) is associated with an increased risk of asthma but the reasons for this remain unclear.
- There has been a lack of joined up work to understand the common or related pathways in the development of both of these conditions
- Furthermore, research on the link between RSV and asthma has not considered the common role of socioeconomic deprivation as a risk factor for both conditions.
- Investigating inequities in hospital admissions for bronchiolitis, in particular, will offer guidance on the planning of future RSV vaccines.

- Through the use of more sophisticated methodology that can disentangle these complex relationships, we can work out how and when to intervene to prevent either or both of the conditions.

2.2. Thesis rationale

Bronchiolitis and asthma are major causes of morbidity among children in the UK and are responsible for substantial burden on the national health care system.^{4,32} Despite this, there are currently no preventative or curative measures for the majority of children that develop bronchiolitis or asthma.^{15,47} There is a large body of evidence investigating environmental, social and biological risk factors for these conditions, but much of this work does not fully consider the ways in which various risk factors interact with each other (and potentially bias results). There is also a myriad of research looking specifically at the link between RSV bronchiolitis and subsequent asthma,^{53,56,264} but precise risk factors leading to both conditions remain elusive. What we do know is that a prevailing facet of both conditions is the inequity in prevalence across socioeconomic groups.^{4,29} Identifying the factors that may mediate or confound the pathway between socioeconomic deprivation, bronchiolitis and asthma will help to unravel this complex picture. This is a pertinent step to working out how and when to best intervene to prevent either or both of the conditions. By focussing on early life risk factors, the aim is to reduce the burden of bronchiolitis and asthma by closing the gap in the incidence/prevalence of these conditions between children born into the most and least deprived socioeconomic groups.

2.3. Thesis aims and objectives

The aim of this project is to describe associations and explore pathways between early life socioeconomic deprivation, infant bronchiolitis and asthma in mid-childhood using administrative health datasets. Overall, this PhD will inform where interventions are best placed to reduce the burden of childhood respiratory ill health by closing the gap between the most and least deprived socioeconomic groups.

Objective 1 To describe the association between seasonality, early life socioeconomic position and bronchiolitis admission rates among infants across England.

Objective 2 To investigate the mediating contribution of maternal smoking in the pathway between early life socioeconomic position and bronchiolitis admission rates in infancy.

Objective 3 To describe typical trajectories of asthma/wheeze among children <10 years in Scotland and their association with early life risk factors, including socioeconomic position.

Objective 4 To investigate the mediating contribution of bronchiolitis in the pathway between early life socioeconomic position and mid-childhood asthma/wheeze trajectory groups.

2.4. Thesis structure

In **Chapter 1**, I presented a conceptual framework based on the Commission on Social Determinants of Health (CSDH) that differentiates between the structural and intermediary determinants of ill health. Within this framework, I have presented the current evidence on risk factors relating specifically to bronchiolitis and asthma in a UK context. I have highlighted areas of policy and practice in England and Scotland that are relevant to preventing and reducing inequity in bronchiolitis and asthma rates.

In **Chapter 2**, I have outlined my rationale as well as the overarching aims and objectives of my PhD.

Chapter 3 introduces the administrative datasets and derived birth cohorts that I use in this thesis. I will outline the various roles I have undertaken as part of my PhD, including applying for research approval and cleaning datasets. Lastly, I will discuss the implications of patients opting out of sharing their data with application to maternity information and research.

In **Chapter 4**, I present the first of two studies that focus on the seasonality of bronchiolitis admissions in England. Using a birth cohort created from administrative hospital data, I explore spatial variation in the seasonality of hospital admissions for bronchiolitis and its association with local demographic characteristics.

Chapter 5 uses the same English longitudinal dataset as in Chapter 4 to quantify the extent to which seasonality of bronchiolitis admissions is modified by socioeconomic position at the

individual level. For a wider understanding of the burden of bronchiolitis on the English NHS, I also calculate the proportion of all infant hospital admissions in England attributable to bronchiolitis by year, month and socioeconomic group.

Using a causal inference framework, **Chapter 6** tackles objective 2. I use a birth cohort constructed from Scottish administrative data to quantify the mediating effect of maternal smoking during pregnancy on the pathway between socioeconomic deprivation and hospital admission for bronchiolitis. To guide the selection of confounders in my model, I apply an evidence synthesis for constructing directed acyclic graphs.

In **Chapter 7**, I model trajectories of asthma/wheeze symptoms among children aged between 2 and 10 in Scotland and describe their association with early-life risk factors. I then use these trajectory groups as the outcome in mediation analysis to explore the mediating contribution of bronchiolitis in the pathway between early life SEP and mid-childhood asthma/wheeze.

Chapter 8 pulls all the work in this thesis together into a summary chapter. I highlight the main findings from my research and discuss the implications of the work, including relevance to public health policy and practice.

Chapter 3. Introduction to thesis datasets

Recap and chapter overview

Chapters 1 and 2 covered an overview of previous literature in the area, the conceptual framework that guides this thesis, current UK policy relating to respiratory health, and the overarching rationale, aims and objectives for this research. This chapter will introduce the data sources used to address my objectives. I will begin by defining administrative data, including the benefits and drawbacks of these types of datasets, before introducing the specific data sources used in my thesis. Next, I focus on an emerging challenge for research that utilises administrative data—patients opting out of sharing their data for research purposes. Throughout this chapter, I will also outline the various research roles I have undertaken as part of my PhD.

3.1. Administrative data

The data used in this thesis comes from administrative datasets, defined here as data that have been generated about citizens through provision and administration of public services.^{265,266} These datasets tend to cover whole populations, since they include all individuals coming in contact with the particular service. They offer information on a large number of people, which is particularly useful for research that compares risk between subgroups or looks at rare events or complex pathways.²⁶⁷ In some instances, multiple datasets can be linked together, enhancing existing information further. The costs associated with administrative data are typically much lower than research involving de novo data collection, particularly at the same scale, and the time usually spent on data collection is saved.²⁶⁶ Administrative data may offer a better degree of accuracy than other retrospective designs as the information is normally recorded at or soon after the time of an event. Given that the primary function of these datasets is administration and provision of often universal public services, the risk of non-response bias and loss to follow-up are also minimised.

On the other hand, as these datasets are not designed for the purpose of research, they also present challenges for researchers.²⁶⁶ Unlike specifically designed cohort studies, researchers generally have no input into data collection and coding practices, and any changes to these over time. Since individuals whose information is contained in administrative databases have not consented to take part in research, a cost and time penalty is associated with applying for

access to these data for research purposes.²⁶⁷ This is especially the case where individual-level data are required due to concerns about meeting legal requirements for use, confidentiality and avoiding individual disclosure. The datasets are often large, complex and received by researchers in a ‘messy’ format, requiring substantial data cleaning before use.²⁶⁸ Data quality may be affected by errors in data collection and, when data linkage is used, when records cannot be linked or are incorrectly linked. These errors can bias the prevalence and distributions of attributes in the dataset, as well as any associational analyses one might perform on the data.²⁶⁷

Nonetheless, administrative records provide great opportunities for epidemiological analysis if limitations are recognised and addressed. For this PhD, I used administrative health data from England and Scotland, which I will now describe in further detail.

3.2. English Hospital Episode Statistics linked to mortality records

3.2.1. Overview

Hospital Episode Statistics Admitted Patient Care (HES APC) is a database of all inpatient admission records, including day cases, to NHS hospitals in England, and all episodes of care provided in independent hospitals but paid for by the NHS.^{269,270} The data available in HES APC include: clinical information (operations and diagnoses); patient information such as sex, ethnicity and age; administrative information including the dates of admission and discharge; and geographical identifiers such as the Lower Super Output Area (LSOA) of the patient’s address.²⁷¹ NHS Digital—the supplier of HES APC data—routinely link admissions to the same patient over time based on NHS number, date of birth, sex, hospital number and postcode.²⁷² For secondary uses of the data, pseudonymisation is applied. That is, a unique identifier called the HESID is assigned to each individual and all personally identifying information removed from the dataset (i.e. NHS numbers and full postcodes).²⁷³ HESIDs are unique to each data extract and are available from the financial year 1997/98 onwards.

Births are recorded as hospital admissions in the English NHS, and HES APC captures approximately 97% of births in England that occur in NHS hospitals.²⁷⁴ There are two separate sources of information on each birth, the delivery episode (maternal record) and the birth episode (baby record), which both contain further details in addition to the main HES APC record. These further details are referred to as the “maternity tail” or “baby tail” and includes information such as birthweight, gestational age, delivery method and birth order (for multiple births).²⁷⁴ Although mothers and baby pairs are not already linked by NHS Digital when

researchers receive HES APC data, it is possible to apply linkage in the pseudonymised data using a mixture of deterministic and probabilistic methods.²⁷³ Deterministic linkage is where records are linked based on the presence of a set of identical values (e.g. date of birth/delivery, sex, hospital and postcode district).²⁷⁵ Probabilistic linkage assigns match weights to all possible record pairs based on the probability that a set of variables agree or disagree.^{275,276} Weights above a certain threshold are then selected as matches. Probabilistic methods are essential for linking mother and baby records in HES APC because errors and missing values in the dataset have been shown to lead to a high proportion of missed matches based on deterministic methods alone.²⁷³

Linking mother and baby records enhances the completeness of the dataset as it is then possible to supplement missing data from one source (e.g. the birth record) with the other (e.g. the delivery record). HES APC allows patients to be tracked via longitudinal hospital records using their HESID. HES APC is routinely linked to the Office for National Statistics (ONS) mortality records, so that the dates and causes of deaths outside NHS hospital settings for people who have a record in HES APC are also available.²⁷⁴ Alongside the enhancements available through data linkage, the quality of data in HES APC has also improved over time. Burns et al. conducted a systematic review of studies that have assessed the accuracy of hospital episode data across England (56% of the papers), Scotland and Wales.²⁷⁷ When compared to case or operative notes, the accuracy of primary diagnosis was found to have risen from 73.8% to 96.0% between 2002 and 2009. Since 2004/05, the dataset serves as the basis for a pay-for-performance system for NHS secondary care, which has led to increased reporting of hospital activity and depth of clinical coding.²⁷⁸ However, for HES APC returns that are not mandated, such as gestational age and birthweight, there remains a high proportion of missing data.²⁷⁹ Further, consistency of information can vary across and within hospital trusts. For example, 11 of 139 NHS trusts in 2009/10 had relatively poor internal consistency on recording of delivery method.²⁸⁰

Another challenge of using HES APC data for research is the proportion of people choosing to opt out of sharing their data for purposes beyond their own care.²⁷² Where opt-outs are applied, no information is received on these patients, which introduces an unknown amount of bias into the dataset and affects the quality of data linkage.²⁸¹ In addition, HES APC does not capture private or charitable hospital activity that is not funded by the NHS.²⁷¹ There is no requirement for these hospitals to centrally record activity and, as such, the precise volume and type of care falling under this remit is not freely available. However, since emergency treatment is not typically covered under private medical insurance in the UK and NHS care is free at the point of need, it is expected that very few children are admitted privately for acute

illnesses like bronchiolitis.²⁸² Therefore, I do not expect this facet to impact on the reliability of findings in this PhD.

3.2.2. Using HES APC in my PhD

Approvals

Approval to use HES APC data was granted by NHS Digital (formally the Health and Social Care Information Centre) for a programme of research within our research unit within the terms of a data sharing agreement, DARS-NIC-393510-D6H1D-v3.2. This agreement is for the use of a de-identified extract of HES APC linked to ONS mortality records for specified objectives. My PhD sits under the second of the three objectives of this work, which is “to determine the risk factors for emergency and recurrent use of secondary care”. As my work fell under current purposes for use, I did not need specific permission from NHS Digital to use these data for my PhD. In addition, as I was using pseudonymised data, my research did not require an ethics review by an NHS Research Ethics Committee.²⁸³

HES APC birth cohort

When I began my PhD, a birth cohort had already been created from linked HES APC and ONS mortality records by Ania Zylbersztejn.²⁸⁴ Table 3.1 gives further details of the datasets that comprise the birth cohort, with examples of variables relevant to my PhD. The HES APC birth cohort included linked mother-baby data from 2003, following the introduction of a programme to provide NHS numbers to babies at birth on 29th October 2002, to 2016. Non-English residents (identified using postcode information) were excluded from the dataset. Multiple births were also excluded because of the high possibility of assignment of the same HESID in same sex individuals from multiple births, thereby creating false matches. This is because the algorithm that creates HESIDs relies on sex, postcode and date of birth in situations where an individual’s NHS number is missing.²⁸⁵

Table 3.1. Description of datasets comprising the English birth cohort used in my thesis^{271,286}

<u>Dataset</u>	<u>Description</u>	<u>Example variables</u>
HES APC	A dataset containing all admissions to NHS-funded hospitals in England. An admission is defined as an activity that requires a hospital bed, including both	HESID, sex, age at admission, admission date, discharge date, diagnoses (up to 20), LSOA of patient address, Index of Multiple

	emergency and planned admissions, day cases, births and deliveries.	Deprivation (based on patient's postcode)
	In addition to general admission details, a tail with 19 variables relevant to the delivery and labour are included in birth (baby) and delivery (mother) records.	Birthweight, gestational age
ONS mortality records	All deaths registered in England among residents and non-residents.	All causes of death (up to 15), date of death

I used cohort data for births from 2011 onwards in my analyses and, when I looked at descriptive data for these years, I noticed a trend towards lower mother-baby linkage rates in the most recent study years, from 96.4% in 2012/13 to 92.0% in 2016/17. This was despite an increase in rates of deterministic linkage (i.e. unique links on several characteristics) between mothers and babies following improvements in hospital recording of data.²⁷⁹ Linkage is particularly important for the geospatial bronchiolitis work in this PhD, as geographical identifiers are more complete in the mother's data. I therefore began investigating the potential source of this unexpected pattern.

Updating the mother-baby linkage

I firstly looked at the methodology used to probabilistically link mothers and babies, which is detailed in Harron et al.²⁷³ This involves calculating match weights for each record pair based on the probability that variables across two records randomly match (U-probabilities) and the probability that there is agreement on a variable given that two records are a true match (M-probabilities). The match weights had been calculated using U- and M-probabilities derived from an earlier extract of the data (≤ 2012) and may not have been as accurate for the newer more complete data. This led me to redo the linkage for HES years 2013/14 to 2016/17 with the aim of improving linkage rates. The full methodology is outlined in Appendix 3.1. However, on completion of this update I still found increasing rates of unlinked babies, as shown in the 'updated linkage' column of Table 3.2. Looking again at the raw number of mothers and babies, I identified that the number of delivery records compared to birth records diverged significantly from 2014/15 ('extra birth records' column, Table 3.2). Even if linkage methods were perfect, 3.3% of babies born in 2016/17 did not have a delivery record available for linkage.

Table 3.2. Number of singleton birth records, proportion unlinked to delivery records in original and updated linkage and available delivery records, by financial year

Year of birth	Birth records (N)	Initial linkage (%)	Updated linkage (%)	Delivery records (N)	Extra birth records* (%)
2012/13	645,937	96.4	96.4	641,080	0.8
2013/14	626,161	96.0	96.4	623,863	0.4
2014/15	613,401	92.4	94.0	596,916	2.8
2015/16	625,065	92.1	93.3	605,643	3.2
2016/17	617,311	92.0	93.6	597,127	3.3

*Compared to the number of delivery records, calculated as (birth records - delivery records) / delivery records.

Impact of patient opt-outs

This led to my second identified issue with the dataset: patient opt-outs. Since January 2014, patients of the NHS in England have been able to opt out of sharing their patient data for research and planning purposes (previously known as a type 2 opt-out). Accordingly, NHS Digital have applied opt-outs to our HES APC data extract before data release for years 2014/15 onwards. In 2017, when we received the data used in this thesis, the average rate of type 2 opt-outs across England was 2.3%.²⁸⁷ This is the average proportion of patients we would expect *not* to appear in our data extract. We do not have details about people who opt out, meaning that it is unclear if pregnant women opt out at the same rate as the general population. However, it is likely that opting out accounts for at least some of the pattern shown in Table 3.2. I discuss the wider impact of patient opt-outs in Section 3.4.

3.3. Scottish administrative health databases

3.3.1. Overview

Health data are organised differently by the NHS in Scotland. Instead of one database with all admitted care, birth records, maternity inpatient cases and other hospital admissions are collected in separate datasets. Patients can be linked across health datasets and over time using their Community Health Index (CHI) number. The CHI number is a unique identification number, equivalent to NHS numbers in England and Wales, that is assigned when a patient first registers with NHS Scotland or at birth.²⁸⁸ There are several Scottish administrative health databases pertinent to this thesis, which I outline in Table 3.3, below. To enable researchers to connect records belonging to the same patient across these datasets, the Electronic Data Research and Innovation Service (eDRIS) team (part of the Information Services Division of

the NHS National Services Scotland)^a assigns a unique, pseudonymised identifier to each individual. For child and maternal datasets, eDRIS provides pre-linked infant and maternal records, with the linkage derived using a combination of deterministic and probabilistic methods. Mother and baby linkage rates reached 98.1% by 2017 in our data extract, emphasising the value of this dataset and the long history of the use of CHI numbers on health records in Scotland.

As noted by Pavis and Morris, “Scotland has some of the best administrative and care data in the world”.²⁶⁵ Electronic national data from the NHS in Scotland is available for all births from 1981 and, in 2015, less than 2% of Scottish residents used private healthcare.²⁶⁵ Pavis and Morris write that public engagement has been key in the Scottish model of administrative health research.²⁶⁵ In response to public attitude and legislative requirements, research datasets are bespoke and project-specific rather than from one permanently linked dataset. Projects are vetted at the local, regional, and national level to ensure there is clear public benefit. Unlike in England, there is no digitalised opt-out service for patients of the Scottish NHS. Instead patients are instructed to opt out by sending “a written request supported by adequate proof of identity” to National Services Scotland—the national NHS Board that provides national strategic support to NHS Scotland.²⁸⁹

Vital to my work on asthma, there is a national individual-level dataset of NHS prescriptions in Scotland that can be linked to other health datasets (see PIS dataset, Table 3.3).²⁹⁰ This information can be used to identify children who require medication to manage asthma, and was not available for England at the time of data request. In 2009 the quality of the prescription data improved drastically, with reimbursed data capturing 87% of patient identifiers, enabling the majority of items to be linked across other ISD datasets.²⁹⁰ This figure rose to 96% of reimbursed prescriptions and almost 100% of prescribed and dispensed items by 2014.

^a**A note on the changing structure of NHS Scotland.** In 2020, National Services Scotland (NSS) began transitioning to Public Health Scotland (PHS). ISD Scotland, which used to sit under NSS, is now part of PHS and has been renamed “data and intelligence”. I use the former names of these teams/departments throughout this section for consistency.

Table 3.3. Description of Scottish health datasets

<u>Dataset</u>	<u>Description</u>
NRS live birth, stillbirth and death registration records	National Records of Scotland (NRS) is the department of the Scottish government that is responsible for vital event registration, the census and national archives. ²⁹¹ Before April 2011 these records were held at the General Register Office for Scotland. ²⁹² We received birth, death, and stillbirth registrations in separate files.
CHI registry (emigration records)	CHI status (no link to CHI, emigrants or untraced, death or currently on CHI) and date of emigration are recorded in a CHI register, which is held by NHS Scotland. ²⁹³ This information is derived from GP records.
SBR and SMR11 birth records	Scottish Birth Records (SBR) are a singular electronic record capturing stillbirths and neonatal care received by babies in Scotland, including home births. ²⁹⁴ It does not include diagnostic information. This dataset replaced the old birth record dataset (SMR11) in 2002/03.
SMR02 delivery records	SMR02 is submitted to ISD Scotland by maternity hospitals and contains inpatient and day case activity relating to obstetrics. ²⁹⁵ Information about the mother, deliveries and new-born babies is available in this dataset.
SMR01 hospital admissions	SMR01 includes all inpatients and day cases discharged from hospitals in Scotland (apart from obstetric wards, SMR02, and psychiatric wards). The records include diagnostic, procedure and operation information, as well as non-clinic information such as location and length of stay.
PIS	The Scottish National Prescribing Information System (PIS) is a dataset of NHS medicines prescribed and dispensed within the community setting in Scotland. ²⁹⁰ Prescriptions dispensed within hospitals are not included. Information includes date and place of prescription and dispensing activity, and drug information including British National Formulary (BNF) codes.

3.3.2. Using Scottish Administrative Health Databases in my thesis

Approvals

Applying for Scottish healthcare data involved submitting a Public Benefit and Privacy Panel for Health and Social Care application form to National Services Scotland (now under Public Health Scotland). My primary supervisor, Pia Hardelid, led this process with input on the content from myself and another colleague. I also contributed to this process by designing a questionnaire to seek the perspectives of patient interest groups specifically on the topic of linking data across databases. We sent this questionnaire to the Great Ormond Street Hospital parent carer advisory group and a stakeholder who replied to an advertisement in Asthma UK's monthly bulletin (the questionnaire is included in Appendix 3.2). The responses revealed a primarily positive view of data linkage for the purposes outlined but raised questions about the lack of wider dissemination of such methods. We received full Public Benefit and Privacy Panel for Health and Social Care approval on 21st November 2018, and the full data extract on 11th December 2019.

Updating the Scottish birth cohort

A birth cohort, including births from 1981 to 2015, had previously been created from Scottish health data by a colleague, Maximiliane Verfürden.^{296,297} I adapted the code created by Maximiliane and other colleagues to clean the new files and create an updated birth cohort including births from 1981 to 2017. I used the NRS birth registration data as the core birth file and added information from other datasets to build the birth cohort as shown in Figure 3.1. In addition, I cleaned and structured the PIS dataset (which had not previously been held by our team), so that information about medication can be easily linked with mothers and children in the main cohort via their mother-ID or child-ID, when needed.

There were 2,203,042 NRS birth registrations in total over the 36-year period. I excluded 12,951 (0.6%) records with an unknown child-ID in the NRS birth registration dataset, as these records cannot be linked across datasets. The number of unknown child-IDs decreased with year of birth and the vast majority (84%) were from births before the year 2000. I excluded 33 stillbirths, identified using the "outcome of pregnancy" variable in SMR02, and 332 infants registered with a non-Scottish address at birth, identified through postcode area. Lastly, 538 duplicate records were excluded from the cohort. I changed improbable values to missing where: maternal age outside of 10 and 60 years ($n=14$); gestational age was outside of 22 and 45 weeks ($n=178$); birthweight was less than 100 grams ($n=100$); and date of death occurred before date of birth ($n=14$). The resultant whole country birth cohort of 2,189,188

infants covered 99.1% of all births in Scotland over this period when compared with official vital event statistics.²⁹⁸

Master file

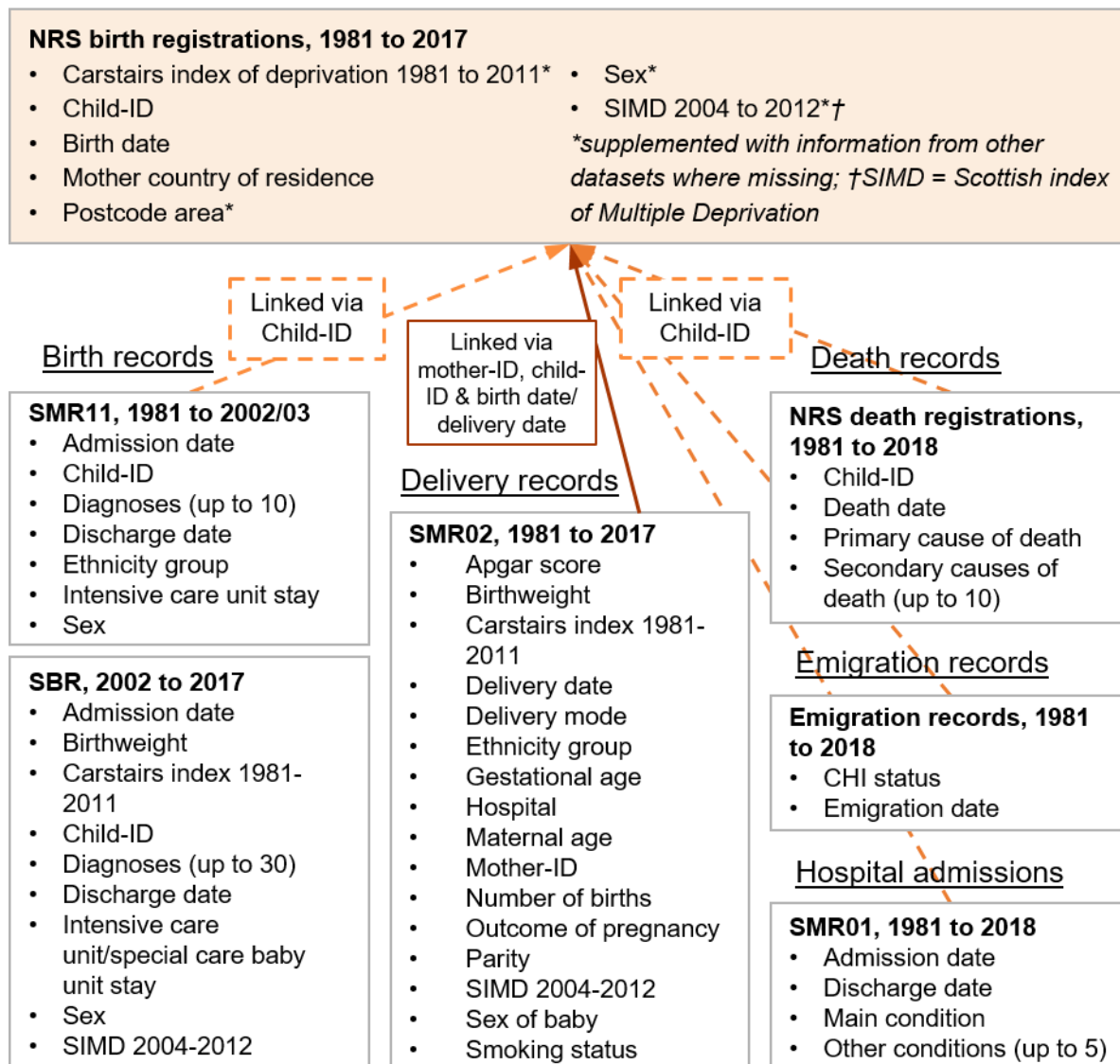


Figure 3.1. Scottish health datasets and relevant variables used to create birth cohort

3.4. National data opt-out programme: consequences for maternity statistics in England

Whilst exploring issues in the HES APC birth cohort, I became interested in understanding the wider potential consequences of patient opt-outs. I noted that the existing discussion of the English NHS patient data opt-out programme mostly neglected a positive view of why sharing data may be for the benefit of population health. I therefore decided to write a commentary article exploring the impact of NHS England patient opt-outs on maternity service statistics. I sought feedback on an early draft of the paper from members of a patient and public working group based at the UCL Institute of Health Informatics (formerly the Farr Institute, see Appendix 3.3). Ten members of the group gave comments that helped to shape the arguments below. The full article, which was published in the *International Journal of Population Data Science*, is available in Appendix 3.4.

3.4.1. Background

NHS Digital, the data capture organisation within the English NHS, collects and stores some of the information recorded when individuals receive health or adult social care in England. The data are used for the purpose of improving both individual care, through improved record sharing across health providers, and the wider population's health and care, through activities such as planning services, monitoring epidemics and researching health conditions and treatments.²⁷² As part of the population-based work, NHS Digital oversees the strictly controlled release of patient information to third parties including NHS providers and commissioners, university researchers, charities and companies that are partnered with the NHS. These organisations go through a lengthy application process to gain access to the data and follow stringent protocols when storing and analysing the data. Further, personal identifiers, such as names and NHS numbers, are removed in all circumstances apart from where specific patient consent is given or where required by law.²⁷²

The quality and completeness of English administrative health data, and the quantity of health research based on these data, have been increasing over time.²⁹⁹ However, a new barrier to the quality of administrative data is emerging in the form of opt outs. That is, patients choosing not to share their information beyond the NHS for anything other than their direct care and treatment. Opting out of NHS data sharing was first made available for patients of the English NHS in January 2014 following recommendations in Dame Fiona Caldicott's 2013 information governance review.^{300,301} In May 2018, a new online consent platform called the National Data Opt-Out service was launched by the NHS. This platform was intended to be simpler and

easier to access than the previous model; however, based on low numbers of uptake four months after its launch, it has been argued that not enough patients are aware of the scheme.³⁰²

As at 1 December 2018, the average national data opt out rate across England was 2.8%.³⁰³ This reasonably low proportion of missing records would be less concerning if the characteristics of those opting out were the same as those not opting out (i.e. missing at random). However, top-level demographic information published by NHS Digital shows that opting out is not uniformly distributed by age, sex or geographical area.³⁰³ Figure 3.2 presents opt out rates by Clinical Commissioning Group (CCG), showing clear variation across the country. Notably, twelve CCGs had opt out rates higher than 5% and one CCG had a rate of 10.1%. At the GP practice level, there were instances where the entire patient population have opted out, which raises significant questions about whether the patients in these practices explicitly opted out for themselves.^{301,303}

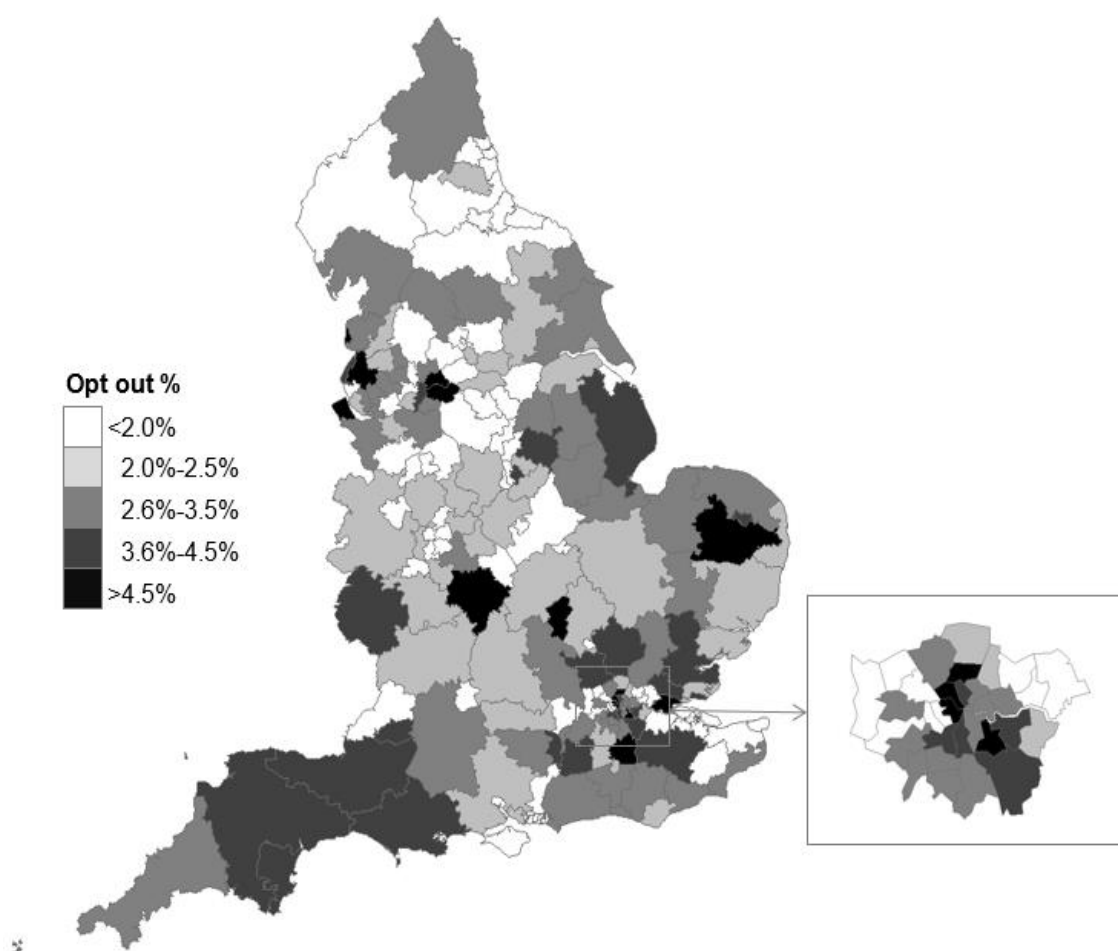


Figure 3.2. Rates of patients opting out, by CCG: England (with inlay map of London CCGs), as at December 2018. Data from NHS Digital.⁴⁰

The 2016 Caldicott review (the third in the series) stated that ‘patients have a right under the NHS Constitution to request that their personal confidential information is not used beyond their direct care’.³⁰⁴ I am not disagreeing with this principle. Legitimate concerns about the security of patient data have been voiced by both the general public and experts in the medical field, particularly in response to the contentious (and discontinued) care.data scheme.^{305,306} However, in parallel with information about this choice, patients should have complete and transparent information about the uses and potential advantages of sharing their information. Using a simple simulation based on publicly available data, the aim of this research is to exemplify the potential impacts of patient opt outs on maternity information and research.

3.4.2. Application to NHS maternity statistics

As introduced in Section 3.2, HES APC is the most comprehensive source of information on all births and deliveries in England.²⁷⁴ Delivery records from this dataset are used for many purposes, including to create resources for parents-to-be, to evaluate and improve maternity care provision and to investigate multiple risk factors for ill-health in mothers and babies.³⁰⁷ Tools, such as the Birth Choices tool from Which?,³⁰⁸ have been created utilising this data to aid expectant parents when making maternity choices. The National Maternity and Perinatal Audit, which was set up in 2016 to evaluate quality in NHS maternity services, uses maternity data from HES linked with data from each maternity unit to provide a range of statistics comparing outcomes at maternity unit level.³⁰⁹ In addition, HES APC has been used for maternal and child health research to examine, for example, factors explaining excess child mortality in England and the safety of surgical procedures during pregnancy.^{284,310} Individual level data are necessary for this type of research, since this allows multiple risk factors for maternal and child outcomes to be taken into account.

To exemplify the impact of biased data on maternal information, I calculated the potential rates of two health indicators had opt outs not been applied to HES delivery data. I downloaded publicly available CCG-level information from Public Health England’s Fingertips Child and Maternal Health Profiles for one common and one rare outcome: the proportion of deliveries with caesarean sections which occurs in approximately 27% of births nationwide; and births with very low birth weight (<1500g), which has a nationwide prevalence of 1.2%.³¹¹ I chose three CCGs for the example, each with a different rate of opt outs as published at 1st December 2018 by NHS Digital:³¹² Bradford City with a 0.3% opt out rate; Merton with a 2.8% opt out rate; and Oldham with a 10.1% opt out rate. Under the assumption that the women in the maternity dataset opted out at the same rate as the whole CCG population, I modelled three scenarios based on the rate of events in the women that opted out: 1) no events; 2) events occurring at the same rate as women who did not opt out; 3) all had an event. Mirroring

methods used by Public Health England,³¹³ 95% confidence intervals were calculated in Stata 15.0³¹⁴ using the Wilson Score method. Microsoft Excel 2013 was used to create graphs.

Table 3.4 displays published and simulated event rates for caesarean sections and infants born with very low birth weight, by CCG. Where patients opting out have the same rate of events as patients not opting out (scenario 2, Table 3.4), published and simulated event rates do not differ. However, the range of potential rates between the extremes of scenario 1 (no new events among women opting out) and scenario 3 (all women opt out had an event) shown in Figure 3.3 highlights the huge potential for misleading information where the frequency of events differs between patients that do and do not opt out. These scenarios show that both the CCGs with average (Merton) and high (Oldham) rates of opting out could be presenting misleading information. For births with very low birth weight (the rarer event), the rate of births with this outcome could potentially be 9-fold higher than the published rate in Oldham. The non-random nature of opt-outs, therefore, has potentially large implications for the outcomes of public health information and monitoring. At an NHS trust level, information bias could produce flawed outputs in audits intending to suggest improvements and highlighting good practice in maternal care and infant outcomes. These biases will also affect the reliability of findings on maternal and children's health research. As shown in this simple simulation, this is particularly the case where less common (but often more serious) outcomes are studied.

Problematically, the bias introduced into datasets with opt-outs applied cannot be treated with the same statistical methods used to treat missing data. Multiple imputation, a method which is commonly applied to deal with missing data, relies on the assumption that the missing information can be explained by differences in the observed data.³¹⁵ However, once patients have opted out, their data is completely removed from the dataset (i.e. complete case removal) meaning that researchers cannot account for systematic differences. Using multiple imputation in these instances may actually add further bias to results.³¹⁵ Methods used in population-based surveys to overcome biases of non-consent, such as weighting adjustments or simulation studies ideally require detailed data on the population (e.g. by gender, age, deprivation level and local area) who have opted out so that correct weights can be derived—information currently not published by NHS Digital. Research to determine whether those who have opted out of sharing data are different in terms of socio-demographic and health characteristics, in specific populations such as expectant mothers, would help in applying such methods to tackle data missing due to opt outs.

Table 3.4. Published and simulated number of events, patients and event rates, by indicator and CCG

Indicator	CCG	Published maternity data ³¹¹			Opt out rate (%) ³¹²	New event rate (95% CI)*		
		Events	Deliveries	Event rate (95% CI)		Scenario 1	Scenario 2	Scenario 3
Very low birth weight	Bradford City	36	1,657	2.17 (1.57, 2.99)		2.17 (1.57, 2.99)	2.17 (1.57, 2.99)	2.47 (1.82, 3.33)
	Merton	39	3,219	1.21 (0.89, 1.65)	0.30	1.18 (0.86, 1.61)	1.21 (0.89, 1.64)	3.93 (3.32, 4.64)
	Oldham	39	3,195	1.22 (0.89, 1.66)	2.74	1.10 (0.80, 1.50)	1.20 (0.90, 1.63)	11.20 (10.2, 12.3)
Caesarean section	Bradford City	478	2,024	23.6 (21.8, 25.5)	10.08	23.5 (21.8, 25.4)	23.6 (21.8, 25.5)	23.8 (22.0, 25.7)
	Merton	948	3,265	29.0 (27.5, 30.6)	0.30	28.2 (26.7, 30.0)	29.0 (27.5, 30.6)	31.0 (29.4, 32.6)
	Oldham	792	2,797	28.3 (26.7, 30.0)	2.74	25.5 (24.0, 27.0)	28.3 (26.8, 30.0)	35.6 (34.0, 37.3)

*Scenario 1 (no new events) = published events/(deliveries + extra deliveries) where extra deliveries= published deliveries/(1-opt out rate)- published deliveries; scenario 2 (average events) = ((extra deliveries*event rate)+published events)/(deliveries + extra deliveries); scenario 3 (max. events) = (events+ extra deliveries)/(deliveries + extra deliveries

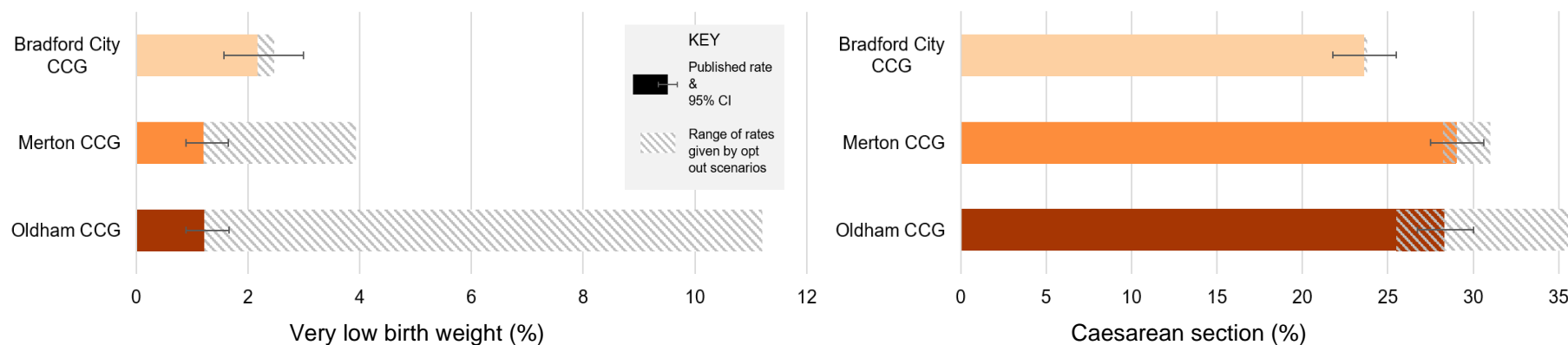


Figure 3.3. Observed and simulated rates of births with very low birth weight (<1500g, 2016/17) and deliveries by caesarean section (2016): by CCG

3.4.3. Are opt outs inevitable?

Public health research in the Nordic countries demonstrates that data sharing can be achieved with buy-in from citizens and at great value to clinical research.³¹⁶ In these countries, residents are assigned a personal identity number from birth, which can be used to track individuals across time and generations. Unlike the UK equivalent (NHS number), which is used only in health and social care settings,³¹⁷ personal identity numbers are used across a multitude of sectors. This means that individual data can be linked across health, education and social security datasets, for example, providing high quality comprehensive information on risk factors for ill-health and other outcomes. The safeguards in place are very similar to those in England. In contrast, however, there is broad public general awareness and acceptance of the use of individual data in research and a long-standing culture of trust in public services and data donation for the good of the population.^{316,318}

There is some evidence from England to suggest that greater knowledge of research processes and safeguards improves the likelihood of acceptance of electronic health records for research being used without explicit consent.³¹⁹ An electronic real-time dataset integrating primary and secondary care was successfully implemented over a decade ago in Salford, Manchester. All patients were sent a letter with information about the project and whether they would like to opt out. Reportedly, less than 0.2% of the nearly quarter of a million patients chose to opt out.³²⁰ In contrast, information about the now withdrawn care.data scheme to integrate primary and secondary care records was disseminated by generic leaflets, which were reportedly not seen by the majority of the population.³⁰⁵ Some additional strategies were used to provide information about the national opt out scheme to the public. A national radio campaign ran for 6-weeks after the launch of the new system and NHS Digital's website now links to *Understanding Patient Data*, a website providing "objective information about how patient data is used" that is run by Wellcome Trust.³²¹ However, given that some of the opt-outs may have been driven by GP practice rather than patient-level decisions,³⁰¹ it is not clear to what extent whether it was the NHS information campaign directed at patients (or indeed the absence of it) that led to the 2.8% opt-out rate, or a lack of buy-in from clinicians.

What is known is that when patients choose to opt out of sharing data beyond their direct care, the reliability of service information and evaluation and wider research based on electronic health records is diminished. NHS Digital is beginning to advertise the benefits of sharing NHS data for research and planning purposes, but there are ways this could be improved. NHS Digital could start by listing examples of how health care data have been used for research, as is available in other NHS held datasets (e.g. CPRD).³²² NHS Digital currently publishes a data release register,³²³ but not a data output register. Other examples of dissemination

include the University of Manchester's citizen's jury on health records and an animation created as part of the *#datasaveslives* campaign by the Farr Institute.^{324,325} However, the impact and wider reach of these schemes are not clear and, arguably, information about the benefit of data sharing can only go so far in raising public confidence.

Evidence from the research literature and reflected in media coverage suggests that there is unease about the potential of commercial entities, such as pharmaceutical and insurance companies, making profit from NHS data. This is in contrast to the largely positive view of university researchers or NHS staff making use of this data.^{320,326} Clear examples of 'acceptable and unacceptable purposes' for which data can and cannot be used, as called for by the Wellcome Trust could help to reassure the public.³²¹ Further, clinicians play a pivotal role in the discourse of patient consent to use NHS data for research. Research to better understand these concerns and how they could be addressed could be an important step to improving information campaigns.

Overall, there needs to be greater transparency through clear and detailed information on: who can apply to use NHS data and for what reasons; the safeguards in place to protect individual information; and, importantly, the wider consequences of opting out on population health research and public health service information. Only with this information can individuals be expected to make an informed decision about opting out of sharing their data.

3.5. Chapter summary

In this chapter, I have outlined the composition of the English and Scottish birth cohorts that will be used to conduct research throughout this thesis. The English birth cohort is a pre-existing dataset derived from HES APC data linked to ONS mortality records. I contributed to this dataset by updating the linkage between maternal and baby records, which led to an increase in the data completeness of the cohort. The Scottish birth cohort, which I updated from an older version of the dataset, is composed of several health databases including national birth and death registrations, hospital admissions and delivery records. The final part of this chapter focuses on the emergent issue of data opt outs and the potential implication of this on maternity research and information.

Chapter 4. Geospatial and seasonal variation of bronchiolitis in England

Recap and chapter overview

Chapters 1 to 3 presented an outline of the research literature and an introduction to the datasets in this PhD. This chapter is the first of two that focus on the seasonality of bronchiolitis admissions. Using a birth cohort created from the English hospital data introduced in Chapter 3, I explore spatial variation in the seasonality of hospital admissions for bronchiolitis and its association with local demographic characteristics. This is the first step in answering thesis objective 1:

To describe the association between seasonality, early life socioeconomic position and bronchiolitis admission rates among infants across England.

An article based on the contents of this chapter was published in Thorax (see Appendix 4.1) and the Stata code for the analyses is published online at <https://github.com/UCL-CHIG>.

4.1. Background

Geographical location is a prominent source of variation in rates of hospital admissions for bronchiolitis in England.^{4,15,45} On average, between 1999 and 2011, a 5.3-fold difference in the rate of hospital admissions for bronchiolitis among infants was reported across the 352 Local Government Areas of England.⁴ Another study that looked at admissions across the 152 Primary Care Trusts in England (the local NHS administrative bodies that were abolished and replaced by Clinical Commissioning Groups, CCGs, in 2013) between 2007/08 and 2010/11, described a 15-fold difference in the rates of admissions for bronchiolitis among children under 2 years of age.³²⁷ An overarching factor proposed to explain this extreme geographical variation lies with socioeconomic differences in the composition of the population at the area-level. However, these studies noted only modest associations between higher socioeconomic deprivation (measured by the index of multiple deprivation) and bronchiolitis rates; $r^2=0.24$ at the Local Government Area-level and $r^2=0.33$ at the Primary Care Trust-level.^{4,327} Instead, the authors of these papers write that it is likely that clinical factors, particularly admission thresholds, are the main driver of the variation in admission rates noted.

However, these geographical studies did not consider the impact of seasonality when assessing spatial variation in admissions for bronchiolitis across England. Seasonality, defined as cyclic changes in disease occurrence,³²⁸ is a central component to the patterns of bronchiolitis admissions observed in England. In temperate maritime climates such as in the UK, respiratory syncytial virus (RSV, the dominant pathogen causing bronchiolitis) epidemics occur during the colder months, with a parallel trend in hospital admissions for bronchiolitis.^{109,110}

By including sine and cosine functions in regression models of times series data (analysis known as *harmonic regression*), generic seasonal patterns can be removed from estimates and underlying differences in trends detected.^{329,330} Harmonic regression analysis also allows for the calculation of additional area-specific disease components, such as the timing of the highest disease rates and the epidemic duration.^{124,331} Researchers in Australia and the USA have used this type of analysis to look at population factors associated with features of RSV epidemics among children.^{124,332} For example, Noveroske et al. found that earlier epidemic timing was associated with areas of Connecticut, USA, that were more densely populated and had a higher proportion of black people in the population (used in this study as an indicator of socioeconomic deprivation).¹²⁴ This type of analysis, yet to be conducted in a UK setting, can be used to plan interventions, such as whether differential timing of palivizumab prophylaxis based on area or population characteristics is necessary.

The aim of this Chapter is to describe the spatial variation in the seasonality of hospital admissions for bronchiolitis across England.

There are two objectives for the Chapter:

Objective 4A To demonstrate how seasonality of bronchiolitis admissions amongst infants varies across region and CCG in England.

Objective 4B To assess how epidemic timing and rates of admissions (after accounting for seasonal differences) are associated with population characteristics between CCGs.

4.2. Methods

4.2.1. Data sources and study cohort

This study uses hospital episode statistics admitted patient care (HES APC) data linked to Office for National Statistics (ONS) mortality records as introduced in Chapter 3.¹⁵ HES APC is a database of all hospital inpatient admissions funded by the English NHS and captures approximately 97% of all births in England.²⁶⁹ NHS Digital, the body that houses and supplies

access to this database, links patient records over time and provides each database extract with a pseudo-anonymised identifier unique to each individual (called a HESID), enabling researchers to create longitudinal patient cohorts. I also used CCG-level data, namely index of multiple deprivation (IMD) scores and population size, which I downloaded from the Ministry of Housing, Communities and Local Government (MHCLG) and ONS websites, respectively.^{333,334}

A birth cohort of children born from 1st January 2011 to 31st December 2016 was derived from the dataset previously created by my colleague, Ania Zylbersztejn.²⁸⁴ This dataset did not contain infants from multiple births or stillbirths. Using the linkage methods developed by Harron et al.,²⁷³ I successfully linked 95.0% of birth records to maternal delivery records (see Appendix 3.1 for further details). This enabled me to complete missing geographical information for infants as recorded at birth/delivery. Non-English resident children were excluded since they cannot be followed up in English hospital data. Our dataset does not contain postcodes or addresses, so non-residency was identified using lower super output area (LSOA). LSOAs not beginning with “E”, which signifies an English LSOA, were used to define non-resident status. Infants with missing geographical data were also excluded from this study.

Children were followed from birth or 1st January 2012, whichever was later, to their 1st birthday, date of death, (estimated) date of emigration or 31st December 2016, whichever came first. This study period ensured that the study had follow-up for children aged up to 12 months in all study years, including the first (2012). Emigration was defined as the presence of a non-English LSOA in any admission during follow up. The emigration date was set as the mid-point between the child’s date of birth and the date at which the non-resident admission occurred. I focused on admissions in the first year of life since more than 90% of bronchiolitis admissions occur in the first year of life and two thirds of admissions are in children less than 6 months old.^{6,261}

4.2.2. Study outcome

My study outcome was the number of hospital admissions for bronchiolitis, recorded as either the primary or one of the 19 secondary diagnoses in HES APC. Diagnoses in HES APC are coded using the International Classification of Diseases, version 10 (ICD-10), and I identified bronchiolitis admissions using ICD-10 J21 codes indicating acute bronchiolitis. All J21 subcategories were included in this definition: J21.0 acute bronchiolitis due to RSV, J21.1 acute bronchiolitis due to human metapneumovirus, J21.8 acute bronchiolitis due to other specified organisms and J21.9 acute bronchiolitis, unspecified. In this study, sequential

admissions for bronchiolitis by the same child were assumed to be associated with the same infection if the start date of a new admission was within 14 days of the previous admission's discharge date. In these instances, only the first of these admissions were included in the number of admissions.

Identifying RSV-related bronchiolitis using HES APC data

HES APC data does not provide reliable information about the aetiology of bronchiolitis cases. As mentioned, there are several sub-categories of J21 available, including one that specifies that the cause of the condition is RSV. However, specific bronchiolitis diagnoses are poorly coded in HES APC because laboratory testing is not routine for children presenting to hospital in England with bronchiolitis.^{45,335} This follows advice from the National Institute for Health and Care Excellence (NICE), who recommend against the routine testing of viral aetiology in their 2015 viral bronchiolitis guidelines for clinicians in England.³ Similar guidance by the American Academy of Pediatrics provides a rationale for this recommendation: "at the individual patient level, the value of identifying a specific viral aetiology causing bronchiolitis has not been demonstrated".^{336 p.e1480} However, other methods have been used to ascertain the proportion of bronchiolitis cases in infants that are attributable to RSV. Reeves et al. jointly modelled of HES APC and laboratory surveillance data over a 5-year period from 2007 and estimated that 82% (95% CI 79% to 87%) of admissions with a primary diagnoses of bronchiolitis among children under 6 months and 70% (95% CI 66% to 75%) in children between 6 and 12 months of age could be attributed to RSV.⁶

4.2.3. Covariates

Year and week of the year at admission were derived from each bronchiolitis admission record. All other covariates were based on each infant's recorded residence at birth.

Government region was used to describe broad patterns across England and CCGs to examine local areas. There are nine government office regions in England, created to delegate functions from national government in 1994, but now used only for administrative and statistical purposes.³³⁷ In ascending order of population size these are the North East, East Midlands, Yorkshire and the Humber, South West, West Midland, East of England, North West, London, and the South East. CCGs, introduced in England in 2013, are NHS bodies responsible for planning and commissioning for local areas in England.³³³ There were 209 CCGs in England over the study period with mid-2016 populations ranging from 68,187 in Corby (Northamptonshire) CCG to 898,025 in Northern, Eastern and Western Devon CCG.

The number of CCGs within each government office region ranged from 11 CCGs in the North East to 39 CCGs in the South East.

I calculated annual population density estimates—the number of residents per square kilometre—for each CCG by dividing annual CCG population sizes by the area of each CCG in square kilometres (estimated using polygon areas in QGIS).³³³ The average population size over the five-year study period was assigned to each CCG. IMD is an overall measure of deprivation experienced by people within a neighbourhood, with a higher score indicating higher levels of deprivation.⁶⁵ The composite score is constructed by combining weighted scores from seven domains of deprivation; income, health and disability, education, skills and training, barriers to housing and services, crime and living environment, with measures mainly derived from the 2011 Census. I used IMD scores for 2015 in this study, summarised across CCGs. These summaries are calculated by averaging the population-weighted IMD scores of the LSOAs contained within a CCG (and are available to download directly from GOV.UK).^{334,338}

4.2.4. Statistical methods

Stata 15.0³¹⁴ was used for data analysis. QGIS 2.18,³³⁹ a free open-source geographic information system, was used to present geospatial data and Microsoft Excel 2013 was used to create graphs. The Stata code for the harmonic regression analysis can be found in Appendix 4.2.

Weekly incidence rates of bronchiolitis admissions for each study year, by region and CCG, were calculated by dividing the number of admissions by person-time at risk. Rates are expressed as admission-based rates per 1000 infant-years.

Objective 4A: Seasonality of bronchiolitis admissions across region and CCG

To explore geographical variation at the regional level I fitted a fixed effects Poisson regression model to the regional count data. I used Poisson regression rather than negative binomial regression because there was no evidence of over-dispersion. I used robust standard errors to protect against model misspecification. The impact of seasonality on rates of admissions at the regional level were modelled using a harmonic function of time in weeks (t) expressed as: $\beta_1(\sin(2\pi t/T)) + \beta_2(\cos(2\pi t/T))$, where T = number of periods within one cycle (i.e. 1 year = 52.14 weeks).^{340,341} I adjusted for year of admission to account for the annual increasing trend in bronchiolitis admission rates.⁴ I included interaction terms between region and the sine and cosine functions in the region-specific analyses to allow for effect modification of seasonality by region, formally tested via a Wald χ^2 test.

Using the estimated harmonic function coefficients, I calculated the following quantities for each region's epidemic curve at reference values of the other covariates:³⁴²

- the amplitude (log) $\hat{\gamma}_j = \sqrt{(\hat{\beta}_1 + \hat{\delta}_{1j})^2 + (\hat{\beta}_2 + \hat{\delta}_{2j})^2}$;
- the phase (in radians) $\hat{\psi}_j = \arctan(\hat{\beta}_1 + \hat{\delta}_{1j}) / (\hat{\beta}_2 + \hat{\delta}_{2j})$;
- and the peak week $\hat{P}_j = 52.14 * \left(\frac{\hat{\psi}_j}{2\pi}\right) + 1$

For descriptive purposes, I also calculated the epidemic duration as the number of weeks between the first of three consecutive weeks with increasing predicted rates and the first of three consecutive weeks with decreasing predicted rates. See Box 4.1 for full model parameterisation.^{124,332}

As outlined in Box 4.1, there were two parts to the CCG-level analyses. In stage 1, I fitted a multilevel (i.e. mixed-effects) Poisson model to the individual admission records. I included the harmonic functions as specified above, and allowed for clustering at the CCG level via random effects for the intercept and the harmonic function parameters.¹²⁴ Using the estimated harmonic function coefficients, I calculated CCG-specific average annual peak week as $\hat{P}_i = 52.14 * (\arctan((\hat{\beta}_1 + \hat{\eta}_{1i}) / (\hat{\beta}_2 + \hat{\eta}_{2i})) / 2\pi) + 1$, where i indicates CCG.^{124,342} The CCG-specific departures from the model intercept (i.e. the random effects for the intercept, η_{0i}) across England were exponentiated to obtain CCG-specific incidence rate ratios (IRRs) relative to the population mean.

Objective 4B: Association between CCG-level epidemic timing, rates of admissions and demographic characteristics

In stage 2, with CCGs treated as units of analysis, multivariable linear regression models were fitted to estimate the mutually adjusted association of population density and deprivation score with, separately, the CCG-specific peak week and IRRs of bronchiolitis admissions predicted in stage 1. Population density was highly positively skewed and therefore first log-transformed to reduce the impact of very high values. I added quadratic terms to allow for non-linear associations with the explanatory variables and included interactions between them, comparing model specifications via likelihood ratio tests. Alongside regression coefficients, I report the proportion of total variance in each dependent variable explained by the model using eta-squared (η^2) and the relative contribution of each model component using partial eta-squared (ηp^2). To assess the impact of ignoring the uncertainty of the predicted values from the first step I computed bootstrapped standard errors from 1,000 bootstraps (from the two

stages of the analysis). I calculated bootstrapped standard errors using a 20% random sample of the data because of computational constraints due to the size of the complete data.

Sensitivity analyses

These models assumed no spatial dependency of the CCG-level residuals. To assess this assumption I calculated Moran's I , which is a measure of how related the values of a variable are based on the locations where they were measured (with -1 indicating perfect dispersion, 0 no autocorrelation and 1 perfect clustering).³⁴³ Due to the large difference in admission rates within London compared with the rest of the country, sensitivity analyses were conducted to assess the robustness of the results when London-based CCGs were excluded from analyses. To assess the impact of repeated events by the same child, I replicated the analyses including only the first admission per child.

Box 4.1. Models' specifications and derived parameters³⁴¹

Region-specific analyses

Let $n_{jk}(t)$, the number of bronchiolitis admissions in region j and year k observed at time t (measured in weeks), follow a Poisson distribution with rate $\lambda_{jk}(t) = E(n_{jk}(t))/N_{jk}(t)$ where $N_{jk}(t)$ denote the person-time at risk in region j and in year k at time t . I modelled this rate, after log-transformation, as a function of region, year and time of admission as follows:

$$\log(\lambda_{jk}(t)) = \beta_0 + \beta_1(\sin(2\pi t/T)) + \beta_2(\cos(2\pi t/T)) + \sum_{j=1}^9 \alpha_j I_{\text{region}=j} \\ + \sum_{j=1}^9 \delta_{1j} \sin(2\pi t/T) I_{\text{region}=j} + \sum_{j=1}^9 \delta_{2j} \cos(2\pi t/T) I_{\text{region}=j} + \sum_{k=2012}^{2016} \theta_k I_{\text{year}=k}$$

Where: T is the length of period within one harmonic cycle (i.e. 1 year = 52.14 weeks); and $I_{X=x}$ is the binary indicator of the variable X taking value x . The parameter β_0 is the intercept, β_1 and β_2 are harmonic function coefficients, δ_{1j} and δ_{2j} are region-specific harmonic function coefficients, and the parameters $\alpha_1, \delta_{11}, \delta_{21}$ and θ_1 are all constrained to be zero to deal with the collinearity of the binary indicators.

CCG-specific analyses: stage 1

Let the number of bronchiolitis admissions in CCG i and year k , observed at time t (in weeks), $n_{ik}(t)$, follow a Poisson distribution with rate $\lambda_{ik}(t) = E(n_{ik}(t))/N_{ik}(t)$, with $N_{ik}(t)$ denoting the relevant person-time at risk. Given the large number of CCGs, the following multilevel model with random intercepts and random harmonic coefficients was specified for the log rate:

$$\log(\lambda_{ik}(t)) = (\beta_0 + \eta_{0i}) + (\beta_1 + \eta_{1i}) \sin(2\pi t/T) + (\beta_2 + \eta_{2i}) \cos(2\pi t/T) + \sum_{k=2012}^{2016} \theta_k I_{year=k}$$

Where: η_{0i} = random effect component of the intercept; η_{1i} = random effect component of the sine parameter; η_{2i} = random effect component of the cosine parameter; and $\theta_1 = 0$ as before. The model's predicted CCG-specific parameters were then used to derive CCG-specific peak time, \hat{P}_i , and CCG-specific amplitude, $\hat{\gamma}_i$.

CCG-specific RSV analyses: stage 2

These predicted values were then related to predictors in two separate linear regression models:

$$\hat{P}_i = \alpha_{10} + \alpha_{11}z_i + \alpha_{12}w_i + \varepsilon_{1i} \quad \hat{\gamma}_i = \alpha_{20} + \alpha_{21}z_i + \alpha_{22}w_i + \varepsilon_{2i}$$

Where: α_{10}, α_{20} = intercepts; α_{11}, α_{21} =IMD coefficients; α_{12}, α_{22} =population density coefficients; z_i =CCG-specific IMD; w_i =CCG-specific log-population density; and $\varepsilon_{1i}, \varepsilon_{2i}$ =random error terms, assumed to be independently distributed.

4.3. Results

4.3.1. Descriptive statistics

Figure 4.1 shows the derivation of the final study cohort. Of the 3,808,247 singleton children born between 2011 and 2016, 81,234 (2.1%) were excluded from the final study cohort. In total, 55,801 (1.5%) infants had missing data for geographical information (region and/or CCG), and the odds of missingness decreased by year of birth (OR 0.92, 95% CI 0.91 to 0.92). The final cohort consisted of 3,727,013 children, of which 34.4% had a London or South East address recorded at the time of birth (Table 4.1) and the average follow-up time per infant was 304 days. Between 1 January 2012 and 31 December 2016, 3.7% of infants (n=139,532) had at least one admission to hospital for bronchiolitis. There were 155,485 admissions for

bronchiolitis by cohort members in total, an average annual admission-based rate of 50.1 per 1000 infant-years (95% CI 49.9 to 50.4).

Bronchiolitis admission rates increased from 46.9 per 1000 infant-years (95% CI 46.4 to 47.4) in 2012 to 58.4 per 1000 (95% CI 57.8 to 59.1) in 2016. This was equivalent to a 5.2% annual increase of admissions, from 29,853 to 36,028, between 2012 and 2016. Annual admission rates were highest in the North West (68.7 per 1000 infant-years, 95% CI 67.9 to 69.5) and North East (63.8 per 1000 infant-years, 95% CI 62.5 to 65.1); both more than twice the rate in London (30.9 per 1000 infant-years, 95% CI 30.4 to 31.3). Nationally, bronchiolitis epidemics peaked once a year, in December (see Figure 4.2). The average weekly variation in observed bronchiolitis admission rates by region are shown in Figure 4.3.

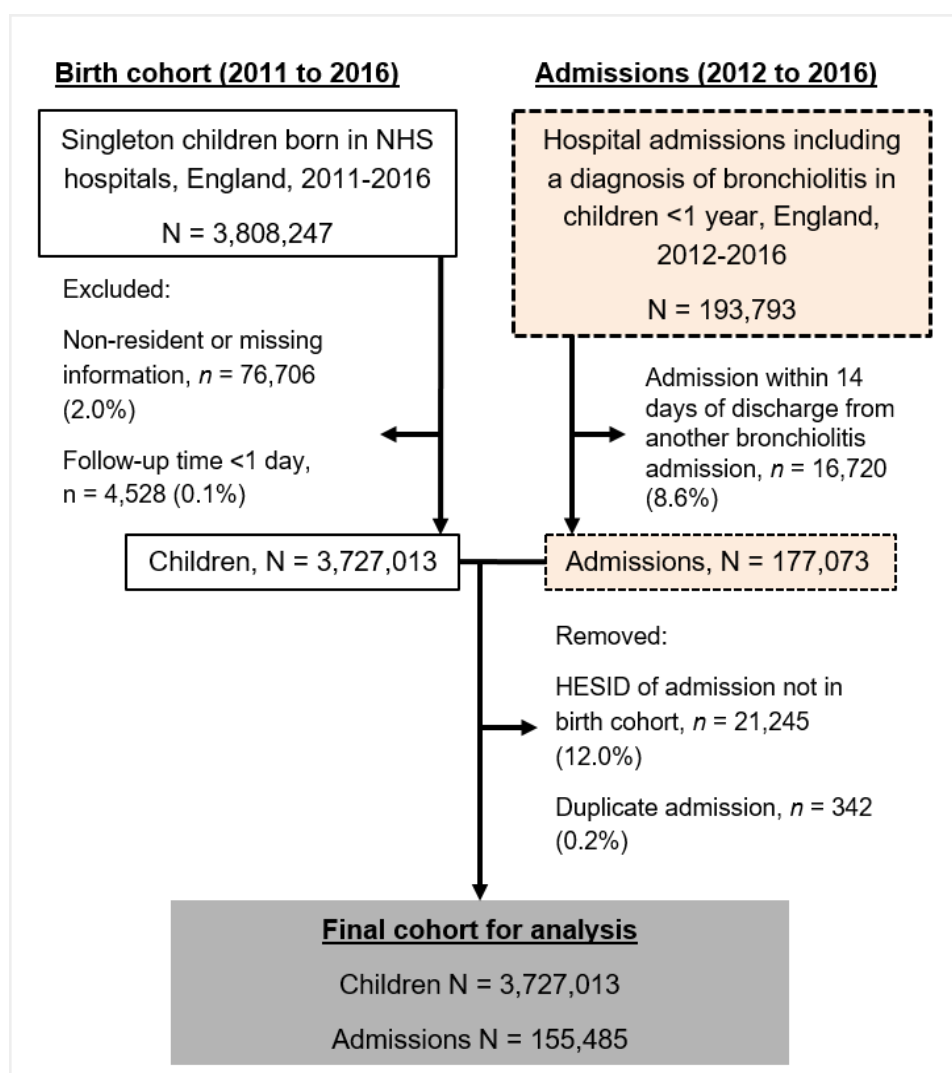
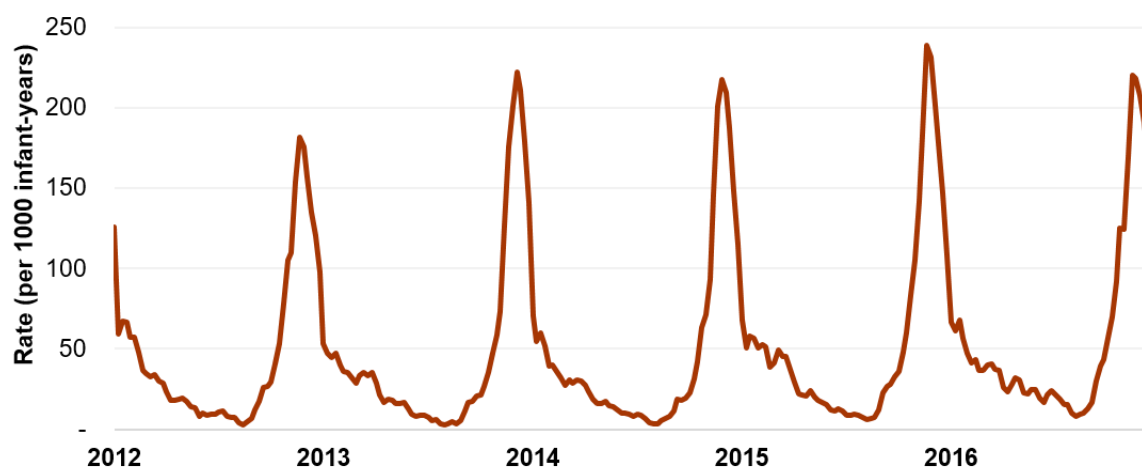


Figure 4.1. Flow diagram showing derivation of final study cohort and bronchiolitis hospital admissions (linked via unique HESID)

Table 4.1. Cohort characteristics and the number and rate (per 1000 infant-years) of bronchiolitis hospital admissions before 1 year of age among cohort members

		<u>All infants</u>		<u>Admitted infants*</u>		
		<u>N</u>	<u>%</u>	<u>N</u>	<u>Rate (95% CI)</u>	
Total		3,727,013	100.0	155,485	50.1	(49.9, 50.4)
Region	North East	176,125	4.7	9,345	63.8	(62.5, 65.1)
	North West	492,007	13.2	28,094	68.7	(67.9, 69.5)
	Yorkshire & the Humber	353,972	9.5	16,838	57.3	(56.4, 58.1)
	East Midlands	289,709	7.8	11,989	49.8	(48.9, 50.7)
	West Midlands	402,365	10.8	19,341	57.9	(57.1, 58.8)
	East of England	403,141	10.8	14,918	44.4	(43.7, 45.1)
	London	705,266	18.9	18,153	30.9	(30.4, 31.3)
	South East	575,762	15.5	22,503	46.9	(46.3, 47.5)
	South West	328,666	8.8	14,304	52.0	(51.2, 52.9)
		<u>Year of birth</u>		<u>Year of admission</u>		
Year	2011	635,325	17.1			
	2012	638,157	17.1	29,853	46.9	(46.4, 47.4)
	2013	621,295	16.7	27,592	44.0	(43.5, 44.6)
	2014	611,217	16.4	28,726	46.6	(46.1, 47.2)
	2015	613,062	16.5	33,286	54.9	(54.3, 55.5)
	2016	607,957	16.3	36,028	58.4	(57.8, 59.1)

*per 1000 infant-years

**Figure 4.2.** Observed weekly rates (per 1000 infant-years) of hospital admissions for bronchiolitis in England, from January 2012 to December 2016

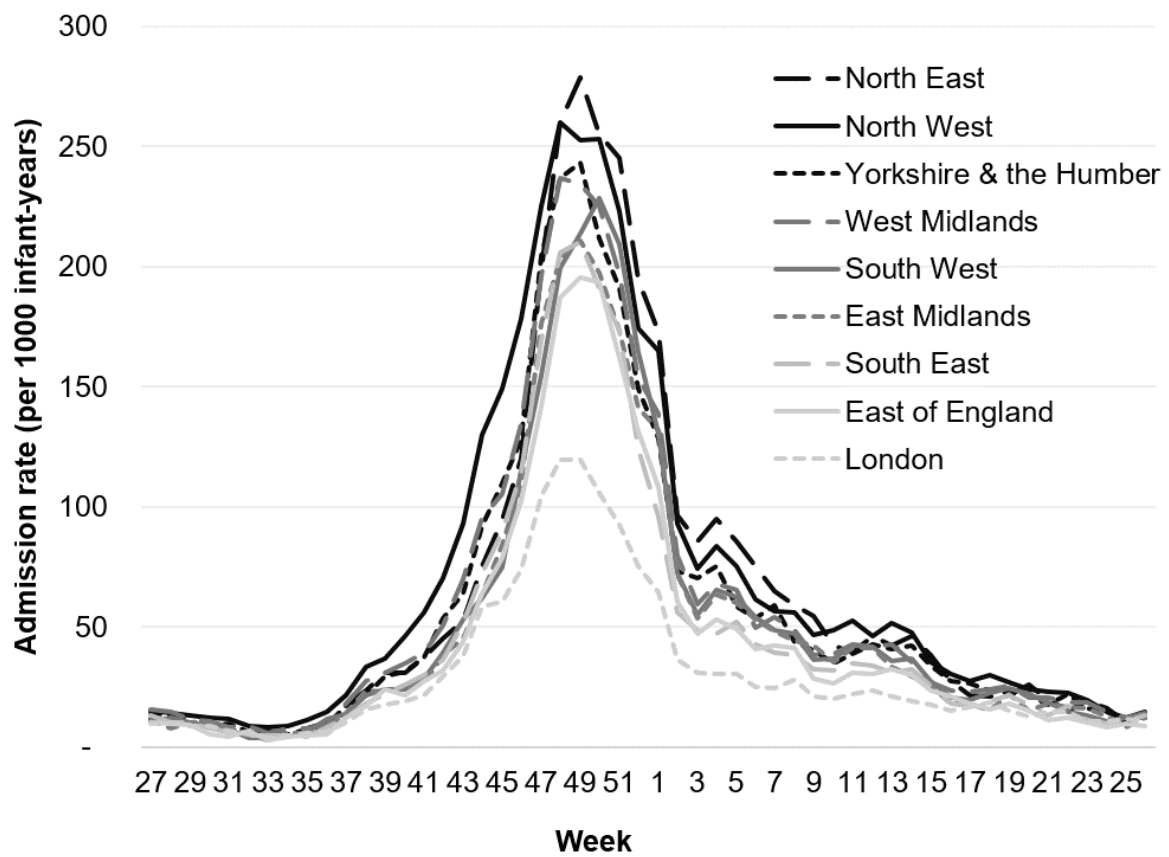


Figure 4.3. Observed weekly rates of hospital admissions for bronchiolitis, by Region: England, 2012-2016 averaged

4.3.2. Objective 4A: Seasonality of bronchiolitis admissions

Regional level

The estimates of region-specific seasonal parameters derived from the harmonic Poisson regression model assessing bronchiolitis seasonality at the regional level are presented in Table 4.2 and predicted weekly admission rates in Figure 4.4. There was some seasonal variation across regions, with the epidemic peak week ranging from week 50.3 in London (95% CI 50.2 to 50.5) to week 51.4 in the North East (95% CI 51.2 to 51.5). The estimated amplitude at peak epidemic week compared with the mean week was lowest in London ($\hat{\gamma}_j=3.41$, 95% CI 3.31 to 3.52) and highest in the North East ($\hat{\gamma}_j=4.95$, 95% CI 4.72 to 5.19). Each region had an epidemic duration of 26 or 27 weeks. At week 51, rates of admission were 2.45 and 2.44 times greater in the North West (95% CI 2.39 to 2.52) and the North East (95% CI 2.36 to 2.52) compared with London.

Table 4.2. Derived average annual seasonal estimates following harmonic Poisson regression (Table A4.1, Appendix 4.3), by region

Region	Amplitude $\hat{\gamma}^{**}$ (95% CI)	Peak week \hat{P} (95% CI)	Phase shift $\hat{\psi}$ (95% CI)
North East	4.95 (4.72, 5.19)	51.4 (51.2, 51.5)	-0.21 (-0.23, -0.19)
North West	4.25 (4.14, 4.36)	50.4 (50.3, 50.5)	-0.34 (-0.35, -0.32)
Yorkshire and the Humber	4.31 (4.17, 4.46)	50.7 (50.6, 50.8)	-0.29 (-0.31, -0.28)
East Midlands	4.61 (4.42, 4.80)	51.1 (50.9, 51.9)	-0.25 (-0.27, -0.23)
West Midlands	4.28 (4.15, 4.42)	50.6 (50.4, 50.7)	-0.31 (-0.32, -0.30)
East of England	4.89 (4.70, 5.08)	50.8 (50.7, 51.0)	-0.28 (-0.29, -0.26)
London	3.41 (3.31, 3.52)	50.3 (50.2, 50.5)	-0.34 (-0.36, -0.32)
South East	4.54 (4.40, 4.68)	50.6 (50.4, 50.7)	-0.31 (-0.32, -0.30)
South West	4.58 (4.41, 4.75)	51.2 (51.1, 51.4)	-0.23 (-0.25, -0.22)
		Duration (weeks)	IRR at week 51 (95% CI)
North East		27 (26 to 01)	2.44 (2.36, 2.52)
North West		26 (25 to 51)	2.45 (2.39, 2.52)
Yorkshire and the Humber		26 (26 to 52)	2.06 (2.00, 2.12)
East Midlands		27 (26 to 01)	1.85 (1.79, 1.91)
West Midlands		27 (25 to 52)	2.08 (2.02, 2.14)
East of England		26 (26 to 52)	1.68 (1.63, 1.73)
London		26 (25 to 51)	Reference.
South East		27 (25 to 52)	1.73 (1.68, 1.73)
South West		27 (26 to 01)	1.93 (1.87, 1.99)

*Adjusted for year of admission, **amplitude exponentiated

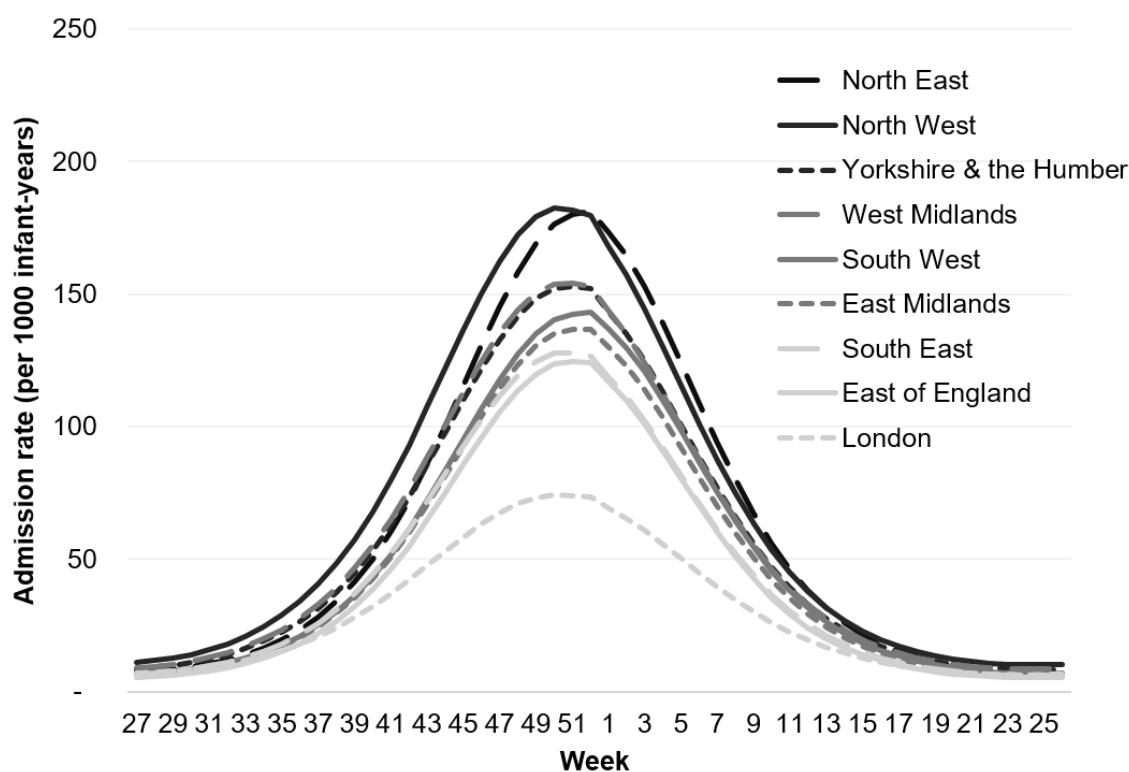


Figure 4.4. Predicted weekly rates of hospital admissions for bronchiolitis, by Region: England, 2012-2016 averaged

CCG-level

Compared with the average rate of hospital admissions for bronchiolitis in England, predicted rates (derived from the multilevel model, Table A4.2, Appendix 4.3) were highest in CCGs based in the North East and North West and lowest in London and South-East CCGs (Figure 4.5). There was a ratio of 5.3 between the lowest (West Kent) and highest (Stoke on Trent) CCG-based incidence rates. Peak epidemic timing by CCG ranged by 2.9 weeks, from week 49.3 to week 52.2. As illustrated in Figure 4.6, the earliest peaks were seen in North West and South London CCGs and CCGs surrounding London.

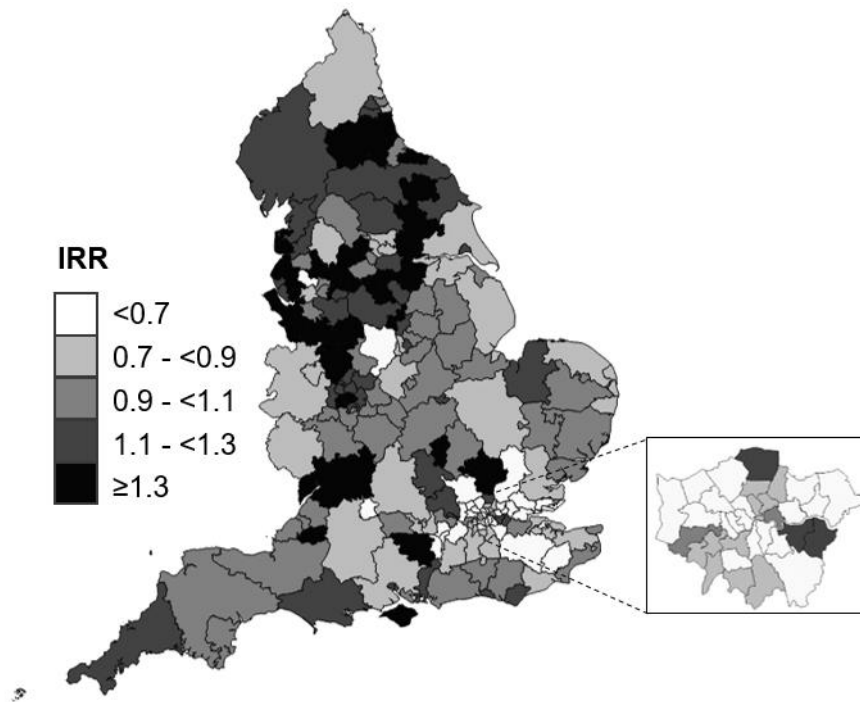


Figure 4.6. Predicted IRRs of bronchiolitis hospital admissions compared with the national average, by CCG: England (with London inset)

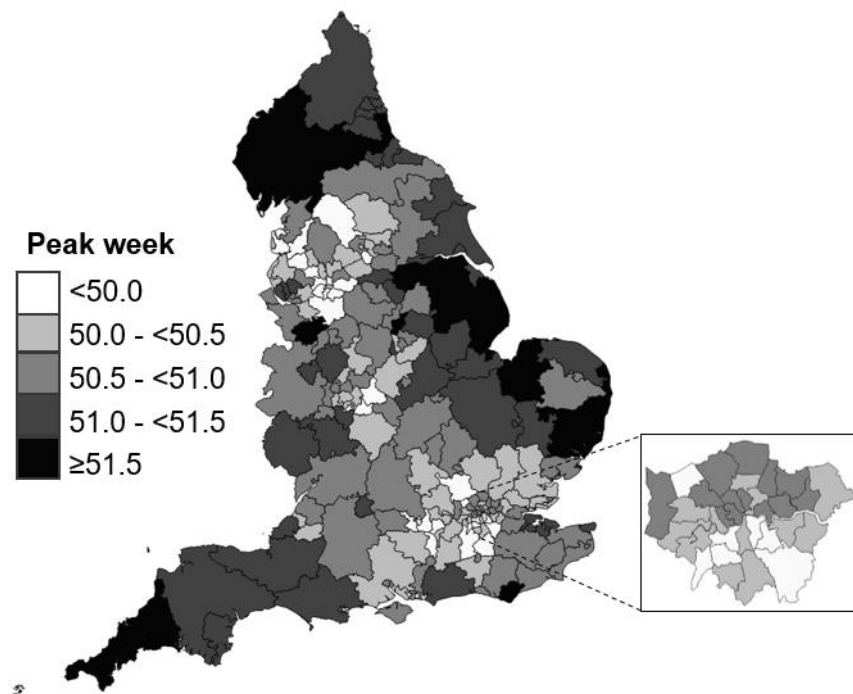


Figure 4.5. Predicted peak timing (weeks) of bronchiolitis admission rates, by CCG: England (with London inset)

4.3.3. Objective 4B: Multivariable regression models

Results of the multivariable regression models for CCG-level rates of admission and peak timings are displayed in Table 4.3. In mutually adjusted models, a greater IMD score was associated with higher rates of admissions at the CCG level, whilst higher log population density was associated with a slightly lower rate of admissions (Figure 4.7). In total, 23% of the variation in incidence rates at the CCG level was explained by this model, $\eta^2(2)=0.23$ (95% CI 0.13 to 0.31). As shown in Figure 4.8, IMD had a quadratic relationship with peak timing, (likelihood ratio test for the quadratic term, $\chi^2(1)=19.6$, $p<0.001$); both low and high IMD scores were associated with earlier peak timing of bronchiolitis admissions. Higher log population density was associated with earlier peak timing. In total, 38.0% of the variation in peak timing at the CCG level was explained by this model, $\eta^2(2)=0.38$ (95% CI 0.27 to 0.46).

Excluding London-based CCGs

With London-based CCGs excluded from the multivariable regression models (Appendix 4.3, Table A4.3), the adjusted association between IMD and both outcomes (peak timing and IRR of admissions compared with the national average) broadly stayed the same (Table 4.3). The association between log population density and rates of admissions weakened considerably, with only a slight negative association now shown (and only 1% of the variation explained by this variable, $\eta^2(2)=0.01$, 95% CI 0.00 to 0.05). The association between log population density and peak timing decreased from -0.25 (95% CI -0.30 to -0.19) to -0.28 weeks (-0.35 to -0.21).

Table 4.3. Multivariable linear regression analyses assessing the effect of population density and IMD score on IRR compared to the national average and peak timing of CCG epidemic

	<u>All CCGs*</u>			<u>Excluding London-based CCGs**</u>			
	Coefficient (95% CI)		Partial η^2 (95% CI)	Coefficient (95% CI)		Partial η^2 (95% CI)	
IRR							
Population density†	-0.09	(-0.12, -0.05)	0.11 (0.04, 0.20)	-0.02	(-0.06, 0.02)	0.01 (0.00, 0.05)	
IMD	0.02	(0.02, 0.03)	0.22 (0.13, 0.31)	0.02	(0.01, 0.02)	0.21 (0.11, 0.31)	
Constant	1.19	(0.98, 1.40)		0.75	(0.53, 0.97)		
	Total η^2		0.23 (0.13, 0.31)	Total η^2		0.24 (0.13, 0.33)	
Peak timing (weeks)							
Population density†	-0.25	(-0.30, -0.19)	0.30 (0.20, 0.39)	-0.28	(-0.35, -0.21)	0.25 (0.15, 0.35)	
IMD	0.11	(0.07, 0.14)	0.17 (0.09, 0.26)	0.10	(0.07, 0.14)	0.16 (0.07, 0.26)	
IMD ²	-0.002	(-0.002, -0.001)	0.10 (0.04, 0.18)	-0.001	(-0.002, -0.001)	0.09 (0.02, 0.17)	
Constant	50.87	(50.40, 51.34)		51.08	(50.49, 51.67)		
	Total η^2		0.38 (0.27, 0.46)	Total η^2		0.35 (0.24, 0.44)	

*IRR and peak week derived from harmonic multilevel mixed-effect Poisson regression analysis displayed in Table A4.2 (Appendix 4.3); ** IRR and peak week derived from harmonic multilevel mixed-effect Poisson regression analysis displayed in Table A4.3 (Appendix 4.3); †Log-scale

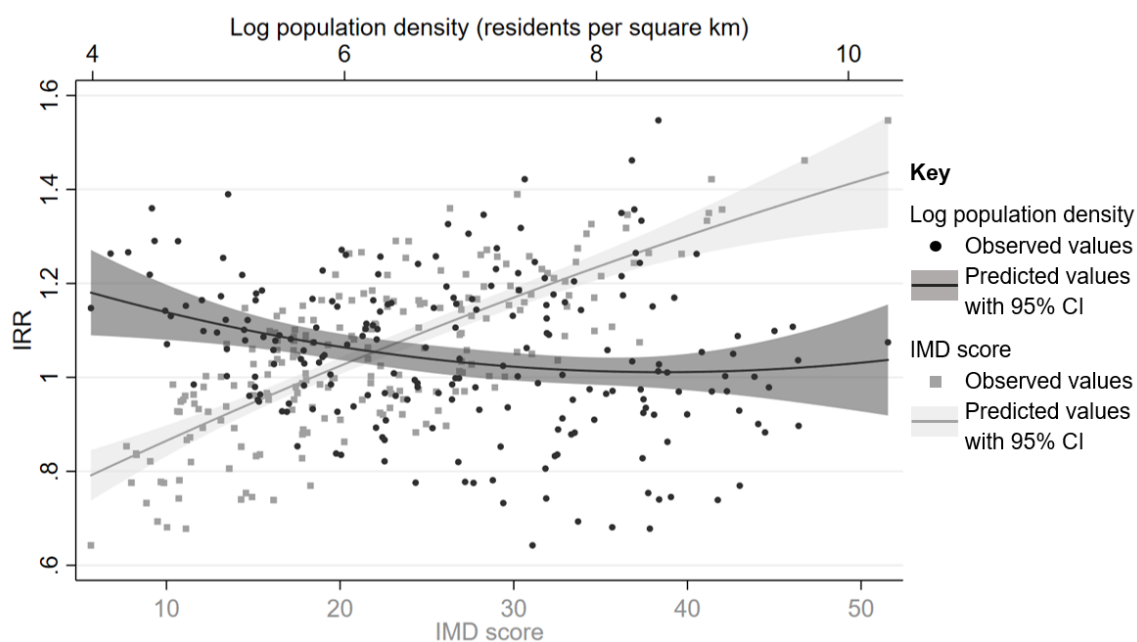


Figure 4.7. Observed and predicted variation in CCG-level rates of bronchiolitis hospital admissions compared with the national average, by IMD score and log population density (including London-based CCGs)

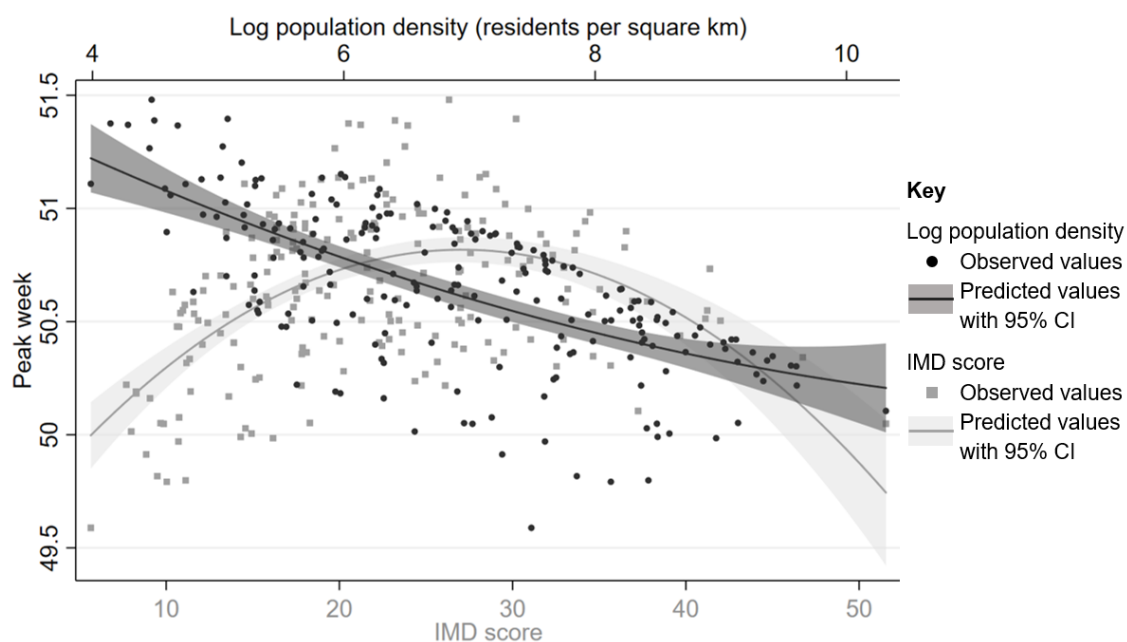


Figure 4.8. Observed and predicted peak timing of CCG-level bronchiolitis hospital admissions, by IMD score and log population density (including London-based CCGs)

4.3.4. Sensitivity analysis

To assess whether the uncertainty of the stage 1 estimation had an impact on inferences at stage 2, I calculated bootstrapped standard errors on 20% of the total sample and compared them to those obtained without bootstrap on the same data. The results showed very minimal differences, supporting the interpretations above, especially given the size of the whole sample (see Table A4.4, Appendix 4.4). The Moran's I values were positive for the IRR residuals ($I=0.28$) and the peak timing residuals ($I=0.37$) of spatially proximal CCGs. When I included only the first admission by each child in the analyses, regional amplitudes increased and peak epidemic timing was approximately 0.2 weeks earlier (Table A4.5, Appendix 4.4). Relative differences in the seasonality of admissions across regions and CCG-level analyses remained broadly the same (Tables A4.6 and A4.7, Appendix 4.4).

4.4. Discussion

I found large variation in the size of the epidemic peak and accompanying admission rates for bronchiolitis, particularly when comparing London to the rest of the country. At week 51 (mid-December) rates of admissions were 2.4 to 2.5 times higher in Northern regions of the country compared with London. There was a 5.3-fold difference in rates across CCGs, with about a quarter of the difference explained by area-level deprivation. The shape of the epidemic curve was similar across regions, with admission rates increasing continually for 26 or 27 weeks of the year and epidemics peaking between week 50 and 51 (2nd and 3rd weeks of December). At the CCG-level the difference in peak week increased to 3 weeks, with earlier peaks tending to occur in areas with higher population densities, such as London and Manchester.

4.4.1. Strengths and weaknesses

The birth cohort was created from HES APC, a national administrative dataset of hospital admissions in England. Data in HES APC are recorded by clinical coders in each hospital and coding consistency may, therefore vary by hospital according to diagnostic protocols and recording practices. Diagnostic codes are rarely entered in accident and emergency departments in England, and there is no national primary care database covering all English general practices.²⁸² Therefore, I could not examine bronchiolitis seasonality in accident and emergency attendances or general practice consultations, which, it could be hypothesised, may peak before hospital admissions. Private hospital activity amongst non-NHS funded patients is not captured in HES APC. However, since emergency treatment is not typically covered under private medical insurance in the UK and NHS care is (for the most part) free at the point of need, I expect that very few children are admitted privately for acute illnesses like

bronchiolitis.²⁸² Overall, with a large number of bronchiolitis admissions across the whole of England, this study provides a contextual understanding of the seasonality of this infection and how it varies across the country

I modelled admissions using harmonic functions, which allowed me to assess variability in the seasonality of bronchiolitis by area of the country. The predicted regional model produced seasonal patterns similar to the observed rates of admissions but underestimated the amplitude of the regional epidemics. Fitting splines may have improved estimation but would not have allowed me to straightforwardly estimate seasonal parameters. Using a single pair of sine/cosine terms in the model mirrored the single annual peak exhibited in the data, and enabled me to estimate the peak epidemic week seen each year in England. I was unable to account for uncertainty in the estimates of incidence rates and peak timing in the second part of the CCG-level model; however, the results were shown to be robust in sensitivity analyses.

4.4.2. Research findings: comparison with other studies

Objective 4A: Seasonality of bronchiolitis admissions across England

As expected, a substantial single epidemic peak in hospital admissions for bronchiolitis occurred each winter in England between 2012 and 2016 in mid-December. A small increase in hospital admissions around the fourth week of year can also be seen in Figures 4.2 and 4.3. It is likely that this increase in cases is driven by the post-holiday return to day care and/or school (by older siblings), which are known sources of infant RSV infection.¹⁸⁹ The 5.3-fold geographical variation in bronchiolitis admission rates by CCG found here matches Green et al.'s analysis by Local Government Area between 1999 and 2011.⁴ Another analysis that looked at Primary Care Trusts estimated that the variation was almost 3 times higher than in my study, although this difference can be explained by their definition of bronchiolitis.³²⁷ Children up to the age of 2 admitted to hospital with a primary diagnoses of bronchiolitis were included in their definition, leading to some area-level rates lower than 5 per 1000 children. I add to these previous findings with information about the epidemic curves, which were found to have a similar duration across regions of England but differ with regards to the peak week at the CCG-level.

Objective 4B: Association between CCG-level epidemic timing, rates of admissions and demographic characteristics

Similar to findings from previous research,^{4,327} I found that one fifth of the variation in admission rates across England was explained by socioeconomic factors, after controlling for population density. Socioeconomic deprivation is a broad facet associated with risk factors for severe

RSV infection including tobacco smoke exposure, housing conditions and overcrowding, amongst others.^{4,45,139} I used a small-area level indicator of deprivation, IMD, to capture socio-economic deprivation of the children contributing to this study. IMD is widely used in studies of health outcomes in England, and has been shown to be associated with bronchiolitis admission rates.^{4,344} Further research is required to establish which particular aspects of deprivation have the highest impact on bronchiolitis admissions.

Converse to expectation, larger population density was associated with lower admission rates; however, sensitivity analyses showed that this relationship was driven by London based CCGs. The difference in emergency admission rates in London compared with the rest of England is stark, but not a new finding—having previously been reported for a range of conditions.^{45,345,346} Plausible reasons for differential admission rates include primary care accessibility and availability, and differences in admission thresholds.⁴ Data published by Public Health England show that, whilst rates of emergency admissions for children aged 0-4 are low in London, accident and emergency (A&E) rates are higher than the national average, potentially reflecting regional differences in admission practices.³⁴⁷ This hypothesis is corroborated by early results from research using English A&E linked to HES APC data led by my colleague, Selina Nath.³⁴⁸

My results suggest that population density is a stronger predictor of epidemic peak timing than of incidence rates, even after adjustment for IMD. Similar to the results of an RSV epidemic study in Connecticut,¹²⁴ I found that higher population density (residents per km²) was associated with an earlier peak timing. Area-level socioeconomic deprivation had an inverted U-shaped association with peak timing, which may be indicative of shared factors between CCGs on the extremes of socioeconomic deprivation. Figure 4.6 illustrates that the earliest peaks are in South London, Manchester and Birmingham—all urban areas with major national and international transport links. High levels of travel, and increased contact and mixing amongst the population likely play a role in viral spread in these areas. Pitzer et al.¹¹¹ undertook detailed modelling of RSV epidemics across USA, noting associations between vapour pressure, temperature, precipitation and the timing of epidemics. Notably, even in a country with much greater climatic variation than the UK, Pitzer et al.¹¹¹ were unable to account for the finding that RSV activity begins in Florida, leading the authors to contemplate the potential role of population mixing.

4.4.3. Implications of study findings

The findings have implications on preparedness for bronchiolitis infections. Given the gap in timing between peak epidemics across the country, messaging systems could be enacted to

relay warnings from CCGs experiencing early peaked admissions to other areas. These systems could inform the distribution of prophylaxis or future vaccinations, and be used to time simple public health messages, such as warning families to avoid taking their young infants into public spaces or encourage frequent handwashing for older siblings.²³⁶ The results also show that surge capacity plans for paediatric intensive care units would likely profit from real-time information on hospital admissions for bronchiolitis across the country.³⁴⁹ This is particularly important given that RSV infections are not currently captured in real-time in the UK.

Green et al. found that admissions for bronchiolitis have risen substantially over the last few decades and my updated analyses show a continuation of this trend.⁴ As described elsewhere,^{4,16} it is likely that changes in admission thresholds and accessibility of primary care services have driven this increase rather than disease severity or increased incidence of bronchiolitis. By 2016, the admission rate among infants in this study had increased to 58.4 per 1000 infant-years (95% CI 57.8 to 59.1), highlighting the need to focus resources on research and interventions aimed at this condition. Evaluations of alternative models of care that reduce the need for hospital admission for acutely ill infants is of particular importance.^{15,350} Future geospatial studies would profit from the inclusion of factors unavailable in the dataset, such as a measure of admission thresholds and population mixing, to examine these factors further. Additional investigation of the incongruent patterns of admissions in London compared with the rest of the country is particularly warranted, as well as examining the impact of ecological factors such as climate and air pollution on the timing of admission rates.

4.5. Chapter summary

This nationwide descriptive analysis of bronchiolitis admissions has shown variation in the size of the epidemic peak across England after accounting for seasonal components of admission rates. At the regional level, predicted rates of admissions were almost 2.5 times greater in the North of England compared to London. The duration of the epidemic was similar across the country, although there was a 3-week difference in the peak week of admissions at the CCG-level. Approximately one quarter of the variation in admission rates and two-fifths of the variation in peak timing of hospital admissions for bronchiolitis were explained by local demographic and socio-economic characteristics. Implementation of an early warning system could help to prepare hospitals for peak activity and to time public health messages.

Chapter 5. Is socioeconomic position associated with bronchiolitis seasonality?

Recap and chapter overview

Chapter 4 investigated how the timing and volume of hospital admissions for bronchiolitis is associated with demographic characteristics, including area-level socioeconomic deprivation, across England. In this Chapter, I use the same English longitudinal dataset to examine the interaction between seasonality in weekly hospital admissions for bronchiolitis and socioeconomic position at the individual level. This work is the second study to address thesis objective 1:

To describe the association between seasonality, early life socioeconomic position and bronchiolitis admission rates among infants across England.

A research paper based on the content of this chapter has been published in the Journal of Epidemiology and Community Health (included in Appendix 5.1).

5.1. Background

As shown in Chapter 4, there is a sharp peak in hospital admissions for bronchiolitis in England in December, indicating an increased burden on health care services during the winter months.^{4,16,351} The main pathogen causing bronchiolitis is respiratory syncytial virus (RSV); a virus that has a diverse pattern of seasonal changes globally, but is far more prevalent during winter months in temperate Northern climates such as the UK.^{6,109,110} Although the precise mechanisms driving this seasonal pattern are yet to be determined, several interacting virus and host (i.e. human) factors have been proposed.^{2,352} In temperate climates, RSV tend to survive longer in the lower temperatures observed in wintertime, with higher relative humidity also thought to play a role.³⁵² In parallel, this is a period where humans spend more time indoors together, thereby increasing the likelihood of virus transmission. It is also thought that humans are more susceptible to infection at this time owing to a combination of factors including respiratory tract dehydration and vitamin D deficiency.^{77,353}

Research conducted in England and Scotland has shown that children from lower socioeconomic groups are at increased risk of admission to hospital for bronchiolitis.^{4,84,354} However, it is not known whether there are differences in the seasonal pattern of bronchiolitis admissions across socioeconomic groups. It is plausible that, due to social patterning of

factors that affect RSV transmission and infection—including damp, overcrowded housing conditions and a greater prevalence of underlying health problems—^{163,200,202,208} there may also be differential seasonal patterns in hospital admissions for bronchiolitis. Understanding these patterns could help to inform preventive interventions to reduce bronchiolitis admissions, including optimal timing of palivizumab prophylaxis prescribing in different population groups, or targeting of maternal RSV vaccination, when this becomes available in the future.³⁵⁵

The aim of this Chapter is to determine how socioeconomic position is associated with the seasonality of hospital admissions for bronchiolitis amongst infants in England.

There are two objectives:

Objective 5A To calculate the proportion of all infant hospital admissions in England between 2012 and 2016 attributable to bronchiolitis by year, month and socioeconomic group.

Objective 5B To quantify the extent to which seasonality of bronchiolitis admissions is modified by socioeconomic position.

5.2. Methods

5.2.1. Data sources and study cohort

The data source for this study was hospital episode statistics admitted patient care (HES APC) linked to Office for National Statistics (ONS) mortality records as presented in Chapters 3 and 4. I used the same birth cohort as in Chapter 4 (see Section 4.2.1). Briefly, this consisted of children born from 1st January 2011 to 31st December 2016, who were followed from birth or the 1st of January 2012, whichever occurred last. Follow up was until their first birthday, date of migration, 31st December 2016 or date of death, whichever occurred first. Using the infant's unique HESID, data on admissions to hospital within the first year of life were added to birth data. Infants were excluded from cohort analyses if they had missing information, were from a multiple birth or were stillborn.^{16,284} Infants with a non-English address recorded at birth (identified through lower super output area, LSOA) were also excluded to prevent potential loss to follow up. Where a non-English LSOA was recorded in a subsequent infant admission, a censoring date was placed halfway between the date of the infant's last known resident admission and the non-resident admission.

5.2.2. Study outcome

I identified hospital admissions with a diagnosis of bronchiolitis in HES APC using the International Classification of Diseases version 10 (ICD-10) code J21 for acute bronchiolitis.

As explained in Chapter 4 (Section 4.2.2), all J21 subcategories were included in this definition because of the low sensitivity of RSV-specific ICD-10 codes.³⁵⁶ All admissions with bronchiolitis recorded as either the primary or secondary diagnosis for an infant during their first year of life were included in the analyses. To calculate admission rates, any admission for bronchiolitis within 14 days of discharge from another bronchiolitis admission was assumed to be associated with the same infection, and therefore only the first of these admissions was included in the analyses.

The numerator for calculating the total burden of infant admissions attributable to bronchiolitis included all admissions with bronchiolitis recorded as either the primary diagnosis or one of the secondary diagnoses where the patient was ≤ 365 days old. Any admission ≥ 1 day after discharge from a previous admission was counted as a separate admission. To ascertain a denominator for the calculation of the total burden of infant admissions attributable to bronchiolitis, I extracted all emergency and planned hospital admissions (excluding birth admissions) between 1st January 2012 and 31st December 2016 where the patient was ≤ 365 days old.

5.2.3. Covariates

I used LSOA-level IMD 2015 to capture socioeconomic position for each cohort member. IMD was derived from mother's recorded LSOA at the time of delivery or the earliest mention of LSOA in infant's hospital admission. IMD quintiles were used for analyses, which the Ministry of Housing, Communities and Local Government (MHCLG) calculates by ranking LSOAs in England in order of IMD score and dividing them into 5 equal groups.^{357,358} In my analyses, the lowest ranked quintile represents the least deprived and the highest ranked quintile the most deprived fifth of LSOAs in England.

Health and disability is one of the seven domains used to calculate IMD 2015 scores. This domain is constructed from four indicators, two of which (acute morbidity and mood and anxiety disorders) include hospital admission statistics.³³⁸ Given that my outcome is also based on hospital admission data, this could be a source of bias in my study.⁷⁶ I therefore also created IMD 2015-minus the health domain-quintiles to look at the difference in results when the health domain is excluded from IMD. The MHCLG have a purposefully created dataset that enables researchers to recalculate IMD with the exclusion of specific domains.⁶⁵ The scores in this dataset have been standardised by ranking, and then transformed to an exponential distribution. Following the guidance issued by the MHCLG,^{65 p.65} I combined the standardised domain scores with all domain weights (shown in Table 5.1) apart from health for each LSOA as follows:

(income deprivation domain score*domain-weight) + (employment deprivation domain score*domain-weight) + (education, skills and training deprivation domain score*domain-weight) + (barriers to housing and services deprivation domain score*domain-weight) + (crime deprivation domain score*domain-weight) + (living environment deprivation domain score*domain-weight)

I then ranked the scores from lowest to highest and divided them into 5 equal groups to create quintiles. The correlation coefficient between IMD scores with and without the health domain was 0.995.

Table 5.1. IMD 2015 domain weightings⁶⁵

Deprivation domain	IMD domain weighting (%)
Income	22.5
Employment	22.5
Health and disability	13.5
Education, skills and training	13.5
Barriers to housing and service	9.3
Crime	9.3
Living environment	9.3

Infant sex (male or female), month and year of birth were extracted from infant's birth record. Infant's age at admission for bronchiolitis was calculated by subtracting the bronchiolitis admission date from the infant's admission date in their birth episode. Age was split into three groups: <3 months, 3 to <6 months and 6 to <12 months. Government Office Region of residence was used to indicate the region of England within which the infant lived at time of birth.³³⁷ HES APC includes information on additional risk factors such as gestational age, mode of delivery and congenital anomalies; however, as these variables are likely on the causal pathway between socioeconomic position and admission for bronchiolitis,^{137,182,359} they were not included in this analyses. Inclusion of mediators in associational analyses would introduce overadjustment bias into the model.³⁶⁰

5.2.4. Statistical methods

Stata 15.0³¹⁴ was used for data analysis and Microsoft Excel 2013 to create graphs. The Stata code for the harmonic regression analysis can be found in Appendix 5.2.

Bronchiolitis admission rates across the cohort were calculated by dividing the number of new admissions for bronchiolitis by person-time at risk for all infants at risk in the birth cohort, and are expressed as annual admission-based rates per 1000 infant-years. I also calculated person-based rates where, for infants with multiple admissions, only the first admission was counted.

Objective 5A: the proportion of infant hospital admissions attributable to bronchiolitis

To calculate the proportion of all infant hospital admissions attributed to bronchiolitis between 2012 and 2016, I divided the number of admissions with a diagnosis of bronchiolitis by the total number of hospital admissions. Admission data was split by year and month of admission and IMD quintile at time of admission. Birth admissions and infants with a non-English LSOA at the time of admission to hospital were excluded from the analysis.

Objective 5B: the extent to which seasonality of bronchiolitis admissions is modified by socioeconomic position

Event rates were modelled using a Poisson regression model with robust standard errors to account for multiple admissions for the same child.³⁶¹ As in Chapter 4, I did not use negative binomial models because there was no evidence of over-dispersion in events. The impact of seasonality on rates of admission was modelled using a harmonic function of time in weeks using the same method described in section 4.2.4.^{340,341} I also included interaction terms of IMD quintile with the sine and cosine regression coefficients (denoted δ_{1j} and δ_{2j}) to assess evidence of effect modification of IMD quintile by seasonality, formally tested using a Wald X^2 test. Year of admission, region, sex, month of birth, age group and interaction terms of age group with the sine and cosine regression coefficients were selected a priori as covariates in the model to increase precision given their known associations with the outcome as identified in the literature.^{4,189,332} See Box 5.1 for the full model specification.^{340,341}

Using the estimated model coefficients, I calculated the amplitude (log) ($\hat{\gamma}_j$), phase ($\hat{\psi}_j$) and peak week of the epidemic curve of each IMD quintile at reference values of the other covariates using the formulae outlined in Section 4.2.4.³⁴² I also calculated the incidence rate ratios (IRRs) of admissions at the average peak week of the top four IMD quintiles relative to the lowest IMD quintile. The delta method, as implemented by the Stata command *lincom*, was used to calculate 95% confidence intervals for each parameter. Epidemic duration for each IMD quintile was calculated using the same method outlined in Chapter 4 (from the first of three consecutive weeks with increasing predicted rates to the first of three consecutive weeks with decreasing predicted rates).

Sensitivity analysis

To assess whether inclusion of hospital admission statistics in the measure of socioeconomic position in this study (IMD) influences the results, I replicated the harmonic regression analysis with IMD 2015-minus the health domain-quintiles.

Box 5.1. Model specifications

Let $n_{jk}(t)$, the number of bronchiolitis admissions in IMD group j and year k observed at time t (measured in weeks), follow a Poisson distribution with rate $\lambda_{jk}(t) = E(n_{jk}(t))/N_{jk}(t)$ where $N_{jk}(t)$ denote the person-time at risk in IMD group j and year k at time t . I modelled this rate, after log-transformation, as a function of year and time of admission as follows:

$$\begin{aligned} \log(\lambda_{jk}(t)) = & \beta_0 + \beta_1(\sin(2\pi t/T)) + \beta_2(\cos(2\pi t/T)) + \sum_{j=1}^5 \alpha_j I_{\text{IMD}=j} \\ & + \sum_{j=1}^5 \delta_{1j} \sin(2\pi t/T) I_{\text{IMD}=j} + \sum_{j=1}^5 \delta_{2j} \cos(2\pi t/T) I_{\text{IMD}=j} + \sum_{k=1}^3 \theta_k I_{\text{age}=k} \\ & + \sum_{k=1}^3 \theta_k \sin(2\pi t/T) I_{\text{age}=k} + \sum_{k=1}^3 \theta_k \cos(2\pi t/T) I_{\text{age}=k} + \sum_{k=1}^K \theta_k I_{x=k} \end{aligned}$$

Where: T is the length of period within one harmonic cycle (i.e. 1 year = 52.14 weeks); and $I_{X=x}$ is the binary indicator of the variable X taking value x ; K indicates covariates included in the model (e.g. age, sex and year categories). The parameter β_0 is the intercept, β_1 and β_2 are harmonic function coefficients, δ_{1j} and δ_{2j} are IMD group-specific harmonic function coefficients, and the parameters $\alpha_1, \delta_{11}, \delta_{21}$ and θ_1 are all constrained to be zero to deal with the collinearity of the binary indicators.

5.3. Results

5.3.1. Objective 5A: the proportion of admissions attributable to bronchiolitis

The average annual proportion of all infant hospital admissions in England with a diagnosis of bronchiolitis was 15.0% over the five study years (Table 5.2). The proportion increased over time, from 14.2% in 2012 to 16.6% in 2016. The same pattern, at a slightly lower percentage, was present when only hospital admissions with a primary diagnosis of bronchiolitis were included (12.9% in 2012, rising to 14.9% in 2016). The burden of bronchiolitis was concentrated in the winter months. In December 2012 to 2016 combined, 39.8% of total infant admissions had a diagnosis of bronchiolitis recorded compared to 2.3% of August admissions over the same period. Including only primary diagnoses reduced this proportion slightly; 36.4% of all infant admissions in December and 2.0% in August had primary diagnoses of bronchiolitis.

Raw numbers of admissions, overall and with a diagnosis of bronchiolitis, were far higher in the more deprived compared to the less deprived IMD quintiles. Of admissions with a recorded IMD score, 30.5% overall and 33.1% admissions with a diagnosis of bronchiolitis were from infants in the most deprived fifth of LSOAs in England. Proportionally, a slight graded pattern in admissions with a bronchiolitis diagnosis was shown between IMD quintiles, from 14.0% in the least deprived to 16.3% in the most deprived group for admissions with any diagnosis of bronchiolitis.

Table 5.2. Postnatal hospital admissions in children <1 year old, any condition and bronchiolitis-related, by year, month and IMD quintile: England, 2012-2016

	<u>Admissions for any condition</u>		<u>Bronchiolitis admissions</u>			
	N (% of total)		Primary or secondary diagnosis N (% of total)		Primary diagnosis N (% of total)	
Overall (2012-2016)	1,292,649	(100.0)	193,793	(15.0)	175,563	(13.6)
Year	2012	262,411 (20.3)	37,136 (14.2)		33,860 (12.9)	
	2013	252,053 (19.5)	34,264 (13.6)		31,089 (12.3)	
	2014	250,454 (19.4)	35,982 (14.4)		32,512 (13.0)	
	2015	255,866 (19.8)	41,408 (16.2)		37,467 (14.6)	
	2016	271,865 (21.0)	45,003 (16.6)		40,535 (14.9)	
Month	January	112,008 (8.7)	24,116 (21.5)		21,717 (19.4)	
	February	104,485 (8.1)	13,787 (13.2)		12,347 (11.8)	
	March	113,802 (8.8)	11,647 (10.2)		10,401 (9.1)	
	April	101,518 (7.9)	8,077 (8.0)		7,213 (7.1)	
	May	100,649 (7.8)	6,238 (6.3)		5,614 (5.6)	
	June	93,729 (7.3)	4,004 (4.3)		3,535 (3.8)	

	July	96,421	(7.5)	3,432	(3.6)	3,040	(3.2)
	August	85,770	(6.6)	1,990	(2.3)	1,707	(2.0)
	September	90,995	(7.0)	5,358	(5.9)	4,767	(5.2)
	October	109,270	(8.5)	14,181	(13.0)	12,823	(11.7)
	November	135,434	(10.5)	41,794	(30.9)	38,268	(28.3)
	December	148,568	(11.5)	59,079	(39.8)	54,122	(36.4)
IMD quintile	1 Least deprived	173,847	(13.4)	24,338	(14.0)	22,239	(12.8)
	2	195,163	(15.1)	27,864	(14.3)	25,415	(13.0)
	3	229,939	(17.8)	33,233	(14.5)	30,164	(13.1)
	4	287,007	(22.2)	43,223	(15.1)	39,185	(13.7)
	5 Most deprived	388,925	(30.1)	63,583	(16.3)	57,215	(14.7)
	Missing	17,318	(1.3)	1,552	(9.0)	1,345	(7.8)

5.3.2. Cohort description

Of the 3,808,247 singleton children born between 2011 and 2016, 81,234 (2.1%) were excluded from the final study cohort as shown in Figure 5.1. This included 20,905 (0.5%) infants excluded due to non-resident status, 55,025 (1.4%) infants with missing information for region, IMD and/or sex and 4,528 (0.1%) infants with follow-up less than 1 day. The final cohort comprised 3,727,789 singleton infants, of which 48.7% were female, 27.2% were in the most deprived quintile of IMD and 18.9% were born to mothers residing in London (Table 5.3). There are 776 more children included in this cohort compared to Chapter 4 due to slight differences in exclusions due to missing data. The number and rate of hospital admissions for bronchiolitis ($N = 155,485$, 50.1 per 1000 infant-years, 95% CI 49.9 to 50.4) and average follow-up time per child (305 days) remained the same. The overall infant-based admission rate was 46.2 per 1000 infant-years (95% CI 46.0 to 46.5).

Alongside later year of admission and residence in the North West or North East of England as discussed in Chapter 4 (see Table 4.1), rates of admissions for bronchiolitis were highest in the most socioeconomically deprived groups, males, and younger infants. The average admission rate was 78.9 per 1000 infant-years (95% CI 77.9 to 80.0) for infants born in October compared to 33.7 per 1000 infant-years (95% CI 33.0 to 34.4) for infants born in March. Figure 5.2 displays observed weekly rates of admissions by IMD, illustrating differential rates of admissions by level of socioeconomic position across the year and a clear annual peak in December.

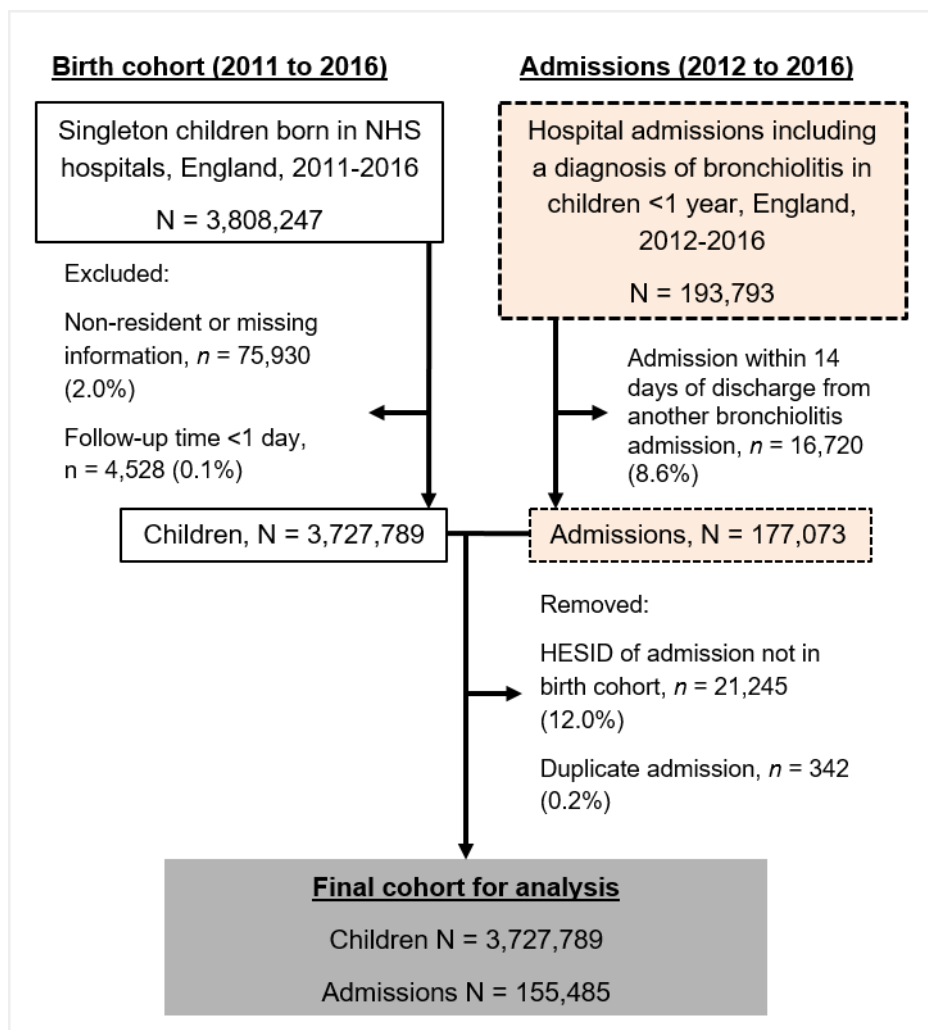


Figure 5.1. Flow diagram to show study participant and hospital admission selection

Table 5.3. Distribution of births and bronchiolitis admissions (per 1000 infant-years) in the cohort

Variable	Infants		Bronchiolitis admissions			
	N	%	Admission-based		Person-based*	
			N	Rate (95% CI)	N	Rate (95% CI)
Total	3,727,789	100.0	155,485	50.1 (49.9, 50.4)	143,206	46.2 (46.0, 46.5)
IMD quintiles						
1 Least deprived	551,740	14.8	18,877	41.1 (40.5, 41.7)	17,585	38.3 (37.7, 38.9)
2	615,852	16.5	22,379	43.6 (43.1, 44.2)	20,842	40.7 (40.1, 41.2)
3	700,171	18.8	27,051	46.4 (45.8, 46.9)	25,023	43.0 (42.4, 43.5)
4	846,766	22.7	35,015	49.7 (49.2, 50.2)	32,089	45.6 (45.1, 46.1)
5 Most deprived	1,013,260	27.2	52,163	62.0 (61.5, 62.5)	47,667	56.7 (56.2, 57.2)
Sex						
Female	1,813,651	48.7	61,349	40.8 (40.4, 41.1)	57,373	38.0 (37.7, 38.4)
Male	1,914,138	51.4	94,134	59.3 (59.0, 59.7)	85,833	54.0 (53.6, 54.3)
Month of birth						
January	310,594	8.3	9,362	36.3 (35.6, 37.1)	8,492	33.0 (32.3, 33.7)
February	284,912	7.6	8,022	33.9 (33.1, 34.6)	7,172	30.3 (29.6, 31.0)
March	305,605	8.2	8,536	33.7 (33.0, 34.4)	7,562	29.9 (29.2, 30.5)
April	298,220	8.0	8,847	35.8 (35.1, 36.5)	7,883	31.9 (31.2, 32.6)
May	316,825	8.5	10,283	39.0 (38.3, 39.8)	9,337	35.5 (34.8, 36.2)
June	310,086	8.3	11,140	43.4 (42.6, 44.2)	10,187	39.7 (38.9, 40.5)
July	325,966	8.7	13,636	50.4 (49.6, 51.3)	12,507	46.3 (45.5, 47.1)
August	319,654	8.6	15,767	59.3 (58.4, 60.2)	14,552	54.8 (53.9, 55.7)
September	322,995	8.7	18,980	70.6 (69.6, 71.6)	17,724	66.0 (65.0, 66.9)
October	322,920	8.7	21,306	78.9 (77.9, 80.0)	20,004	74.2 (73.2, 75.3)
November	304,734	8.2	17,568	69.0 (68.0, 70.1)	16,585	65.3 (64.3, 66.3)
December	305,278	8.2	12,038	46.9 (46.1, 47.8)	11,201	43.7 (42.9, 44.5)
Age at event						
< 3 months			61,926	80.2 (79.5, 80.8)	60,500	78.4 (77.8, 79.0)
3 to < 6 months			45,907	59.3 (58.7, 59.8)	41,785	54.0 (53.5, 54.5)
6 to < 12 months			47,652	30.7 (30.4, 30.9)	40,921	26.4 (26.1, 26.6)
Month of admission						
January			19,154	72.7 (71.7, 73.7)	17,331	65.8 (64.8, 66.8)
February			10,968	45.4 (44.6, 46.3)	9,249	38.3 (37.5, 39.1)
March			9,303	35.3 (34.6, 36.1)	7,599	28.9 (28.2, 29.5)
April			6,536	25.6 (25.0, 26.3)	6,497	25.5 (24.9, 26.1)
May			5,094	19.3 (18.8, 19.9)	4,774	18.1 (17.6, 18.6)
June			3,246	12.8 (12.3, 13.2)	2,910	11.4 (11.0, 11.9)
July			2,801	10.6 (10.3, 11.0)	2,470	9.4 (9.0, 9.8)
August			1,575	6.0 (5.7, 6.3)	1,401	5.3 (5.1, 5.6)
September			4,419	17.4 (16.9, 17.9)	3,938	15.5 (15.0, 16.0)
October			11,593	44.1 (43.3, 44.9)	10,761	41.0 (40.2, 41.7)
November			33,838	133.8(132.4,135.2)	32,123	127.0(125.6,128.4)
December			46,958	177.7(176.1,179.3)	44,153	169.7(168.1,171.3)

*where only the first of multiple hospital admissions for bronchiolitis by the same child is counted.

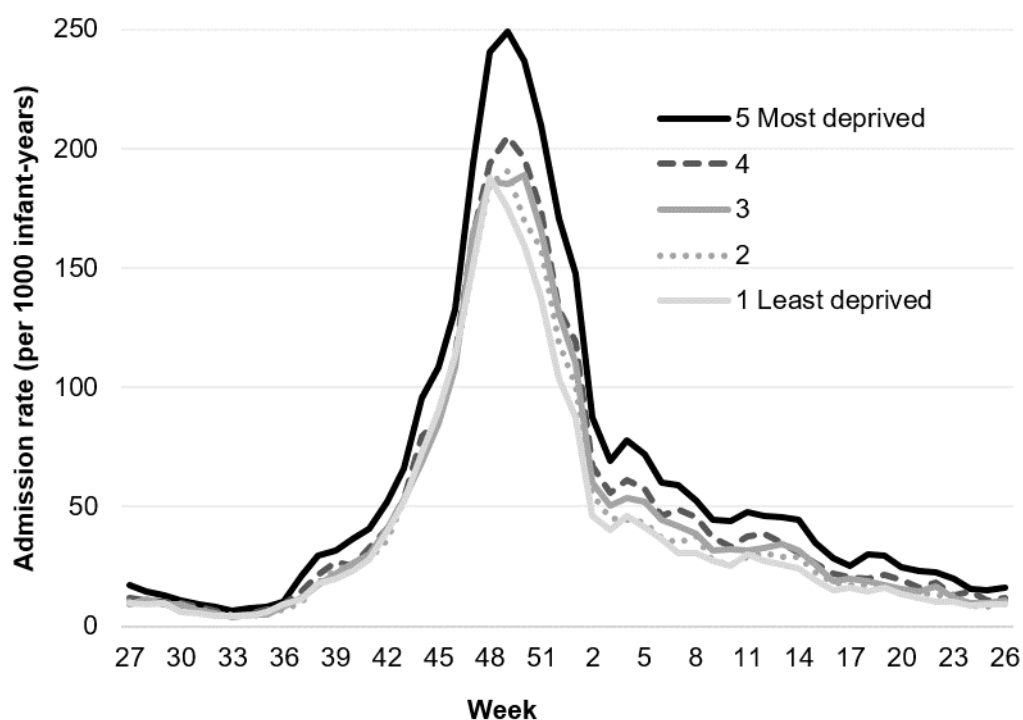


Figure 5.2. Observed weekly rates of hospital admission for bronchiolitis, by IMD group: England 2012-2016

5.3.3. Objective 5B: effect modification by socioeconomic position

The multivariable model that included interaction terms between IMD and the two harmonic functions fitted the data better than the model without interactions, Wald test, $\chi^2(8) = 239.2$, $p < 0.001$ (Table A5.1, Appendix 5.3). Fitted values and seasonal estimates from the model are presented in Figure 5.3 and Table 5.4. The amplitude (that is, the difference in rates of bronchiolitis hospital admissions between the peak and trough of the epidemic curve) was smallest in the most deprived IMD quintiles and largest in the least deprived quintiles. This points to less seasonal variation in the rates of admissions in the more disadvantaged groups. The average peak timing of the annual epidemic varied marginally across IMD quintiles, from week 49.4 (95% CI 49.2 to 49.5) in the least deprived quintile to week 50.3 (95% CI 50.1 to 50.4) in the most deprived quintile. The estimated epidemic duration was 26 weeks across all IMD quintiles. After adjustment for covariates, infants in the most deprived group had a 40% (IRR 1.40, 95% CI 1.37 to 1.43) greater risk of admission to hospital for bronchiolitis at week 50 compared to infants in the least deprived group.

5.3.4. Sensitivity analysis: excluding the health domain from IMD

There was very little difference between the seasonal estimates when the analysis was replicated using IMD-excluding health-quintiles. The amplitude was almost identical across quintiles and the phase shift and peak week estimates remained the same. The point estimates of the IRR of admission rates at week 50, compared to the least deprived quintile, became slightly smaller (moving from a 40% to 37% increased risk in the most extreme example) but confidence interval overlapped considerably.

Table 5.4. Average annual seasonal estimates derived from harmonic Poisson regression models (see Appendix 4.3), by IMD quintile

IMD quintiles	Amplitude* (95% CI)	Phase shift (95% CI)	Peak week (95% CI)	IRR at week 50 (95% CI)
Full IMD				
1-Least deprived	3.97 (3.81, 4.12)	-0.45 (-0.47, -0.44)	49.4 (49.2, 49.5)	Reference
2	3.83 (3.70, 3.97)	-0.41 (-0.43, -0.39)	49.7 (49.6, 49.9)	1.07 (1.04, 1.10)
3	3.75 (3.62, 3.87)	-0.37 (-0.38, -0.35)	50.1 (50.0, 50.2)	1.16 (1.13, 1.19)
4	3.54 (3.44, 3.64)	-0.36 (-0.37, -0.34)	50.2 (50.1, 50.3)	1.25 (1.22, 1.28)
5-Most deprived	3.35 (3.27, 3.43)	-0.35 (-0.36, -0.33)	50.3 (50.1, 50.4)	1.40 (1.37, 1.43)
IMD excluding health				
1-Least deprived	3.96 (3.81, 4.11)	-0.45 (-0.47, -0.44)	49.4 (49.2, 49.5)	Reference
2	3.85 (3.71, 3.98)	-0.42 (-0.43, -0.40)	49.7 (49.6, 49.8)	1.07 (1.04, 1.10)
3	3.71 (3.59, 3.83)	-0.37 (-0.38, -0.35)	50.1 (50.0, 50.2)	1.14 (1.11, 1.17)
4	3.60 (3.50, 3.70)	-0.36 (-0.37, -0.34)	50.2 (50.1, 50.3)	1.23 (1.20, 1.26)
5-Most deprived	3.33 (3.25, 3.41)	-0.35 (-0.36, -0.33)	50.3 (50.1, 50.4)	1.37 (1.33, 1.40)

*amplitude exponentiated

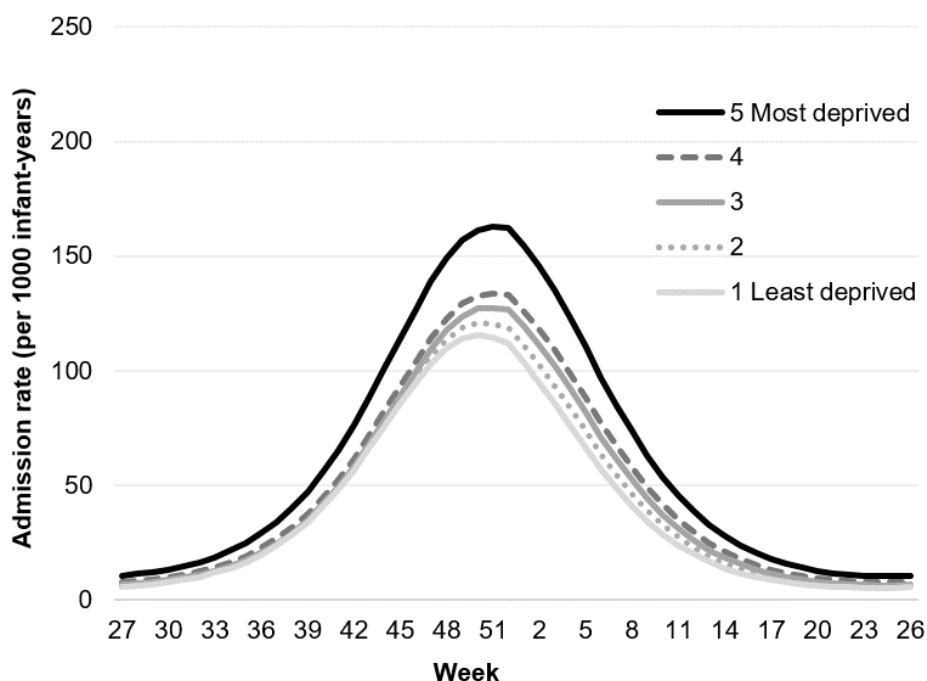


Figure 5.3. Predicted weekly rates of hospital admission for bronchiolitis, by IMD: England 2012-2016

5.4. Discussion

This study adds to the evidence of the burden of bronchiolitis presented in Chapter 4, showing that in December in England between 2012 and 2016 40% of all infant hospital admissions included a bronchiolitis diagnosis. There is a clear socioeconomic gradient to these admissions, and at week 50 (peak admission week) infants born into the highest level of socioeconomic deprivation had a risk of admission 40% greater than the lowest group. The results suggest that the association between seasonality and bronchiolitis admission rates is marginally moderated by level of socioeconomic position, with increasing socioeconomic deprivation associated with less seasonal variation and a slightly delayed epidemic peak.

5.4.1. Strengths and weaknesses

The study data were derived from a national hospital administrative dataset, which enabled me to create a representative birth cohort with minimum selection bias and follow patients up over time. I was able to link the majority of infants to maternal records, which was crucial for acquiring socioeconomic data as this information was missing for infant records before 2014 (but available in maternal records). However, the probabilistic method used is likely to have included erroneous links, introducing some bias into the dataset.³⁶² Using the clinical diagnosis

of bronchiolitis without data on laboratory test results for specific viruses meant that I could not attribute the bronchiolitis admissions to particular pathogens, although previous work from England shows that 80% of bronchiolitis admissions in infants (defined using ICD-10 code J21) are due to RSV.^{4,6} More broadly, as in Chapter 4, this study is a measure of hospital use and therefore does not measure community or primary care burden of bronchiolitis.

I added harmonic functions to the regression model to account for sinusoidal patterns in a highly seasonal infection;³⁶³ however, the model underestimates the size of the amplitude across socioeconomic groups. Improved estimation could have been achieved with the addition of more harmonic pairs or by using splines; however, as mentioned in Chapter 4, these methods would have added undue complexity to the estimation of seasonal parameters required by my research aim.^{332,342} The likeness in the relative difference between amplitudes by IMD across observed and fitted rates (Figures 5.1 and 5.2) gives me confidence in the results.

5.4.2. Research findings: comparison with other studies

Objective 5A: the proportion of infant hospital admissions attributable to bronchiolitis

Overall, between 2012 and 2016, more than 1 in 7 of all hospital admissions in England among infants included a diagnosis of bronchiolitis. During December, this figure rose to 1 in every 2.5 admissions, highlighting the substantial toll this condition takes on hospitals in England during winter. In absolute numbers, infants from the lowest fifth of deprivation made up 33% of the 193,793 hospital admissions with a diagnosis of bronchiolitis between 2012 and 2016. In comparison, 13% of admissions were among children in the highest fifth.

It is important to note that more children in the general population are in more deprived groups; therefore, this measure may be less illustrative than relative measures (that take into account the overall size of the groups). When considered as a proportion of all admissions within the group, the proportions are far less stark; bronchiolitis comprised 16.3% of all admission in the most deprived group compared to 14.0% in the least deprived group. Taking into account the time spent at risk of hospital admission, there was a 1.5-fold difference between the most and least deprived groups (59.8 versus 41.0 per 1000 infant-years).

Objective 5B: the extent to which seasonality of bronchiolitis admissions is modified by socioeconomic position

This work presents a small, but arguably inconsequential, difference in epidemic timing—the two most extreme IMD groups had a 0.9 week difference in predicted peak—and no difference in the relative duration of the epidemic. Research from Western Australia found no difference in the seasonality of RSV positive specimens among aboriginal children (a population that experience high levels of socioeconomic deprivation) compared to non-aboriginal children.³³² USA-based research, on the other hand, found a negative association between the proportion of the population from a black ethnic group (correlated with socioeconomic position of the population) and seasonal peak timing at the ZIP-code level.¹²⁴

Less seasonal variation in lower socioeconomic groups reflects the continued (but low level) admissions for bronchiolitis during warmer months of the year amongst these infants. My study found a greater relative difference in the rates between infants at the lowest point of the epidemic curve (mid-June), where the predicted admission rate was 10.2 per 1000 infant-years in the most deprived compared to 5.2 per 1000 infant-years in the least deprived IMD quintile. This may reflect an increased divide between risk factors for RSV transmission during summer months (such as time spent indoors and overcrowding) in combination with the influence of non-RSV pathogens at this time. Understanding precise viral aetiology of bronchiolitis by time of the year and socioeconomic group may help to delineate seasonal differences further. A study in New Zealand found that non-RSV bronchiolitis was more common in infants from areas of higher deprivation.³⁶⁴ Future work would benefit from linkage to surveillance datasets, such as Public Health England's Second Generation Surveillance or Respiratory Datamart Systems, to assess these factors.^{365,366}

5.4.3. Implications of study findings

In the context of current UK guidelines for administering Palivizumab, the results present little evidence to support differential timing of interventions for bronchiolitis amongst socioeconomic groups in England at the current time. The work does, however, present substantial difference in admission rates across population subgroups irrespective of season. This study highlights that, at the very least, infants from poorer backgrounds should be considered a priority group for future interventions. Based on previous research, I hypothesise that a combination of risk factors, such as tobacco smoke exposure, premature birth, congenital heart and chronic lung disease, outdoor air pollution and damp, overcrowded housing conditions, contribute to these inequities.^{4,45,139,367} Further work using formal methods to establish pathways through which

these factors affect the risk of bronchiolitis is needed to guide the most appropriate interventions.

5.5. Chapter summary

In this chapter, I have highlighted the substantial burden caused by bronchiolitis on English hospitals in winter. This is particularly apparent in December when, on average between 2012 and 2016, almost 40% of all infant admissions included a diagnosis of bronchiolitis. Using harmonic analysis, I found only a marginal moderation in the seasonality of these admissions by level of socioeconomic position. The results suggest that differential timing of interventions may not be necessary for different socioeconomic groups, but, with a social gradient in the volume of admissions, highlight the continued inequity in admissions rates for bronchiolitis more broadly. Moving forward, investigating the precise viral aetiology of bronchiolitis by season and socioeconomic group, as well as the risk factors mediating the link between socioeconomic deprivation and bronchiolitis rates, will help to further unpick the reasons for inequities in rates of hospital admissions for bronchiolitis in the UK.

Chapter 6. Socioeconomic position and bronchiolitis— mediation through maternal smoking

Recap and chapter overview

In chapters 4 and 5, I explored the seasonality of annual bronchiolitis epidemics across England, and associated risk factors for bronchiolitis focussing on socioeconomic deprivation. This work showed a clear socioeconomic gradient to the rates of hospital admissions for bronchiolitis amongst infants in England. Using a dataset constructed from Scottish administrative data, this chapter will look more closely at one modifiable risk factor that sits on the pathway between socioeconomic deprivation and bronchiolitis admissions, answering thesis objective 2:

To establish to what extent maternal smoking mediates the association between early life socioeconomic position and bronchiolitis admission rates in infancy.

I discuss the causal inference framework used to guide the mediation analysis, and employ a novel method to examine plausible causal assumptions of my study. I then discuss the available mediation estimands, before choosing and applying the most appropriate method to answer my research question.

6.1. Background

A social gradient is present in the incidence rates of hospitalisation for bronchiolitis.⁴ My research presented in Chapter 5 shows that children born in England between 2011 and 2016 into the most deprived quintile of IMD had an admission rate of 62.0 per 1000 child-years (95% CI 61.5 to 62.5) compared to 41.1 per 1000 child-years (95% CI 40.5 to 41.7) among children in the least deprived quintile. I hypothesised that a combination of risk factors including tobacco smoke exposure, premature birth, congenital heart and chronic lung disease, outdoor air pollution and damp, overcrowded housing conditions, contribute to these inequities.^{4,45,139,367} This hypothesis is based on pre-existing research where the association between socially patterned risk factors and bronchiolitis admission rates has been examined. However, whilst these associational studies help to describe patterns of bronchiolitis by infant characteristics, they do not break down the pathways through which socioeconomic circumstances influence risk of infection.

Understanding the mechanisms through which an exposure is linked to an outcome is an important component of epidemiological research. Specifically, quantifying the relative magnitude of the impact of intermediate variables could guide options amenable to intervention (as well as the best timing for those interventions).³⁶⁸ This is particularly helpful in instances where intervening on the exposure is less feasible (as is arguably the case with socioeconomic position).³⁶⁹ However, traditional formulations of mediation analyses, which used product of coefficients and difference in coefficients estimators,^{370,371} have been shown to rely on very strong parametric assumptions and cannot be generalised beyond simple linear regression models.^{372,373} The latter are only suitable for settings where all relationships are linear and there are no exposure-mediator interactions, nor non-linearities or multiple interacting mediators.³⁷⁴ In many applied settings, the focus is on more than one mediator, either because the mediating mechanisms of interest are complex, or because the mediator of interest is influenced by earlier mediators (an example of mediator-outcome intermediate confounding).³⁷⁵ These methodological challenges may explain the lack of research on the complex pathways between socioeconomic position and bronchiolitis.

Advances in the causal inference literature, which draws on the counterfactual framework, has produced some solutions to the problem of decomposing the effects of mediators in complex settings.³⁷⁰ Several causal mediation estimands are now available to researchers—their choice dependent on the research question at hand and the plausibility of the required assumptions.³⁷⁶ As I will elucidate in this chapter, implementation of mediation analysis (even when focussing on a single mediator) is complex, requiring careful consideration of the causal structure and associated assumptions of the examined pathways.^{371,377} There are further intricacies specific to my study because of the complexities associated with defining the exposure, socioeconomic position, and the number of mediator-outcome confounders that are associated with this exposure.³⁷⁰ Drawing upon the causal inference framework, the aim of this chapter is:

To establish the extent to which maternal smoking during pregnancy mediates the association between parental socioeconomic position and bronchiolitis hospital admission rates in infancy.

I am focussing on maternal smoking because it is an established risk factor for bronchiolitis that is strongly socioeconomically patterned and, in comparison to other potential mediators such as birthweight and gestational age, presents a pathway that has practical implications in terms of interventions.^{131,138} There are 4 objectives underpinning this aim:

Objective 6A Outline the basic principles and assumptions of causal inference.

Objective 6B Use an evidence based approach to create a causal diagram to inform the analyses.

Objective 6C Choose the appropriate mediation estimand(s) with consideration to both the complexities of the process and the available data.

Objective 6D Based on findings from objective 6A-6C, establish the extent to which maternal smoking explains the socioeconomic disparity in hospital admission for bronchiolitis in infancy.

6.2. Causal inference framework

6.2.1. Basic principles

Typically, observational studies have been used to look at associations—showing the strength of relationship between two or more variables.³⁷⁸ The previous two studies in this thesis, for example, were concerned with the association between the time of year, population factors and hospital admissions for bronchiolitis. This type of analysis enabled me to describe the patterns, and estimate the risk, of bronchiolitis events across England under specified conditions. Judea Pearl writes that standard statistical analysis, which uses samples drawn from a population to estimate parameters, allows researchers to infer associations among variables where conditions remain constant.³⁷⁹ However, the same approaches cannot be applied for situations where we wish to infer risks under changing conditions, such as when an intervention is or is not implemented. To understand risks under these conditions, the focus needs to shift to understanding why the associations exist or, to put it another way, determining that a cause leads to an effect (i.e. causal inference).

Causal interpretation of observational associations can be unconvincing because, unlike in randomised trials, the exposure (or treatment) is not randomly assigned to individuals.³⁸⁰ Another way of putting this, is that association is largely not taken to mean causation in observational studies because the exposed and unexposed are not generally exchangeable.³⁸¹ This is unlike randomised experiments, where randomisation is applied to ensure the exposed and unexposed *are* exchangeable.³⁸⁰ This means that the outcomes observed in one of the randomised groups can be taken to represent the outcomes from the other group if they had received the same exposure/treatment. However, as written by Pearl, the move from association to causation is achievable by including information provided through causal assumptions that “identify relationships that remain invariant when external conditions change”.³⁷⁹ Understanding these causal assumptions, detailed below, can be aided by the use of counterfactual reasoning.

Using counterfactual statements, research questions can be reframed to draw contrasts between outcomes given different sets of conditions. In this chapter, I adopt the potential outcomes framework to define causal parameters of interest (causal estimands),^{370,382–384} e.g. the total causal effect (TCE) on Y (the outcome) of a change in A (the exposure) from a^* to a is defined as $TCE = E[Y(a) - Y(a^*)]$, where E stands for expectation in the population of interest, and $Y(a)$ is the potential outcome had the exposure A been set to take the value a , possibly contrary to the fact, and $Y(a^*)$ is the potential outcome had exposure A been set to take the value a^* , where $a \neq a^*$. More specifically, ‘what would the risk of bronchiolitis be if all infants were born into low versus high levels of socioeconomic deprivation?’ would be expressed as the TCE where a^* and a stand for either low or high socioeconomic deprivation. This type of reasoning can help researchers to consider the effects of an intervention without it being enacted, and highlight where limitations in reasoning and methods exist.^{385,386}

6.2.2. Key assumptions

Hernán and Robins describe conditions that, if met, may informally allow an observational study to be conceptualised as a conditionally randomized experiment.³⁸⁰ I will now define four assumptions, labelled no interference, consistency, positivity and exchangeability, that are usually invoked to derive causal effects with application to the current study.

No interference

No interference assumes that an individual’s potential outcome does not depend on other individuals’ exposure values.^{380,387} The presence of interference would mean that the counterfactual outcomes have more complex definitions and, thus, causal inference also becomes more complex.³⁸⁷ In my study, interference would be present if the risk of being admitted to hospital for bronchiolitis for one infant was dependent on the SEP of other infants. As shown in Chapters 3 and 4, there is a high concentration of hospital admissions for this condition in winter, driven by the seasonality of RSV. Children from lower SEPs do make up a higher proportion of those admitted; however, the continual rise in hospital admission rates for bronchiolitis over time suggests that infants with bronchiolitis are not turned away due to capacity issues. Al-Mahtot and colleagues, for example, showed that the odds of an emergency admission for hospital in Scotland with a diagnosis of bronchiolitis was 24% greater in 2010-13 relative to 2000-03.³⁵⁴ It is suggested that duration of stay may be influenced by bed capacity;³²⁷ however, this is not the outcome I am looking at in my study. For these reasons, I assume that the assumption of no interference is upheld in my study.

Consistency

Under the consistency assumption, write Hernán and Robins, “the values of treatment under comparison correspond to well-defined interventions that, in turn, correspond to the versions of treatment in the data”.³⁸⁰ Put another way, the exposure definition must have enough precision that any variation in that exposure does not lead to a different outcome.³⁶⁹ There are two components to consider here in relation to my exposure: whether SEP can actually be intervened upon; and, if so, whether it is defined well enough to meet the consistency assumption. As discussed at the beginning of Chapter 1, intervening on SEP is not an impossible goal, but neither is it an easy one; requiring major and sustained social change.³⁸⁸ Even after putting the argument about intervening directly upon SEP aside, however, Rehkopf et al.³⁶⁹ argue that composite measures (such as the Scottish IMD, used in this study) are not well defined interventions and, for this reason, violate the consistency assumption. They recommend avoiding composite measures altogether. Other non-composite measures of SEP such as education or income lend themselves better to the constraints of the consistency assumption. However, as explained from the outset of this thesis, IMD is advantageous in the comprehensive and nationally relevant measurement of inequity it provides.

By focussing on the mechanisms underpinning inequities, there are ways to circumvent the strong requirements needed to meet the consistency assumption. Naimi and colleagues show how mediation methods can be used to consider how upstream risk factors for health are explained by proximal risk factors by allowing the statistical association between exposure and outcome to carry “relevant disparity-related (but not necessarily causal) information”.³⁷⁰ This is the approach I will adopt in my study and is discussed in further detail in Section 6.4.2.

The positivity assumption

Positivity means that all individuals have a probability greater than 0 (a positive probability) of being assigned all values of the exposure variable, in every stratum defined by the covariates C that allow for the next assumption to be met, i.e. the confounders.³⁸⁰ In the current study, there must be infants from low and high socioeconomic groups in each stratum of the confounder variables to meet this assumption. Given the large size of the birth cohort to be used in my study, it is likely that this assumption will be met; however, this may be dependent on the number and rarity of confounders. I will revisit this assumption after these aspects have been examined in my dataset.

The exchangeability assumption

Exchangeability, as applied in the current study, assumes that infants born into high and low socioeconomic position groups (i.e. the unexposed and exposed) are comparable in every respect apart from the exposure. If this is not the case, then it cannot be expected that the observed outcome (hospital admission for bronchiolitis) for the exposed is the same as the potential outcome under exposure of the unexposed. As exchangeability is unlikely to occur in observational data, we can instead aim for achieving *conditional* exchangeability in order to elicit a valid causal interpretation, conditional on a set of covariates C . Thus, we assume that there are no uncontrolled confounding variables, which means that within strata defined by these confounders, the exchangeability condition that is guaranteed in randomised controlled trials is met. Hernán and Robins describe this as “the conditional probability of receiving every value of treatment, though not decided by the investigators, depends only on measured covariates $[C]$ ”.³⁸⁰ Causal diagrams (introduced next, Section 6.3) help to identify the confounders necessary to condition on to achieve conditional exchangeability.

6.3. Causal diagrams

The identification of the set of confounding variables C needed to meet the conditional exchangeability assumption can be aided by the use of directed acyclic graphs (DAGs). DAGs are a simple method of graphically representing the assumptions regarding how the variables in a study and their common precursors are causally related to one another.³⁸⁹ DAGs consist of nodes (representing variables) and unidirectional edges (or arrows, representing potential causal effects) that are either present or missing between nodes. Causal DAGs are DAGs that include all common causes of any nodes in the diagram and are drawn starting from the main exposure-outcome relationship. Causal DAGs can be used as a tool to illustrate both the assumptions made in causal analysis and the wider context of the causal questions being asked.³⁹⁰ Such diagrams help to identify: whether there is confounding in the exposure-outcome relationship and which variables need to be controlled for, or not controlled for, when estimating the effects of interest.^{391,392} A missing edge represents the strongest assumption in a DAG: that there is no direct causal effect between two variables.³⁹³

There is a lack of best practice guiding the specification of DAGs and, where studies have used them, the diagrams have been criticised for being overly simplistic and altered to fit available data.^{391,394} Further, DAGs aren't always presented in studies that claim to use them.³⁹⁵ For these reasons, Ferguson and colleagues created an evidence synthesis for constructing DAGs (ESC-DAG).³⁹¹ The ESC-DAG primarily serves to guide DAG development from background knowledge and direct primary data analysis, whilst examining plausible causal assumptions.

Evidence synthesis for constructing directed acyclic graphs (ESC-DAG) application

To apply the ESC-DAG to my research question I started by conducting a search of the relevant literature. An initial search for studies that had looked at all three components of my research question (socioeconomic position, smoking and bronchiolitis) led to just one result, which was a study that looked at the second-hand smoking after birth.³⁶⁴ I therefore broadened my search question to 'how does early life socioeconomic position influence rates of bronchiolitis in early childhood?' (Table 6.1). This enabled me to include more studies and consider other pathways beyond smoking, which is necessary to deal with the potential confounding of the relationship between the mediator (maternal smoking) and the outcome (bronchiolitis admissions).

Table 6.1. Search strategy for literature used in ESC-DAG

Research question	How does early life socioeconomic position influence rates of bronchiolitis in early childhood?
<u>Search strategy</u>	
Databases	Medline, Embase, Maternity and Infant care database (using OvidSP) and PubMed
Search terms	OvidSP: (Social determinants of health/ OR Social status/ OR Social structure/ or Socioeconomics/) AND (bronchiolitis OR human respiratory syncytial virus/). PubMed: (Socioeconomic factors [MeSH term] OR Social class [MeSH term]) AND (bronchiolitis OR human respiratory syncytial virus)
Inclusion criteria	<ul style="list-style-type: none"> • English language • Human • High-income countries • Study participants <2 years • Published ≥2000 • Outcome is bronchiolitis cases or rates • Longitudinal, cohort or prospective research
<u>Identified papers</u>	
Initial number	72 studies
Removed	17 duplicates 46 not relevant (from title/abstract) 73 not relevant (from paper)
Final number	9 studies

Using the list of identified and relevant studies, ordered descending by year of publication, I carried out the 3-stage process outlined below. The process closely follows the ESC-DAG protocol described by Ferguson and colleagues,³⁹¹ with some adjustments that are explicitly noted in the text below:

Stage 1. Firstly, I noted the main aspects of the study, including the study design, statistical methods, study exposure(s), study outcome(s) and control variables, in a decision log (Table A6.1, Appendix 6.1). I used this information to map the conclusions of each study to an initial graph called an implied graph. For each implied graph, the following process was followed: a directed edge was drawn from the exposure to the outcome; all control variables were entered into the graph; mediators were mapped out based on the study conclusions; and, finally, the implied graph was saturated by drawing edges between all nodes. To reduce the high volume of assessments required, I excluded variables at this initial stage that were not applicable to a UK context (and included a reason for the exclusion in the decision log). For example, presence/type of health insurance was used in Inagaki et al.'s study of bronchiolitis in New York State as one of two indicators of SEP,¹⁰³ but is less applicable in the context of the publicly funded NHS in the UK.

Stage 2. The second stage, translation, involved applying causal theory to each relationship in the implied graph, including consideration of temporality, face-validity and prior theory/research. I then applied a counterfactual thought experiment; hypothesising the potential outcome that would occur if all individuals were set to receive the same counterfactual exposure (for example, if all infants were born female). At this stage, a directed edge could be retained, reversed or removed. Importantly, it is a stronger assumption to have a line removed than to let it remain. Again, the theory/evidence was considered in relation to the UK rather than the country where the study was conducted. Each step of the translation was recorded in an Excel spreadsheet (not shown) and summarised in the resultant DAG for each assessed study (Table A6.1, Appendix 6.1). Relationships that had already been assessed in other studies were not re-assessed.

Stage 3. When all papers were assessed, a final index of all directed edges was created (Table 6.2). Similar nodes were combined at this stage to avoid repetition, as well as for purposes of parsimony. For example various socioeconomic measures, including household income, neighbourhood SEP and social exclusion risk,^{103,396,397} were combined into the overarching parental SEP. This index was used to create the integrated-DAG representing the relationship between parental SEP and infant hospital admission for bronchiolitis, shown in Figure 6.1, below. I will use the integrated-DAG to guide confounder selection in the mediation analysis in section 6.5 in this chapter and again in Chapter 7.

DAGitty,³⁹⁸ an online open source graphical tool, was used to create the DAGs presented in Table A6.1 and Figure 6.1, and to determine minimally sufficient adjusted sets for the main analyses.

Table 6.2. Final directed edge index

#	Edge originates from	Edge terminates at	#	Edge originates	Edge terminates
1	Parental SEP	Bronchiolitis	76	Rural/urban	Bronch. dysplasia*
2	Child ethnicity	Bronchiolitis	77	Congenital anomaly	Bronchiolitis
3	Sex	Bronchiolitis	80	Parental SEP	Congenital anomaly
4	Delivery method	Bronchiolitis	81	Sex	Congenital anomaly
5	Child birth season	Bronchiolitis	82	Congenital anomaly	Gestational age
6	Child year of birth	Bronchiolitis	83	Congenital anomaly	Birth weight
7	Child ethnicity	Parental SEP	84	Parity	Congenital anomaly
8	Child ethnicity	Delivery method	85	Congenital anomaly	Delivery method
9	Parental SEP	Delivery method	86	Maternal smoking	Congenital anomaly
10	Child year of birth	Parental SEP	87	Maternal asthma	Congenital anomaly
11	Child year of birth	Delivery method	88	Maternal age	Congenital anomaly
12	Maternal smoking	Bronchiolitis	89	Gestational age	Delivery method
13	Maternal age	Bronchiolitis	90	Maternal smoking	Gestational age
14	Breastfed	Bronchiolitis	91	Maternal asthma	Gestational age
15	Maternal birth country	Bronchiolitis	92	Maternal age	Gestational age
16	Parity	Bronchiolitis	93	Parity	Birth weight
17	Parental SEP	Maternal smoking	94	Birth weight	Delivery method
18	Maternal age	Maternal smoking	95	Maternal smoking	Birth weight
19	Maternal smoking	Breastfed	96	Maternal asthma	Birth weight
20	Maternal birth country	Maternal smoking	97	Maternal age	Birth weight
21	Parity	Maternal smoking	98	Parity	Delivery method
22	Maternal smoking	Delivery method	99	Maternal age	Parity
23	Parental SEP	Maternal age	100	Maternal asthma	Delivery method
24	Parental SEP	Breastfed	101	Maternal age	Delivery method
25	Maternal birth country	Parental SEP	102	Maternal age	Maternal smoking
26	Parental SEP	Parity	103	Maternal age	Rural/urban
27	Maternal age	Breastfed	104	Pre-eclampsia	Bronchiolitis
28	Maternal birth country	Maternal age	105	Gestational diabetes	Bronchiolitis
29	Maternal age	Parity	106	Breech presentation	Bronchiolitis
30	Maternal age	Delivery method	107	Pre-eclampsia	Delivery method
31	Maternal birth country	Breastfed	108	Gestational diabetes	Delivery method
32	Parity	Breastfed	109	Breech presentation	Delivery method
33	Delivery method	Breastfed	110	Maternal smoking	Delivery method
34	Maternal birth country	Parity	111	Maternal asthma	Delivery method
35	Maternal birth country	Delivery method	112	SGA	Delivery method
36	Parity	Delivery method	113	Maternal age	Delivery method
36	Ambient air pollution	Bronchiolitis	114	Parental SEP	Pre-eclampsia
37	Region	Bronchiolitis	115	Parental SEP	Gestational diabetes
38	Maternal asthma	Bronchiolitis	116	Parental SEP	Breech presentation
39	Region	Ambient air pollution	117	Gestational diabetes	Pre-eclampsia
40	Child year of birth	Ambient air pollution	118	Breech presentation	Pre-eclampsia
41	Parental SEP	Region	119	Pre-eclampsia	Gestational age
42	Parental SEP	Maternal asthma	120	Maternal smoking	Pre-eclampsia
43	Maternal ethnicity	Region	121	Maternal asthma	Pre-eclampsia
44	Maternal ethnicity	Maternal asthma	122	Sex	Pre-eclampsia
45	Maternal ethnicity	Maternal smoking	123	Pre-eclampsia	SGA
46	Maternal ethnicity	Maternal age	124	Parity	Pre-eclampsia

47	Maternal asthma	Maternal smoking	125	Maternal age	Pre-eclampsia
48	Child year of birth	Maternal asthma	126	Child year of birth	Pre-eclampsia
49	Maternal age	Maternal asthma	127	Gestational diabetes	Breech presentation
50	Child year of birth	Maternal smoking	128	Gestational diabetes	Gestational age
51	Child year of birth	Maternal age	129	Maternal smoking	Gestational diabetes
52	Gestational age	Bronchiolitis	130	Maternal asthma	Gestational diabetes
52	Birth weight	Bronchiolitis	131	Sex	Gestational diabetes
53	SGA	Bronchiolitis	132	Gestational diabetes	SGA
54	Bronch. dysplasia	Bronchiolitis	133	Parity	Gestational diabetes
55	Parental SEP	Gestational age	134	Maternal age	Gestational diabetes
56	Parental SEP	Birth weight	135	Child year of birth	Gestational diabetes
57	Parental SEP	SGA	136	Gestational age	Breech presentation
58	Parental SEP	Bronch. dysplasia	137	Maternal smoking	Breech presentation
59	Parental SEP	Rural/urban	138	Maternal asthma	Breech presentation
60	Gestational age	Birth weight	139	Sex	Breech presentation
61	Gestational age	SGA	140	SGA	Breech presentation
62	Sex	Gestational age	141	Parity	Breech presentation
63	parity	Gestational age	142	Maternal age	Breech presentation
64	Gestational age	Bronch. dysplasia	143	Child year of birth	Breech presentation
65	Birth weight	SGA	144	Maternal smoking	SGA
66	Sex	Birth weight	145	Maternal asthma	SGA
67	parity	Birth weight	146	Sex	SGA
68	Birth weight	Bronch. dysplasia	147	Maternal age	SGA
69	Sex	SGA	148	Child year of birth	SGA
70	parity	SGA	149	Child year of birth	Parity
71	SGA	Bronch. dysplasia			
72	Rural/urban	SGA			
73	Sex	Bronch. dysplasia			
74	parity	Bronch. dysplasia			
75	Rural/urban	parity			

**bronch. dysplasia = bronchopulmonary dysplasia*

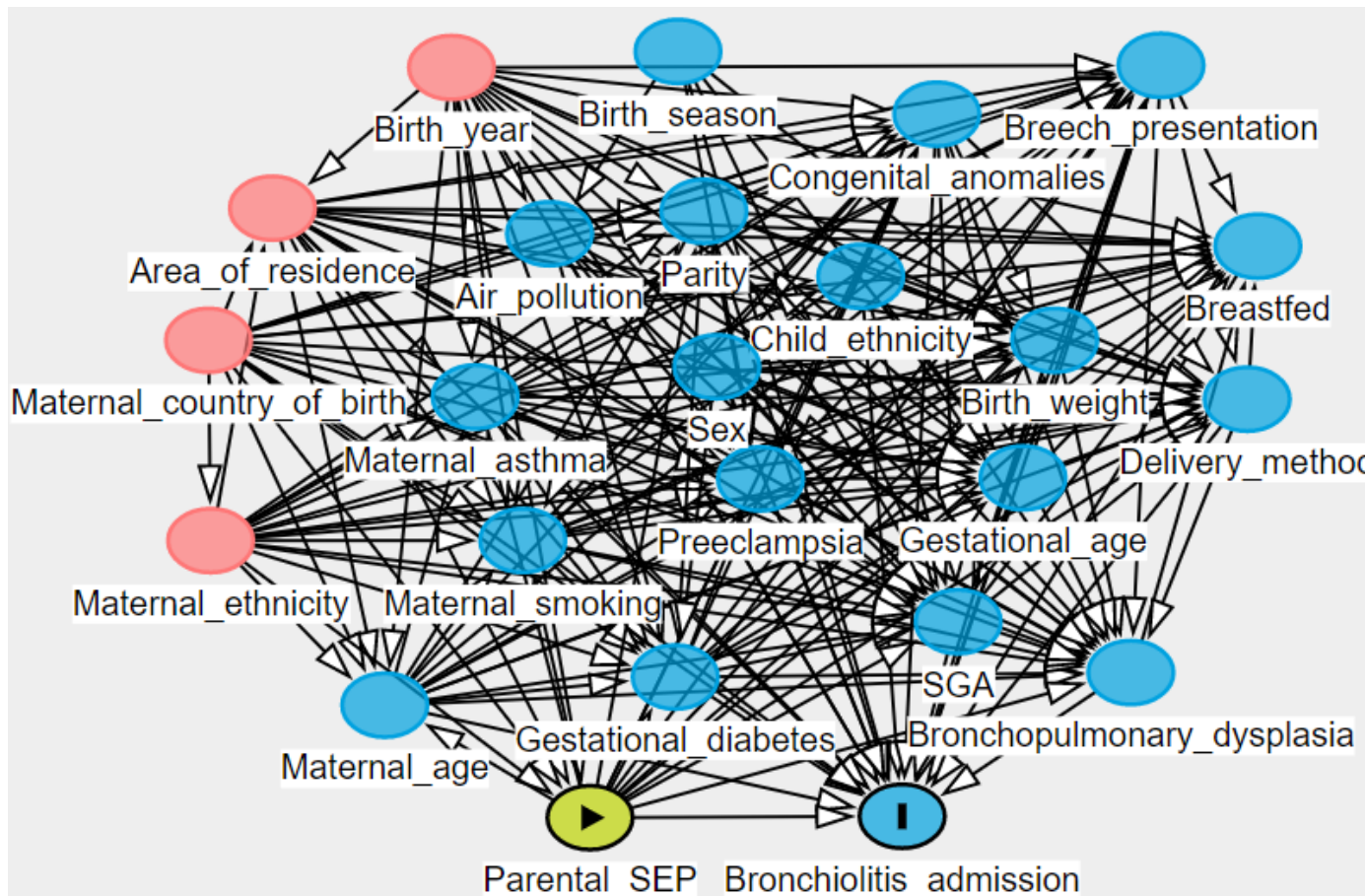


Figure 6.1. Overall integrated-DAG created through the ESC-DAG protocol*

*DAGitty identified the minimally sufficient adjustment set for estimating the direct effect of parental SEP on hospital admission for bronchiolitis as:

- area of residence,
- birth year,
- maternal country of birth
- and maternal ethnicity

These are indicated by red nodes in the diagram.

The yellow node is the exposure, the blue node with a black outline is the outcome and all other blue nodes are ancestors of the outcome (and not confounders of the exposure-outcome relationship).

6.4. Causal mediation analysis

6.4.1. Controlled, natural and interventional (in)direct effects

There are different approaches to study mediation in the causal literature, some of which attempt to separate the total effect of an exposure on an outcome into effects that act directly from the exposure or via a particular mediator (or set of mediators), respectively, the direct and indirect effects. In my study, the indirect effect of interest is the effect of socioeconomic position on the risk of hospital admission for bronchiolitis that involves maternal smoking. I will now present the main approaches to mediation analysis, discussing the applicability of each to my research question.

Controlled direct effects

Controlled direct effects (CDEs) quantify the effect of the exposure when the mediator is set to a specific value for everyone in the population.³⁸³ The CDE of A on Y , not mediated via M when expressed as a linear contrast, is defined as $CDE(m) = E[Y(a,m) - Y(a^*,m)]$, where $Y(a,m)$ is the potential outcome when A is set to take the value a and M the value m . This is the contrast between the potential outcomes under exposure $A=a$ versus $A=a^*$, where the mediator is set at the same fixed level for everyone in the population ($M=m$).³⁷⁷ The $CDE(m)$ may vary with values of m . Moreover the CDE could be expressed on different scales, for example as ratios of expected potential outcomes (when the outcome is binary).

CDEs are described as relevant to interventional settings where policy evaluation is of interest, in particular when the mediator is binary or categorical, such as maternal smoking in my study.^{383,399,400} However, it is difficult to ascertain effect decomposition with CDEs, with Naimi et al. describing the controlled indirect effect as “some contrast between the total and controlled direct effects in the absence of exposure-mediator interactions”.^{399, p.1657} This is particularly problematic where more than one mediator is present.³⁷³ Alongside the two technical assumptions of no interference and consistency for the outcome, the two main identification assumptions required to ascertain CDEs are:⁴⁰⁰

1. no uncontrolled (or unmeasured) confounding of the exposure-outcome relationship
2. no uncontrolled (or unmeasured) confounding of mediator-outcome relationship

Natural (in)direct effects

Another causal mediation strategy quantifies effects based on exposure and mediator values that naturally occur in the population.³⁸³ Pearl writes that these “natural effects” are a descriptive and more subtle interpretation of effects compared to the CDEs.³⁸³ The natural direct effect (NDE) of A on Y , not mediated via M (when expressed as a linear contrast), is defined as $NDE = E[Y(a, M(a^*)) - Y(a^*, M(a^*))]$, where $M(a^*)$ is the potential mediator had the exposure been set to take the baseline value a^* . This is the contrast between potential outcomes under exposure $A=a$ versus $A=a^*$, where the mediator is set to the value it would have taken under a reference (i.e. “natural”) exposure value ($M=a^*$). These values are specific to each individual, unlike the mediator value m used in the CDE definition. The natural indirect effect (NIE) of A on Y that operates via the mediator M is defined as $NIE = E[Y(a, M(a)) - Y(a, M(a^*))]$. This is the contrast between the potential outcomes where A is set to the same exposed value ($A=a$), but the mediator is set to take the value it would have taken had the exposure been set to exposure ($A=a$) or no exposure ($A=a^*$).^{383 399}

These definitions allow for effect decomposition, with the NDE and NIE summing to the TCE. Natural effects are seen as particularly appropriate for providing information on the actions of mechanisms.^{401,402} However, the identification assumptions required for natural effects are considerably stronger than for CDEs, making this method also unsuitable for my research question.⁴⁰⁰ Alongside the two conditionally exchangeability assumptions listed for CDEs (no uncontrolled confounding of the exposure-outcome nor the mediator-outcome relationships), and consistency for mediator and outcome, estimating natural effects additionally requires that:

3. there is no uncontrolled confounding of the exposure-mediator relationship,
4. the potential outcome under exposure and set mediator value, $Y(a, m)$, and the potential mediator under no exposure, $M(a^*)$, need be independent conditionally on the baseline confounders.

This fourth assumption is known as the “cross-world independence assumption” because it leads us to simultaneously contemplate worlds where an individual is exposed (the potential outcome under exposure) and unexposed (the potential mediator under no exposure).⁴⁰³ In my study, this would require imagining an infant born to parents with high socioeconomic deprivation and, at the same time, the same infant born to parents with low socioeconomic deprivation. As this “recanting witness” exposure assignment could not occur in real life,⁴⁰⁴ the reliance of causal mediation analysis on this assumption has been criticised as being empirically unverifiable.⁴⁰⁵ This assumption however implies that there are no variables that

are downstream from the exposure and that confound the mediator-outcome relationship. This may be a defensible assumption in certain settings where there is a small time-gap from exposure to mediator.⁴⁰² For example, if an educational video about smoking (exposure) led to a quick change in attitude (mediator), which, in turn, led to smoking cessation (outcome). In our application, however, this assumption is less defensible. Indeed, it is argued that this assumption diminishes the public health relevance of the NDE and NIE.⁴⁰³

Randomised interventional analogues of the NDE and NIE

Alternative mediation estimands have been introduced that can be identified under weaker assumptions than the NDE and NIE, for example, the randomised interventional analogues (RIA) of the NDE and NIE.³⁸⁴ The definitions for these estimands are the same as for the NDE/NIE, above, but differ in respect to the values the mediators are set. RIA-NDE= $E[Y(a, M_{1c}^*) - Y(a^*, M_{1c}^*)]$, a quantity that captures the direct effect of A on Y not involving M , when the contrast is between the potential outcomes under exposure $A=a$ versus $A=a^*$, where the mediator is set to take a value that is randomly drawn from its conditional counterfactual distribution under no exposure, conditional on the confounders C . It is interpreted as the effect of A that would remain if the distribution of the mediator M could be changed to be the same as the distribution of the mediator had A been set to a^* . RIA-NIE= $E[Y(a, M_{1c}) - Y(a, M_{1c}^*)]$, captures the effect of A on Y operating via the mediator M , where the value of the mediator is set to take a value randomly drawn from its conditional counterfactual distribution under $A=a$ versus a value drawn from its conditional counterfactual distribution under $A=a^*$, while the exposure is set to remain fixed at the same value a .

I present an example where these mediational effects are estimated in the presence of a single mediator, gestational age, in Box 6.1.³⁸⁴ I estimated the RIA-NIE of gestational age as the counterfactual difference in bronchiolitis admission rates between children born in the high versus low deprivation group. Here, gestational age (in weeks) is drawn at random from the distribution of gestational age among infants in the low socioeconomic deprivation group, conditionally on confounders, whilst levels of socioeconomic deprivation remained high. The effects do not lead to a decomposition of the TCE but, as reflected by Daniel and De Stavola, there is trade-off to be made between “how closely an estimand corresponds to what we truly wish to learn about, and the strength of the assumptions needed to identify it”.^{373 p15} Despite not providing insight into all pathways, this method is arguably more applicable to real world and policy-relevant scenarios.⁴⁰⁶

Box 6.1. Could socioeconomic inequity in bronchiolitis admissions be reduced through intervening on gestational age?*

Aim To establish the extent to which socioeconomic inequity in hospital admissions for bronchiolitis would be reduced if socioeconomic disparities in gestational age were eliminated. In this study, socioeconomic inequity refers to the difference in outcomes between infants born to parents in high versus low levels of socioeconomic deprivation.

Methods Children born in English NHS hospitals between 2011 and 2016 were followed up for one year using HES APC records ($N=3,948,819$). The exposure, socioeconomic deprivation, was measured using the Index of Multiple Deprivation 2010 score assigned to the infant's registered lower super output area at birth (split into high or low levels of deprivation). Gestational age (in weeks) and sex (male or female) were also retrieved from hospital records at birth. Presence of congenital anomalies (yes/no), based on the Harlelid UK chronic condition ICD-10 code list,¹ were identified in infant hospital admissions or death certificate up to age two. The study outcome was any emergency admissions to hospital with a diagnoses of bronchiolitis in the first year of life (identified in hospital records by ICD-10 code J21). I estimated the RIA-NIE of gestational age using g-computation (as implemented in Stata using the gformula command), while controlling for the confounders of the mediator and outcome association—congenital abnormalities and sex. I also estimated the total causal effect (TCE) of socioeconomic deprivation on hospital admission rates for bronchiolitis, controlling for the same confounders.

Findings On average, between 2012 to 2016, the bronchiolitis admission rate was 41.0 per 1000 infant-years (95% CI 40.6 to 41.4) among infants in the low deprivation group compared to 51.6 per 1000 infant-years (95% CI 51.2 to 52.0) in the high deprived group. Assuming no unmeasured confounding, the estimated TCE indicates that, if socioeconomic disparities could be removed, the rate of admissions overall would decrease by 9.68 per 1000 infant-years (95% CI 9.17 to 10.18). The estimated RIA-NIE indicates that if socioeconomic disparity remained but the distribution of gestational age were set to that of the low deprivation group, the rate of admissions would decrease by 1.40 per 1000 infant-years (95% CI 1.31 to 1.42). This is the equivalent to removing 14.0% (95% CI 13.2 to 14.8) of the difference in admission rates due to socioeconomic deprivation.

**An earlier version of this abstract was published as a meeting abstract in the Lancet (included in Appendix 6.2).*

6.4.2. Multiple mediators

The methods and example presented above consider only a single mediator. However, as Steen and colleagues argue, even where researchers are interested in studying a single mechanism, the presence of other mediators cannot simply be ignored.³⁷⁵ Vanderweele describes an informal approach to multiple mediators whereby mediators are assessed one at a time and the proportion mediated summed to get the total indirect effect.³⁷¹ However, for this method to produce reliable estimates, mediators must be conditionally independent given the exposure and baseline confounders.³⁷⁴ As shown in Figure 6.1, mediators on the pathway between SEP and bronchiolitis, such as maternal asthma and maternal age, are not causally independent from maternal smoking, even after controlling for possible confounders. The conditional independence assumption is therefore broken, which means that I cannot simply replicate the analysis in Box 6.1 with alternative mediators using the RIA approach. This would likely produce estimates that together do not meaningfully correspond to the total association.

A potential solution to the complex issue of intermediate confounding lies with counterfactual disparity measures (CDMs).³⁷⁰ By setting the mediator to a predefined value (e.g. $m=0$), CDM captures the proportion of outcome disparity due to the exposure that would remain if a mediator (i.e. maternal smoking) were intervened upon and set to a chosen value m , without intervening on A .⁴⁰⁷ Formally defined, $CDM(m)=E[Y(m)|A=1]-E[Y(m)|A=0]$, where $Y(m)$ is the potential outcome had the mediator M been set to take the value m , and where “ $|A=1$ ” and “ $|A=0$ ” indicate respectively that the expectation is in the exposed/unexposed subpopulation. CDMs have two major benefits: they can still be identified in cases where there are confounders of the mediator-outcome relationship that are on the causal pathway from the exposure (such as maternal asthma and maternal age), as long as they are controlled for; and the identifiable assumption of no uncontrolled confounding of the exposure-outcome and exposure-mediator relationships are not required.^{370 408} This is because we wish only to characterise the overall disparity with this measure and therefore we are not concerned with whether the exposure-outcome association is confounded. The assumption of no unmeasured mediator-outcome confounding still applies. Although confounders of the exposure-outcome relationship are not required when estimating the CDMs, they can be included produce comparisons adjusted for these variables.³⁷⁰

To address my aim, and given the difficulties in meeting the assumptions to be invoked to estimate natural effects (which would allow me to establish the extent to which maternal smoking mediates the association between early life socioeconomic position and bronchiolitis admission rates in infancy), I will instead focus on estimating the CDM of setting the mediator to be equal to non-smoking. This requires a reformulation of objective 6D, which becomes

“establish the magnitude of socioeconomic disparity in infant hospital admission for bronchiolitis that would remain if no women smoked during pregnancy”, defined as $CDM(m=0) = E[Y(m=0)|A=a] - E[Y(m=0)|A=a^*]$, where A = socioeconomic deprivation (a if high/medium, a^* if low), M = maternal smoking (1 if yes, 0 if no) and Y = hospital admission for bronchiolitis within the first year of life (1 is yes, 0 if no). This captures the contrast between the potential risks of hospital admission for bronchiolitis under high compared to low levels of socioeconomic deprivation, if all mothers were non-smokers.

In the following analyses, I will first identify, and then adjust for, confounders of the relationships between socioeconomic deprivation and smoking, and smoking and bronchiolitis cases.

6.5. Study cohort and variable selection

6.5.1. Study DAG

I adapted Figure 6.1 to look at confounders of the exposure-outcome (parental SEP-bronchiolitis admission) and mediator-outcome (maternal smoking-bronchiolitis admission) relationships. The final study DAG is shown in Figure 6.2. To satisfy the assumption of no unmeasured mediator-outcome confounding, the minimal adjustment set is area of residence, maternal ethnicity (or child ethnicity), maternal country of birth, year, maternal age, maternal asthma, and parity. Confounders of the exposure-outcome relationship are area of residence, year, maternal ethnicity and maternal country of birth. Other covariates in the ESC-DAG (Figure 6.1) are not included here because they are all downstream from parental SEP and maternal smoking so are therefore not confounders in either the exposure-outcome or mediator-outcome relationships.

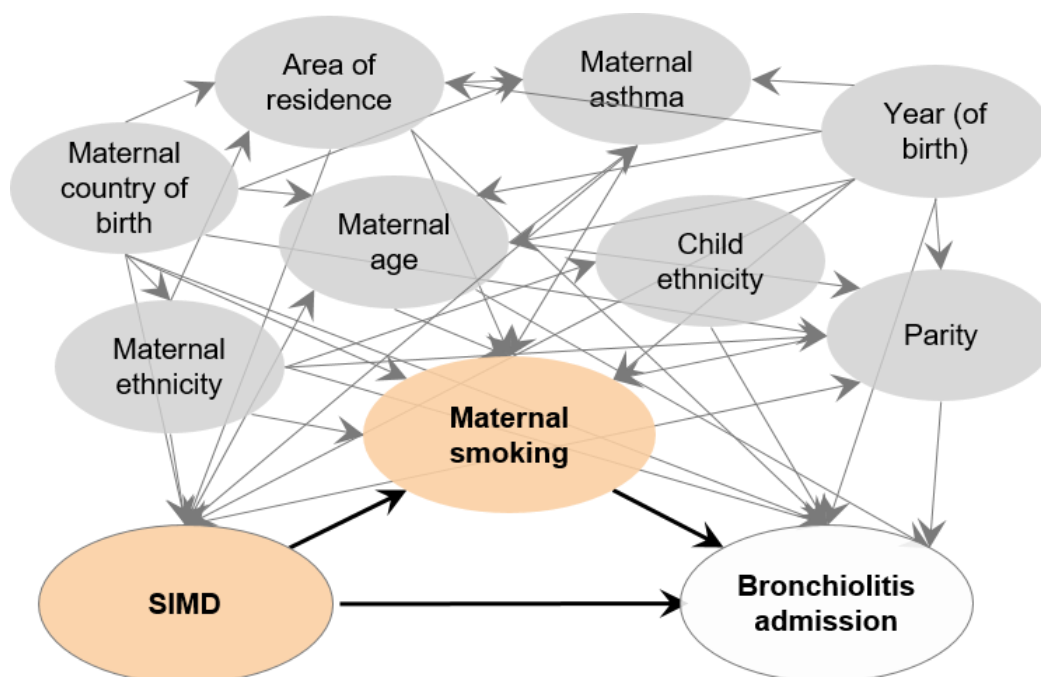


Figure 6.2. Assumed causal DAG*

*DAG estimating the study exposure-outcome (parental SEP-bronchiolitis admission) and mediator-outcome (maternal smoking-bronchiolitis admission) associations;

SIMD (high versus medium and low groups) indicates socioeconomic inequity; maternal smoking refers to the period during pregnancy.

6.5.2. Data sources and study cohort

I used a subsection of the birth cohort of Scottish children described in Chapter 3 as the cohort for this study. I created this cohort using several linked administrative health datasets obtained via the Electronic Data Research and Innovation Service (eDRIS), who assigned a unique, pseudonymised identifier to each individual and deterministically linked children to their mothers via SMR-02. This allowed me to link records belonging to the same child and to mother-child pairs throughout the datasets. The linked datasets that formed the cohort are shown in Table 6.3.

Births are required by law to be registered with NRS within 21 days of the event and this register was used to form the core spine of the cohort.²⁹¹ The birth registration data were supplemented by the mother's delivery record, which was retrieved from the SMR02, a dataset submitted by maternity hospitals on inpatient and day case activity relating to obstetrics. Information about deaths was retrieved from NRS death registrations and emigrations from

the CHI registry. Non-birth related hospital admissions in children were retrieved from SMR01, which includes information on inpatients and day cases discharged from hospitals in Scotland.⁴⁰⁹ SMR01 records include diagnostic, procedure and operation information, as well as non-clinical information such as location and length of stay. Data were also extracted from PIS, which contains information on medicines prescribed and dispensed within the community setting in Scotland (not including hospital administered/dispensed medication).²⁹⁰

The cohort is defined by children born between 1st January 2012 and 31st December 2016 in Scotland. Their follow-up began at birth and continued until the child's 1st birthday. Children born to mothers of non-Scottish residency (identified through mother's country of residency in the NRS birth registrations) were excluded from the dataset to prevent potential systematic loss to follow up. Children had to be alive for at least one day to be included in the cohort. One child from each non-singleton birth was randomly selected to be included in the study.

Table 6.3. Scottish datasets used to create the study cohort for this study

Acronym	Dataset	Information	Variables
NRS	National Records of Scotland birth & death registrations	Child	Postal area, SIMD, date of birth, date of death, mother country of birth, mother country of residence
CHI registry	Community Health Index registry	Child	Date of emigration
SMR02	Scottish Morbidity Records – obstetric related hospital admissions	Mother	Smoking status, age at delivery, parity, SIMD
SMR01	Scottish morbidity records—general hospital admissions	Child and mother	Diagnoses, date of admission, date of discharge
PIS	Scottish national prescribing information system	Mother	Medication name, date dispensed

6.5.3. Study variables

Outcome

The study outcome was any emergency hospital admission with a main or secondary diagnosis of bronchiolitis in the first year of life (yes/no). Bronchiolitis admissions were identified in SMR01 by the ICD-10 code J21 as a primary or any of the five secondary diagnosis codes.

Exposure

I used Scottish Index for Multiple Deprivation (SIMD) 2012 to capture socioeconomic position for each cohort member. SIMD is a relative measure of deprivation based on small areas called data zones (the Scottish equivalent of lower super output areas).⁶⁴ SIMD combines scores of data zone-level deprivation across seven domains: income, employment, health, education, access to services, crime and housing. SIMD 2012 deciles (which are based on the data zone of residential addresses) were retrieved from NRS birth registration files and supplemented from SBR or SMR02 where missing. Deciles are constructed by ranking data zones by SIMD scores and dividing these into ten equal groups. SIMD deciles were split into three groups for analyses indicating high (top 30% rank of SIMD scores), medium (middle 40% rank of SIMD scores) and low levels of socioeconomic deprivation (bottom 30% rank of SIMD scores). Three groups as opposed to five were used because of the complications involved in estimating and interpreting results from the mediation analyses. I could not construct exact SIMD tertiles from the data because raw SIMD scores or data zones (to assign SIMD scores manually) were not available in my dataset.

Mediator

Mothers smoking status, stored in the SMR02, was recorded by midwives at first antenatal appointment (at around 8-12 weeks gestation). A binary variable, maternal smoking (yes/no), was created from this information with yes signifying that traditional cigarettes (i.e. non-e-cigarettes) were smoked at any point during their pregnancy and no for no smoking during pregnancy. E-cigarette smokers and women who smoked prior to their pregnancy were placed in the non-smoking category.

Confounders

Date of birth and postcode area were retrieved from each child's birth registration file (NRS). There are 16 postcode areas in Scotland, which were grouped to make 5 regions with similar population sizes as shown in Figure 6.3. SMR02 provided mother's age at delivery, which was split into four groups (≤ 19 , 20-29, 30-39 and ≥ 40 years) and older siblings (dichotomised into yes or no). Mother's country of birth also was retrieved from the NRS database and dichotomised into UK or non-UK born. Maternal asthma was defined using SMR01 hospital admission and PIS data. Mothers were recorded as having asthma during pregnancy (yes/no) if, in the year preceding their child's birthdate, they had:

- ≥ 1 hospital admission that included a diagnosis of asthma (ICD-10 code J45, including all subcategories);
- and/or ≥ 2 asthma medications dispensed. Medication was defined using the British National Formulary (BNF) and included adrenoreceptor agonists (BNF codes 3.1.1), theophyllines (BNF codes 3.1.3), inhaled corticosteroids (BNF codes 3.2), cromoglicates and related therapies (BNF codes 3.3.1) and leukotriene antagonists (BNF codes 3.3.2).⁴¹⁰

Ethnicity was not available in infant birth records and maternal ethnicity was missing in SMR02 for >25% of the cohort so this variable was not included in the main analysis.

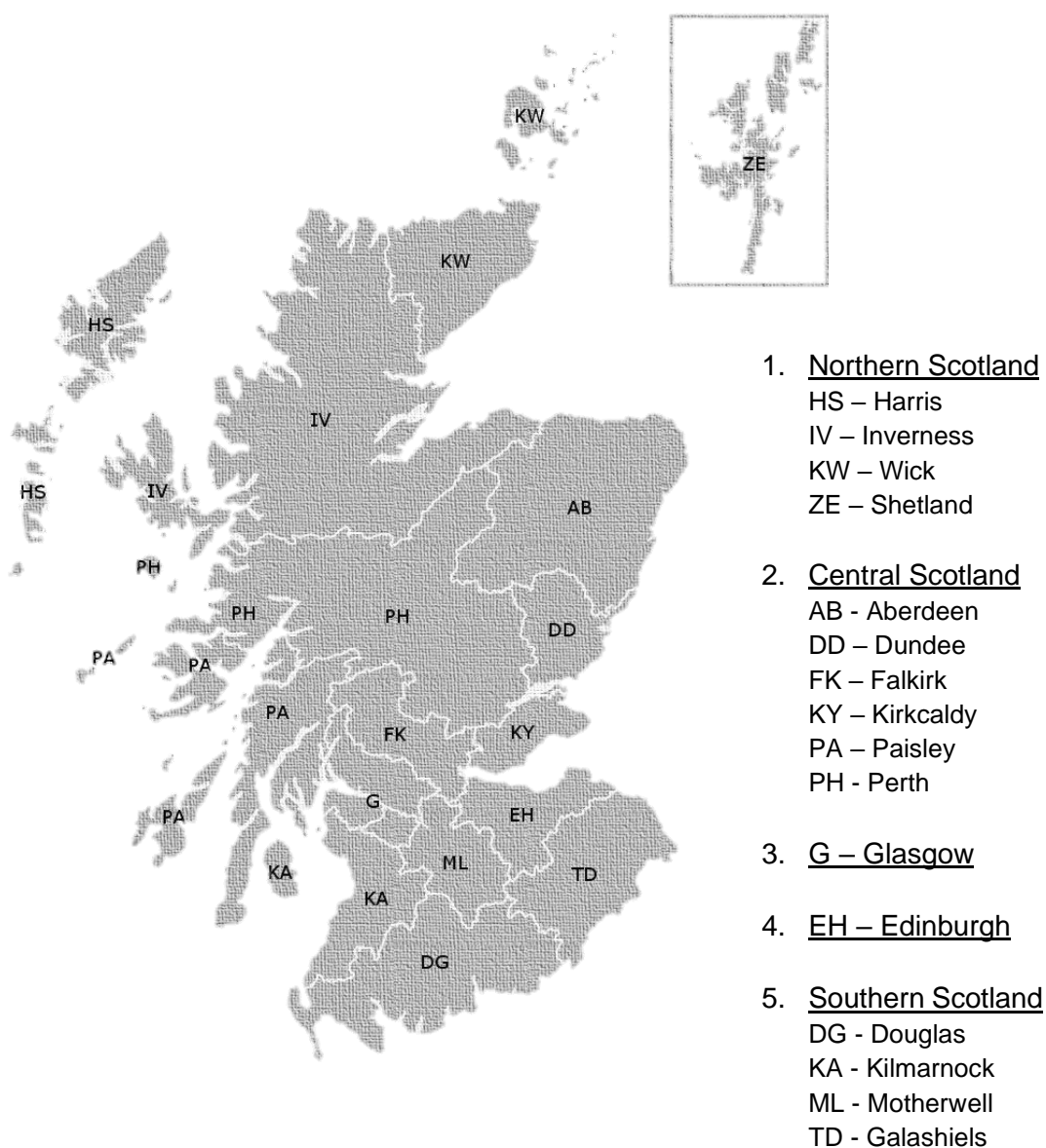


Figure 6.3. Map of Scotland, by postal area; key showing 5 grouping used in this study

6.6. Statistical methods

Cohort characteristics and missing data were firstly described using descriptive statistics. The association between covariates and the probability of missingness in at least one study variable was explored using multivariable logistic regression. These findings were used to decide how to treat missing data in this study.

6.6.1. Admission rates

Replicating the procedure for the English birth cohort (Chapters 3 and 4), I calculated the incidence rates of bronchiolitis admissions per 1000 infant-years. All admissions (i.e. not just the first admission) with bronchiolitis recorded as either the primary or secondary diagnosis for an infant during their first year of life were included in the analyses. To calculate admission rates, any admission for bronchiolitis within 14 days of discharge from another bronchiolitis admission was assumed to be associated with the same infection, and therefore only the first of these admissions was included in the analyses. Children were followed from birth or 1st January 2012, whichever was later, to their 1st birthday, date of death, date of emigration or 31st December 2017, whichever came first. Incidence rates were calculated by dividing the number of admissions by person-time at risk and expressed as admission-based rates per 1000 infant-years for each of the key study variables.

6.6.2. Disparities in bronchiolitis admissions

I used two approaches to estimate the magnitude of socioeconomic disparity in hospital admission for bronchiolitis that would remain if smoking during pregnancy were eliminated (the $CDM(m=0)$). Namely, inverse probability weighting (IPW) of marginal structural models (MSMs) and g-estimation of structural nested models. Both are part of the family of methods known as Robins' generalised methods or g-methods, which can provide estimates of contrasts of potential outcomes with less of the identification conditions required by standard regression models.⁴¹¹ Alongside the $CDM(m=0)$, I quantified the marginal risk of bronchiolitis by SIMD group and the marginal risk difference between SIMD groups. The $CDM(m=0)$ and risk difference were used to calculate the proportion of the socioeconomic disparity in bronchiolitis that can be explained by maternal smoking. I computed bootstrap confidence intervals for these estimates using 1000 bootstrap replications. Stata 15.0 was used for data analysis,³¹⁴ the code for which can be found in Appendix 6.3.

In the following two approaches:

- the CDM($m=0$) estimates capture the potential risk differences between high and low SIMD group, and medium and low SIMD group, in ≥ 1 hospital admission with a diagnosis of bronchiolitis during infancy per 1,000 live births if no mothers smoked during pregnancy;
- confounders of the mediator-outcome relationship were year of birth, maternal age, maternal country of birth, maternal asthma and area of residence;
- and confounders of the exposure-outcome relationship were maternal country of birth, year of birth and area of residence (which were included to calculate average effects across these characteristics).

Inverse probability weighting of marginal structural models

IPW of MSMs can be used to estimate CDMs on the assumption that the mediator model is correctly specified as a function of all mediator-outcome confounders (in addition to assumptions of no interference, consistency, positivity and conditional exchangeability, see Section 6.2.2).³⁷⁰ To calculate the CDM($m=0$) using this approach, I firstly derived two sets of weights (see also Box 6.2). The first set of weights were the inverse of the predicted propensity from a multinomial logistic regression model for the (categorical) exposure, SIMD groups, adjusted for exposure-outcome confounders. The second set of weights were the inverse of the predicted propensity from a logistic model for the (binary) mediator, maternal smoking, adjusted for mediator-outcome confounders (and exposure-outcome confounders, although in this instance the confounders are already included as mediator-outcomes confounders). The two sets of weights were multiplied together to create the final inverse probability weights. I then fitted a weighted logistic regression model of the outcome on the exposure, the mediator and the interaction between the exposure and mediator using the inverse probability weights. The estimated coefficient for the mediator in this model represents the estimated CDM($m=0$).

G-estimation of structural nested models

This method specifies associational models for the outcome, the exposure and the mediator and is doubly robust against invalid inferences because estimators are consistent if either the outcome or the mediator model is correctly specified.⁴¹² G-estimation of structural nested models can be used to estimate CDM($m=0$) as these can be viewed as a component of structural nested models. The steps, adapted from Naimi et al.,³⁷⁰ are outlined in Box 6.3. A logistic regression model including exposure-outcome and mediator-outcome confounders was used to predict the mediator and a multinomial logistic regression model including exposure-outcome confounders was used to predict the exposure.

Marginal risk difference

To estimate the marginal risk difference, I quantified the potential risk of at least one bronchiolitis admission per 1000 live births by SIMD group and took the absolute difference in these potential risks. I present differences between the high and low SIMD groups and high and medium SIMD groups. As with the CDM($m=0$), these quantities were estimated using two different approaches.

To complement the CDM($m=0$) estimated using the IPW of MSMs approach, a logistic regression model of the outcome against the exposure was estimated using IPW for the exposure via Stata's *binreg* command (with the risk difference option specified). The weights were found by logistic regression modelling of the exposure that includes the exposure-outcome confounders. The estimated constant from this model can be interpreted as the marginal potential risk of bronchiolitis admission if everyone was assigned to the baseline (low deprivation) SIMD group and the estimated coefficient for each level of the exposure as the estimate of the relevant risk difference.

To complement the g-estimation of structural nested model method, logistic regression models of the outcome, adjusted for exposure-outcome confounders, were fitted for each level of the exposure and standardised across the cohort. The mean of the predicted estimates give the standardised potential risk of admission for bronchiolitis for each SIMD group. The risk differences are derived by taking one mean from another.

Proportion of inequity explained by maternal smoking

Using the estimates derived above for the CDM($m=0$) and marginal risk difference, I calculated the total proportion of the inequity between SIMD groups that is explained by maternal smoking as:

$$\text{disparity reduction (\%)} = (\text{risk difference} - \text{CDM}(m=0)) / \text{risk difference} * 100$$

Two estimates of the disparity reduction were calculated, one for each approach. The two approaches rely on different parametric assumptions (when using IPW that the propensity score models for the exposure and mediator are correctly specified, when using standardisation that the outcome and mediator models are correctly specified) and the use of both therefore allows me to assess the robustness of the estimates.

Box 6.2. Steps for estimating CDM($m=0$) using IPW estimation of MSMs with a categorical exposure and binary mediator

Step 1a. Obtain predicted exposure probabilities from a multinomial model for X , unadjusted and adjusted for the exposure-outcome confounders, $p(X = x)$ and $p_{adj}(X = x)$, respectively.

Step 1b. Generate the weights corresponding to each level of the exposure

$$w(X = x) = \frac{p(X = x)}{p_{adj}(X = x)}$$

Step 2a. Obtain predicted mediator probabilities (unadjusted and adjusted for confounders),

$$p(M = 1) = \{1 + \exp[-\alpha_0]\}^{-1}$$

$$p_{adj}(M = 1) = \left\{ 1 + \exp \left[-\alpha_0 - \sum_{j=2}^3 \alpha_{1j} \hat{I}_{X=j} - \alpha'_2 \mathbf{C}_{XY} - \alpha'_3 \mathbf{C}_{MY} \right] \right\}^{-1}$$

where $I_{X=j}$ is an indicator of whether the exposure X takes value j , \mathbf{C}_{XY} are exposure-outcome confounders, \mathbf{C}_{MY} are mediator-outcome confounders, bold indicates matrices and dashes (') indicate transposition of matrices.

Step 2b. Generate the weights for the mediator using the predicted values from these models, where the weights depend on the observed level of the mediator, e.g.

$$w(M = 0) = \frac{p(M = 0)}{p_{adj}(M = 0)}$$

Step 3. Generate overall weights by multiplying the exposure and mediator weights:

$$ipw = w(X = x)w(M = m)$$

Step 4. Fit a weighted regression model of the outcome against the exposure, the mediator and the interaction between the exposure and mediator, using the ipw weights:

$$E_{ipw}(Y) = \theta_0 + \theta_1 X + \theta_2 M + \theta_3 XM$$

Box 6.3. Steps for estimating CDM($m=0$) using g-estimation of structural nested models with a categorical exposure and binary mediator

Step 1a. Obtain predicted mediator probabilities from a logistic regression model for M , adjusted for the exposure-outcome and mediator-outcome confounders.

$$\hat{p}(M = 1) = \left\{ 1 + \exp \left[-\delta_0 - \sum_{j=2}^3 \delta_{1j} I_{X=j} - \delta'_2 \mathbf{C}_{XY} - \delta'_3 \mathbf{C}_{MY} \right] \right\}^{-1}$$

Step 1b. Derive mediator residuals, $\hat{r}(M) = M - \hat{p}(M)$ where $\hat{p}(M)$ is the estimate of the propensity for $M = 1$.

Step 2. Regress the outcome against mediator residuals, the interaction between exposure and mediator residuals, and mediator-outcome confounders using ordinary least squares,

$$E(Y|X, M, \mathbf{C}_{XY}, \mathbf{C}_{MY}) = \beta_0 + \gamma_2 \hat{r}(M) + \sum_{j=2}^3 \gamma_{3j} I_{X=j} \hat{r}(M) + \beta'_1 \mathbf{C}_{XY} + \beta'_2 \mathbf{C}_{MY}$$

Step 3. Create a transformed outcome by removing the estimated mediator effect from the observed outcome, $\tilde{Y} = Y - \hat{\gamma}_2 M - \sum_{j=2}^3 \hat{\gamma}_{3j} I_{X=j} M$

Step 4a. Obtain predicted exposure probabilities from a multinomial model for X , controlling for the exposure-outcome confounders, \mathbf{C}_{XY} .

Step 4b. Create exposure residuals for the levels of the exposure that are not the reference (baseline) value, $\hat{r}_2(X) = I_{X=2} - \hat{p}(X == 2)$ and $\hat{r}_3(X) = I_{X=3} - \hat{p}(X == 3)$

Step 5. Regress transformed outcome against predicted exposure residuals and exposure-outcome confounders using ordinary least squares to estimate counterfactual disparity measure, $E(Y) = \theta_0 + \psi_2 \hat{r}(X) + \hat{r}_2(X) + \hat{r}_3(X) + \theta_2 M + \theta \mathbf{C}_{XY}$

6.7. Results

6.7.1. Missing data

After exclusions were applied, there were 274,444 infants included in the birth cohort (Figure 6.4). Of these, 8,946 (3.3%) had missing information for at least one study variable. Table 6.4 shows the distribution of cohort characteristics by infants with and without missing data. Missingness was driven by maternity records, which were unavailable (i.e. unlinked) for 5,913 (2.2%) of infants. The highest proportion of missing information was for maternal smoking, for which 3.3% of infants did not have a record. Maternal age, maternal asthma and parity all had 2.2% of records missing. In the adjusted logistic regression model for the odds of record incompleteness, missing data on ≥ 1 variable was associated with non-northern areas of residence (particularly Glasgow), earlier year of birth, non-UK maternal country of birth and, to a lesser extent, lower SIMD groups (see Table 6.4). Using simulation studies, White and Carlin illustrate that in circumstances where missingness is independent of the outcome given the covariates, complete case analysis has negligible bias.⁴¹³ Analysis in this study was therefore performed using the 96.7% of infants with complete data.

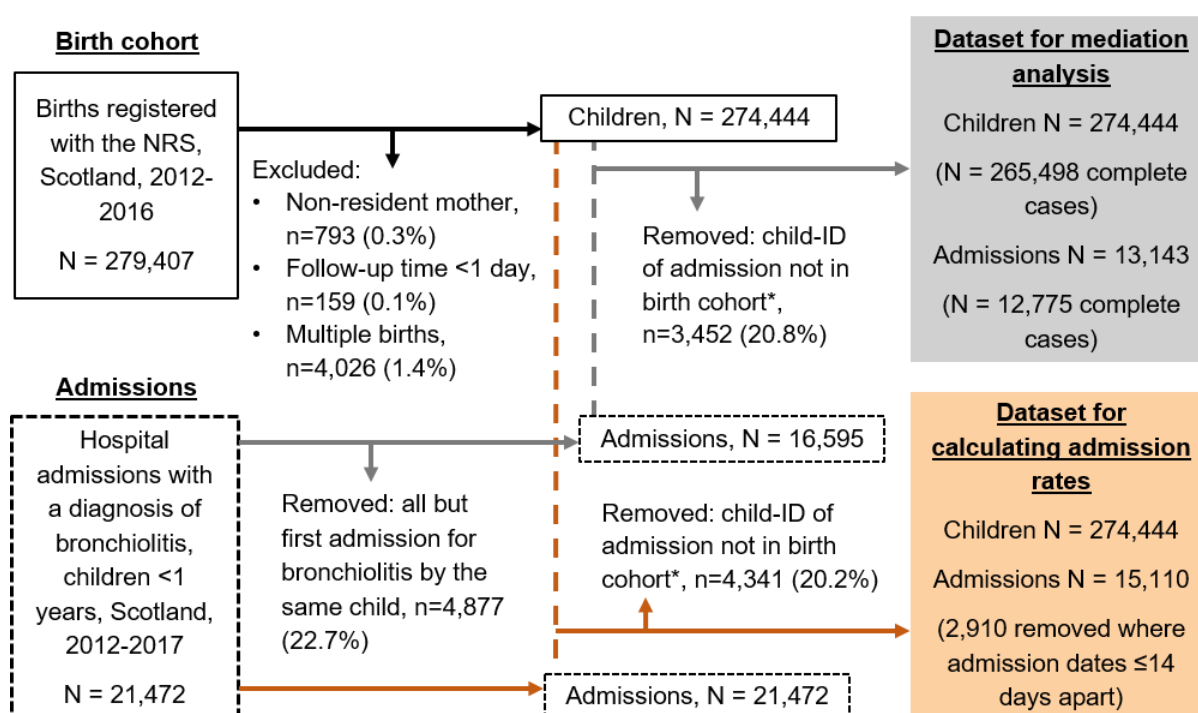


Figure 6.4. Flow diagram showing derivation of final study cohorts for mediation analysis and calculations of admission rates

**Child-IDs belonging to children born in 2017 or excluded from birth cohort at earlier stage of diagram*

Table 6.4. Cohort characteristics, missing data and admission rates

		Infants				Odds of ≥1 missing value aOR** (95% CI)	Admissions per 1000 infant-years Rate (95% CI)
		All		Complete cases			
		N	%	N	%		
Total		274,444	100.0	265,498	100.0		55.5 (54.6, 56.4)
≥1 adm. for bronchiolitis	No	261,311	95.2	252,732	95.2	1.09 (0.98, 1.21)	
	Yes	13,133	4.8	12,766	4.8	Reference	
Year of birth	2012	56,835	20.7	54,475	20.5	1.67 (1.56, 1.79)	48.5 (46.7, 50.3)
	2013	54,839	20.0	52,814	19.9	1.47 (1.37, 1.58)	47.9 (46.1, 49.8)
	2014	55,414	20.2	53,896	20.3	1.08 (1.00, 1.16)	54.7 (52.8, 56.7)
	2015	53,946	19.7	52,267	19.7	1.23 (1.15, 1.33)	58.5 (56.5, 60.6)
	2016	53,410	19.5	52,046	19.6	Reference	68.8 (66.6, 71.0)
SIMD	Low	67,585	24.6	65,486	24.7	1.05 (0.99, 1.11)	45.0 (43.4, 46.6)
	Medium	106,268	38.7	102,827	38.7	1.13 (1.07, 1.18)	52.5 (51.1, 53.9)
	High	100,591	36.7	97,185	36.6	Reference	65.7 (64.1, 67.3)
Area of residence	Northern	34,884	12.7	34,192	12.9	Reference	41.4 (41.4, 45.8)
	Central	78,119	28.5	76,280	28.7	1.19 (1.09, 1.30)	52.4 (50.8, 54.0)
	Glasgow	64,671	23.5	60,785	22.9	3.18 (2.93, 3.46)	45.0 (43.4, 46.7)
	Edinburgh	48,450	17.7	47,117	17.7	1.36 (1.24, 1.49)	58.6 (56.5, 60.8)
	Southern	48,320	17.6	47,124	17.7	1.30 (1.18, 1.43)	79.9 (77.4, 82.5)
Maternal birth country	Non-UK	43,728	15.9	41,865	15.8	1.35 (1.28, 1.43)	35.5 (33.7, 37.3)
	UK	230,716	84.1	223,633	84.2	Reference	59.3 (58.3, 60.3)
Maternal smoking	No	215,704	78.6	215,704	81.2		49.7 (48.7, 50.6)
	Yes	49,794	18.1	49,794	18.8		82.4 (80.0, 85.0)
	Missing	8,946	3.3				47.7 (43.3, 52.5)
Maternal age (years)	≤19	12,062	4.4	11,744	4.4		68.0 (63.5, 72.8)
	20-29	119,111	43.4	117,713	44.3		72.6 (70.1, 75.2)
	30-39	127,095	46.3	125,902	47.4		49.2 (47.7, 50.7)
	≥40	10,263	3.7	10,139	3.8		42.9 (39.8, 46.3)
	Missing	5,913	2.2				48.8 (42.4, 54.8)
Maternal asthma	No	248,526	90.6	245,709	92.5		53.4 (52.5, 54.3)
	Yes	20,005	7.3	19,789	7.5		84.0 (80.1, 88.1)
	Missing	5,913	2.2				48.8 (43.4, 54.8)
Older siblings	No	113,783	41.5	111,548	42.0		36.4 (35.3, 37.5)
	Yes	154,748	56.4	153,950	58.0		69.7 (68.4, 71.0)
	Missing	5,913	2.2				48.8 (43.4, 54.8)

** (aOR=adjusted odds ratio) multivariable logistic regression comparing the odds of having at least one missing value compared with no missing values, adjusted for bronchiolitis admission, year of birth, socioeconomic deprivation, country of birth and area of residence

6.7.2. Cohort description (complete cases)

There were 265,498 infants in the final cohort of which 12,766 (4.8%) had at least one hospital admission with a diagnosis of bronchiolitis in the first year of life. The proportion of children with at least one admission for bronchiolitis increased over time, from 4.3% for children born in 2012 to 5.8% for children born in 2016. Overall, 36.6% of infants were in the high SIMD group and 18.8% of mothers smoked during their pregnancy. When split by deprivation group, there are differences in the distribution of key risk factors across the cohort (Table 6.5). Notably, 29.2% of mothers in the high SIMD group reported smoking during pregnancy compared to 7.2% in the low SIMD group. 5.7% of infants in the high SIMD group were admitted to hospital with bronchiolitis at least once in the first year of life compared to 3.9% of infants in the low deprivation group. There were also marked differences by area of residence (30.3% of infants in the high SIMD group are resident in greater Glasgow at birth), maternal age (lower in the low SIMD group) and maternal asthma (highest in the high SIMD group).

6.7.3. Admission rates (all infants)

As shown in Table 6.4, the overall incidence rate of admissions for bronchiolitis over the study period was 55.5 per 1000 infant-years (95% CI 54.6 to 56.4). The rate of admissions increased over time, from 48.5 per 1000 infants-years (95% CI 46.7 to 50.3) in 2012 to 68.8 per 1000 infant-years (95% CI 66.6 to 71.0) in 2016, and was particularly high among children with a southern Scottish residence listed at birth (79.9 per 1000 infant-years, 95% CI 77.4 to 82.5). The rate of admissions increased with level of socioeconomic deprivation, with infants in the most deprived SIMD group having an admission rate of 65.7 per 1000 infant-years (95% CI 64.1 to 67.3) compared to 45.0 per 1000 infant-years (95% CI 43.4 to 46.6) in the least deprived SIMD group over the 5-year period. UK maternal country of birth, maternal smoking during pregnancy, younger maternal age, maternal asthma and having older siblings were all risk factors associated with higher rates of admissions.

Table 6.5. Final cohort characteristics, by socioeconomic deprivation group, and ≥ 1 hospital admission for bronchiolitis in infancy

		<u>SIMD group</u>						<u>≥ 1 bronchiolitis admission</u>	
		<u>Low</u>		<u>Medium</u>		<u>High</u>		<u>N</u>	<u>% of row</u>
		<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>		
≥ 1 bronchiolitis admission	No	62,932	96.1	98,200	95.5	91,645	94.3	-	-
	Yes	2,554	3.9	4,730	4.6	5,540	5.7	2,776	100.0
Deprivation	Low	65,486	100.0	-	-	-	-	2,564	3.9
	Medium	-	-	102,827	100.0	-	-	4,682	4.6
	High	-	-	-	-	97,185	100.0	5,520	5.7
Year of birth	2012	13,097	20.0	21,080	20.5	20,214	20.8	2,325	4.3
	2013	12,835	19.6	20,565	20.0	19,437	20.0	2,232	4.2
	2014	13,359	20.4	20,771	20.2	19,729	20.3	2,585	4.8
	2015	13,097	20.0	20,154	19.6	19,048	19.6	2,623	5.0
	2016	13,097	20.0	20,257	19.7	18,757	19.3	3,001	5.8
Maternal smoking	No	60,771	92.8	86,066	83.7	68,807	70.8	9,301	4.3
	Yes	4,715	7.2	16,761	16.3	28,378	29.2	3,465	7.0
Area of residence	Northern	7,072	10.8	17,789	17.3	9,427	9.7	1,308	3.8
	Central	23,837	36.4	29,306	28.5	23,130	23.8	3,472	4.6
	Glasgow	13,097	20.0	18,303	17.8	29,447	30.3	2,442	4.0
	Edinburgh	15,389	23.5	18,509	18.0	13,314	13.7	2,306	4.9
	Southern	6,156	9.4	19,023	18.5	21,964	22.6	3,237	6.9
Maternal age (years)	≤ 19	851	1.3	3,907	3.8	6,997	7.2	703	6.0
	20-29	19,056	29.1	45,141	43.9	53,452	55.0	6,352	5.4
	30-39	41,846	63.9	49,768	48.4	34,306	35.3	5,335	4.2
	≥ 40	3,733	5.7	4,010	3.9	2,430	2.5	376	3.7
Maternal asthma	No	61,360	93.7	95,321	92.7	89,021	91.6	11,400	4.6
	Yes	4,126	6.3	7,506	7.3	8,164	8.4	1,366	6.9
Maternal birth country	Non-UK	10,281	15.7	15,321	14.9	16,327	16.8	1,295	3.1
	UK	55,205	84.3	87,506	85.1	80,858	83.2	11,471	5.1
Older siblings (parity)	No	27,308	41.7	44,524	43.3	39,749	40.9	3,609	3.2
	Yes	38,178	58.3	58,303	56.7	57,436	59.1	9,157	5.9

6.7.4. Mediation analyses

As shown in Table 6.6, after accounting for exposure-outcome confounders, the potential risk of at least one hospital admission for bronchiolitis in the first year of life rose with each level of deprivation (SIMD). The two approaches used to calculate potential risk and risk differences produced almost identical results and I present the IPW of MSMs-derived estimates here. After accounting for year of birth, maternal country of birth and area of residence by IPW estimation, the potential risk of bronchiolitis admission in infancy would be 40.8 per 1000 live births (95% CI 39.2 to 42.3) if everyone were assigned to the low SIMD group at birth. This compares to 57.1 per 1000 live births (95% CI 55.6 to 58.6) if all infants were assigned the high SIMD group. Overall for every 1000 live births, if all children were assigned to the high SIMD group, there would be 16.3 additional infants with (at least one) admission for bronchiolitis per 1000 live births (95% CI 14.4 to 18.2). The additional potential risk would be 4.6 per 1000 infants were all infants assigned to the medium SIMD group (95% CI 2.4 to 6.8).

IPW of MSMs suggested that the counterfactual disparity between the high and low SIMD groups would be reduced to 13.0 per 1000 live births (95% CI 10.0 to 16.0) if maternal smoking were eliminated. This translates into a 20.3% disparity reduction in admissions (95% CI 12.1 to 28.4). A slightly lower reduction in counterfactual disparity was estimated using g-estimation of structural nested models (13.4 potential risk difference, 95% 11.0 to 15.7; 18.0% disparity reduction, 95% CI 10.3 to 25.6). The estimated remaining disparity in the medium compared to the low SIMD group was 3.3 per 1000 live births (95% CI 1.2 to 5.5) using the IPW of MSMs method. This would be a disparity reduction of 27.2% (95% CI 9.0 to 45.4) if maternal smoking were eliminated. The g-estimation of MSMs method suggested a remaining disparity of 4.0 admissions per 1000 live births (95% CI 1.8 to 6.1), which translated to noticeably smaller disparity reduction of 14.5% (95% CI -9.1 to 38.2) between the medium and low SIMD groups.

Table 6.6. Potential risk of ≥ 1 admission to hospital for bronchiolitis during infancy, estimated counterfactual disparity in admissions and corresponding disparity reduction explained by maternal smoking, by g-method and SIMD group

SIMD group	Marginal risk per 1000 live births* (95% CI)	Marginal risk difference per 1000 live births* (95% CI)	CDM($m=0$) per 1000 live births (95% CI)**	% disparity reduction (95% CI)**†
IPW MSMs				
Low	40.8 (39.2, 42.3)	Reference	Reference	Reference
Medium	45.4 (44.1, 46.7)	4.6 (2.4, 6.8)	3.3 (1.2, 5.5)	27.2 (9.0, 45.4)
High	57.1 (55.6, 58.6)	16.3 (14.4, 18.2)	13.0 (10.0, 16.0)	20.3 (12.1, 28.4)
G-estimation of SNMs				
Low	40.7 (39.1, 42.3)	Baseline	Baseline	Baseline
Medium	45.4 (44.2, 46.5)	4.6 (2.5, 6.8)	4.0 (1.8, 6.1)	14.5 (-9.12, 38.2)‡
High	57.0 (55.2, 58.9)	16.3 (14.1, 18.5)	13.4 (11.0, 15.7)	18.0 (10.3, 25.6)

*Risk and absolute risk difference after adjusting for differences in year of birth, maternal country of birth and area of residence; **95% CIs are calculated using the bootstrap with 1000 replications; †disparity reduction (%) was estimated as = (risk difference-CDM($m=0$))/risk difference*100; ‡bootstraps of the g-estimation procedure produced some estimates of the CDM that were larger than the TCE, leading to some negative percentages

6.8. Discussion

Using a population-based Scottish birth cohort, I employed a counterfactual causal inference approach to establish the extent to which maternal smoking explains the socioeconomic disparity in the risk hospital admission for bronchiolitis in infancy. The overall relationship between SIMD and bronchiolitis admission was 16.3 (95% CI 14.4 to 18.2) more children with at least one hospital admission for bronchiolitis in infancy for every 1000 infants born into high compared to low levels of SIMD. Using the CDM approach, it was estimated that the elimination of maternal smoking would reduce about one fifth of the disparities in the proportion of children who had at least one hospital admission for bronchiolitis in the first year of life.

6.8.1. Strengths and weaknesses

Linked Scottish health datasets

The coverage and completeness of the birth cohort used in this study is a major strength of this work. The core cohort was derived from vital records held by NRS, meaning that almost all infants born in Scotland between 2012 and 2016 were included in the study. Infant information was supplemented with maternal delivery records, dispensing information and hospital admission data, creating a comprehensive dataset. This enabled me to capture

important risk factors in the development of offspring respiratory illness that are not available in the equivalent English dataset, including indicators of maternal asthma and maternal smoking during pregnancy. However, there are also drawbacks to this dataset.

Whilst variables recorded in NRS birth registrations were 100% complete, 2.2% of children's records were not linked to maternal records, meaning that they were missing information on several key variables. Maternal ethnicity was particularly affected by missingness and was therefore not included in the main analysis despite being identified as a potential confounder between maternal smoking and offspring bronchiolitis. Omission of ethnicity in Scottish hospital records is a long noted problem,⁴¹⁴ although recording of this characteristic has improved markedly since 2010 when the Equality Act was brought in (from 42% completeness in 2010 to 82% in 2016). Other studies using Scottish administrative data have benefitted from linkage to census data for higher completeness of ethnicity.⁴¹⁵⁻⁴¹⁷ I did include a related variable (maternal country of birth) as a confounder, which likely partially captured the confounding effect of maternal ethnicity. Although I used a structured method to derive study confounders, there may also be additional unmeasured confounding, which would violate the conditional exchangeability assumption.

Measurement error is likely present in the study mediator. Maternal smoking is self-reported in this study and, as a factor known to be affected by social desirability bias,⁴¹⁸ is likely underreported. A study of pregnant women in 2003/04 in the West of Scotland found that self-reported smoking status from SMR02 (as used in this study) underestimated smoking by 25% when compared to an objective measurement of smoking, and women from more affluent areas were more likely to misreport their smoking status.⁴¹⁹ In addition, having a singular binary indicator of maternal smoking measured at one time point (during pregnancy) means that the full complexities of this risk factor, including the amount and timing of tobacco smoke exposure, could not be studied. In particular, many women who smoke during pregnancy are likely to continue after birth, so it is difficult to differentiate between the effects during pregnancy and smoking exposure after birth. Conversely, there will be some non-smokers who live with someone who smokes in the house. Linkage to datasets that capture early exposure to second hand smoke, such as the Child Health Surveillance Programme Pre-School,⁴²⁰ would allow for more nuanced analyses of the effect of smoking on early respiratory health.

The PIS database does not include diagnosis codes and very few mothers had hospital admissions with a diagnosis of asthma, meaning that the indicator of maternal asthma used in this study relied primarily on dispensed medication data, which could have resulted in disease misclassification. Other studies have identified chronic obstructive pulmonary disease

as a source of misclassification with asthma;⁴²¹ however as this condition is relatively uncommon in people younger than 40,⁴²² this is unlikely to be a large source of bias in this study. Given that the pathway between maternal asthma and child bronchiolitis is far from determined—and likely includes a mixture of a genetic predisposition to respiratory illness and shared environmental factors—the use of an indicator that may have captured a broader group of maternal respiratory ill-health (as in this study) is unlikely to be of detriment to the findings.

ESC-DAG (Objective 6B)

The ESC-DAG protocol helped me to follow a systematic method to search the literature and create a study DAG. This evidence-based DAG construction and its application to DAGitty (which identifies causal and biasing paths) gives me confidence in the adjustment strategy for my mediation analysis.⁴²³ The ESC-DAG process gave me the opportunity to re-examine how previous research describing inequities in bronchiolitis have considered covariates in their analyses. Commonly, I reversed the directed edges between pathways that were treated as confounders in the original research but were likely mediators. This highlights a common issue of over-control for mediators in these studies, which may have led to an underestimation of the impact of socioeconomic deprivation on bronchiolitis rates.

There were downsides to using an ESC-DAG approach. Prominently, it was a particularly time intensive undertaking. Despite a restrictive exclusion criteria for the literature search, which resulted in the inclusion of 9 studies, I still had to undertake 552 directed edges assessments. I used a formal method to assess directed edges, considering temporality, face-validity, prior theory/research and counterfactual outcomes, but this did not completely eliminate subjectivity in the process. Consider, for example, the pathway from area of residence to number of siblings. Family size does vary by area in the UK,⁴²⁴ but it is unclear whether there is any direct link between the two factors or whether this association could be completely explained by alternative correlated factors such as socioeconomic position and ethnicity. For the most part, I erred on the side of caution for these debatable cases; keeping the directed edge rather than removing it (which would be a stronger assumption). The reliance on previous research to guide variable selection meant that understudied research areas, particularly paternal risk factors, were not included in the ESC-DAG.

Future applications may benefit from independent assessment of decisions using a secondary reviewer. In addition, a simplified version of the ESC-DAG process – for example, by combining related covariates into one node for assessment or excluding covariates unrelated to the mediator at an earlier stage in the process – is advisable where time is constrained. The

process took many days and may only be feasible for large projects and/or scenarios where there is a very simple and specific research question.

Causal inference approach (Objectives 6A & 6C)

I outlined the main mediation estimand(s) available and chose CDM as the most appropriate to answer my research question, given the available data and complexities of estimation. This estimand enabled me to characterise the overall level of disparity and capture the proportion of the disparity in the percentage of children with at least one bronchiolitis admission due to the impact of socioeconomic deprivation that would remain if maternal smoking were intervened upon. A major advantage of this approach is that, in comparison to other estimands, fewer assumptions were required to obtain unbiased estimates of the CDM. Several assumptions were still required, however, including no unmeasured mediator-outcome confounding, no interference, positivity and conditional exchangeability. I clarified how most of these assumptions were met in this study in section 6.2.2. The small number of confounders and broad categories used to define them (e.g. UK/non-UK for maternal country of birth) gives me confidence that the positivity assumption was met.

However, there are also some drawbacks of the CDM approach. In order to apply this approach I needed to alter my original research question to specify that the disparity measure was based on the scenario of no women smoking during pregnancy. This assumes an intervention that leads to a 100% stopping smoking rate that, based on the outcome of current interventions, is arguably unrealistic. One stop smoking service for pregnant women in a deprived area of Scotland, for example, achieved a 32% stop smoking rate at 4 weeks after the participants' quit dates.⁴²⁵ Other estimands, such as the RIA-NIE that allows the mediator to be set to take a value that is randomly drawn from a distribution, arguably offer a more policy relevant estimate. However, this method could not be applied with the assurance of unbiased estimates.

I used two approaches to estimate the magnitude of socioeconomic disparity in infant hospital admission for bronchiolitis that would remain if no women smoked during pregnancy. The similarity in results for the high compared to low groups across the two methods gives me assurance of the robustness of these estimations. Both confidence intervals for these estimates are wide, which may be expected when the calculation involves the ratio of two risk differences; however, a percentage disparity estimate that is lower than 0% (as was the lower bound of the g-estimation of structural nested models method) is less plausible. It has been noted elsewhere that MSMs cope well with difference outcome variables, including logistic

regression (as was applied in this study), whereas there are issues applying structural nested models to binary data.⁴²⁶

6.8.2. Research findings and implications (Objective 6D)

Overall, this work shows that hospital admission rates for bronchiolitis have continually increased in Scotland. From 2012 to 2016, the rate of admissions increased from 48.5 per 1000 infant-years to 68.8 per 1000 infant-years. A similar difference is apparent between socioeconomic groups; infants from the least deprived socioeconomic group had an admission rate of 45.0 per 1000 infant-years compared to 65.7 per 1000 infant-years in the most deprived group across the whole study period. These findings continue to highlight the importance of work to tackle this condition and unravel the social determinants of inequity in admission rates.¹⁵

Looking beyond descriptive statistics, the main purpose of this study was to apply formal mediation methods to look at the specific pathway of maternal smoking during pregnancy in the association between socioeconomic position and hospital admission for bronchiolitis. Previous studies in this field have looked at the adjusted association of several risk factors on early-life RSV-related illnesses without focussing on elucidating socioeconomic pathways.^{44,137,138} I estimate that about one fifth of the difference in being admitted to hospital for bronchiolitis between infants in high and low socioeconomic groups in Scotland can be attributed to maternal smoking. My results therefore show the continuing need to reduce rates of maternal smoking specifically in the most socioeconomically deprived populations. These reductions would substantially reduce the social gradient in bronchiolitis hospital admissions.

With the rapid advances in methodology in the field of causal mediation analysis, future studies will likely benefit from methods that can better account for follow-up time and multiple admissions by the same child. In addition, future studies using similar formal methods but focussing on other policy-relevant mediators, such as outdoor air pollution and indoor conditions, such as damp and mould, are extremely important to deciphering the development of bronchiolitis and positioning interventions with the greatest impact.

6.9. Chapter summary

In this chapter, I applied a counterfactual causal inference approach to quantify the role of maternal smoking in bronchiolitis hospital admission disparities. To achieve this I have reviewed a broad range of mediation estimands recently proposed in the literature and discussed the potential application of these to my research question. I also employed a novel

and evidence-based approach to inform the appropriate adjustment strategy for my study to minimise bias. I included maternal country of birth, year, maternal age, maternal asthma, parity and area of residence as confounders of the mediator-outcome relationship. I found that an intervention aimed at eliminating maternal smoking during pregnancy could reduce the disparities in hospital admissions for bronchiolitis between the most and least deprived infants by up to one fifth.

Chapter 7. Trajectories of asthma/wheeze symptoms and their association with early-life risk factors

Recap and chapter overview

Chapters 4 to 6 explored socioeconomic inequities in bronchiolitis admissions among infants in England and Scotland. I used harmonic regression analyses to investigate the association between the seasonality of bronchiolitis admissions and socioeconomic deprivation at both the area and individual level. I then applied a counterfactual causal inference approach to quantify the role of maternal smoking in bronchiolitis hospital admission disparities.

In this chapter, I move my focus to childhood asthma; a common and chronic respiratory condition that has been linked to early-life bronchiolitis infection. Applying latent growth modelling techniques to the Scottish health datasets, I identify trajectories of asthma/wheeze symptoms among children aged less than 10 years. The association between early risk factors and asthma/wheeze group membership are then explored. Lastly, drawing from the counterfactual disparity work presented in Chapter 6, I investigate the mediating contribution of bronchiolitis in the pathway between socioeconomic position and asthma/wheeze. The final two thesis objectives are addressed in this chapter:

To describe typical trajectories of asthma/wheeze among children <10 years in Scotland and their association with early life risk factors, including socioeconomic position

To investigate the mediating contribution of bronchiolitis in the pathway between early life SEP and mid-childhood asthma/wheeze

7.1. Background

In 2018, 8% of children younger than 15 years old and 17% of those aged 16 to 24 years in Scotland reported a current doctor diagnosis of asthma.⁸⁹ Despite this high prevalence, substantial contribution to morbidity and associated health-care burden across the world, the precise causes of asthma are relatively unknown.^{32,47} Moreover, the last decade has seen a stagnation in improvements of clinical outcomes, including mortality rates, for people with the condition.⁴⁷ Asthma experts from a multitude of medical disciplines, writing in a 2018 Lancet

commission, argue that the lack of distinction between the different disease phenotypes under the umbrella term of “asthma” is a major contributor to this stagnation.⁴⁷ These experts placed strong emphasis on the need to move beyond the current over-simplistic definition of asthma, which is currently based on common physiology and symptoms, despite encompassing conditions with pathobiologically distinct mechanisms.

Cohort studies in the UK, USA, the Netherlands and Taiwan, as well as in other countries, have been used to investigate different patterns of symptoms (often referred to as phenotypes) that comprise childhood asthma.^{21,25,26,197,427–429} Using statistical models that allow researchers to identify typical sub-groups of symptoms in longitudinal cohort data, between 3 and 6 trajectory groups have been identified based on the prevalence of asthma and wheeze symptoms at different ages.¹⁹⁷ When defined by symptom onset and longevity, these groups are broadly characterised as:¹⁹⁷

- never/infrequent asthma and/or wheeze;
- early-transient asthma/wheeze, beginning before age 3 and subsiding after a few years;
- early-persistent asthma/wheeze, beginning before 3 and persisting into early adolescence (or the length of study follow-up);
- intermediate-onset asthma/wheeze, beginning between 3 and 5 years and persisting beyond study follow-up;
- and late-onset asthma/wheeze, beginning between 6 and 8 years and persisting beyond study follow-up.

The precise number of groups identified varies between studies, depending on factors such as the length of follow-up, definitions of asthma/wheeze employed and the sample size. As these groups are not directly observable but, rather, constructed from the data, they are described as latent variables. However, this does not mean that these different groupings are not recognised clinically. This data driven approach to classification has found to be comparable with clinical phenotypes defined by patient histories, diagnostic work-up and treatment responses.⁴³⁰

Once trajectory groups have been identified, the probability of a study participant following each trajectory can then be associated with risk factors commonly associated with childhood respiratory ill-health, such as prenatal smoking exposure and early respiratory infections.^{21,429} Differentiating between asthma/wheeze groups and understanding trajectory-specific risk factors can better inform prognosis, and may be critical to identifying for which children and at which time early interventions may be best place to halt the progression of this disease into adulthood.¹⁹⁷ Further, such analyses may be key to untangling the complex relationship

between socioeconomic deprivation and asthma, particularly the pathway relating bronchiolitis to asthma. As discussed in Chapter 1, this relationship is complex because of the likely interplay of environmental and biological risk factors occurring prenatally, in-utero and in early life on the development of an infant's respiratory functions.⁴⁷

Limitations have been noted in asthma trajectory studies to date, however. A 2020 systematic review and meta-analysis of childhood wheeze trajectory-specific risk factors identified high risk of information bias in all of the 13 included cohort studies.¹⁹⁷ Child or parental reports, rather than clinical validation of symptoms, limited sample sizes and loss to follow-up are noted problems that lead to low power to detect associations with risk factors.⁴³¹ Overcoming many of these previous shortcomings, Sbihi et al.²⁴ used administrative health data to identify different asthma trajectories in a sample of more than 65,000 children in British Columbia. Results of Sbihi et al.'s study demonstrate the unique opportunity administrative data offers to study asthma trajectories, overcoming logistic and financial constraints and self-report biases inherent to purposefully designed longitudinal cohort studies based on questionnaires or interviews.²⁴

The aim of this Chapter is to identify trajectories of asthma/wheeze symptoms in a Scottish context and to study whether and how they are socially patterned.

The objectives for this chapter are threefold:

Objective 7A To identify typical trajectories of asthma/wheeze in children aged <10 years old using Scottish administrative health data.

Objective 7B To examine how the identified asthma/wheeze trajectory groups vary according to social and biological risk factors operating early in the life course.

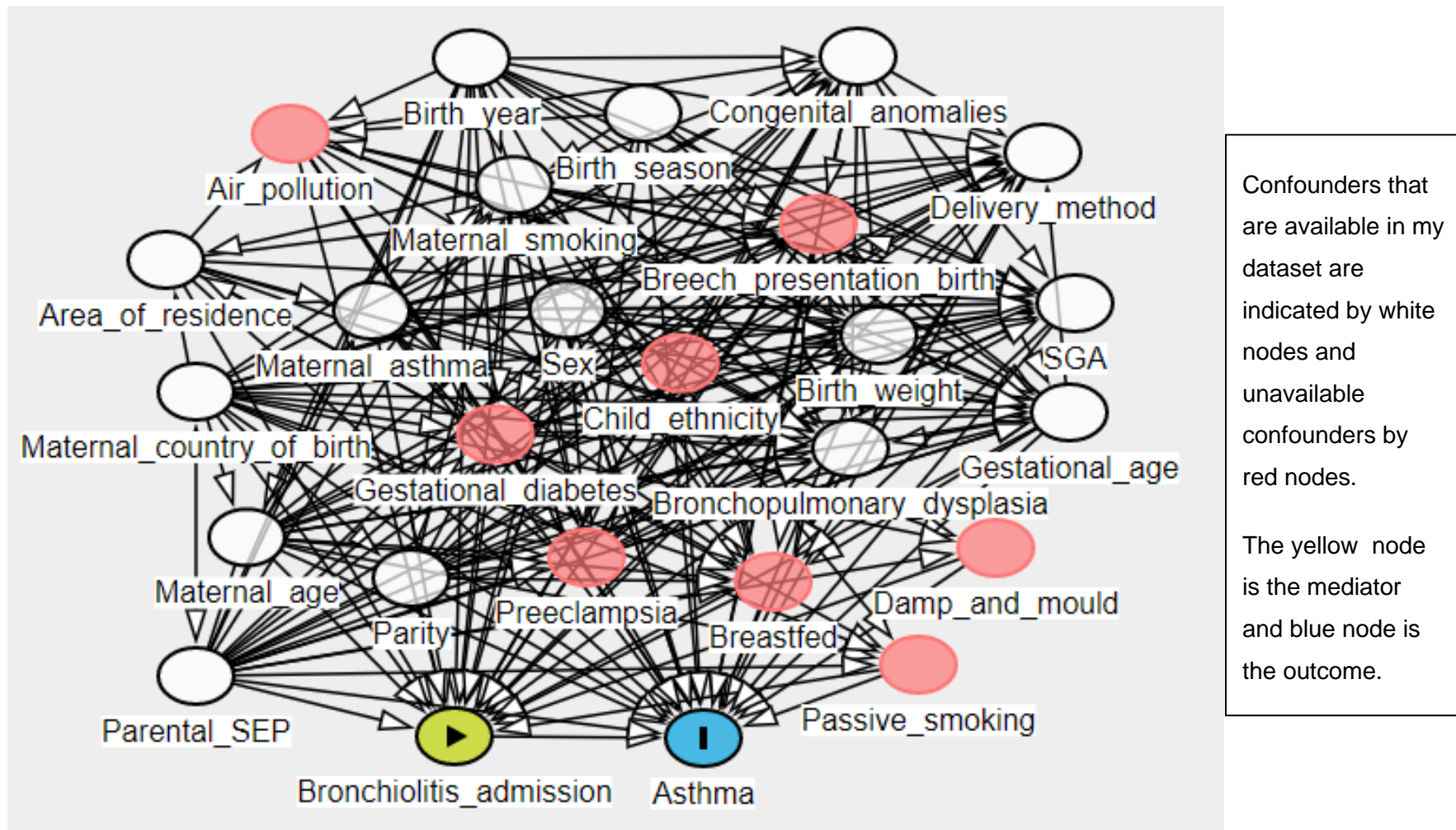
Objective 7C To quantify the mediating role of bronchiolitis in the pathway between socioeconomic position and the asthma/wheeze trajectory groups.

7.2. Study cohort and variable selection

7.2.1. Study DAG for mediation analyses

To identify confounders of the mediator-outcome (bronchiolitis admissions-asthma/wheeze) relationship, I adapted the diagram that emerged from the ESC-DAG procedure in Chapter 6 (Figure 6.1). Using the research presented in my introductory chapter, I determined that all identified risk factors for bronchiolitis were also associated with asthma and therefore, using

DAGitty, I drew a line from each node to asthma. To this, I added other risk factors for both bronchiolitis and asthma identified through my literature review in Chapter 1: passive smoking, and mould and damp. The assumed DAG for the mediator-outcome relationship is presented in Figure 7.1. This presents the minimal adjustment set to satisfy the assumption of no unmeasured mediator-outcome confounding: air pollution, area of residence, birth season, birth weight, birth year, breastfed, breech presentation birth, bronchopulmonary dysplasia, congenital anomalies, delivery method, child ethnicity, gestational age, gestational diabetes, maternal age, maternal asthma, maternal country of birth, maternal smoking, parity, preeclampsia, child sex and small for gestational age (SGA). As shown using pink nodes, some of these confounders are unavailable in my dataset. The implications of this are reviewed further in the discussion.



Confounders that are available in my dataset are indicated by white nodes and unavailable confounders by red nodes.

The yellow node is the mediator and blue node is the outcome.

Figure 7.1. Assumed causal DAG for the mediator-outcome (bronchiolitis admission-chronic asthma) association

7.2.2. Data sources and study cohort

I used a subsection of the birth cohort of Scottish children described in Chapter 3 as the cohort for this study. The linked datasets that formed the cohort as well as covariates relevant to this study are shown in Table 7.1. In brief, information on births and deaths were obtained from NRS, with the birth register forming the core spine of the cohort. The birth and death registration data were supplemented with information from SBR, a singular electronic record capturing all the neonatal care received by babies in Scotland,²⁹⁴ and information on emigrations from the CHI registry. Further information about the birth was contained in the mother's delivery record, which was retrieved from SMR02—a dataset submitted by maternity hospitals on inpatient and day case activity relating to obstetrics. Subsequent hospital admissions among children in the cohort were retrieved from SMR01, which includes information on all inpatients and day cases discharged from hospitals in Scotland.⁴⁰⁹ Data were also extracted from the PIS, which contains information on medicines prescribed and dispensed within the community setting in Scotland.²⁹⁰

Table 7.1. Scottish datasets used to create the study cohort in this study

Acronym	Dataset	Information	Key variables
NRS	National records of Scotland birth & death registrations	Child	Postal area, SIMD, date of birth, date of death, child sex, date of death, country of birth (mother)
CHI registry	Community Health Index registry	Child	Date of emigration
SBR	Scottish birth register	Child	SIMD, birthweight, diagnoses
SMR02	Scottish morbidity records – obstetric related hospital admissions	Mother	Smoking status, age at delivery, parity, SIMD, method of delivery, birthweight, gestational age, maternal age
SMR01	Scottish morbidity records—general hospital admissions	Child and mother	Diagnoses, data of admission, date of discharge
PIS	Scottish national prescribing information system	Child	Medication name, date dispensed

The initial study population was all live births between 1st January 2007 and 31st June 2008 in Scotland. Follow-up began at birth and continued until the child's 10th birthday. I excluded children who died or emigrated out of Scotland by 2 years of age. This step was taken to ensure that all children in the cohort had a recording of the study outcome (i.e. presence or no presence of asthma/wheeze symptoms) for at least one time point. Children born to

mothers of non-Scottish residency (identified through mother's country of residency in the NRS birth registrations) were also excluded from the dataset to prevent potential systematic loss to follow up. One child from each non-singleton birth was randomly selected to be included in the study.

7.2.3. Study variables

Definition of asthma/wheeze

The outcomes in this study are trajectory groups of asthma/wheeze, which are determined through the latent growth curve modelling described in section 7.3.2, below. In order to model trajectories, I firstly needed to define instances of asthma/wheeze at specific time points. I used a multidimensional definition, with expert input from Steve Cunningham, Professor of Paediatric Respiratory Medicine at the University of Edinburgh, to ensure this classification reflected medical practice. For each year of age between 2 and 9, inclusive, children were defined as having asthma and/or wheeze if they met any of the following criteria in the 12-month period between birthdays:

- ≥ 1 hospital admission in SMR01 with a main diagnosis, as defined by International Classification of Diseases, 10th revision (ICD-10) codes, for asthma (J45 including all subcategories), status asthmaticus (J46X) or wheezing (R06.2);
- or ≥ 4 dispensed prescriptions for any specified asthma medication (see Table 7.2 for medications and relevant age exclusions);
- or asthma mentioned in death records as primary or secondary reason of death.

Asthma medication was defined using the British National Formulary for Children (BNFC),⁴³² with age of child at each asthma/wheeze event was retrieved from the relevant dataset (SMR01, PIS or death record). I selected age 2, rather than an earlier age, as the start of symptom follow-up to reduce the risk of capturing instances of bronchiolitis (the study mediator) in the outcome.

Diagnosis of asthma is not recommended in children younger than 5, in part due to the difficulty of using standardised tests in this population.³ Moreover, recurrent symptoms of wheeze (typically defined as a high pitched sound with a musical quality that probably results from turbulence through narrowed tubes)⁴³³ is common in young children and is not necessarily an indicator of later asthma.⁴³⁴ As such, including symptoms between ages 2 and 4 in this study is intended to capture early wheeze patterns rather than asthma per se. Conversely, a child may have asthma, but currently be asymptomatic (thus not presenting with symptoms of

wheeze).⁴³⁵ It is for these reasons that I use the broader term “asthma/wheeze” rather than singular terms of asthma or wheeze to describe trajectories of symptoms in this study.

Table 7.2. Drugs that are used to treat asthma, defined using the BNFC⁴³²

BNF Chapter section	Conditions treated in children	Inclusions (age restriction applied)	Exclusions (age restriction applied)
3.1 Bronchodilators			
3.1.1.1 <i>Selective beta₂ agonists</i>	Short acting beta2 agonists: Reversible airway obstruction Long acting beta2 agonists: Croup – single dose of corticosteroids	Salbutamol, Terbutaline Formoterol (age 6+), Salmeterol	
3.1.1.2 <i>Other adrenoceptor agonists</i>	Acute allergic and anaphylactic reactions, angioedema, cardiopulmonary resuscitation, severe croup	Orciprenaline sulfate (withdrawn from the market in 2010)	Adrenaline (epinephrine), Ephedrine (hydrochloride)
3.1.2 <i>Antimuscarinic bronchodilators</i>	Reversible airway obstruction, rhinitis	Ipratropium bromide	Acidinium bromide, Glycopyrronium, Indacaterol, Tiotropium, Umeclidinium
3.1.3 <i>Theophylline</i>	Reversible airway obstruction, bronchospasm associated with chronic bronchitis, neonatal apnoea	Theophylline, Aminophylline	
3.1.5 <i>Peak flow meters, inhaler devices and nebulisers</i>		Standard range peak flow meter (age 6+), Low range peak flow meter (age 6+), Inhaler devices, Spacer devices	
3.2 Corticosteroids			
3.2 <i>Corticosteroids</i>	Croup, bronchopulmonary dysplasia	Beclometasone dipropionate (age 2+), Budesonide (age 6+), Ciclesonide (age 12+), Fluticasone propionate (age 4+), Mometasone furoate (age 12+)	
3.3 Cromoglicate and related therapy and leukotriene receptor antagonists			
3.3.1 <i>Cromoglicate and related therapy</i>	Food allergy, allergic conjunctivitis, allergic rhinitis	Sodium cromoglicate (age 5+), Nedocromil Sodium (age 5+)	
3.3.2 <i>Leukotriene receptor antagonists</i>	Allergic rhinitis (15-19 years)	Montelukast	Zafirlukast (12+ years only)
3.4.2 <i>Allergen Immunotherapy</i>		Omalizumab (6+)	

6.3.2 Glucocorticoid therapy

6.3.2 <i>Glucocorticoid therapy</i>	Croup, inflammation, ulcerative colitis, Crohn's disease	Hydrocortisone, Dexamethasone, Prednisone, Prednisolone (in conjunction with hospital admission)	Betamethasone, Deflazacort, Methylprednisolone
--	--	--	--

Exposure

I used Scottish Index for Multiple Deprivation (SIMD) 2006 to capture socioeconomic position for each cohort member. SIMD 2006 deciles were split into three groups for analyses indicating high (top 30% rank of SIMD scores), medium (middle 40% rank of SIMD scores) and low levels of socioeconomic deprivation (bottom 30% rank of SIMD scores). This is the same exposure as used in Chapter 6, except SIMD 2006 deciles rather than SIMD 2012 deciles were used (because these are nearest indices to the birth years of cohort members). Again, SIMD deciles were retrieved from NRS birth registration files and supplemented from SBR or SMR02 where missing.

Mediator

I identified children with ≥ 1 hospital admission with a primary or secondary diagnosis of bronchiolitis during the first year of life in SMR01 records using ICD-10 code J21 for acute bronchiolitis. This was the same definition of bronchiolitis as used in Chapter 6.

Confounders

The selection of variables in this section are based on the confounders identified in the study DAG for mediation analysis (Figure 7.2, below). I refer to these as risk factors in the multivariable logistic regression models and as confounders in the mediation analysis.

Child sex (female/male), date of birth, mother's country of birth (dichotomised into UK or non-UK born) and postcode area were retrieved from the child's birth registration file. Year and month of birth, used as an indicator of birth season, were derived from date of birth. The 16 postcode areas in Scotland were grouped to make 5 broader areas with similar population sizes: northern Scotland, central Scotland, Glasgow, Edinburgh and southern Scotland (see Figure 6.3). I created an indicator of chronic conditions, including congenital anomalies and bronchopulmonary dysplasia, using the list of ICD-10 codes previously defined by Kristensen and colleagues (see Table A7.1, Appendix 7.1).¹⁶³ Children were categorised as having a chronic condition (yes or no) if they had any of these diagnoses recorded in their birth record,

death record or any hospital admission before 6 months of age. The mother's delivery record provided method of delivery (categorised into vaginal, elective caesarean and emergency caesarean), maternal age, parity, mother's smoking status during pregnancy, gestational age in weeks and birthweight in grams. Birthweight was supplemented with information from the child's birth record where missing.

Maternal age was coded into a four-category variable (< 20, 20–29, 30–39 and ≥40 years) and parity (count of previous births) was used to create a number of older siblings variable that was split into groups (0, 1 and 2 or more siblings). Mothers self-reported smoking status, recorded at around 8-12 weeks gestation, was treated as a binary variable with yes signifying mothers smoked traditional cigarettes (i.e. non-e-cigarettes) during their pregnancy. Gestational age was categorised as preterm (<37 weeks) or non-preterm (≥37 weeks). SGA (yes/no) was defined by birthweights falling below the 10th centile of the sex-, and gestational age-specific distributions (using LMS tables).⁴³⁶ I created an indicator of severe maternal asthma (yes/no), for mothers who had one or more hospital admission before their child's birth that included a diagnosis of asthma (ICD-10 code J45, including all subcategories).

Additional covariates

These additional variables were extracted from the datasets to be included as predictors in the imputation models for missing data (see motivation for this in Sections 7.3.3 and 7.4.1). The 9-class National Statistics Socio-Economic Classification (NS-SEC), based on father's occupation for married/cohabiting couples and on mother's occupation for all other birth registration types,⁴³⁷ was retrieved from NRS birth registration files. A binary intensive care stay variable was generated from information in the child's SBR file, with yes indicating any postnatal intensive or high dependency care. Any child admitted to hospital diagnosed with any of the following ICD-10 codes in the first few months of life were recorded as having a prematurity-related complication: H35.1 (retinopathy of prematurity); P01.1 (newborn affected by premature rupture of membranes); P07.0 (extremely low BW <1000g) and P07.2 (extreme prematurity, <28 w); P27.0, P27.1, P27.8 and P27.9 (chronic respiratory diseases originating in the perinatal period); P59.0 (neonatal jaundice associated with preterm delivery); and P61.2 (anemia of prematurity). This indicator was used in addition to the chronic conditions indicator discussed above specifically for imputation purposes. Birth hospital was retrieved from the child's birth file and supplemented with the mother's delivery file where missing.

7.3. Statistical methods

7.3.1. Overview of the analytical steps to address the 3 objectives

During the process of the four analytical steps in this chapter, I switch between statistical software. To illustrate the order in which these steps are carried out I have included a diagram of the process in Figure 7.2, below. Step 1 involved modelling asthma/wheeze trajectories (addressing objective 7A). In step 2, I used multiple imputation by chained equations (MICE) to impute missing values in risk factors (a process conducted independently of step 1). The results of step 1 and 2 were then combined to carry out multinomial logistic regression in step 3 (to examine characteristics of asthma/wheeze trajectory groups, objective 7B). For the mediation analysis (step 4), the classes derived from the final latent growth model were merged with the original (unimputed) dataset, then imputation, bootstrap inference and mediation analysis carried out together. I used Stata 15³¹⁴ for descriptive data analysis, multiple imputation and mediation analysis and Mplus version 8.2⁴³⁸ to implement latent growth modelling and univariable multinomial logistic regression.

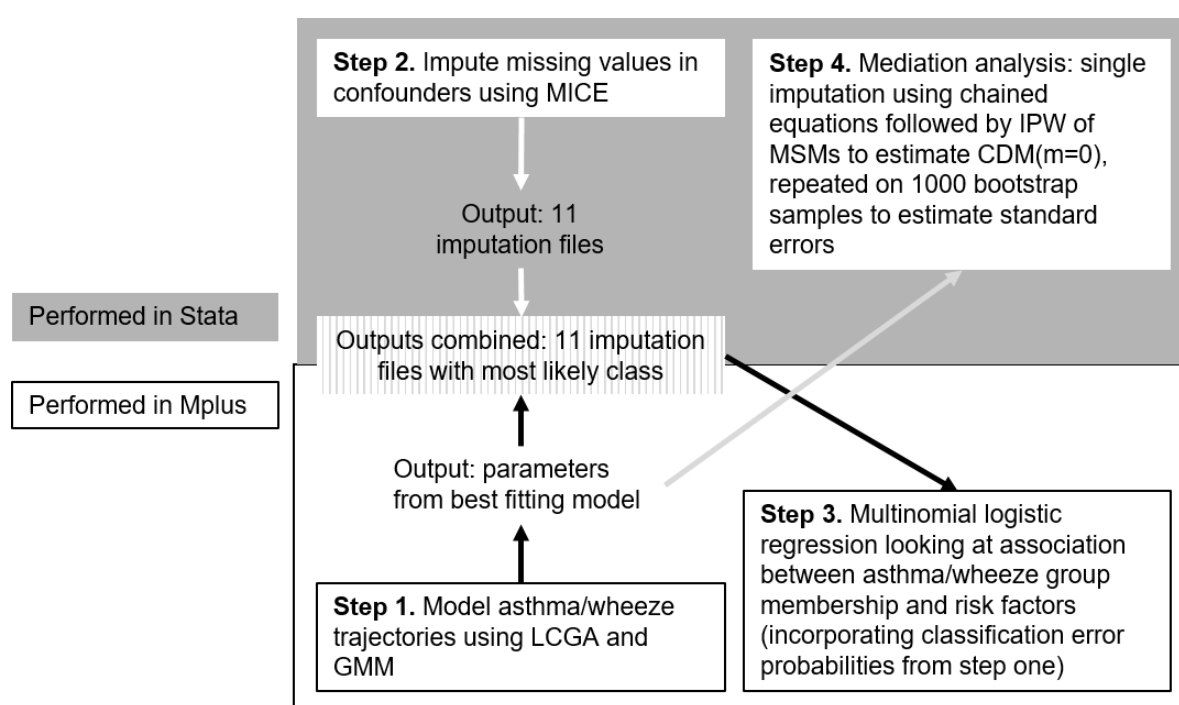


Figure 7.2. Analysis steps in this Chapter, split by software

7.3.2. Typical trajectories of asthma/wheeze (objective 7A)

To model asthma/wheeze trajectories, I used two approaches latent growth curve modelling: latent class growth analysis (LCGA), which is also known as group-based trajectory modelling, and growth mixture modelling (GMM). Both methods offer statistical approaches to detecting classes of individuals with a similar developmental trajectory,⁴³⁹ but differ in whether within-class heterogeneity is allowed. LCGA assumes there is no variation within a class, meaning that all individual trajectories within a class are represented by a single curve.^{439–442} GMM, on the other hand, allows variation across individuals within classes and produces a class of trajectories with a common mean and shape.^{442,443} As described by Muthén,⁴⁴⁴ p.624 GMM therefore offers “a more realistic representation of complex data” and, unlike with LCGA, the resultant classes may (cautiously) be interpreted as underlying phenotypes.⁴⁴³ However, GMM is also far more computationally intensive and noted for having convergence issues because of maximum likelihood estimation problems.⁴³⁹

The purpose of latent growth curve modelling in this study was to model the probability of asthma/wheeze symptoms over time, allowing for different trajectories for these probabilities corresponding to the latent classes. I fitted both LCGA and GMMs. In each case, probabilities were modelled using a mixture of logistic distributions for the dichotomous outcomes of asthma/wheeze presence (yes/no) at the ages from 2 to 9 years, independently of other covariates. For GMMs, the logistic models have additional terms capturing the within class variability. The outcome, asthma/wheeze, was set to missing for all follow-up years following the year of emigration or death (for children with these emigration or death recorded before age 9). Mplus assumes that this missing data in the outcome is missing at random and uses information held in the repeated outcome observations to predict the missing values.⁴⁴⁵ Model selection requires comparing goodness-of-fit statistics obtained from models fitted with varying numbers of trajectory groups and shapes. I began my analyses with LCGA before using the more computationally demanding GMM.⁴³⁹

Model selection for latent class models

Based on the number of classes identified in previous research, I applied 3 groups in a LCGA model initially and added more groups (up to a maximum of 6) in a step-wise manner. I compared models with linear growth (intercept and slope) and non-linear growth (intercept, linear, quadratic and cubic parameters) in turn. All LCGA estimations were run with 500 random starting value sets to avoid local solutions, a problem described by Jung and Wickrama as occurring:

“where during curve estimation a largest value (maximum) or smallest value (minimum) that a function takes is identified for only a given area on that curve, but that is not necessarily the largest or smallest value for the entire curve (i.e., the global minimum or maximum)”.^{439 p.203}

I fitted GMM models in a similar step-wise fashion, but used the estimated class-specific growth factor (e.g. intercept, slope and polynomial) means from the LCGA models (with the same number of classes and growth factors) as starting values. This is one of two strategies offered by Muthén and Muthén to try to bypass the common problem of convergence in GMM and, thus, find global maximum solutions.^{446 p.521} The other strategy is to use the estimated class-specific growth coefficient means and standard deviations from a k class GMM model to compute starting values for a $k + 1$ class model (by, for example, adding half the standard deviation to the mean).

I used the Bayesian information criterion values (BIC; a lower score indicating better model fit) and the bootstrapped likelihood ratio test (BLRT; obtained using the TECH14 option in Mplus) to guide model selection. In simulation studies, the BIC and BLRT have been shown to be the two best performing indices/tests to determine the number of classes in mixture modelling.^{447,448} The null hypothesis of the BLRT is that the model with $k - 1$ classes fits the data as well as the model with k classes, thus a low p -value points towards the k class solution being a better fit of the model. I also required the final model to have latent classes that included at least 1% of the sample and an entropy value >0.8 .^{449,450} The entropy value, which ranges from 0 to 1, indicates separation of classes.⁴⁵¹ A lower entropy value implies poorer separation of classes which, in this case, indicates that there is uncertainty around membership of trajectory groups.

To choose the final model, I compared the results of the final LCGA and GMM fitted models with the same number of classes and shape using the log likelihood values, with a higher value indicating a better fit.⁴⁴⁴ For the final selected model, I produced a plot showing the probability of symptoms at each time point for each trajectory group based on posterior distributions of asthma/wheeze prevalence. An example of Mplus code used to derive asthma/wheeze trajectories groups is included in Appendix 7.2.

7.3.3. Risk factors by asthma/wheeze trajectory group (objective 7B)

Firstly, I described cohort characteristics and missing data for risk factors using descriptive statistics (independent of the latent growth curve modelling described above). The association between variables and the probability of missingness in at least one study variable was

explored using multivariable logistic regression. There was a relatively high level of missing data across the cohort (10.3% of children had a missing value for at least one risk factor) and, prominently, I found that missing data was strongly associated with a particular birth hospital. Therefore, I imputed missing data before carrying out further analyses in this study.

Multiple imputation

Multiple imputation by chained equations (MICE) was used to impute missing values in risk factors for the multivariable logistic regression models. MICE generates imputed values using imputation models for each variable with missing values specified as a function of all other variables involved in the model of interest for the outcome and any other additional variable that may influence why values are missing. This approach allows the handling of different variable types (continuous, binary, unordered categorical and ordered categorical) that may be affected by missingness.⁴⁵² It operates under the assumption that data are missing at random, conditional on the variables included in the imputation models. In other words, the probability that a value is missing depends on the observed variables that are being conditioned on in the imputation.⁴⁵³

To meet this assumption I conditioned on all covariates used in the multivariable logistic regression models (introduced below), as well as: birth at the identified hospital with high levels of missingness (yes or no), NS-SEC, preterm-related complication (yes or no) and postnatal intensive care stay (yes or no). A binary indicator of any asthma/wheeze symptoms between age 2 and 9 was also included. Partially observed covariates were imputed using linear regression (for birthweight, gestational age, maternal age and number of older siblings), logistic regression (for severe maternal asthma and smoking status) and multivariable logistic regression (for number of older siblings and delivery method) models. Preterm birth, maternal age groups, siblings groups and SGA were defined using the relevant variables after imputation. Stata's *mice* command was used to implement MICE in Stata 15.³¹⁴

Characteristics of children by asthma/wheeze trajectory group

Firstly, the distribution of risk factors, including missing data, by each child's most likely asthma/wheeze trajectory group were described in percentages. I then used multinomial logistic regression on the imputed datasets to estimate the probability of belonging to a particular asthma/wheeze trajectory group, relative to the never or infrequent group, based on these risk factors. These analyses are all univariable (i.e. unadjusted for other risk factors). This is because rather than a prediction model, I wanted to describe the most likely composition of the identified trajectory groups. Further, these variables are so closely

correlated that any understanding of possible causal relationships would require a different focus in my analyses and a new investigation of the plausible pathways (including a new DAG).

I report results of the multinomial logistic regression models as estimated relative risk ratios (RRRs), which are the ratios of the relative probability of being in a trajectory group over the probability of being in the never or infrequent asthma/wheeze group. To account for error introduced through potential misclassification of children into trajectory groups via the latent class models, I employed what is known as the refined three-step procedure.⁴⁵⁴ This refined three-step procedure is an improvement to the standard three-step procedure often used with latent class modelling because it uses the probability of each individual being a member of all classes rather than just their most likely class; it has been shown to lead to less biased results than simpler three-step techniques.^{443,455,456} Univariable multivariate logistic regression analyses were replicated on all imputed datasets and the final estimates derived using Rubin's rules (implemented in Mplus). An example of Mplus code used to implement the multivariate logistic regression is included in Appendix 7.2.

7.3.4. Mediation analysis (objective 7C)

Counterfactual disparity measures

Drawing from my work in chapter 6, I will use the CDM to capture the proportion of socioeconomic disparity in asthma/wheeze that would remain if severe (i.e. hospitalisation for) bronchiolitis were intervened upon in infancy.³⁷⁰ For this analysis, I created a binary outcome indicating chronic asthma from the trajectory groups derived through LCGA. Cohort members with early-persistent and intermediate-onset asthma/wheeze as their most likely trajectory group in LCGA analyses were defined as having chronic asthma, and early-transient and no or infrequent class members were defined as not having chronic asthma. I initially tried to run this analysis using a categorical outcome (i.e. the four trajectory groups), but encountered problems of stability of the estimates due to the small numbers in some subgroups. I focussed on chronic asthma (as opposed to any asthma/wheeze) to delineate between symptoms that may be indicative of early wheeze rather than asthma.

I reformulated objective 7C to reflect the estimate of mediation that can be derived using the CDM estimand and the new binary outcome. The new objective was "to establish the magnitude of socioeconomic disparity in the presence of chronic asthma trajectory groups that would remain if no child had an admission for bronchiolitis during infancy". This was defined as: $CDM(m=0)=E[Y(0)|A=a]-E[Y(0)|A=a^*]$, where A = SIMD group (a if high/medium, a* if low),

where M is ≥ 1 hospital admission for bronchiolitis in the first year of life (1 if yes, 0 if no) and Y is chronic asthma trajectory group (1 is yes, 0 if no). This captures the contrast between the potential risks of chronic asthma under high/medium compared to low levels of socioeconomic deprivation, were no infants to have a hospital admission for bronchiolitis in infancy.

As explained in Chapter 6, the assumptions usually required to identify and estimate causal effects from observational data are no interference, consistency, positivity and conditional exchangeability. In the context of both types of CDMs, the assumption of no interference would mean that one child's potential outcome (of asthma/wheeze) was dependent on another child's level of SEP, or another child's hospitalization.^{380,387} Given that the causes of asthma are complex and manifold, but mostly dependent on one's own levels of exposure, it is unlikely that interference is violated in this study. Consistency, another condition that is usually invoked to identify causal effects, requires the study exposure to be well-defined. However, as outlined in Section 6.2.1, this assumption is not required to be upheld when estimating the CDM. Through this approach we accept that the association between exposure and outcome carry "relevant disparity-related (but not necessarily causal) information".³⁷⁰

With regards to the positivity assumption, this would require that within each stratum defined by the mediator-outcome confounders there is a non-zero probability that a child is hospitalised for bronchiolitis. Based on the large number of children in this study and small number of strata (particularly with a binary outcome) it is justified to assume that the positivity assumption is upheld.³⁸⁰ Through my study DAG for the relationship between bronchiolitis admission and asthma/wheeze group (see Figure 7.2), I have outlined the many confounders required for the conditional exchangeability assumption (i.e. exchangeable conditional on a set of covariates C) to be met. Some of these covariates are unavailable in the study dataset, and the implications of this on the conditional exchangeability assumption will be considered in the discussion.

Inverse probability weighting of marginal structural models

I used IPW of MSMs to estimate $CDM(m=0)$ for children in high and medium SIMD groups, compared to the low (deprivation) SIMD group. The steps to this approach, outlined in Box 6.2, began by generating weights from the exposure model using multinomial logistic regression and from the mediator model using logistic regression. I then fitted a weighted regression model of the outcome against the exposure, the mediator and the interaction between the exposure and mediator. The estimated coefficient for the mediator in this model represents the estimated $CDM(m=0)$. Area of residence, year and maternal country of birth were included in the propensity score models that generated the weights as confounders of

the exposure-outcome relationship. Area of residence, birth season birth year, chronic condition, delivery method, gestational age, maternal age, severe maternal asthma, maternal country of birth, maternal smoking, number of older siblings, SGA and child sex were included as confounders of the mediator-outcome relationship. Birthweight, which was available in the dataset and identified as a potential confounder, was not included due to its high correlation with gestational age and SGA.

The estimated $CDM(m=0)$ is displayed as the risk difference between high/medium and low SIMD groups in the risk of chronic asthma in childhood per 1,000 children if no child had hospital admission for bronchiolitis during infancy. I also quantified the absolute risk and risk difference of chronic asthma by SIMD group and proportion of this inequity explained by bronchiolitis admissions. As discussed in Chapter 6, SEP is not suitable for causal investigations as such, so I quantified the extent of the exposure-outcome association by estimating the marginal risk of chronic asthma by SIMD group using a logistic regression model of the outcome against the exposure using IPW. The weights used for this are found by logistic regression modelling of the exposure that includes the exposure-outcome confounders, year of birth, maternal country of birth and area of residence. As in Chapter 6, the proportion of inequity explained by bronchiolitis admissions was estimated as: $(\text{risk difference} - CDM(m=0)) / \text{risk difference} * 100$.

Bootstrap inference method

To calculate confidence intervals for the above estimates of $CDM(m=0)$, TCE and proportion of disparity explained, whilst addressing the presence of missing data in my dataset, I used single stochastic imputation using chained equations with 10 burn-in iterations as outlined by Micali et al.⁴⁵⁷ Imputation was directly followed by the mediation analysis and both processes were repeated on 1000 bootstrap samples to estimate standard errors. I carried out the imputation under the assumption that missingness was at random, conditional on the variables included in the imputation models. To meet this assumption I conditioned on all covariates from the substantive model, alongside birth at the identified hospital with high levels of missingness (yes or no), NS-SEC, preterm-related complication (yes or no) and postnatal intensive care stay (yes or no). Partially observed covariates were imputed using linear regression (for birthweight, gestational age, maternal age and number of older siblings), logistic regression (for severe maternal asthma and smoking status) and multivariable logistic regression (for number of older siblings and delivery method) models. Following each imputation, preterm birth, maternal age groups, siblings groups and SGA were defined using the relevant variables. The Stata code for this analyses is available in Appendix 7.3.

7.4. Results

7.4.1. Missing data

As displayed in the flow chart (Figure 7.3), 2.7% of the original cohort of births between 1 January 2007 and 31 June 2008 were dropped from the dataset based on the predefined exclusion criteria. Of the remaining 83,853 children in the cohort, 10.3% (8,632) children had missing data for at least one variable. Missingness was driven by infants without linked maternal records ($n=4,353$, 5.2% of the cohort) and missing data on smoking status during pregnancy, which was missing for an additional 3,807 of the infants who were linked to mother's records (9.7% missing in total cohort).

As shown in table 7.3, missingness in any variable was strongly associated with area: 45.9% of children with missing data had a Glasgow postcode area. Further investigation revealed that one hospital in Glasgow in particular contained a high proportion of the missing information on maternal smoking status and I, therefore, included this hospital as a binary predictor when performing imputations. In multivariable logistic regression, infants born at this hospital had a 9-fold greater odds of being in the missing data group compared to the non-missing data group (aOR 8.99, 95% CI 8.33 to 9.70). Other notable predictors of missingness were: earlier birth year, spring birth, postal areas not in northern Scotland, mother born outside the UK and intensive care stay after birth. Based on the proportion of missingness in the dataset,⁴⁵⁸ 11 imputation datasets were generated through MICE.

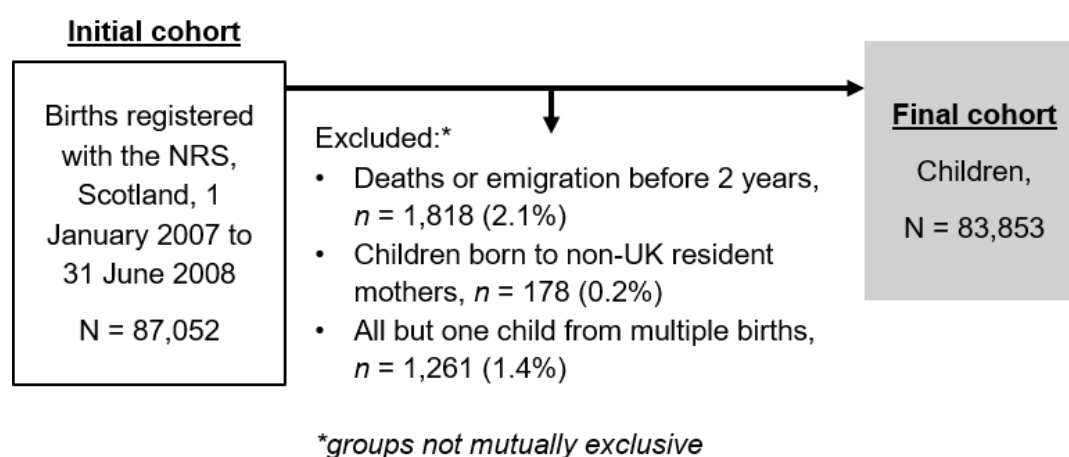


Figure 7.3. Flow diagram showing derivation of final study cohort

Table 7.3. Characteristics of final study cohort (pre-imputation), number of children with missing data and estimated adjusted odds ratios (aORs) for missingness

		Whole cohort		Missing data on ≥ 1 risk factor		
		N	%	N	%	aOR** (95% CI)
Total		83,853	100.0	8,632	100.0	
SIMD	Low	21,991	26.2	2,115	24.5	Reference
	Medium	31,738	37.8	2,974	34.5	1.03 (0.97–1.10)
	High	30,124	35.9	3,543	41.0	0.97 (0.90–1.04)
Any asthma/wheeze*	No	76,034	90.7	7,819	90.6	Reference
	Yes	7,819	9.3	813	9.4	0.99 (0.92–1.08)
Bronchiolitis	No	81,178	96.8	8,404	97.4	Reference
	Yes	2,675	3.2	228	2.6	0.92 (0.80–1.07)
Birth season	Spring	27,967	33.4	3,045	35.3	Reference
	Summer	19,180	22.9	1,787	20.7	0.80 (0.75–0.86)
	Autumn	14,083	16.8	1,479	17.1	0.90 (0.83–0.96)
	Winter	22,623	27.0	2,321	26.9	0.91 (0.85–0.96)
Sex	Female	40,972	48.9	4,212	48.8	Reference
	Male	42,881	51.1	4,420	51.2	0.99 (0.95–1.04)
Chronic condition	No	80,498	96.0	8,236	95.4	Reference
	Yes	3,355	4.0	396	4.6	1.26 (1.13–1.41)
Area	Northern	10,903	13.0	566	6.6	Reference
	Central	23,785	28.4	1,865	21.6	1.52 (1.37–1.67)
	Glasgow	19,516	23.3	3,966	45.9	1.42 (1.28–1.58)
	Edinburgh	14,254	17.0	1,144	13.3	1.55 (1.39–1.72)
	Southern	15,395	18.4	1,091	12.6	1.22 (1.10–1.36)
Maternal county of birth	Non-UK	9,061	10.8	1,053	12.2	Reference
	UK	74,792	89.2	7,579	87.8	0.88 (0.82–0.95)
Maternal age	<20	6,019	7.2	560	6.5	
	20-29	36,099	43.1	1,970	22.8	
	35-39	33,071	39.4	1,523	17.6	
	>40	4,311	5.1	226	2.6	
	missing	4,353	5.2	4,353	50.4	
Severe maternal asthma	No	77,518	92.4	4,177	48.4	
	Yes	1,982	2.4	102	1.2	
	missing	4,353	5.2	4,353	50.4	
Maternal smoking	No	59,093	70.5	371	4.3	
	Yes	16,600	19.8	101	1.2	
	Missing	8,160	9.7	8,160	94.5	
Siblings	0	37,158	44.3	2,869	33.2	
	1	26,701	31.8	863	10.0	
	2+	15,337	18.3	243	2.8	
	missing	4,657	5.6	4,657	54.0	
Delivery method	Vaginal	59,403	70.8	2,951	34.2	
	Elective caesarean	8,202	9.7	425	4.9	
	Emergency caesarean	11,893	14.2	901	10.4	
	Missing	4,355	5.2	4,355	50.5	
Premature birth	No	74,302	88.6	3,820	44.3	
	Yes	5,140	6.1	401	4.6	
	Missing	4,411	5.3	4,411	51.1	
SGA*	No	71,596	85.4	3,330	38.6	
	Yes	7,442	8.9	487	5.6	
	Missing	4,815	5.7	4,815	55.8	

*Children meeting any asthma/wheeze definition between 2 and 9 years; **(aOR = adjusted odds ratio) multivariable logistic regression comparing the likelihood of being in the missing data group compared to the non-missing data group. Also included in the missingness model and, for simplification, not shown in table: birth at hospital with high levels of missing data, aOR 8.99 (95% CI 8.33 to 9.70); later birth year, aOR 0.87 (95% CI 0.82 to 0.92); intensive care stay, aOR 1.18 (95% CI 1.08 to 1.29); prematurity-related complication, aOR 1.01 (95% CI 0.87–1.18); and NS-SEC analytic classes.

7.4.2. Cohort characteristics

The final study cohort comprised 83,853 children of which 48.9% were female and 40.3% were registered with a Glasgow or Edinburgh postal area at birth (table 7.3). At least 6.1% infants were born before 37 weeks' gestation and 8.9% were classified as SGA. Just under a quarter of births with a delivery method recorded were born via a caesarean section and 16,600 mothers (19.8%) reported smoking traditional cigarettes during their pregnancy (among those for whom delivery method and smoking data were available). Of the cohort, 3.2% (2,675) were admitted to hospital at least once with a diagnosis of bronchiolitis in the first year of life and 9.3% (7,819) met the study definition for asthma/wheeze at one time point or more between the ages 2 and 9. Information from the mother's delivery record was available (i.e. linked) for 79,500 (94.8%) of children and the average follow-up time per child was 9.87 years. Of children in the cohort at age 2 years, 2,325 (2.8%) either emigrated (according to their CHI registry) or died by their 9th birthday (see Table 7.4).

Table 7.4 Observed completeness and frequency of asthma/wheeze symptoms, by age of child

Age	Completeness of asthma/wheeze indicator			Asthma/wheeze symptoms <i>N</i>	% of cohort (<i>N</i> = 83,853)
	Complete child record <i>N</i>	Child died or emigrated <i>N</i>	% of cohort (<i>N</i> = 83,853)		
2	83,853	0	0.0%	2,212	2.64
3	83,298	555	0.7%	2,771	3.30
4	82,783	1,070	1.3%	3,209	3.83
5	82,431	1,422	1.7%	3,190	3.80
6	82,182	1,671	2.0%	3,074	3.67
7	81,959	1,894	2.3%	3,093	3.69
8	81,735	2,118	2.5%	3,018	3.60
9	81,528	2,325	2.8%	3,151	3.76

7.4.3. Growth modelling

Table 7.5 presents model fit indices for the LCGA and GMM growth modelling. Only results for two versions of GMM models are presented; the best fitting one is a 3-class linear model, which had a high entropy value (0.98) and a lower BIC value compared to the equivalent 3-class linear LCGA solution. However, a global maximum solution could not be obtained for this 3-class GMM model and one of its classes contained only 0.5% of the sample. GMM models with non-linear growth failed to converge, even after trying both strategies suggested by Muthén and Muthén (using estimated class-specific growth factors means/coefficients from LCGA/smaller GMM models) to determine starting values.⁴⁴⁶

The LCGA model that identified 4 groups with cubic growth was selected as the most suitable model overall. This decision was guided by previous research, a low BIC value relative to other solutions, as well as groups each with >1% of the cohort (see Table 7.5). Figure 7.4 displays average latent class probabilities (classes defined by colour, see key), grouped by most likely class membership. The mean membership probabilities of classes in this 4-class cubic LCGA solution were between 0.89 and 0.99, and the overall entropy value was 0.97 indicating a well-classified model. The posterior distributions of asthma/wheeze prevalence at each time point by class are displayed in Figure 7.5.

The 4 trajectory groups are not necessarily true phenotypic groups, but represent descriptions of the variation in individual trajectories, where the prevalence of each group is based on estimated posterior probabilities. The groups can be summarised as:

1. Never or infrequent asthma/wheeze: 93.3% of the cohort had little to no asthma/wheeze symptoms at any time point.
2. Early-transient asthma/wheeze: 2.6% of the cohort had asthma/wheeze symptoms that had begun by age 2, had a peak prevalence at age 3 and dissipated by age 7.
3. Intermediate-onset asthma/wheeze: 2.2% of the cohort had asthma/wheeze symptoms that had begun by age 4 and had a peak prevalence at age 8, a year before follow-up ended.
4. Early-persistent asthma/wheeze: 1.8% of the cohort had asthma/wheeze symptoms that had begun by age 2, had a peak prevalence at age 6 and were ongoing at the end of follow-up.

Table 7.5. Model fit indices for LCGA and GMM on asthma/wheeze symptoms (selected model highlighted)

		LL	BIC	Entropy	VLMR (p)	BRLT (p)	Sample size per class (based on estimated posterior probabilities)											
							Class 1		Class 2		Class 3		Class 4		Class 5		Class 6	
							N	%	N	%	N	%	N	%	N	%	N	%
LCGA																		
3-class	Linear	-62,051	124,193	0.96	<0.001	<0.001	77,943	93.0	4,087	4.9	1,824	2.2						
	Quadratic	-61,670	123,465	0.95	<0.001	<0.001	77,768	92.7	4,189	5.0	1,898	2.3						
	Cubic	-61,593	123,345	0.95	<0.001	<0.001	77,789	92.8	4,134	4.9	1,931	2.3						
4-class	Linear	-60,925	121,974	0.97	<0.001	<0.001	78,230	93.3	2,296	2.7	1,802	2.1	1,526	1.8				
	Quadratic	-60,147	120,464	0.97	<0.001	<0.001	78,232	93.3	2,216	1.9	1,850	2.6	1,555	1.9				
	Cubic	-60,014	120,244	0.97	<0.001	<0.001	78,294	93.3	2,180	2.6	1,838	2.2	1,541	1.8				
5-class	Linear	-60,620	121,400	0.92	<0.001	<0.001	75,204	89.7	4,681	5.6	1,851	2.2	1,141	1.4	978	1.2		
	Quadratic	-59,938	120,091	0.97	<0.001	<0.001	78,393	93.5	1,892	2.3	1,519	1.8	1,401	1.7	649	0.8		
	Cubic	-59,782	119,836	0.97	<0.001	<0.001	78,147	93.2	2,214	2.6	1,472	1.8	1,198	1.4	823	1.0		
6-class	Linear	-60,433	121,059	0.94	0.007	0.008	76,063	90.7	3,505	4.2	1,754	2.1	1,309	1.6	613	0.7	611	0.7
	Quadratic	-59,745	119,751	0.96	0.010	0.009	77,903	92.9	1,609	1.9	1,543	1.8	1,123	1.3	927	1.1	748	0.9
	Cubic	-59,594	119,518	0.97	<0.001	<0.001	78,150	93.2	1,682	2.0	1,527	1.8	1,107	1.3	723	0.9	666	0.8
GMM																		
3-class	Linear*	-60,235	120,596	0.98	<0.001	<0.001	82,413	98.2	1,053	1.3	388	0.5						
	Quadratic	<i>Model did not converge</i>																
	Cubic	<i>Model did not converge</i>																
4-class	Linear*	-60,230	120,619	0.79	0.004	<0.001	64,379	76.8	18,453	22.0	646	0.8	376	0.4				

L=log likelihood, BIC=Bayesian Information Criteria, VLMR=Vuong-Lo-Mendell-Rubin likelihood ratio test, BLRT= bootstrap likelihood ratio test;

*model did not reach global solution

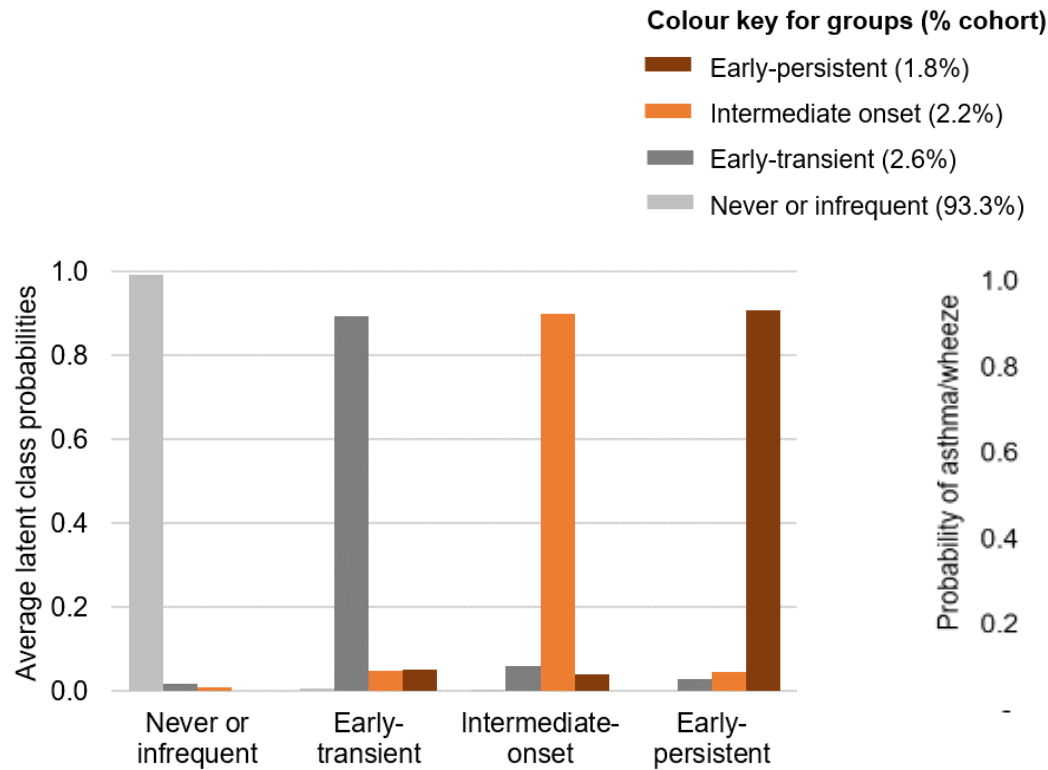


Figure 7.5. Average trajectory group probabilities*

*Classes are defined by colour (see key) and grouped by most likely class membership: results derived from the 4-class LCGA model with cubic growth

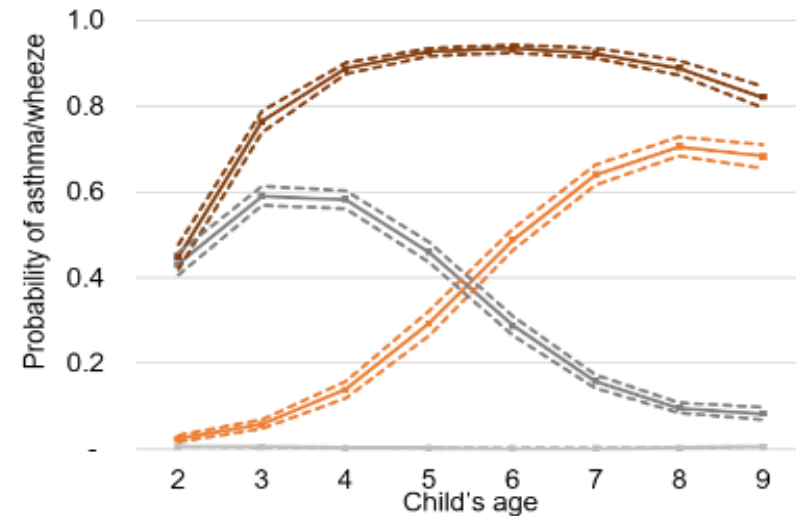


Figure 7.4. Probability and 95% CI of symptoms at each time point for latent trajectory groups*

*Based on posterior distributions of asthma/wheeze prevalence, derived from the 4-class LCGA model with cubic growth. Never or infrequent asthma/wheeze group has a consistent near 0 probability of reporting asthma/wheeze symptoms.

7.4.1. Association between asthma/wheeze trajectories and risk factors

Table 7.6 and Figure 7.6 show the results of multinomial logistic regression models for class membership by asthma/wheeze risk factors. The results presented in this section are based on these results, which take into account error introduced through potential misclassification of children into trajectory groups. I have also included raw risk factor distribution by asthma/wheeze trajectories, including missing data, in Table A7.2 (Appendix 7.4). The results in this table are based on most likely class membership (i.e. not accounting for error) in the pre-imputation dataset. A notable take away from Table A7.2 is the difference in raw numbers and percentage of children with ≥ 1 hospital admission for bronchiolitis by trajectory group: 2.8% ($n = 2,206$) in the never or infrequent group, 10.8% ($n=200$) in the early-transient group, 9.0% ($n=139$) in the early-persistent group and 5.5% ($n=90$) in the intermediate-onset group.

As displayed in Figure 7.6, a graded effect of SIMD was observed across all asthma/wheeze groups relative to the never or infrequent group, with the most deprived SIMD group having the highest risk of asthma/wheeze. The most pronounced association was in the early-persistent group: children born into the most deprived SIMD group had a 65% greater relative risk (RRR 1.65, 95% CI 1.45 to 1.87) and children in the medium SIMD group had a 32% greater relative risk (RRR 1.32, 95% CI 1.15 to 1.50) of membership of this latent asthma/wheeze group than the never or infrequent group, compared to those born into the least deprived group. As illustrated in Figure 7.6, hospital admission for bronchiolitis and severe maternal asthma have the strongest unadjusted association with asthma/wheeze group membership. One or more hospital admission with a diagnosis of bronchiolitis in the first year of life increased the relative risk of early-transient asthma/wheeze group membership by 4.44 (95% 3.86 to 5.10), early-persistent asthma/wheeze group membership by 3.47 (95% 2.94 to 4.09) and intermediate-onset group membership by 1.83 (95% CI 1.47 to 2.28), compared to the never or infrequent group. Severe maternal asthma was most strongly associated with early-persistent asthma/wheeze (IRR 3.33, 95% CI 2.77 to 4.00), followed by the intermediate-onset group (RRR 2.71, 95% CI 2.22 to 3.31), then the early-transient group (RRR 1.96, 95% CI 1.56 to 2.45).

Other risk factors shown to have positive associations with group membership across all 3 asthma/wheeze groups, compared to no or infrequent asthma/wheeze, include: chronic condition (46-95% greater relative risk), male sex (46-66% greater relative risk); UK maternal country of birth (22-30% greater relative risk); maternal smoking during pregnancy (23-37% greater relative risk); emergency caesarean delivery method (9-30% greater relative risk compared to vaginal delivery); and SGA (11-33% greater relative risk). Notably, the strongest association for all these risk factors was with the early-persistent group. This was again the

case for preterm birth, which increased the relative risk of early-persistent group membership by 2.08 (95% CI 1.80 to 2.40), early-transient group membership by 1.90 (1.65 to 2.19) and intermediate-onset by 1.23 (95% CI 1.03 to 1.48). There was an S-shaped pattern to maternal age across asthma/wheeze groups, with slightly stronger associations in the youngest and oldest age groups. The presence of older siblings had a negatively graded association with the risk of membership of the early-transient and intermediate-onset groups and no clear association with early-persistent asthma/wheeze. A Glasgow or southern Scottish compared to an Edinburgh postal area at birth was associated with membership of the early-transient and early-persistent group.

Table 7.6. Univariable RRRs (95% CI) of belonging to asthma/wheeze groups compared to the never or infrequent asthma/wheeze group, by risk factor

Risk factor		Early-Transient	Early-Persistent	Intermediate-onset
SIMD	Low	Reference	Reference	Reference
	Middle	1.27 (1.13–1.44)	1.32 (1.15–1.50)	1.24 (1.10–1.41)
	High	1.59 (1.42–1.79)	1.65 (1.45–1.87)	1.42 (1.25–1.60)
Bronchiolitis	Yes vs. no	4.44 (3.86–5.10)	3.47 (2.94–4.09)	1.83 (1.47–2.28)
Season of birth	Autumn	Reference	Reference	Reference
	Winter	1.05 (0.92–1.20)	1.10 (0.94–1.27)	1.03 (0.89–1.19)
	Spring	0.88 (0.77–1.00)	1.13 (0.98–1.31)	1.14 (0.99–1.30)
	Summer	1.02 (0.89–1.18)	1.14 (0.98–1.33)	1.06 (0.91–1.23)
Sex	male vs. female	1.54 (1.41–1.68)	1.66 (1.51–1.83)	1.46 (1.33–1.60)
Area	Edinburgh	Reference	Reference	Reference
	Northern	1.18 (0.99–1.41)	1.19 (1.00–1.43)	0.99 (0.83–1.17)
	Central	1.31 (1.13–1.51)	1.11 (0.95–1.30)	0.91 (0.79–1.05)
	Glasgow	1.42 (1.23–1.65)	1.31 (1.12–1.54)	1.16 (1.01–1.34)
	Southern	1.47 (1.26–1.72)	1.52 (1.30–1.78)	1.06 (0.91–1.24)
Maternal birth country	UK vs. Non-UK	1.22 (1.05–1.42)	1.30 (1.10–1.54)	1.25 (1.07–1.47)
Maternal age	<20	1.21 (0.90–1.63)	1.19 (0.85–1.66)	1.24 (0.94–1.63)
	20-29	1.25 (0.97–1.62)	1.25 (0.93–1.66)	0.94 (0.74–1.20)
	30-39	0.90 (0.69–1.17)	1.02 (0.77–1.37)	0.79 (0.62–1.00)
	≥40	Reference	Reference	Reference
Maternal asthma	Yes vs. no	1.96 (1.56–2.45)	3.33 (2.77–4.00)	2.71 (2.22–3.31)
Maternal smoking	Yes vs. no	1.23 (1.11–1.37)	1.37 (1.23–1.53)	1.19 (1.07–1.33)
Older siblings	0	Reference	Reference	Reference
	1	0.94 (0.85–1.03)	1.01 (0.91–1.12)	0.94 (0.84–1.04)
	≥2	0.80 (0.71–0.91)	1.02 (0.90–1.16)	0.84 (0.74–0.96)
Chronic condition	yes vs. no	1.76 (1.48–2.10)	1.95 (1.63–2.34)	1.46 (1.19–1.79)
Caesarean delivery	Elective vs. vaginal	0.98 (0.85–1.14)	1.10 (0.93–1.29)	1.06 (0.91–1.24)
	Emergency vs. vaginal	1.18 (1.04–1.33)	1.30 (1.15–1.48)	1.09 (0.95–1.24)
Preterm birth	yes vs. no	1.90 (1.65–2.19)	2.08 (1.80–2.40)	1.23 (1.03–1.48)
SGA	yes vs. no	1.11 (0.96–1.29)	1.33 (1.15–1.54)	1.13 (0.97–1.31)
Year	2008 vs. 2007	0.91 (0.83–1.00)	0.99 (0.90–1.10)	0.96 (0.88–1.06)

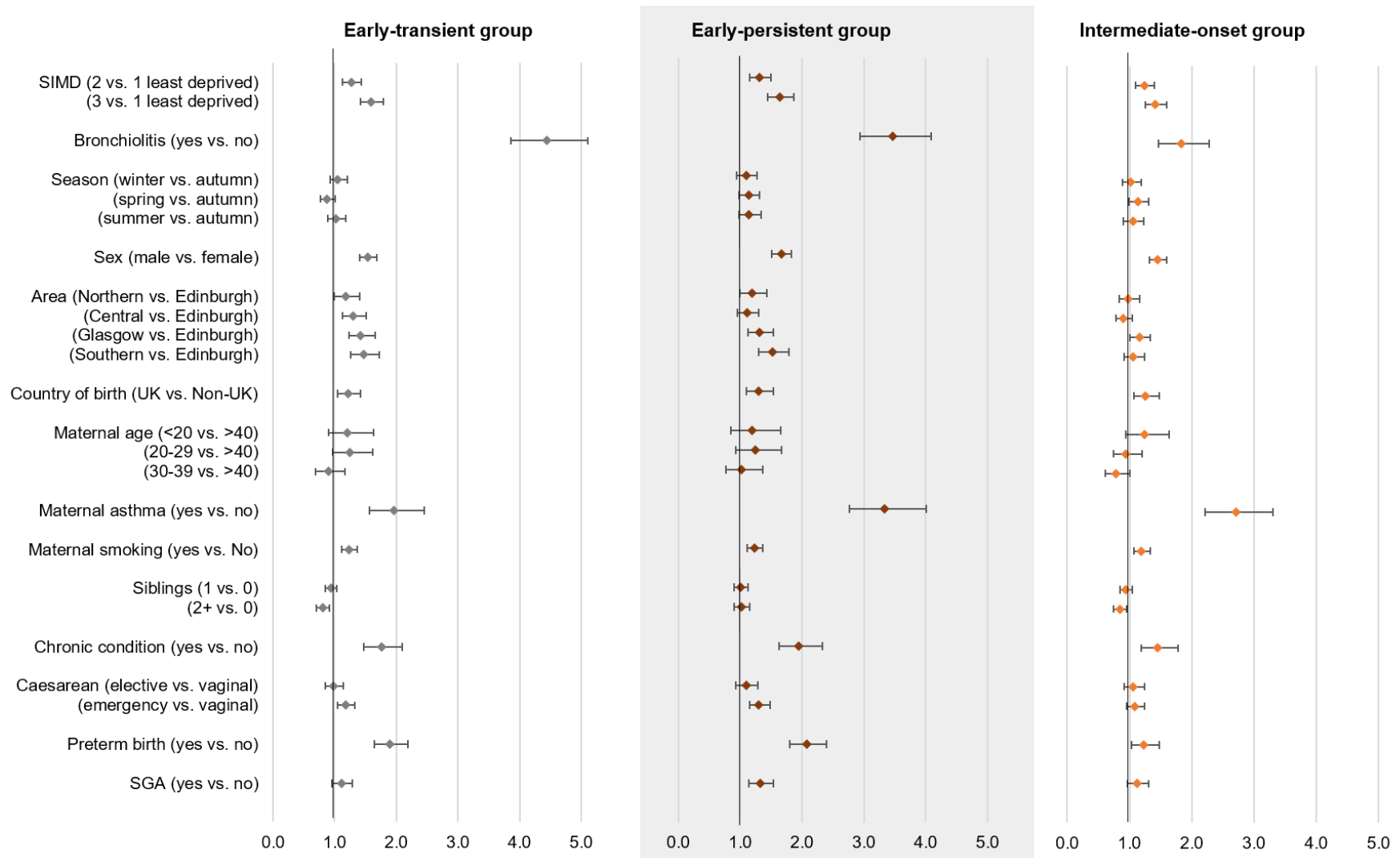


Figure 7.6. Forest plots showing univariable RRR estimates with 95% CI for membership of asthma/wheeze groups compared to the never or infrequent asthma/wheeze group, by risk factor

7.4.2. Mediation analysis

As shown in Table 7.7, After accounting for exposure-outcome confounders, the potential risk of chronic asthma by age 10 would be 30.8 per 1000 children (95% CI 28.3 to 33.3) if all children were assigned to the low SIMD group at birth. This compares to 37.7 per 1000 for the medium SIMD group (95% CI 36.2 to 39.3) and 43.7 per 1000 for the high SIMD group (95% CI 41.3 to 46.1). The CDM($m=0$), calculated using IPW of MSMs, suggests that if no children had a hospital admission for bronchiolitis in infancy, the counterfactual disparity between children in the high compared to low SIMD group would be reduced from 13.0 chronic asthma cases per 1000 children (95% CI 9.6 to 16.4) to 10.7 per 1000 (95% CI 7.3 to 14.0). This is an estimated disparity reduction of 17.9% (95% CI 10.4 to 25.4). A larger proportional disparity reduction was estimated between the medium and low SIMD groups (21.2%, 95% CI 4.9 to 37.5). This reflects a larger relative (but lower absolute) change in potential risk difference, from 7.0 chronic asthma cases per 1000 (95% 3.7 to 10.3) before to 5.5 per 1000 (95% CI 2.3 to 8.8) after counterfactual disparity in bronchiolitis admission is removed.

Table 7.7. Potential risk of chronic asthma* by age 9, estimated counterfactual disparity in risk and corresponding disparity reduction involving bronchiolitis admissions, by SIMD group*

SIMD group	Marginal risk per 1000 children** (95% CI)	Marginal risk difference per 1000 children** (95% CI)	CDM($m=0$) per 1000 children (95% CI)†	% disparity reduction (95% CI)†‡
IPW MSMs				
Low	30.8 (38.3, 33.3)	Reference	Reference	Reference
Medium	37.7 (36.2, 39.3)	7.0 (3.7, 10.3)	5.5 (2.3, 8.8)	21.2 (4.9, 37.5)
High	43.7 (41.3, 46.1)	13.0 (9.6, 16.4)	10.7 (7.3, 14.0)	17.9 (10.4, 25.4)

*binary outcome created from most likely trajectory group membership (early-persistent and intermediate-onset indicating yes); ** Risk and absolute risk difference after accounting for differences in year of birth, maternal country of birth and area of residence between groups using IPW; †95% CIs calculated using the bootstrap with 1000 replications; ‡disparity reduction (%) was estimated as = (risk difference-CDM($m=0$))/risk difference*100

7.5. Discussion

Latent growth modelling applied to longitudinal observations of asthma and wheeze symptoms retrieved from administrative health records led to the identification of four different asthma/wheeze trajectories between ages 2 and 9 years. I labelled the groups as follows:

never or infrequent (93.3% of the cohort), early-transient (2.6%), early-persistent (2.2%), and intermediate-onset asthma/wheeze (1.8%) trajectory groups. There was a socioeconomic gradient to group membership across all asthma/wheeze groups compared to the never or infrequent group, with children in the most deprived group having between 42-65% increased risk of asthma/wheeze compared to children in the least deprived group. At least one admission to hospital with bronchiolitis was strongly associated with asthma/wheeze group membership, particularly the early-transient and early-persistent groups, which had a 4.4 times and 3.5 times greater risk respectively. I estimated that if bronchiolitis admissions in infancy were eliminated then up to 18% of the difference in the risk of chronic asthma between children from the most and least deprived 30% of SIMD would be removed.

7.5.1. Strengths and weaknesses

Using administrative data, this study was able to follow up more than 80,000 children until their tenth birthday to determine asthma/wheeze trajectories. I am aware of only one other study, based in Canada, that has modelled asthma trajectories on such a large population-based level.²⁴ A range of clinical definitions were used to describe asthma/wheeze in this study with input from a clinician, in contrast to the commonly used self- or parental-reported measures used elsewhere.¹⁹⁷ This approach, paired with a cautious methodology to defining symptoms by dispensed medication (≥ 4 prescriptions in a 12 month period), means there are likely to be less false positives captured in this study.⁴⁵⁹ This is mirrored by the high proportion of children classified as most likely belonging to the none or infrequent asthma/wheeze group: over 93% of the study population compared to 82-83% in studies using the Millennium Cohort Study.^{21,25}

On the other hand, some children (likely with milder symptoms) may have been misclassified as having no symptoms in this study. The addition of primary care records in future administrative data studies may improve identification of these cases. Using only indicators of asthma and wheeze presumes that only these latent variables underlie the heterogeneity in how these conditions evolve over time.⁴⁴³ Other studies have included measures of atopy in modelling of asthma trajectories;^{21,460} however, overlap between medication used to treat asthma and atopic conditions meant that the current dataset would not have allowed for this nuanced differentiation. In addition, although I used predetermined criteria to determine trajectories, these methods still have a degree of subjectivity.²¹ Other studies have invoked avoidance of classes $< 5\%$ of the study population;⁴⁴³ however, this seemed unnecessarily restrictive in a sample where 5% of the sample is more than 4,000 children.

I had to take other several pragmatic steps in this study based on constraints of the datasets. I restricted the sample to children born between 2007 and mid-2008 to obtain a 10-year follow

up period that included PIS data (only available from 2009 onwards) from age 2 for all cohort members, which was necessary to model asthma outcomes. The lack of complete PIS data before 2009 meant that I could not use dispensed medication data to define maternal asthma prior to their child's birth. I instead relied on diagnoses during hospital admissions, leading to an indicator that would have mostly captured more severe and poorly managed cases of asthma. Given that poorly controlled symptoms are more prevalent in the most deprived socioeconomic groups, this limited definition of maternal asthma likely conflates the association between SEP and maternal asthma. I dichotomised the outcome for the mediation analysis into a binary indicator of chronic asthma/wheeze. I considered examining the four classes, but encountered problems of stability of the estimates due to the small numbers in some subgroups. In the future, cohort studies from Scottish administrative data with births spanning several years will enable longer follow up periods and a greater sample size for more nuanced analyses.

LCGA, used to define trajectory groups in this study, assumes zero variance within classes and may not be the best statistical representation of the underlying population. Successive cubic LCGA models (with more classes) led to an increasingly improved fit according to BIC and BLRT values, and the 4-class solution was chosen because of small class sizes in the 5-class model. Berlin and colleagues write that, rather than signifying a better fitting model, the pattern towards increasing classes may actually be the result of a poorly-specified model in data that has substantial within-class variability.⁴⁶¹ I also modelled trajectories in the data using GMM, which offers a more flexible approach to variability within classes. However, non-linear GMM models failed to converge—a known problem with these models.^{439,461} As my aim was to replicate classes found in previous studies within a Scottish context (and use these to look further at pathways to inequities in asthma), rather than produce new trajectories, the resemblance of classes in this study to those found in numerous other studies gives me confidence with the results presented.

I used a three-step approach to growth modelling, which meant that I firstly created the latent classes and then used these classes in further analyses. Asparouhov and Muthén write that one-step models, which combine latent class modelling with the secondary model, can be flawed as the additional variables may affect latent class formation and result in unintended changes to the latent class meaning.⁴⁵⁴ I included classification errors in analyses looking at the association between risk factors and trajectories to account for potential latent class misclassification introduced in latent class analysis.¹⁹¹ I was unable to do this in the mediation analysis, which has the potential to affect the precision of estimates, although may not be an issue in light of the well-defined classes in this study (see Figure 7.4).

There were several confounders identified in previous research that could not be included in this study, which may have led to an overestimation of the mediating effect of bronchiolitis. This includes breastfeeding, which has broadly shown to be a protective factor for bronchiolitis and asthma, and associated with lower levels of deprivation,^{194,196,197} and outdoor air pollution and damp/mould, which are positively associated with socioeconomic deprivation and respiratory ill-health.^{116,125,153,202} The confounding effect of some unmeasured confounders may be partially captured by other confounders included in the model. For example, passive smoke exposure after birth by maternal smoking, and ethnicity by maternal country of birth. The pathways with which the final three unmeasured confounders (gestational diabetes, pre-eclampsia and breech presentation birth) are thought to influence offspring respiratory are also included as confounders e.g. delivery method and preterm birth.^{182,462}

Future studies utilising administrative data could be enhanced with linkage to other datasets, for example the Department for Environments Food and Rural Affairs' modelled background pollution data,⁴⁶³ which will enable more of these confounders to be accounted for and bias to be eliminated from calculations. Alternatively, purposefully designed cohort studies offer detailed information on many variables that are not typically available in administrative data sources. For example, the Millennium Cohort Study has been used to examine the complex causal pathways between socioeconomic circumstances and adolescent mental health.⁴⁶⁴

7.5.2. Comparison to other studies

The classes identified in this study are similar to those found in multiple other studies, showing four trajectories with distinct timing in the onset and longevity of symptoms. As noted above, the size of the classes were smaller (proportionally) than those found in other studies; however, the relative size between asthma/wheeze trajectory groups is similar. I will now compare my results to those of other studies, drawing particularly from the 2020 systematic review and meta-analysis of childhood wheeze trajectory-specific risk factors by Owora and Zhang.¹⁹⁷ These authors note that there are few risk factor associations that are trajectory specific across the literature (i.e. there are similar associations for all trajectory groups compared to the never or infrequent group), which is broadly the finding of my study.

Early-transient asthma/wheeze

Asthma/wheeze symptoms that begin early, but cease after the first few years of life have been identified and defined as early transient wheeze across the literature.¹⁹⁷ Membership of this group is strongly associated with viral respiratory infection. In my study, hospital admission for bronchiolitis in infancy was associated with a 4.4 times increased risk of early-transient

asthma/wheeze group membership. This trajectory group has been found to have a low risk of atopy and of future asthma, but to still exhibit diminished airway function in mid-childhood.^{427,465} Greater number of siblings was associated with a low relative risk of membership of this group in my study. In contrast, Owora and Zhang found a 47% greater odds of early-transient wheeze group membership among children with siblings in their meta-analysis (pooled OR 1.47, 95% CI 1.06 to 2.04).¹⁹⁷ The exclusion of symptoms under 2 years in my study outcome, the ages where children are most sensitive to infections and thus the influence of virus transmission from older siblings,⁴⁴ may contribute to these findings.

Early persistent asthma/wheeze

Prolonged asthma/wheeze symptoms that begin in early life are associated with the poorest prognosis in term of lung function out of the asthma groups identified in other research.¹⁹⁷ In my study, membership of this group had the strongest association with almost all risk factors, including socioeconomic position, relative to other asthma/wheeze groups. This group appears to mirror non-atopic asthma groups found in other studies; however, as a measure of atopy was not captured in my study this cannot be confirmed. In a trajectory analysis of asthma and atopy symptoms using the Millennium Cohort Study, children who were in the poorest households at 9 months, had the highest likelihood of being in the group with high wheeze and no atopy at ages 3, 5 and 7. Conversely, the most advantaged group were most likely to be in the low wheeze but high atopy group.²¹ As in Owora and Zhang's meta-analysis, family of history was most predictive of early-persistent wheeze (maternal asthma was associated with more than 3-fold increased relative risk compared to no or infrequent asthma/wheeze in my study).

Intermediate-onset asthma/wheeze

Two asthma/wheeze trajectories that begin at different points in mid-childhood have been described in other studies, labelled intermediate-onset (beginning between 3 and 5 years) and late onset (beginning between 6 and 8 years) by Owora and Zhang.¹⁹⁷ I identified only one of these groups my final 4-class model, perhaps owing to my relatively strict definition of asthma/wheeze and the lack of follow up beyond the age of ten. In future work with a longer follow up, differences in the number of identified classes by stringency of asthma/wheeze definition could be explored. In my analysis, the intermediate-onset group had broadly similar but weaker associations with the included risk factors compared to the early-transient and early-persistent groups (versus the never/infrequent group). This group appears to mirror an atopic asthma phenotype found in other studies,^{427,430} although this association cannot be ascertained from my study. Risk factors operating in early childhood (as opposed to the pre-

and perinatal risk factors included in my study), such as growing up in a non-farming household and a lack of pet ownership, may partially explain the development of later onset atopic childhood asthma/wheeze.^{430,466}

The contribution of bronchiolitis to chronic childhood asthma

There has been a longstanding debate on the link between early RSV-related lower respiratory tract infections (including bronchiolitis) and the development of asthma, although recent research point towards a non-causal relationship.⁴⁶⁷ Two systematic reviews of the literature on the link between early RSV-related lower respiratory tract infections and subsequent pulmonary function were published in 2020.^{468,469} One of these concluded that the evidence was “not overwhelming with conflicting results between studies”.^{469 p.1567} The other review reached a firmer conclusion: “Our findings, limited to exposure and immunoprophylaxis studies, do not support basing policy decisions on an assumption that prevention of RSV-LRTI will reduce recurrent chronic wheezing illnesses.”^{468 p.795} Notably, both reviews identified a high risk of confounding bias in observational studies.^{468,469}

Novelly, I have used causal inference methods, with explicit discussions of the involved assumptions, to reach the estimates in the current observational study. This is the first time the pathways between socioeconomic position, bronchiolitis and asthma have been explored using these methods. In my study, I estimate that up to 18% of the disparity between socioeconomic groups (at the extreme 30 percentiles of SIMD) may be attributable to hospital admission for bronchiolitis during infancy. Moreover, it shows that there is at least 82% of the association unexplained by bronchiolitis and broadly highlights the need to further investigate the inequities in admissions using causal inference methods. Notably, as part of this work, I have recognised some key confounders that are missing from my analyses, including air pollution, a lack of breastfeeding and damp/mouldy housing conditions. Research including information on other identified confounders is needed to clarify these pathways further.

7.6. Chapter summary

In this chapter, I have used trajectory analysis methods to define groups of asthma/wheeze within a Scottish population. I identified four distinct trajectories of symptoms in early life, labelled no or infrequent, early-transient, early-persistent and intermediate-onset asthma/wheeze, which is similar to those found in other studies. I looked at the association of membership of these groups with several early life risk factors and found notable associations with bronchiolitis and severe maternal asthma. There was also a clear socioeconomic gradient

to group membership. Using counterfactual disparity measures, I quantified the role of bronchiolitis admissions in the relationship between socioeconomic position and chronic asthma. I estimated that up to 18% of the inequity between groups was explained by hospital admissions for bronchiolitis in infancy, although unmeasured confounding in my analyses limits the causal interpretation of this finding.

Chapter 8. Discussion

Recap and chapter overview

I began this thesis by presenting an overview of research in the area of socioeconomic inequity, bronchiolitis and asthma among paediatric populations in the UK. I outlined the administrative data sources that enabled me to conduct my research and discussed the emerging issue of data opt outs to research using these types of data sources. In total, I carried out four main studies to look at the association between socioeconomic deprivation and bronchiolitis and asthma, two common and highly burdensome respiratory conditions. This final chapter draws on information from all these chapters to highlight the overall findings, as well as the strengths and limitations of my thesis. I will highlight the unique contribution of my work, before outlining the main implications of the key findings on future policy, practice and research. Lastly, I will re-assess my study implications in the context of the SARS-CoV-2 pandemic, which was ongoing at the time of writing.

8.1. Critical assessment of my PhD

8.1.1. Restatement of thesis rationale and aim

Bronchiolitis and asthma are major causes of morbidity among children in the UK and are responsible for substantial burden on the national health care system.^{4,32} Despite this, there are currently no preventative or curative measures for the majority of children that develop bronchiolitis or asthma.^{15,47} There is a large body of evidence investigating environmental, social and biological risk factors for these conditions, but much of this work does not fully consider the ways in which various risk factors interact with each other (and potentially bias results). There is also a myriad of research looking specifically at the link between RSV bronchiolitis and subsequent asthma,^{53,56,264} but precise risk factors leading to both conditions remain elusive. What we do know is that a prevailing facet of both conditions is the inequity in prevalence across socioeconomic groups.^{4,29} Identifying the factors that may mediate or confound the pathway between socioeconomic deprivation, bronchiolitis and asthma will help to unravel this complex picture. This is a pertinent step to working out how and when to best intervene to prevent either or both of the conditions. By focussing on early life risk factors, the aim is to reduce the burden of bronchiolitis and asthma by closing the gap in the incidence/prevalence of these conditions between children born into the most and least deprived socioeconomic groups.

To guide the structure of my work, I drew from the WHO's multifaceted conceptual framework developed for the Commission on Social Determinants of Health (CSDH).⁶⁰ As displayed in Figure 8.1, this framework differentiates between the levels of factors that likely influence health outcomes (in this case, the incidence of bronchiolitis and prevalence of asthma). According to the CSDH, at the outmost level are the wider social, economic and political context of society, which, in turn, stratify the population into socioeconomic positions (SEP). I do not include this outer level in my diagram, choosing to instead focus my work on the factors that sit between SEP, bronchiolitis and asthma because these pathways arguably represent the more pragmatic and achievable approach to intervention. My results add to the substantial evidence base showing the early origins of socioeconomic inequities in disease, which highlights that a broader approach to tackling the unequal distribution of health outcomes is warranted.

I focused broadly on SEP as the structural determinant of health in my project rather than more specific strata, which can include income, education, occupation, ethnicity and gender. I bring in specific components of the physical and built environment that likely interact with SEP and then, in turn, influence the presence of intermediary factors to shape inequities in health outcomes. For bronchiolitis and asthma, intermediary factors operating in the pre-natal and early post-natal period include tobacco smoke and air pollution exposure, premature birth, chronic conditions and poor housing conditions (shown in Figure 8.1). My project focuses in on specific components of this framework to better explain how inequities in bronchiolitis and asthma arise.

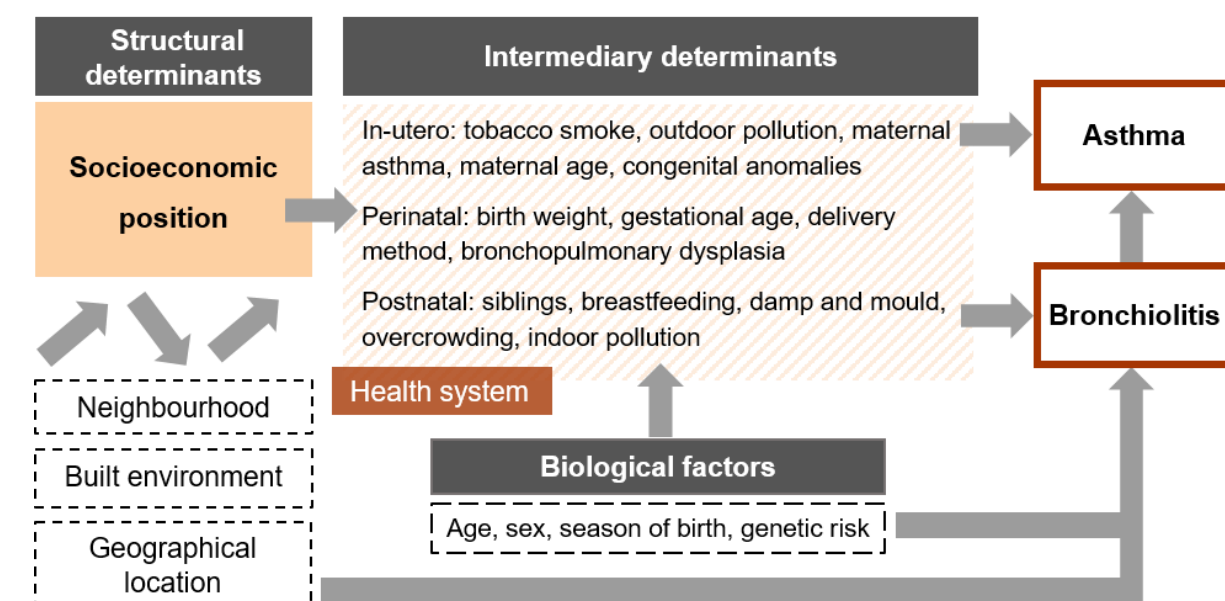


Figure 8.1. Framework for pathways to socioeconomic inequities in the risk of bronchiolitis and paediatric asthma

Specifically, the aim of this PhD was to describe associations and explore pathways between early life socioeconomic deprivation, infant bronchiolitis and asthma in mid-childhood.

Focussing on specific pathways involved careful consideration of covariates, statistical methods and related assumptions, which I clearly outlined in each study. Relatedly, the datasets for this work are national birth cohorts created from English and Scottish administrative health and vital event datasets that, although extremely beneficial to epidemiological analyses, bring their own set of complications. Underlying my project aim are four major objectives, and I will now outline the key findings for each in turn.

8.1.2. Key findings

Thesis objective 1 To describe the association between seasonality, socioeconomic position and bronchiolitis admission rates among infants across England

Local area-level (Chapter 4)

- There was a 5.3-fold difference in rates of bronchiolitis admissions across Clinical Commissioning Groups (CCGs) in England, with about one quarter of the variation explained in rates by area-level IMD scores
- The peak epidemic week of hospital admissions for bronchiolitis varied by 3 weeks across CCGs and tended to be earlier among CCGs with higher population densities, such as those located around/within London and Manchester

Individual level (Chapter 5)

- The seasonality of admissions varied marginally by individual-level socioeconomic group; the predicted peak epidemic week was 49.4 among infants in the least deprived IMD quintile and 50.3 in the most deprived IMD quintile
- Socioeconomic disparities in admissions to hospital for bronchiolitis persisted after taking seasonality into account, with a 40% greater risk of admissions among infants in highest compared to lowest quintile of IMD at week 50 (peak admission week)

Thesis objective 2 To investigate the mediating contribution of maternal smoking in the pathway between early life socioeconomic position and bronchiolitis admission rates in infancy (Chapter 6)

- Among children born in Scotland between 2012 and 2016, I estimated that the potential risk of bronchiolitis hospital admission in infancy would have been 57.1 per 1000 live births if all children were in the most deprived 30% of SIMD at birth
- This was 16.3 more children with admissions for bronchiolitis for every 1000 infants were all children in the least deprived 30% of SIMD at birth
- Using two methods to calculate the counterfactual disparity measure, I estimated that the elimination of maternal smoking would reduce about one fifth of the disparity in the proportion of children who had at least one hospital admission for bronchiolitis in the first year of life

Thesis objective 3 To describe typical trajectories of asthma/wheeze among children <10 years in Scotland and their association with early life risk factors, including SEP (Chapter 7)

- I identified four trajectories of asthma/wheeze symptoms from the Scottish dataset, which I labelled never or infrequent (93.3% of the cohort), early-transient (2.6%), early-persistent (2.2%), and intermediate-onset asthma/wheeze (1.8%)
- There was a socioeconomic gradient to group membership across all asthma/wheeze groups compared to the never/infrequent group, with children in the most deprived 30% of SIMD having between 42-65% increased risk of asthma/wheeze compared to children in the least deprived 30% of SIMD
- At least once hospital admission for bronchiolitis in infancy was strongly associated with asthma/wheeze group membership in unadjusted analyses, particularly the early-transient and early-persistent groups, which had a 4.4 times and 3.5 times greater risk respectively

Thesis objective 4 To investigate the mediating contribution of bronchiolitis in the pathway between early life SEP and mid-childhood asthma/wheeze trajectory groups (Chapter 7)

- Among infants born in Scotland between January 2007 and June 2008, I estimated that the prevalence of chronic asthma by age 9 would have been 4.4% if all children were in the most deprived 30% of SIMD at birth compared to 3.1% if all children were in the least deprived 30% of SIMD at birth
- I estimated that the elimination of bronchiolitis requiring hospital admission could reduce up to 18% of the disparity in the proportion of children with chronic asthma
- I identified air pollution, housing conditions and breastfeeding as key risk factors to be incorporated into future research utilising similar causal inference methods

8.1.3. Strengths and limitations

I have discussed the strengths and limitations specific to each study within the relevant chapter. Here, I bring together some of the major themes running through this thesis. Namely, the use of administrative data sources, the measurement of my exposure (SEP) and outcomes (bronchiolitis and asthma), and the statistical methods employed to address my thesis objectives.

Administrative data sources

Using administrative data sources in my thesis meant that I was able to follow up almost the whole population of children born in England and Scotland over the specified years of study.²⁷⁴ These large datasets (approximately 620,000 births in England and 53,000 in Scotland per year) meant that my I could explore complex pathways that included relatively rare events and subgroups of the population.²⁶⁷ There have been very few studies that have looked at the mechanisms of the seasonality of RSV bronchiolitis using epidemiological methods,⁴⁷⁰ and none in a UK setting. Using nationwide hospital admission data in England, I was able to do this for bronchiolitis (a common but distressing condition) in a vulnerable population (infants), whilst focussing on the pressing issue of socioeconomic inequities. I created a Scottish birth cohort with 9 years of follow-up, enabling the exploration of longitudinal patterns of asthma/wheeze symptoms identified using objective measures.

However, I also encountered several drawbacks to using these data sources in my research. This was particularly apparent for the Scottish dataset as I was involved in the process of obtaining, cleaning and using the data from the beginning to end. Obtaining access to this data was a complex and time intensive process (taking 13 months to receive the data following approval of the data request). As part of the data request, we had to specify the exact variables needed for analyses and the reasons why. This meant that I needed to know the desired variables for analysis at the beginning of the process (meaning some later identified clinical risk factors, for example, gestational diabetes and pre-eclampsia, were not available to me). There were issues with the completeness of data. As shown in Chapter 6, patient records from one hospital in Glasgow in particular had a 9-fold greater odds of missing data. I also found that the completion of ethnicity information, a factor that is integral to fully understanding disparities in health outcomes, to be particularly poor (although, improving over time).⁴¹⁴ There are wider ethical issues to the use of these datasets (also, for the most part, beyond the influence of researchers) and I identified patient opt-outs as a particular problem in my work.²⁷²

Socioeconomic position

In Chapter 1, Section 1.2.2, I write about the benefits and disadvantages of IMD (the measure of SEP in my research) at some length. Listed benefits include its national applicability, composite construction and broad use across other health inequity research (aiding comparability of results).^{4,68,69} However, I also describe the potential issue of mathematical coupling in health inequity research where IMD (composed of 7 domains, one of which is health based) is used as the exposure.⁷⁶ I investigated this issue in Chapter 5 by replicating my analysis with the health deprivation and disability domain excluded from IMD and found only slight differences in my results, giving me confidence that this did not bias my inferences. The components within the health domain are all based on adult health (including premature mortality and hospital admissions in adults)³⁵⁸ and therefore may still pose greater issues when looking at health outcomes in non-paediatric populations.

Another problem identified in Chapter 1 with using IMD as an indicator of SEP is its lack of specificity in comparison to other measures. This issue was revisited in Chapter 6 in regards the consistency assumption for causal inference. Here, I described how the exposure definition must have enough precision that any variation in that exposure does not lead to a different outcome.³⁶⁹ There was another measure of SEP available in the Scottish dataset based on occupation, the UK National Statistics Socio-Economic Classification (NS-SEC), which may have been better suited to meet the assumption of consistency. However, occupational categorisations like NS-SEC have been criticised for not fully capturing contemporary occupational structures, particularly given its basis of male-orientated job classifications.⁶⁷ Instead, I used a method of causal inference (the counterfactual disparity measures (CDMs)), which requires less stringent assumptions for estimation and have been careful in the interpretation of my findings.

Defining bronchiolitis and asthma

Both health outcomes studied in this thesis rely on subjective clinical judgement for diagnosis,^{3,17} so a certain amount of imprecision in their identification is likely. The use of clinical definitions are particularly advantageous in comparison to parent-reported cases, which may be influenced by recall biases.^{459,471} I identified bronchiolitis through ICD-10 diagnosis codes in hospital admission records, limited to admissions where the patient was less than 1 year of age (as 90% of admissions for this condition occur in infancy).^{6,261} I did not have information on other (non-hospital admission) health care contacts for bronchiolitis, which means the captured cases likely represent the most severe cases of infection. It also means that capturing bronchiolitis cases is dependent on admission thresholds, highlighted in Chapter 4 where London was shown to have far lower rates of admissions (likely stemming from higher threshold for admission from accident and emergency departments).³⁴⁸ Alongside

admission and death records, I used dispensed medication information to define asthma/wheeze (with clinical input). Asthma is an ambulatory care sensitive condition, meaning that it ought to be preventable by interventions in primary care.²¹⁴ Reliance on hospital admission records for this outcome would have biased my research in favour of children from lower socioeconomic groups, who are far more likely to have emergency treatment for this condition.⁶⁹

8.2. Implications of my research findings

8.2.1. Continuing high and unequal burden of disease

As outlined in Section 1.8.3 (Chapter 1), respiratory diseases, particularly asthma, have been identified as an important focus of public health policy in England and Scotland.^{255,257} My findings add weight to the need to focus resources on preventing both bronchiolitis and asthma as particularly burdensome conditions in the paediatric population. I showed that between 2012 and 2016 almost four in every ten infant admissions to hospitals in England during December included a diagnosis of bronchiolitis. More than 9% of children born between 2007 and mid-2008 in Scotland had symptoms of asthma and/or wheeze at least once that were severe enough to meet my stringent definition.

Prominently, I identified a socioeconomic gradient to the incidence of bronchiolitis admissions, and in the prevalence of all three identified trajectories of asthma/wheeze. This is not a new finding, although the detailed identification inequities across three asthma trajectories in a Scottish context is. Policies in England and Scotland have clearly highlighted the role of socioeconomic inequities in the development of respiratory ill health. The NHS England Long Term Plan, published in 2019, specifically refers to the higher incidence of respiratory disease in “disadvantaged groups and areas of social deprivation”.²⁵⁷ p.66 However, my findings highlight the importance of retaining focus on the social determinants of health and applying interventions in a way that brings down this gradient (i.e. proportionate universalism).⁷⁵ This is particularly pertinent in restating in the context of asthma, for which policy has traditionally focused on controlling severe symptoms and reducing hospital admissions, rather than the prevention of disease.²⁵¹

8.2.2. Clinical practice and policy

By focusing on particular aspects of the framework in Figure 8.1, my work begins to untangle some of the ways in which preventative measures and interventions might be best applied to alleviate the burden of bronchiolitis and asthma.

Prevention of respiratory syncytial virus (RSV) infections

The first two studies in my thesis looked at the interplay of socioeconomic position and wider ecological factors on rates of hospital admissions for bronchiolitis (the left hand side of the framework displayed in Figure 8.1). Patterns of admissions can be viewed as a proxy for the circulation of RSV, which has been shown to cause the vast majority of bronchiolitis cases that lead to hospitalisation.⁶ I found differences in the timing of the epidemic peak at the CCG-level. The peak tended to occur earlier in CCGs that have higher population densities and are within/near major cities, such as London and Manchester. This highlights the potential role of population mixing, and possibly international travel, in the early spread of RSV. These findings have implications on preparedness for bronchiolitis infections. An early warning system could be set up for hospitals in these cities to alert Public Health England and the NHS at the annual onset of admissions for bronchiolitis. These systems could trigger surge planning in paediatric departments in other hospitals, inform the distribution of prophylaxis or future vaccinations, and be used to time simple public health messages, such as warning families to avoid taking their young infants into public spaces or encourage frequent handwashing for older siblings.

I found no clear differences in the timing of admissions for bronchiolitis by SEP at the individual level. This suggests that differential timing of administering palivizumab prophylaxis—the one currently available preventative measure³—may not be the best course of action to tackle inequities in the incidence of this condition. It also suggests that information campaigns to reduce the spread of RSV, such as handwashing, are also unlikely to reduce inequities (as would perhaps be the case had I found that RSV spreads quicker among more socioeconomically deprived families). Instead, I found continued higher levels of admissions for bronchiolitis among infants born into the most socioeconomically deprived households, even after accounting for seasonal patterns. This presents evidence that to reduce inequities in admissions for bronchiolitis, the focus should be on tackling the chronic exposures that make children in more deprived groups susceptible to severe illness when infected with RSV (see below).

It is likely that a vaccine for RSV for pregnant women, which aims to protect infants during the first few months of life, will become available within the next 5-10 years.⁴² The evidence of unequal uptake of vaccines for children across socioeconomic groups,²⁴⁷ paired with a greater reluctance for vaccination among pregnant women,²⁴⁴ means that a vaccine for RSV in pregnancy has the potential to widen inequities in rates of bronchiolitis further. Embedding maternal vaccination within routine antenatal care (as opposed to a separate visit to a primary care centre) has the potential to improve uptake.²⁴⁸ Targeted campaigns, such as those being

implemented to boost Covid-19 vaccine uptake among specific ethnic minority groups in the UK,⁴⁷² may also prove to be beneficial in reducing inequities in maternal vaccine uptake.

Prevention of bronchiolitis and chronic asthma

The majority of children become infected with RSV in early life, but only some infants develop bronchiolitis. In turn, only around 15% of children who are hospitalised with bronchiolitis have underlying conditions that are known to contribute to severity of infection (and an even smaller proportion of these are eligible for palivizumab prophylaxis against RSV infection).^{44,45} My third study concentrated in on one risk factor shown in the intermediary determinants section of Figure 8.1 (maternal smoking) to try to precisely estimate its contribution to inequities in admissions for bronchiolitis. I estimated that maternal smoking contributes to one fifth of the inequities in hospital admissions for bronchiolitis between infants. My fourth study was concerned with the relationship shown at the far right hand side of Figure 8.1. Again using a structured approach I precisely estimated the contribution of admission for bronchiolitis in the relationship between socioeconomic position and chronic childhood asthma. I show that targeting the prevention of bronchiolitis could reduce inequities in chronic asthma by almost one fifth. This further emphasises the importance of concentrating on interventions at the earliest point to prevent inequities in bronchiolitis and, in turn, childhood asthma.

Maternal smoking has long been identified as a risk factor for children's respiratory ill health, and it continues to be on the public health policy agenda. Public Health England's (PHE) strategy for 2020-25 specifically sets out specialist smoking cessation services for pregnant women to tackle social disadvantage in health.²⁵⁹ My findings add evidence to the importance of such policies that focus on maternal smoking (as well as other family members). On the other hand, my research also shows that 80% of the disparities in bronchiolitis admissions are not due to maternal smoking. More research is needed to understand the precise components of this 80%; however, reductions in exposure to air pollution and improving housing conditions in early life likely represent the most important avenues for early intervention. These risk factors are recognised in public health strategies for England and Scotland as contributing broadly to health inequities. PHE's 2020-25 strategy specifically commits to lower air pollution to prevent health inequalities,²⁵⁹ and NHS Health Scotland's 2017 strategy commits to improving suitable housing options. Evaluations of the implementation of these commitments and quantifying their impact on specific outcomes, including bronchiolitis and asthma, are an important next step.

Clinical interventions for bronchiolitis

My work also demonstrates that the previously observed increase in bronchiolitis admissions in England has continued through to 2016 (and likely beyond).⁴ In 2016, the rate of admission for bronchiolitis reached 58.4 per 1000 infant-years in England (95% CI 57.8 to 59.1). I show that these rates are even higher in Scotland (68.8 per 1000 infant-years in 2016, 95% CI 66.6 to 71.0). It has been suggested that hospital and healthcare policies such as changes in admission thresholds and accessibility of primary care services are an important driver of this increase, rather than changes in the transmissibility or severity of RSV.^{4,45} Hospital admission data are influenced not just by disease incidence and severity, but also by clinical decision-making, parental expectations and national healthcare policy (e.g. the availability of out of hours primary care services).^{216,473} Evaluations of alternative models of care for acutely ill infants, which reduce the need for hospital admission, could therefore also contribute to the reduction in bronchiolitis admissions.^{15,473} As argued by the British Lung Foundation (in reference to respiratory admissions on paediatric wards over winter) “much of this pressure could be eased by the development of decision support that would improve care across conditions”.^{350 p.26} The NHS Long Term outlines plans to address accident and emergency attendances in children using locally designed holistic services,²⁵⁷ which may have a beneficial effect on admission for bronchiolitis.

8.2.3. Future research

There are many avenues of further research to be exploited from the findings in this PhD. Many of these suggestions could be achieved through current research initiatives involving linkage between the administrative health data that I used in my thesis to cohort studies or other administrative and environmental datasets.

Geographical and socioeconomic variation in rates of bronchiolitis

Future investigations into the variation in rates of hospital admissions for bronchiolitis across England would profit from the inclusion of clinical factors unavailable in my datasets, such as a measure of admission thresholds. Including information on pathways to hospital admission would also be advantageous, particularly to understanding whether models of care for bronchiolitis differ by SEP. The incongruent patterns of admissions in London compared with the rest of the country is particularly warranted (and research on this is underway).³⁴⁸ Other important factors to consider in future investigations of geographical variation of RSV bronchiolitis epidemics include weather patterns and levels of common air pollutants. Studying the underlying pathogens that cause bronchiolitis will help understand whether the

socioeconomic gradient observed for all bronchiolitis admissions (mostly caused by RSV) are similar for other pathogens. This would require linkage to the Second Generation Surveillance System held by Public Health England or Electronic Communication of Surveillance in Scotland held by Public Health Scotland.⁴⁷⁴ This linkage is underway as part of the PICNIC study, which is a national birth cohort study on air pollution, housing and respiratory tract infections in children in England.⁴⁷⁵

Pathways between infant bronchiolitis and development of asthma

I introduce statistical methods (and include the code for implementation) that can be used in future observational studies to investigate pathways between socioeconomic position and bronchiolitis and asthma. As causal inference methods advance and the potential to meaningfully investigate multiple pathways expands, teasing out the precise contribution of multiple mediators will become easier. Replication of the asthma/wheeze trajectory analyses over many study years is important for advancing understanding of trajectory-specific effects. Several identified confounders were not available in my studies, which ought to be included in future research on the association between bronchiolitis and asthma are outdoor air pollution, housing conditions (including damp and mould, overcrowding and indoor air quality) and breastfeeding.^{196,199,476} These risk factors should additionally be investigated as mediators in the relationship between SEP and asthma to further inform the focus and timing of interventions. This is a particular priority for air pollution, a risk factor with wide public awareness and an ever growing evidence base,^{149,477} but for which the precise early-life contribution to long-term disease risk (particularly asthma) remains to be quantified.⁴⁷⁸

8.3. The impact of the SARS-CoV-2 pandemic

At the time of writing my discussion, the SARS-CoV-2 pandemic is ongoing. The hospital admission and death rate due to this virus and the resulting disease, Covid-19, are particularly high in the UK relative to most other countries around the world. As at 1 March 2021, the UK has the fifth highest recorded number of Covid-19 deaths of any country in the world.⁴⁷⁹ The effects of the pandemic on public health more broadly in the UK will take some time to fully emerge; however, there are some immediate factors that ought to be acknowledged in relation to the findings in my thesis.

At the same time as SARS-CoV-2 continues to circulate, albeit at lower levels than in the autumn and winter of 2020, other respiratory viruses had far lower circulation compared to other years. Surveillance information collated through the Respiratory Datamart system and

published by Public Health England, showed that RSV positivity was 0.0% in week 50 (2020) compared to 13.5% in the same week a year before.⁴⁸⁰ Research published from Western Australia shows a 98% decrease in RSV detections in children through winter 2020.⁴⁸¹ Recent Hospital Episode Statistics for Admitted Patient Care and Outpatient Data (HES APC) statistics published by NHS Digital do not include information on diagnoses, but point to the implications on hospital admissions for bronchiolitis. This provision data show a 55.9% decrease in emergency hospital admissions among 0-4 year olds in December 2020 ($N=61,707$) compared to December 2019 ($N=27,234$).⁴⁸²

These changes in admission patterns for respiratory viruses in children strongly highlight the utility of simple behavioural measures in preventing infectious disease, and has the promise of changes to the transmission of viruses in the future (if sustained). Hand washing is a prominent example, which is a very effective measure to prevent respiratory infection in children.²³⁶ As it is recognised by paediatricians that bronchiolitis places a major burden on NHS paediatric wards and intensive care units between October and March, a larger group of at risk children were recommended to receive palivizumab prophylaxis in 2020/21 to keep them out of hospital.⁴⁸³ However, these patterns of RSV also have the potential to have negative repercussions in future years. It is hypothesised that, partially due to a larger proportion of immunologically naïve children, the next RSV epidemic may be larger than previous years, leading to greater morbidity from infection.⁴⁸¹ Prominently, this period of sudden change offers a natural experiment to examine whether early bronchiolitis actually leads to asthma.⁴⁸⁴

8.4. Concluding remarks

The stark socioeconomic gradient in the incidence of bronchiolitis and the prevalence of asthma are major concerns for child public health in the UK. Intervening on maternal smoking during pregnancy continues to offer an important avenue for intervention, particularly to reduce disparities in infant bronchiolitis infection. I have also demonstrated, using novel application of causal inference methodology, that preventing severe bronchiolitis could reduce the socioeconomic gradient in chronic asthma prevalence by nearly one fifth. This thesis offers a guide to the use of causal inference methods in the study of inequities in health using complex, linked administrative health data. Future work including indicators of air pollution and housing conditions is particularly warranted in the complex unravelling of pathways to bronchiolitis and subsequent childhood asthma.

APPENDIX 0: RESEARCH OUTPUTS

Appendix 0.1. Publications during PhD period

Research papers

Lewis K, De Stavola B, Hardelid P. Is socioeconomic position associated with bronchiolitis seasonality? A cohort study. *J Epidemiol Community Health* Published Online First: 03 September 2020. doi: 10.1136/jech-2019-213056

Lewis K. M., Parekh S. M., Ramnarayan P., et al. Emergency paediatric critical care in England: describing trends using routine hospital data. *Archives of Disease in Childhood*. Published Online First: 22 May 2020. 10.1136/archdischild-2019-317902

Lewis K. M., De Stavola, B. & Hardelid, P. Geospatial and seasonal variation of bronchiolitis in England: a cohort study using hospital episode statistics. *Thorax*, 2020; 75:262-268. [Chosen as Editor's choice for the issue]

Lewis K. M. & Hardelid, P. (2019). National data opt out programme: consequences for maternity statistics in England. *IJPDS*, (2020) 5:1:03. doi.org/10.23889/ijpds.v5i1.1126

Lewis KM, Pikhart H, Morrison J. Does adiposity mediate the relationship between socioeconomic position and non-allergic asthma in childhood? *J Epidemiol Community Health* 2018; 72:390-396.

Lewis, K.M., Ruiz, M., Goldblatt, P. et al. Mother's education and offspring asthma risk in 10 European cohort studies. *Eur J Epidemiol* (2017) 32: 797. doi.org/10.1007/s10654-017-0309-0

Peer reviewer

November 2020 Scientific Reports

May 2019 The Journal of Epidemiology and Community Health

May 2019 BMC Pulmonary Medicine

Appendix 0.2. Conferences and presentations

12-13 September 2019 UCL-Birkbeck MRC DTP retreat. Poster entitled 'Defining asthma/wheeze groups using symptom trajectories'

3-6 September 2019 Joint European Congress of Epidemiology & Society of Social Medicine and Population Health Annual Scientific Meeting (including ECR event on 3rd September), Cork (Ireland). Rapid fire presentation entitled 'National data opt out programme: Consequences for maternal and child health research in England'. Awarded the *Rapid Fire – Pregnancy & Infant, Child & Adolescent Health prize*.

25 February & 9 April 2019 3-minute thesis (3MT) competition, heats 1 & 2. Presentation entitled 'Childhood asthma in the UK: An understated and unequal problem'. Progressed from departmental to Institute wide competition.

23 November 2018 Lancet Public Health Science conference, Belfast. Poster presentation entitled 'Could socioeconomic inequity in bronchiolitis admissions be reduced through intervening on gestational age? Causal inference using a cohort study'.

7 November 2018 Annual ICH Open Day and Poster Competition. Poster entitled 'Seasonal and geographical variation of severe bronchiolitis across England'. Awarded 2nd prize in the Population, Policy and Practice department category.

15-19 October 2018 UK China Workshop on Childhood Respiratory Disease, Chongqing (China). Poster presentation entitled 'Seasonal and geographical variation of severe bronchiolitis across England'. Funded via the Newton Fund.

Appendix 0.3. Secondment

January to July 2020 Fact checker on the editorial team, Full Fact

During this 6-month secondment at the UK's independent, non-partisan, fact checking charity, I researched and published articles on the misreporting of health research in the mainstream media, on social media and by public figures. The majority of these focussed on misinformation about Covid-19 and included several longer reads on topics related to my PhD. For example, [how does the new coronavirus affect pregnancy?](#) and [what do we know about Covid-19 inequalities among people from minority ethnic groups?](#)

APPENDIX - CHAPTER 1

Appendix 1.1. EJE publication: Mother's education and offspring asthma risk in 10 European cohort studies

Eur J Epidemiol (2017) 32:797–805
DOI 10.1007/s10654-017-0309-0



RESPIRATORY DISEASES

Mother's education and offspring asthma risk in 10 European cohort studies

Kate Marie Lewis¹ · Milagros Ruiz¹ · Peter Goldblatt² · Joana Morrison¹ · Daniela Porta³ · Francesco Forastiere³ · Daniel Hryhorczuk⁴ · Oleksandr Zvinchuk⁵ · Marie-Josephe Saurel-Cubizolles⁶ · Sandrine Lioret⁷ · Isabella Annesi-Maesano⁸ · Martine Vrijheid^{9,10,11} · Maties Torrent^{11,12} · Carmen Iniguez^{11,13} · Isabel Larranaga^{14,15} · Margreet W. Harskamp-van Ginkel¹⁶ · Tanja G. M. Vrijkotte¹⁶ · Jana Klanova^{17,18} · Jan Svancara¹⁸ · Henrique Barross^{19,20} · Sofia Correia^{19,20} · Marjo-Riitta Jarvelin^{21,22,23,24} · Anja Taanila²⁴ · Johnny Ludvigsson²⁵ · Tomas Faresjo²⁶ · Michael Marmot^{1,2} · Hynek Pikhart¹

Received: 31 May 2017 / Accepted: 1 September 2017 / Published online: 19 September 2017
© The Author(s) 2017. This article is an open access publication

Abstract Highly prevalent and typically beginning in childhood, asthma is a burdensome disease, yet the risk factors for this condition are not clarified. To enhance understanding, this study assessed the cohort-specific and pooled risk of maternal education on asthma in children aged 3–8 across 10 European countries. Data on 47,099 children were obtained from prospective birth cohort

studies across 10 European countries. We calculated cohort-specific prevalence difference in asthma outcomes using the relative index of inequality (RII) and slope index of inequality (SII). Results from all countries were pooled using random-effects meta-analysis procedures to obtain mean RII and SII scores at the European level. Final models were adjusted for child sex, smoking during

✉ Hynek Pikhart
h.pikhart@ucl.ac.uk

- ¹ Research Department of Epidemiology and Public Health, University College London, 1-19 Torrington Place, London WC1E 6BT, UK
- ² Research Department of Epidemiology and Public Health, UCL Institute of Health Equity, University College London, London, UK
- ³ Department of Epidemiology, Lazio Regional Health System, Rome, Italy
- ⁴ Center for Global Health, University of Illinois College of Medicine, Chicago, IL, USA
- ⁵ Institute of Pediatrics, Obstetrics, and Gynecology, National Academy of Medical Sciences of Ukraine, Kiev, Ukraine
- ⁶ Institut National de la Santé et de la Recherche Médicale (INSERM), UMR 1153, Obstetrical, Perinatal and Pediatric Epidemiology Research Team (Epopé), Center for Epidemiology and Statistics, Sorbonne Paris Cité, DHU Risks in Pregnancy, Paris Descartes University, Paris, France
- ⁷ Institut National de la Santé et de la Recherche Médicale (INSERM), UMR 1153, Early Origin of the Child's Health and Development Team (ORCHAD), Center for Epidemiology and Statistics, Sorbonne Paris Cité, Paris Descartes University, Paris, France

- ⁸ Pierre Louis Institute of Epidemiology and Public Health (iPLESP), UMR 1136, Epidemiology of Allergic and Respiratory Diseases (EPAR), Paris, France
- ⁹ ISGlobal, Center for Research in Environmental Epidemiology (CREAL), Barcelona, Spain
- ¹⁰ Universitat Pompeu Fabra (UPF), Barcelona, Spain
- ¹¹ Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Barcelona, Spain
- ¹² IB-Salut Menorca Health Area, Balearic Islands, Spain
- ¹³ Joint Research Unit of Epidemiology and Environmental Health, FISABIO – Universitat Jaume I – Universitat de València, Castellón de la Plana, Spain
- ¹⁴ Public Health Department of Gipuzkoa, Gipuzkoa, Spain
- ¹⁵ BIODONOSTIA Health Research Institute, San Sebastian, Spain
- ¹⁶ Department of Public Health, Amsterdam Public Health Research Institute, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands
- ¹⁷ Research Centre of Toxic Compounds in the Environment (RECETOX), Faculty of Science, Masaryk University, Brno, Czech Republic
- ¹⁸ Institute of Biostatistics and Analyses (IBA), Masaryk University, Brno, Czech Republic

pregnancy, parity, mother's age and ethnicity. The higher the score the greater the magnitude of relative (RII, reference 1) and absolute (SII, reference 0) inequity. The pooled RII estimate for asthma risk across all cohorts was 1.46 (95% CI 1.26, 1.71) and the pooled SII estimate was 1.90 (95% CI 0.26, 3.54). Of the countries examined, France, the United Kingdom and the Netherlands had the highest prevalence's of childhood asthma and the largest inequity in asthma risk. Smaller inverse associations were noted for all other countries except Italy, which presented contradictory scores, but with small effect sizes. Tests for heterogeneity yielded significant results for SII scores. Overall, offspring of mothers with a low level of education had an increased relative and absolute risk of asthma compared to offspring of high-educated mothers.

Keywords Asthma · Children · Cohort studies · Maternal education · Socioeconomic position · Disease risk

Introduction

Asthma is a chronic disease of the bronchial tubes in the lungs, which typically develops in childhood, and is characterised by wheezing, breathlessness, chest tightness and coughing [1–3]. Children with this condition face increased absence from school, reduced family life participation and lower quality of life compared to non-asthmatic peers [4]. Asthma is one of the most common diseases amongst children worldwide [1] and is particularly prevalent in Western Europe (9.7% asthma ever in 6–7 year olds) [5]. With the appropriate measures in place most cases of asthma can be controlled, but no proven prevention or cure

currently exist [1, 6]. With financial costs for asthma in EU countries estimated at €72.2 billion per annum, the case for improved understanding is strong [7].

The causes of childhood asthma are multifaceted, with genetic and environmental exposures increasing susceptibility to the development of this condition [5, 8]. Known early life risk factors include preterm birth, infant weight gain and adiposity, in utero and post-natal tobacco smoke exposure, household damp or mould, lower respiratory infections and pollutants such as NO_x, sulphur dioxide and particulate matter [8–12]. As socioeconomically deprived groups face increased exposure to these health damaging risk factors, grasping how social determinants impact on asthma represents a broader framework for establishing preventative measures. Prospective cohort studies in Sweden, the UK and the Netherlands have previously linked parental income, occupation and education to the increased development and severity of childhood asthma [13–17]; however, this evidence is restricted to a handful of European countries and has limited comparability due, in part, to the use of inconsistent socioeconomic position indicators [18].

This research contributes to current understanding through analysis of maternal education, arguably the most important social determinant of early child health [19], as a risk factor for early childhood asthma across Europe. Europe is highly diverse, with wide variation in social, economic and health policies and outcomes between and within countries [20, 21]. Comparing differences in asthma inequities using a common indicator across the continent will help to decipher contextual factors and enable further consideration of the pathways to asthma development. Early childhood is a crucial developmental period of life and given that asthma trajectories can be set in place as early as 6/7 years old, an optimal time for intervention [17, 18]. We hypothesised that social inequities in asthma risk would differ between European countries, but broadly follow an inverse relationship with mother's education level.

Methods

Data sources

Drawing on research by the DRIVERS for Health Equity research programme [22], the current study used data from 10 European birth cohort studies (Box 1). In total, complete data was available for 47 099 singleton children born between June 1985 and January 2008. The sample comprised of cohorts (in descending order of size) from: England and Wales (UK-MCS, $n = 13\ 829$); Southeast Sweden (SE-ABIS, $n = 8308$); Oulu and Lapland, Finland

¹⁹ EPIUnit – Institute of Public Health, University of Porto, Porto, Portugal

²⁰ Department of Clinical Epidemiology, Predictive Medicine and Public Health, University of Porto Medical School, Porto, Portugal

²¹ Department of Epidemiology and Biostatistics, MRC Health Protection Agency (HPE), Centre for Environment and Health, School of Public Health, Imperial College London, London, UK

²² Biocenter Oulu, University of Oulu, Oulu, Finland

²³ Unit of Primary Care, Oulu University Hospital, Oulu, Finland

²⁴ Faculty of Medicine, Center for Life Course Health Research, University of Oulu, Oulu, Finland

²⁵ Division of Pediatrics, Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

²⁶ Department of Medicine and Health, Community Medicine/ General Practice Faculty of Health Sciences, Linköping University, Linköping, Sweden

Box 1 Cohort study locations, names and cohort members' years of birth

CZ-CELSPEC:	Czech Republic-the European longitudinal study of pregnancy and childhood, 1991–1992
FI-NFBC:	Finland-Northern Finland birth cohort 1985/1986 study
FR-EDEN:	France-the mother-child study of pre- and postnatal determinants of child growth, development, and health, 2003–2006
IT-GASPII:	Italy-the gene and environment prospective study on infancy in Italy, 2003–2004
NL-ABCD:	The Netherlands-Amsterdam born children and their development study, 2003–2004
PT-G21:	Portugal-the generation XXI study, 2005–2006
ES-INMA:	Spain-the environment and childhood project, 1997–2008
SE-ABIS:	Sweden-all babies in Southeast Sweden, 1997–1999
UA-FCOU:	Ukraine-the family and children of Ukraine study, 1992–1996
UK-MCS:	United Kingdom-Millennium cohort study, 2000–2001

(FI-NFBC, $n = 6892$); Porto, Portugal (PT-G21, $n = 6486$); Amsterdam, the Netherlands (NL-ABCD, $n = 3387$); Brno and Znojmo, Czech Republic (CZ-CELSPEC, $n = 2801$); Gipuzkoa, Menorca, Sabadell and Valencia, Spain (ES-INMA, $n = 1843$); Kyiv, Dneprodzerzhinsk (now Kamianske) and Mariupol, Ukraine (UA-FCOU, $n = 1757$); Nancy and Poitiers, France (FR-EDEN, $n = 1159$); and Rome, Italy (IT-GASPII, $n = 637$).

Ethical approval and participant consent had been approved by each cohort study before data collection, and no further ethical approval was required for the current study. Anonymisation and extraction of the relevant data was performed by researchers from each cohort study before combining with other datasets for analyses. For further cohort details, see previous research [22].

Study variables

The definition of childhood asthma varied across the 10 datasets. SE-ABIS retrieved doctor diagnoses directly from primary care and hospital records at age 5 (ICD-10 code J45). All other studies determined asthma presence (yes/no) by mothers' answer to a survey question, which asked whether their child had: had a doctor diagnosis of asthma (FR-EDEN age 5, UA-FCOU age 7, PT-G21 age 4/5); asthma as 'a long term illness' (FI-NFBC age 7/8) or an 'allergic disease' (CZ-CELSPEC age 3); ever had asthma (NL-ABCD age 7/8, IT-GASPII age 7, UK-MCS age 5); or had asthma in the last 12-months (ES-INMA age 3/4). Each cohort study ascertained all other information relevant to the current analyses through mother's self-report prior to their child's birth, shortly after birth or when the child was 9-months old (UK-MCS only).

Maternal education level was determined with a variety of questions, including years (FI-NFBC; FR-EDEN; PT-G21; NL-ABCD) and stage (ES-INMA; IT-GASPII; SE-ABIS; UA-FCOU) of schooling completed and highest academic qualification (UK-MCS). CZ-CELSPEC and

UA-FCOU enquired about highest education level out of the parents and grandparents. Education levels across countries were converted to the same scale using the international Standard Classification of Education (ISCED 97) [23]. We banded the seven levels of the ISCED 97 into three: low, up to lower secondary education (ISCED 0-2); medium, upper secondary education (ISCED 3); and high, post-secondary education and higher (ISCED 4-6).

Several potential confounding factors were present across all cohorts and were therefore included as covariates in the final models. Maternal age at child's birth (grouped into $<30/\geq 30$ years old). Sex of the child (male/female). Parity, which we dichotomised into yes (nulliparous) or no (multiparous) by mother's report of whether the current birth was their first live-born child. Maternal ethnicity, dichotomised into prevalent ethnicity of the country/region—defined as Caucasian White (FR-EDEN; IT-GASPII), White (UK-MCS), European (CZ-CELSPEC) or by country of birth (NL-ABCD; ES-INMA)—or minority ethnicity for all women not meeting the definition. Smoking during pregnancy (yes/no) based on the following definitions: \geq one cigarette a day during pregnancy (ES-INMA); smoking throughout pregnancy (UK-MCS); smoking regularly in last 6 (CZ-CELSPEC) or 12 months (UA-FCOU); smoked in 1st or 3rd trimester (PT-G21); smoked at the beginning of pregnancy (FI-NFBC); or smoked at all during pregnancy (IT-GASPII; FR-EDEN; NL-ABCD; SE-ABIS). Non-singleton children ($n = 1399$) were excluded from further analyses.

Data analysis

All analyses were conducted in Stata 14 [24]. Missing variables data ranged from 3.2% (sex) to 27.7% (asthma) and missingness in the outcome variable was significant related to lower education level and smoking in pregnancy ($p < 0.001$). To enable comparison across cohorts, cases without complete data for all study variables were excluded

from analyses. Sampling and attrition survey weights, available for UK-MCS, and generic weights for non-MCS cohorts (probability weight = 1) were applied to all analyses [18]. The distribution of mothers' education for each cohort was age-standardised using the WHO European Standard Population [25].

Prevalence of asthma by mother's education was calculated for each cohort. The χ^2 test for trend assessed linearity across educational groups with the exception of UK-MCS, to which linear regression was applied to survey weighted data (and compared to unweighted χ^2 test results). The Relative and Slope Indices of Inequality (RII/SII) were used to compute cohort-specific and total differences in asthma outcome [22, 26, 27]. RII represents the relative change in asthma outcome between highest and lowest education group (the prevalence ratio), whilst the SII is the absolute change (the prevalence difference). A score higher than 1 (RII) and 0 (SII) indicates inequity, and the higher the score the greater the magnitude of the inequity.

RII and SII scores, adjusted for child sex, first birth, mother's age at birth, mother's ethnicity and maternal smoking status during pregnancy were determined for each cohort. We pooled the results from each cohort using meta-analysis procedures to obtain mean RII and SII scores at the European level. Assuming between-cohort heterogeneity, we applied a random effects model. The I^2 measure was used to test the extent to which heterogeneity was present. Lastly, we conducted sensitivity analysis comparing the results of fixed and random effects upon the models.

Results

Table 1 presents characteristics of the final sample. Of the 47 099 children, 47.9% were female, 50.9% of mothers were younger than 30 at their child's birth and for 46.3% this was their first live birth. UK-MCS and ES-INMA had the highest rates of smoking in pregnancy, 33.0 and 29.7% respectively, compared to 7.8% in NL-ABCD and 8.8% in SE-ABIS. The proportion of mothers from ethnic minorities ranged from 0% (FI-NFBC; SE-ABIS; UA-FCOU) to 14.5% (NL-ABCD). The gap between proportion of mothers in low and high ISCED education levels was largest in IT-GASPII (13.2 vs. 36.0%) and PT-G21 (47.2 vs. 25.4%). Asthma was identified in 3737 (7.9%) children with the prevalence ranging from 1.7% (UA-FCOU) to 14.4% (UK-MCS) across cohorts. As shown in Table 2, the unadjusted trend of increased asthma prevalence with lower maternal education level was strongest in the UK-MCS ($p < 0.001$), NL-ABCD ($p < 0.001$) and FR-EDEN ($p = 0.008$) cohorts. A weaker trend was noted for SE-ABIS ($p = 0.031$). IT-GASPII showed a reverse pattern, with asthma more prevalent in the highest education level; however, this results showed no significant trend ($p = 0.60$). There was little to no evidence of an educational trend across all other cohorts, which, in some cases, may be due to a lack of statistical power to detect an effect.

As displayed in Fig. 1a, the inverse relative risk of asthma with maternal education is strongest in cohorts FR-EDEN (RII 2.07 95% CI 1.10, 3.89), UK-MCS (RII 1.73 95% CI 1.44, 2.09) and NL-ABCD (RII 1.65 95% CI 1.00, 2.72). This pattern is reflected in absolute inequity values (Fig. 1b); Children in the FR-EDEN cohort have a 9.0% risk difference of asthma, UK-MCS a 7.3% difference and NL-ABCD a 5.9% difference. CZ-CELSPAC, FI-NFBC, ES-INMA, PT-G21, SE-ABIS and UA-FCOU have

Table 1 Characteristics of study sample, overall and cohort-specific

	Total	CZ	FI	FR	IT	NL	PT	ES	SE	UA	UK ^a
Sample size (N)	47,099	2801	6892	1159	637	3387	1843	6486	8308	1757	13 829
Asthma: yes (%)	7.9	1.9	3.2	11.7	8.2	9.1	2.7	4.3	6.7	1.7	14.4
Child's sex: female (%)	47.9	47.0	49.1	47.0	49.1	48.9	48.6	49.2	48.3	47.0	48.9
Mother's ethnicity: minority (%)	4.3	0.7	0.0	2.1	2.0	14.5	4.4	2.5	0.0	0.0	8.9
Mother's age at birth: <30 (%)	50.9	80.4	66.0	46.4	20.7	24.5	35.3	50.4	43.7	84.3	48.6
Firstborn: yes (%)	46.3	48.2	34.3	46.4	58.6	58.0	54.8	57.1	40.2	67.2	43.2
Maternal smoking: yes (%)	21.0	20.4	18.5	19.3	11.0	7.8	29.7	22.5	8.8	10.8	33.0
Mother's education											
Low (%)	24.1	35.6	25.2	4.2	13.2	8.3	29.7	47.2	33.7	5.2	12.2
Medium (%)	43.5	39.1	51.2	34.9	50.9	20.4	38.9	27.5	32.1	78.4	57.3
High (%)	32.4	25.3	23.7	60.9	36.0	71.3	31.4	25.3	34.2	16.3	30.5

^aWeighted proportions

Table 2 Cohort-specific asthma prevalence, by maternal education level (ISCED levels: low 0-2; medium 3; high 4-6)

	Total asthma		ISCED			Test for trend	
	%	(n)	Asthma %			z	p
			Low	Medium	High		
CZ-CELSPAC	1.93	(54)	2.40	1.46	1.97	0.85	0.396
FI-NFBC	3.18	(219)	3.51	3.15	2.88	1.05	0.294
FR-EDEN	11.65	(135)	10.20	15.59	9.49	2.65	0.008
IT-GASPII	8.16	(52)	8.33	7.41	9.17	-0.53	0.597
NL-ABCD	9.12	(309)	13.83	11.87	7.79	4.35	<0.001
PT-G2I	4.33	(281)	4.54	4.88	3.35	1.52	0.128
ES-INMA	6.67	(554)	3.10	2.09	2.94	0.14	0.890
SE-ABIS	2.66	(49)	7.32	6.82	5.88	2.16	0.031
UA-FCOU	1.71	(30)	2.17	1.76	1.40	0.54	0.591
UK-MCS ^a	15.24	(2108)	19.95	15.17	10.59	8.78*	<0.001

^aWeighted proportions and analysis (unweighted n)

*T-value shown. Test for trend similar when replicated using unweighted values

positive RII and SII values, but with smaller statistically insignificant effect sizes. IT-GASPII displays a contrary pattern; low education level is associated with 34% lower odds (RII = 0.66 95% CI 0.24, 1.83) and 4.68% decreased risk difference (SII = -4.68 95% CI -12.92, 3.56) of asthma compared to high education. However, these relationships are not statistically significant.

Across all European cohorts, children born to mothers in the lowest tertile of education were 46% more likely to develop asthma than children born to mothers in the highest tertile (95% CI 1.26, 1.71). The pooled estimate of SII was 1.90 (95% CI 0.26, 3.54), indicating a 1.9% mean risk difference of asthma between children born to mothers in the low compared to high educational group. An I^2 value of 9.2% ($p = 0.36$) in the RII model gives evidence of low heterogeneity in relative scores between cohorts; however, absolute scores (SII) were moderately to highly heterogeneous ($I^2 = 72.0%$, $p < 0.001$) [28]. Fixed effects meta-analysis techniques slightly increased relative estimates (RII = 1.50 95% CI 1.32, 1.71), whilst decreasing absolute estimates (SII = 1.47, 95% CI 0.70, 2.34).

Discussion

This study uniquely combined prospective data of educational inequalities in childhood asthma from 10 European cohort studies. Overall, we found relative and absolute asthma risk inversely related to maternal education, which is broadly consistent with previous research [8, 14, 29–32]. However, the magnitude of difference varied across countries and substantial heterogeneity was present in

absolute differences. The highest prevalence of asthma and most pronounced inequity were observed in Western European countries.

France, the UK and the Netherlands cohorts had asthma rates between 9 and 14%, compared to less than 2% of the Czech Republic and Ukraine cohorts. East-to-West geographic variations have been noted previously, although are now reportedly diminishing [1]. Whilst climatic risk factors such as low altitude, temperature variations and relative humidity may explain some of the difference, it is argued that the urban 'western lifestyle' is a major contributor—and the adoption of this by Eastern Europe the reason for the closing gap [33, 34]. Western lifestyle factors include increased exposure to vehicle emissions and a diet high in calories, fat and sugar, but low in fruit and vegetables.

Maternal education affects child health through several mechanisms. Lower levels of education are linked to lower employment positions, fewer material resources and social support, as well as limited access to health information (and a diminished ability to act on this information) [19, 35, 36]. Prominently, this leads to greater exposure to health damaging risk factors for the mother and consequently their child. Although our analyses were adjusted for several potential risk factors, data constraints meant that many other risk factors were not included. It is likely that a large proportion of inequity may be explained by differences in exposure to risk factors such as tobacco smoke after birth, infant weight gain and adiposity, household damp, infections, and pollution [8–12]. Family history of asthma and atopy, unmeasured in this study, is one of the strongest risk factors for childhood asthma [8]. However, this does not fully explain why the impact of

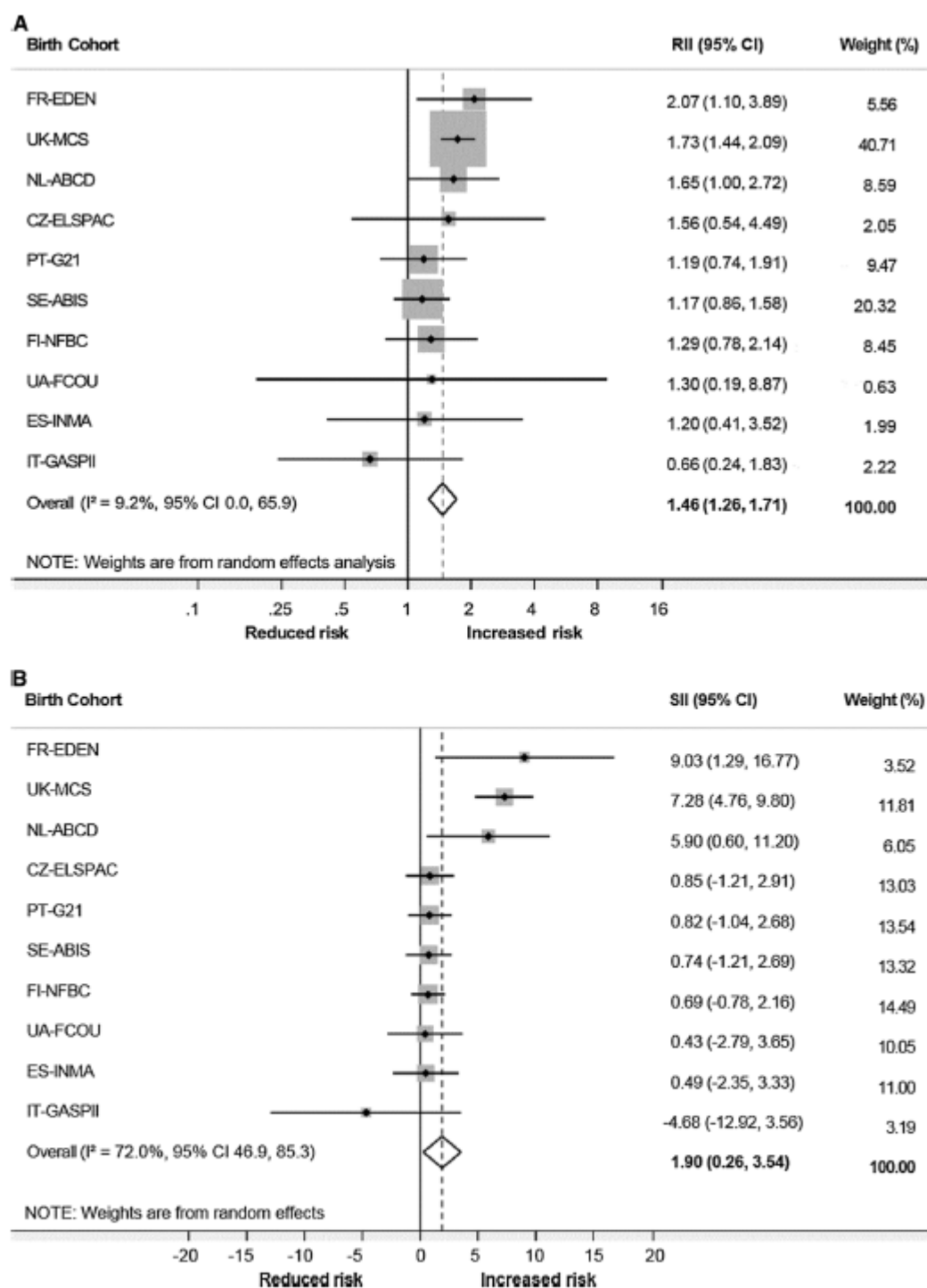


Fig. 1 Forest plots of the RII (a) and SII (b) in childhood asthma across 10 European cohorts (squares) and combined (diamond), adjusted for child sex, smoking during pregnancy, parity, mother’s age and ethnicity; values to the right of the *undashed vertical line* indicate greater inequality for low compared to high maternal education groups

maternal education is greater in some countries compared to others.

Explaining difference in inequity across countries requires us to look at the structural facets that constrain and disadvantage people with lower education [37]. It is proposed that the intergenerational impact of education is additionally conflated by the broader socio-political and cultural landscape within countries [38]. The results from Western Europe suggest that whilst access to health care and social security is important, it is not enough to counter inequities [38]. Universal provision in healthcare, in particular, does not mean universal uptake [29]. For example, disadvantaged groups in the UK are less likely to avail of health promotion and prevention services [39]. Anxo et al.'s [40] review found that the impact of women's educational attainment on employment integration is more pronounced in France, Italy and Spain compared to Sweden [40]. Higher employment rates and macrosocial policies that value women and families in the Nordic countries may partially explain lower levels of inequity compared to Western Europe [41, 42]. However, high rates of adiposity amongst children of parents with low education persist in Sweden and Finland despite egalitarian policies [37, 43]. The Mediterranean diet may protect against the impact of maternal education in Southern Europe [44].

Birth cohort studies are an essential and unique source of information on the multifaceted predictors of chronic disease [45]. Combining 10 European studies has enabled a comparison of more than 47,000 children over several years of life. The prospective design addresses issues of reverse causation inherent to cross-sectional studies. In addition, as the results were not reliant on published data, this study avoided publication bias. However, several limitations must be acknowledged when considering the current study. Firstly, rates of asthma and the inequities shown are not necessarily indicative of the country or area in which each cohort study were based. Many cohorts sampled only one major urban area and we may expect regional differences in rates of education in rural areas, for instance.

Combining multiple datasets brings difficulties and, as highlighted by the heterogeneous absolute rates of inequity, there is unexplained difference across the included cohorts. Unmeasured confounding, such as the risk factors mentioned above, is likely to contribute to this. In addition, several study aspects were not consistent across cohorts. Age at which asthma was defined ranged from 3 to 8 years old, and the timespan in which to consider a diagnosis from the last 12 months to at any point across the child's lifetime. Symptomatology of asthma, as well as the impact of inequity on the disease outcome, can differ across age in childhood [1, 14, 46]. Birth year ranged from 1985 to 2008 across the cohorts, adding time-varying contextual factors

to the findings. For example, the Spanish cohort study recruited children over a decade, encompassing a period of economic growth and better perinatal outcomes than in the period of economic recession that followed [19].

The proportion of mothers with low education was inconsistent across studies. Both low education and young mothers appear underrepresented in the Italian sample, which may imply limited study power and explain the contradictory results shown by this cohort. Previous cross-sectional research in Rome found a significant trend between lower paternal education and increased risk of asthma, and a similar but insignificant trend of maternal education [47]. The proportion of mothers with low education was also low in France, Ukraine and the Netherlands; however, these are similar to national averages for the time [22]. Highest education level in the Ukraine and Czech Republic cohorts were based on several family members, which likely conflated the education level of some mothers. Almost all cohorts relied on maternal reports of asthma rather than doctor's diagnosis, which likely led to imprecise estimations of the condition. Cesaroni et al. [47] speculate that parents from higher socioeconomic positions are more likely to report asthma especially where symptoms are mild. Missing data for the asthma outcome was significantly related to lower maternal education across the dataset, further indicating a socially patterned bias to the results.

Despite limitations, this comparative analysis across European cohorts demonstrates that maternal education is implicated in offspring asthma risk. The difference in magnitude of inequity across countries, suggests contextual factors affect the translation between maternal education and offspring health outcomes and, importantly, that reduction in inequity is possible. Given that early disadvantage in respiratory health puts into place a negative trajectory that can be difficult to modify [17], these results reinforce the need for early, if not pre-birth, interventions to offset the burden of this disease [18, 38]. Universal and high-quality perinatal care including information about asthma risk that is clearly and respectively articulated is a necessity [48, 49]. Both improving the educational attainment of young women and combatting systematic barriers faced by women with lower education, such as access to employment, is essential to ensure that these offspring are not further disadvantaged [40, 42].

Future cross-European analyses could be improved with the inclusion of known environmental risk factors for asthma and by the use of standardised measures (such as the ISAAC for asthma) to increase comparability and enable pathway analyses [5, 17]. Exclusion of genetic predispositions using genetic methods, such as twin and family studies, will clarify environmental causes with more precision. Differentiating between asthma phenotypes and

additional indicators of socioeconomic position, such as income, will help further untangle cross-country differences [29, 50].

Acknowledgements All phases of this study were supported by a European Union's Seventh Framework Programme grant, 278350, as part of The Determinants to Reduce Health Inequity Via Early Childhood, Realising Fair Employment, and Social Protection (DRIVERS) research programme. The Czech ELSPEC Study (CZ-ELSPAC) was supported by the Ministry of Education of the Czech Republic: CETOCOEN plus project (CZ.02101/00/00/15_003/0000469) and RECETOX Research Infrastructure (LM2015051). The Northern Finland Birth Cohort (FI-NFBC8586) received financial support from the Academy of Finland; Biocenter, University of Oulu, Finland; the European Commission EUROBLCS, Framework 5 Award QLGI-CT-2000-01643; EU FP7 EurHEALTHAgeing-277849; the Medical Research Council, UK (PrevMetSyn/SALVE); and the MRC Centenary Early Career Award. The Amsterdam Born Children and their Development Study (NLABCD) received funding from the Netherlands Organization for Health Research and Development (ZonMw) Grant (TOP, 40-00812-98-11010). The All Babies in Southeast Sweden Study (SE-ABIS) has received financial support from the Juvenile Diabetes Research Foundation, Swedish Child Diabetes Foundation (Bamdiabetesfonden), The Research Council of South-east Sweden (FORSS), Swedish Research Council K2005-72X-11242-11A, and ALF/County Council of Östergötland. The INMA study was funded in part by grants from the European Union (FP7-ENV-2011-282957 and HEALTH.2010.2.4.5-1), Spain (Instituto de Salud Carlos III and The Ministry of Health), the Conselleria de Sanitat of the Generalitat Valenciana, Department of Health of the Basque Government, the Provincial Government of Gipuzkoa, and the Generalitat de Catalunya-CIRIT. Family and Children of Ukraine (UA-FCOU) study was supported by US NIH Fogarty International Center and National Academy of Medical Sciences of Ukraine. KML is funded by a Medical Research Council UK doctoral training studentship.

Compliance with ethical standards

Conflict of interests The authors declare that they have no conflict of interests.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Gibson GJ, Loddenkemper R, Sibille Y, Lundbäck B. Childhood asthma. European Lung White Book. Sheffield: European Respiratory Society; 2013.
- de Nijs SB, Venekamp LN, Bel EH. Adult-onset asthma: is it really different? *Eur Respir Rev Off J Eur Respir Soc.* 2013;22(127):44–52.
- Asthma (Fact sheet No. 307) Geneva: World Health Organisation. 2013. <http://www.who.int/mediacentre/factsheets/fs307/en/>.
- Braman SS. The global burden of asthma. *Chest.* 2006;130(1 Suppl):4S–12S.
- Lai CK, Beasley R, Crane J, Foliaki S, Shah J, Weiland S, et al. Global variation in the prevalence and severity of asthma symptoms: phase three of the International Study of Asthma and Allergies in Childhood (ISAAC). *Thorax.* 2009;64(6):476–83.
- Unknown. An outcomes strategy for chronic obstructive pulmonary disease (COPD) and asthma in England: Department of Health. 2011. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216139/dh_128428.pdf.
- Gibson GJ, Loddenkemper R, Sibille Y, Lundbäck B. The economic burden of lung disease. European Lung White Book. Sheffield: European Respiratory Society; 2013.
- Strina A, Barreto ML, Cooper PJ, Rodrigues LC. Risk factors for non-atopic asthma/wheeze in children and adolescents: a systematic review. *Emerg Themes Epidemiol.* 2014;11(1):5.
- Galobardes B, Granell R, Sterne J, Hughes R, Mejia-Lancheros C, Davey Smith G, et al. Childhood wheezing, asthma, allergy, atopy, and lung function: different socioeconomic patterns for different phenotypes. *Am J Epidemiol.* 2015;182:763–74.
- Civelek E, Cakir B, Orhan F, Yuksel H, Boz AB, Uner A, et al. Risk factors for current wheezing and its phenotypes among elementary school children. *Pediatr Pulmonol.* 2011;46:166–74.
- McCormack MC, Breyse PN, Matsui EC, Hansel NN, Peng RD, Curtin-Brosnan J, et al. Indoor particulate matter increases asthma morbidity in children with non-atopic and atopic asthma. *Ann Allergy Asthma Immunol.* 2011;106(4):308–15.
- Sonnenschein-van der Voort AM, Arends LR, de Jongste JC, Annesi-Maesano I, Arshad SH, Barros H, et al. Preterm birth, infant weight gain, and childhood asthma risk: a meta-analysis of 147,000 European children. *J Allergy Clin Immunol.* 2014;133(5):1317–29.
- Ruijsbroek A, Wijga AH, Kerkhof M, Koppelman GH, Smit HA, Droomers M. The development of socio-economic health differences in childhood: results of the Dutch longitudinal PIAMA birth cohort. *BMC Public Health.* 2011;11:225.
- Gong T, Lundholm C, Rejino G, Mood C, Langstrom N, Almqvist C. Parental socioeconomic status, childhood asthma and medication use—a population-based study. *PLoS ONE.* 2014;9(9):e106579.
- Almqvist C, Pershagen G, Wickman M. Low socioeconomic status as a risk factor for asthma, rhinitis and sensitization at 4 years in a birth cohort. *Clin Exp Allergy.* 2005;35(5):612–8.
- Wickman M, Melén E, Berglind N, Lennart Nordvall S, Almqvist C, Kull I, et al. Strategies for preventing wheezing and asthma in small children. *Allergy.* 2003;58(8):742–7.
- Panico L, Stuart B, Bartley M, Kelly Y. Asthma trajectories in early childhood: identifying modifiable factors. *PLoS ONE.* 2014;9(11):e111922.
- Pillas D, Marmot M, Naicker K, Goldblatt P, Morrison J, Pikhart H. Social inequalities in early childhood health and development: a European-wide systematic review. *Pediatr Res.* 2014;76(5):418–24.
- Juarez S, Revuelta-Eugercios BA, Ramiro-Farinas D, Viciano-Fernandez F. Maternal education and perinatal outcomes among Spanish women residing in southern Spain (2001–2011). *Matern Child Health J.* 2014;18(8):1814–22.
- Jakab Z, Marmot M. Social determinants of health in Europe. *The Lancet.* 2012;379(9811):103–5.
- Mackenbach JP, McKee M. A comparative analysis of health policy performance in 43 European countries. *Eur J Pub Health.* 2013;23(2):195–201.
- Ruiz M, Goldblatt P, Morrison J, Kukla L, Svancara J, Riitta-Jarvelin M, et al. Mother's education and the risk of preterm and small for gestational age birth: a DRIVERS meta-analysis of 12 European cohorts. *J Epidemiol Community Health.* 2015;69(9):826–33.

23. Organization IL. International Standard Classification of Education (ISCED-97). LABORSTA 1996-2010.
24. StataCorp. Stata statistical software: release 14. College Station, TX: StataCorp LP; 2015.
25. European health for all database (HFA-DB). World Health Organisation Regional Office for Europe 2016.
26. Regidor E. Measures of health inequalities: part I. *J Epidemiol Community Health*. 2004;58(10):858–61.
27. Mackenbach JP, Kunst AE. Measuring the magnitude of socioeconomic inequalities in health: an overview of available measures illustrated with two examples from Europe. *Soc Sci Med*. 1997;44(6):757–71.
28. Higgins J, Thompson S, Deeks J, Altman D. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557–60.
29. Basagana X, Sunyer J, Kogevinas M, Zock JP, Duran-Tauleria E, Jarvis D, et al. Socioeconomic status and asthma prevalence in young adults: the European Community Respiratory Health Survey. *Am J Epidemiol*. 2004;160(2):178–88.
30. Cruz AA, Bateman ED, Bousquet J. The social determinants of asthma. *Eur Respir J*. 2010;35(2):239–42.
31. Kozyrskyj AL, Kendall GE, Jacoby P, Sly PD, Zubrick SR. Association between socioeconomic status and the development of asthma: analyses of income trajectories. *Am J Public Health*. 2010;100(3):540–6.
32. Spencer N. Maternal education, lone parenthood, material hardship, maternal smoking, and longstanding respiratory problems in childhood: testing a hierarchical conceptual framework. *J Epidemiol Community Health*. 2005;59(10):842–6.
33. Beasley R, Semprini A, Mitchell EA. Risk factors for asthma: is prevention possible? *The Lancet*. 2015;386(9998):1075–85.
34. Papoutsakis C, Chondronikola M, Antonogeorgos G, Papadaku E, Matziou V, Drakouli M, et al. Associations between central obesity and asthma in children and adolescents: a case-control study. *J Asthma Off J Assoc Care Asthma*. 2015;52(2):128–34.
35. Prickett KC, Augustine JM. Maternal education and investments in children's health. *J Marriage Fam*. 2016;78(1):7–25.
36. Mangrio E, Hansen K, Lindstrom M, Kohler M, Rosvall M. Maternal educational level, parental preventive behavior, risk behavior, social support and medical care consumption in 8-month-old children in Malmo, Sweden. *BMC Public Health*. 2011;11:891.
37. Mackenbach JP, Stirbu I, Roskam A-JR, Schaap MM, Menvielle G, Leinsalu M, et al. Socioeconomic inequalities in health in 22 European countries. *N Engl J Med*. 2008;358(23):2468–81.
38. Marmot M, Allen J, Goldblatt P, Boyce T, McNeish D, Grady M, Geddes I. Fair society, health lives (The Marmot Review). Great Britain: department of Health; 2010.
39. Goddard M, Smith P. Equity of access to health care services: theory and evidence from the UK. *Soc Sci Med*. 2001;53(9):1149–62.
40. Anxo D, Fagan C, Cebrian I, Moreno G. Patterns of labour market integration in Europe—a life course perspective on time policies. *Socio-Economic Review*. 2007;5(2):233–60.
41. Putnam S, Galea S. Epidemiology and the macrosocial determinants of health. *J Public Health Policy*. 2008;29(3):275–89.
42. Borrell C, Palencia L, Muntaner C, Urquia M, Malmusi D, O'Campo P. Influence of macrosocial policies on women's health and gender inequalities in health. *Epidemiol Rev*. 2014;36:31–48.
43. Magnusson M, Sørensen TIA, Olafsdottir S, Lehtinen-Jacks S, Holmen TL, Heitmann BL, et al. Social inequalities in obesity persist in the nordic region despite its relative affluence and equity. *Current Obesity Reports*. 2014;3(1):1–15.
44. Lang JE. Obesity, nutrition, and asthma in children. *Pediatric allergy, immunology, and pulmonology*. 2012;25(2):64–75.
45. Bousquet J, Anto J, Sunyer J, Nieuwenhuisen M, Vrijheid M, Keil T, et al. Pooling birth cohorts in allergy and asthma: european Union-funded initiatives—a MeDALL, CHICOS, ENRIECO, and GA(2)LEN joint paper. *Int Arch Allergy Immunol*. 2013;161(1):1–10.
46. Chen E, Martin AD, Matthews KA. Socioeconomic status and health: do gradients differ within childhood and adolescence? *Soc Sci Med*. 2006;62(9):2161–70.
47. Cesaroni G, Farchi S, Davoli M, Forastiere F, Perucci CA. Individual and area-based indicators of socioeconomic status and childhood asthma. *Eur Respir J*. 2003;22(4):619–24.
48. Beasley R. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. *The Lancet*. 1998;351(9111):1225–32.
49. Renfrew MJ, McFadden A, Bastos MH, Campbell J, Channon AA, Cheung NF, et al. Midwifery and quality care: findings from a new evidence-informed framework for maternal and newborn care. *The Lancet*. 2014;384(9948):1129–45.
50. Civelek E, Cakir B, Orhan F, Yuksele H, Boz AB, Uner A, et al. Risk factors for current wheezing and its phenotypes among elementary school children. *Pediatr Pulmonol*. 2011;46(2):166–74.

APPENDIX - CHAPTER 3

Appendix 3.1. Linking Mother and Infant Data in HES APC

The Hospital Episode Statistics Admitted Patient Care (HES) dataset contains routinely collected maternity records from NHS hospitals across England.⁴⁵ There are two types of HES maternity records available, the (maternal) delivery record and the (baby) birth record.²⁷³ Maternal and baby records in HES do not share a unique identifier, however, and are not linked together in the dataset we receive from NHS Digital. Linking delivery and birth records leads to a more complete medical record for the mother and child, improving reliability and statistical power of secondary research using this data. To connect mother and baby records for HES years 2013/14 to 2016/17, I used deterministic and probabilistic linkage methods as outlined by Harron et al.²⁷³

Deterministic linkage is relatively straightforward. It involves applying a set of rules to classify pairs, in this case mothers and infants, as links or non-links.⁴⁸⁵ For this dataset, GP practice, infant sex, birthweight, gestational age, mother's age and birth order were used to deterministically match pairs (see Table A3.1 for more details). Improvement in hospital recording of data means that rates of deterministic linkage between mothers and babies have grown over the years;⁴⁵ however, in 2016/17 this type of linkage still remained below 50%. For the remaining mother and infant records, probabilistic linkage was applied, where the likelihood of records belonging together is represented by a match weight. This method is suitable for matching records that contain several non-unique identifiers, which may, in combination, uniquely identify an individual.²⁷³ These variables are referred to as 'quasi-identifiers'⁴⁸⁶ and are listed for the current dataset in Table A3.1.

Steps for probabilistic linkage

Two probabilities are needed to calculate match weights:

- U-probabilities - the likelihood that (non-missing) variables across two records randomly match.
- M-probabilities are the probability there is an agreement on a (non-missing) variable given that two records are a true match.

I calculated U-probabilities by selecting a random sample of 5000 unlinked mother and baby records within each HES year, and running pairwise comparisons of each variables across all these records (i.e. 25,000,000 comparisons).²⁷³ M-probabilities were calculated using the

deterministically linked records (as outlined above). Weights were then calculated for each variable across record pairs using the formula, $\log_2(M\text{-probability}/U\text{-probability})$, and summed to get an overall match weight for each pair.

Match weights across record pairs were reviewed using plots and manual review to select a single threshold. All pairs above this values were deemed to be true matches and these were placed into the deterministically linked dataset. The process of calculating M-probabilities and match weights was repeated on the new dataset. I carried out three repetitions of these steps in total to ensure match weights had stabilised. The final iteration involved applying the now stabilised M-probabilities to all non-deterministically linked pairs to get the final match weights. Pairs below the final threshold were placed into a separate dataset. Pairs above the threshold were ordered by descending match weight and the highest-weighted mother for each baby was selected, followed by the highest weighted baby for each delivery record (mothers may have more than one singleton baby in a year).

Table A3.1. Details of variables used in linkage

Variable	HES name	Description and linkage type (D = Deterministic and P = probabilistic)	
GP practice	gpprac	The code of the general practice where the patient's registered GP works	D & P
Provider code	procode3	The first 3 characters of the organisation acting as the health care provider	D
Infant sex	sexbaby	Sex of the baby	D & P
Birth order	birordr	Position in sequence of birth (i.e must be 1 for singleton births)	D & P
Birth weight	birweit	Weight of the baby in grams immediately after birth	D & P
Gestational age	gestat	Length of gestation in weeks	D & P
Mother's age	matage	Mother's age at delivery	D & P
Delivery place	delplac	Delivery place [2013/14 only]	P
Postcode district	postdist	First 3/4 characters of postcode [2014/15 onwards]	P
Postcode area	postdist	First character of postcode	P
Gestation period	anagest	Gestation period in weeks at first antenatal assessment	P
Delivery place (intended)	delinten	The intended place of delivery	P
Delivery method	delmeth	defines the method used to deliver a baby that is a registrable birth	P

Delivery method (onset)	delonset	Defines the method used to induce (initiate) labour	P
Anaesthetic (post)	delposan	Anaesthetic given post labour or delivery	P
Anaesthetic (during)	delprean	Anaesthetic given during labour or delivery	P
Delivery status	delstat	Status of person conducting the delivery (e.g. hospital doctor, midwife)	P
Resuscitation method	biresus	Method to get the baby breathing	P
Birth status	birstat	Live or still birth	P
Ethnicity	ethnos	Ethnicity of the patient	P

Appendix 3.2. Data linkage feedback form

This form was sent to the Great Ormond Street hospital parent carer advisory group and advertised through Asthma UKs monthly bulletin for parents and carers.

10th November 2017

Using data to understand risk factors for asthma in children

Who am I and what is my research?



I am a research (PhD) student from UCL Great Ormond Street Institute of Child Health.

My research project is focussed on understanding the early risk factors leading to different types of childhood asthma.



To do my research I will use electronic health data from hospitals in England and Scotland.

Using data on the whole childhood population in these countries means that I get a more complete picture of hospital admissions for asthma and the early risk factors that lead to this. For example, early respiratory infection and premature birth.

What do I want to do?

To identify more children with asthma, we want to link hospital admission data to drug dispensing data from pharmacies. I am interested in asthma medication in particular. This will improve my research by including more children with different types and severities of asthma.

FIGURE 1, below, shows the linked data required for my research project. This type of linkage is only available for children in Scotland at the moment.

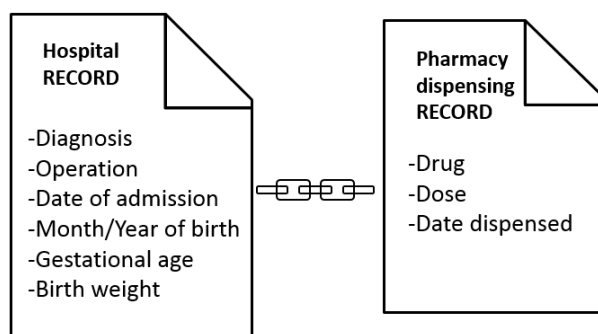


FIGURE 1: Proposed data linkage

3.0 How can you help?

I would like you to consider this research project and the data linkage proposed above. The process of secure data linkage is explained in **FIGURE 2**, on the next page. Then I would like your feedback, by answering the question below.

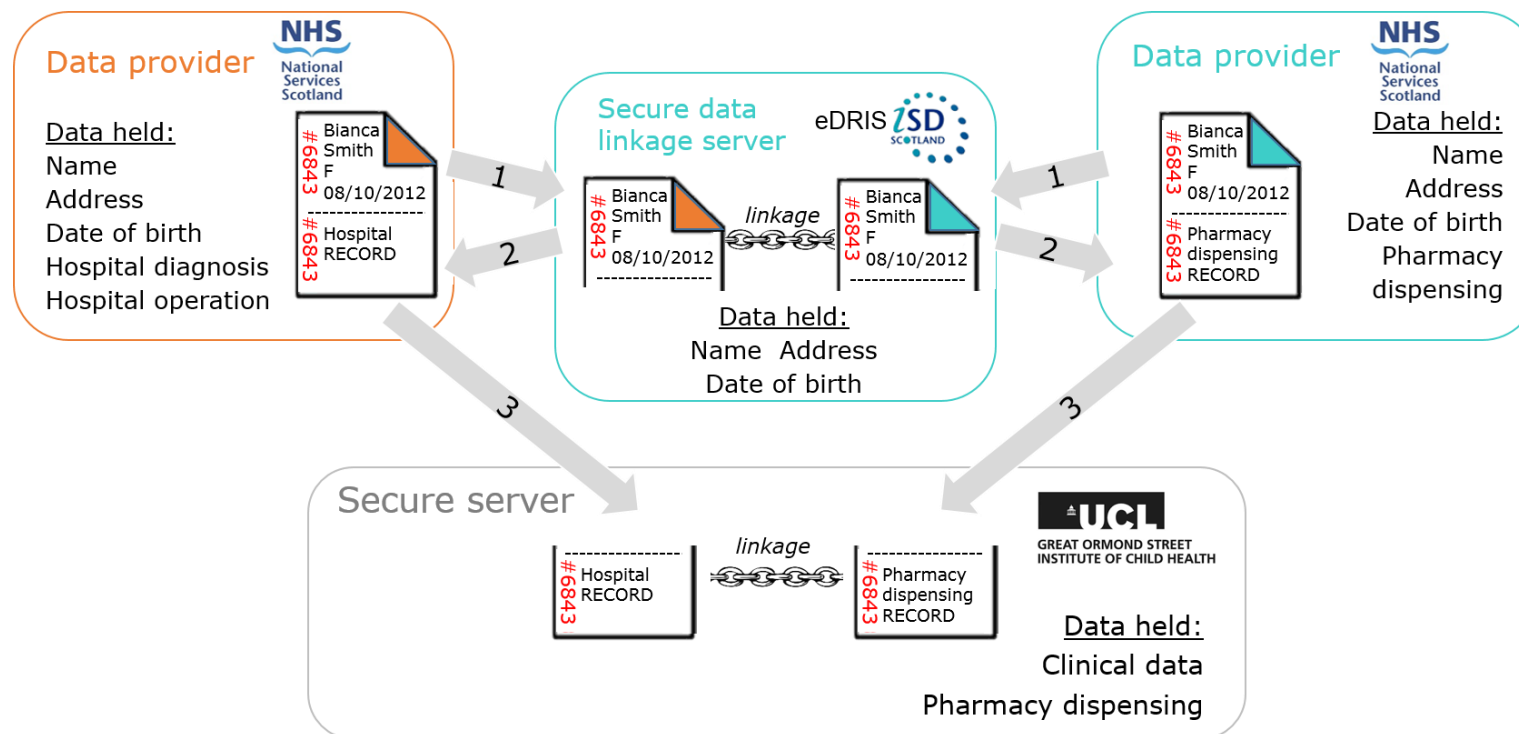
Question for you: Do you have any thoughts or concerns on combining these sources of data for this project and if so why?

Any other questions for me?

If you would like to find out more about this research, email [REDACTED]

Thank you for taking the time to answer these questions , your help is invaluable to my work.

FIGURE 2: Process of secure data linkage



Steps to data linkage:

1. Personal information sent to linkage server
2. Personal information linked and returned
3. Linked data, not personal information, sent to the researcher (me!)

What about consent?

We do not receive information that can directly identify a patient, like names, postcode or NHS number. We require special permissions from the NHS and the Office for National Statistics to allow these data to be linked. We can only use the data to answer specific research questions. We are using the data without specific consent.

Appendix 3.3. Feedback request sent to the Farr Institute Public Panel

The following email was sent to members of the Farr Institute Public Panel alongside an earlier draft of the patient opt out commentary article presented in Appendix C. We received 10 responses from the panel, which helped shape the points within the article.

4th July 2018

I am a PhD student looking into the impact of deprivation on childhood respiratory illness. To do this work I use hospital data from babies (linked to their mother's data) across England. This provides me with information on various risk factors for respiratory illness including preterm birth, deprivation, area of the country and early childhood illness. The linked maternal data is very important as it enables me to look at a wider range of risk factors than baby data alone would provide, such as rates of pollution in the local area of residence and delivery type. It is very important to have complete whole population data in this work so that I can understand the full picture and look at differences across the country. However, in the newest years of hospital data there are more and more instances where there is no maternal information available at all, which means that critical information (such as pollution and deprivation) are not available for some babies.

After speaking with colleagues, we realised that the increase in missing data is largely due to patients choosing to 'opt-out' of sharing their data for purposes other than clinical care (such as for research). Research on this topic shows that there is little information provided for the public to understand how and why their data may be used, and the safeguards in place to protect their information. This finding led me to write an article to explain these problems, with specific reference to the implications for maternal statistics. The British Medical Journal have expressed interest in this topic; however, before I submit the article, I would like to get different perspectives on what I have written.

How can you help?

I would like you to read the attached draft of the article and let me know what you think of it. Specifically if you could answer the following questions:

- Do you think the article fairly represents general public perception?
- Is there anything missing? Or not covered in enough detail?
- Do you have any other comments?

Appendix 3.4. IJPDS publication: National data opt out programme

International Journal of Population Data Science

Journal Website: www.ijpds.org



National data opt out programme: consequences for maternity statistics in England

Lewis, KM^{1*} and Hardelid, P¹

Submission History

Submitted:	14/05/2019
Accepted:	09/08/2019
Published:	30/01/2020

¹UCL Great Ormond Street Institute of Child Health, London, UK

Abstract

Electronic health records offer great potential for individual care, service improvement and, when collated, the health of the wider population. Datasets composed of these types of records have been invaluable to our understanding of risk factors for maternal and infant ill-health. However, a potential barrier to data quality in England is emerging where patients choose to opt out of sharing their information beyond the NHS. Focussing on maternity statistics, we will present the importance of population level health data for monitoring NHS services, and the potential consequences for patients of opting out. Evidencing the success of similar systems in Nordic countries, we argue that the English population must be better informed of the implications of opting out of sharing NHS data for research and the safeguards in place to protect patient information.

Background

Patient records across healthcare providers in England are being digitalised—a process whereby clinical data are stored in digital form and shared with authorised users—to the benefit of patients [1]. When compared to handwritten notes, computerised records improve the detail, completeness and reliability of patient data [1]. Electronic patient record systems improve communication between health professionals within and between different providers [2]. Patients who access their health records online report improved self-care, greater satisfaction with the communication from their doctor and, in some instances, improved safety through patient identification of medication errors [2]. The benefits of electronic health records are not limited to the care of individual patients. When collated, these data are a cost effective way to advance the health of the population through improved knowledge of healthcare services and the aetiologies and treatments of health conditions [3, 4]. Electronic health records in the UK have been successfully used, for example, to investigate: neonatal impact of antibiotic prescription during pregnancy; success of different interpregnancy intervals on pregnancy following miscarriage; and pregnancy complications after caesarean section at first birth [5–7].

However, just as the use of electronic health for research and planning becomes more common place [8], new barriers to the quality of the data are also surfacing. In particular, some patients are choosing not to share their information beyond the NHS for anything other than their direct care. Opting out (see Piel et al. [9] for details about the different types of opt outs) was first made available for patients of the English NHS in January 2014 in response to a recommendation

by Dame Fiona Caldicott in her 2013 information governance review [10]. This review was published amid severe concerns, voiced by both the general public and experts in the medical field, about the security of patient data following the release of individual-level health records to profit-making companies [11, 12]. A new consent model intended to be simpler and easier to access (through an online platform) launched in May 2018; however, initial figures suggest that very few patients know about this scheme [13]. As stated in the 2016 Caldicott review [14], ‘patients have a right under the NHS Constitution to request that their personal confidential information is not used beyond their direct care’. We argue that, in parallel with information about this choice, it is imperative that patients have complete and transparent information about the uses and potential advantages of sharing their information.

NHS data sharing and safeguards

The data capture organisation within the English NHS, NHS Digital, collects and stores some of the information recorded when individuals receive health or social care in England. This includes records of diagnoses and operations recorded during hospital admissions. The data are used for a variety of reasons including planning NHS services and monitoring patient safety [15]. Strictly controlled release of some patient information may be shared with NHS providers and commissioners, university researchers, charities and companies that are partnered with the NHS. Where permission is granted, all organisations must follow stringent protocols when storing and analysing the data. Personal identifiers, such as names and NHS numbers are removed in all circumstances apart from where specific

*Corresponding Author:

Email Address: kate.lewis.14@ucl.ac.uk (KM Lewis)

patient consent is given or where required by the law. NHS Digital states that 'we make sure data is only used for the good of health and care', and all organisations go through a lengthy application process to ensure this is the case.

NHS maternity statistics

The most comprehensive source of information on all births and deliveries in the NHS in England is Hospital Episode Statistics (HES); a dataset that includes all admissions to NHS or NHS-funded hospitals in England [3]. Delivery information, such as the place of delivery, baby's sex, birthweight, gestational age and method of delivery, is used for many purposes, including to create resources for parents-to-be, to evaluate and improve maternity care provision and to investigate multiple risk factors for ill-health in mothers and babies [16]. Tools have been created utilising this data to aid expectant parents when making maternity choices. One such resource, the Birth Choices tool from Which? [17] recommends considering essential maternity statistics when planning which hospital to give birth in, such as variation in caesarean sections, induction and other medical interventions. Further, the National Maternity and Perinatal Audit (NMPA) was set up in 2016 to evaluate quality in NHS maternity services [18]. Using maternity data from HES linked with data from each maternity unit, the NMPA provides a range of statistics comparing outcomes at maternity unit level, including induction of labour and caesarean section rates. In addition, HES has been used for maternal and child health research to examine, for example, factors explaining excess child mortality in England and the safety of surgical procedures during pregnancy [19, 20]. Individual level data are necessary for this type of research, since this allows multiple risk factors for maternal and child outcomes to be taken into account.

Opting out and the effect on the quality of NHS maternity data

As at 1 December 2018, the average national data opt out rate across England was 2.8% [21]. Top-level demographic information published by NHS Digital shows that rates of opt outs are higher in older people and females [21]. This is in keeping with findings from surveys and qualitative research that the characteristics of people who are less willing to share their health data with researchers differ from those who are willing to consent (although these characteristics have not been consistent across studies) [22, 23]. For example, a study of mothers in the UK Millennium Cohort Study found that the proportion consenting to link survey data with their child's NHS records differed by country of residence, age, ethnicity, lone parenthood status and education [24]. As shown in Figure 1, opt out rates in the general population are also not uniformly distributed across geographical area in England. Twelve Clinical Commissioning Groups (CCGs) have opt out rates higher than 5% and, strikingly, one CCG has a rate of 10.1% [21]. At the GP practice level, there are instances where the entire patient population have opted out [21]. As remarked by Piel et al. [9], this raises significant questions about whether the patients in these practices explicitly opted out for themselves.

To exemplify the impact of biased data on maternal information, we simulated outcome rates for two health indicators had opt outs not been applied to the data. We downloaded publicly available CCG-level information from Public Health England's Fingertips Child and Maternal Health Profiles for one common and one rare outcome, the proportion of deliveries with caesarean sections (occurring in approximately 27% of births nationwide) and births with very low birth weight (<1500g, 1.2% nationwide) [25]. We chose three CCGs for our example, each with a different rate of opt out, as published at 1st December 2018 by NHS Digital [26]. Under the assumption that the women in the maternity dataset opted out at the same rate as the whole CCG population, we modelled three scenarios based on the rate of events in the women that opted out: 1) no events; 2) events occurring at the same rate as women who did not opt out; 3) all had an event. Mirroring methods used by Public Health England [27], 95% confidence intervals (CIs) were calculated in Stata 15.0 [28] using the Wilson Score method. Microsoft Excel 2013 was used to create graphs.

Table 1 displays published and simulated event rates for caesarean sections and infants born with very low birth weight for Oldham (10.1% opt outs), Merton (2.8% opt outs) and Bradford City (0.3% opt outs) CCGs. In scenario 2, where patients opting out have the same rate of events as patients not opting out, published and simulated event rates do not differ. Figure 2 and 3 show the possible range of rates between the extremes of scenario 1 (no new events) and scenario 3 (all opt outs have an event). These scenarios show that both the CCGs with average (Merton) and high (Oldham) rates of opting out could be showing misleading information. For the rarer event, births with very low birth weight, the rate of births with this outcome could potentially be 9-fold higher than the published rate.

The non-random nature of opt-outs has potentially large implications for the outcomes of public health monitoring, research studies and clinical audits. At an NHS trust level, information bias could produce flawed outputs in audits intending to suggest improvements and highlighting good practice. These biases will also affect the reliability of findings on maternal and children's health research. As shown in our simulation, this is particularly the case where less common (but often more serious) outcomes are studied. Problematically, the bias introduced into datasets with opt outs applied cannot be treated with the same statistical methods used to treat missing data. Multiple imputation, a method which is commonly applied to deal with missing data, relies on the assumption that the missing information can be explained by differences in the observed data [29]. However, once patients have opted out, their data is completely removed from the dataset (i.e. complete case removal) meaning that we cannot account for systematic differences. Using multiple imputation in these instances may actually add further bias to results [29]. Methods used in population-based surveys to overcome biases of non-consent, such as weighting adjustments or simulation studies ideally require detailed data on the population (e.g. by gender, age, deprivation level and local area) who have opted out so that correct weights can be derived - currently not published by NHS Digital. Research to determine whether those who have opted out of sharing data are different in terms of socio-demographic and health characteristics, in specific population

Lewis, KM et al. / International Journal of Population Data Science (2020) 5:1:03

Figure 1: Rates of patients opting out, by CCG: England (with inlay map of London CCGs), as at December 2018. Data from NHS Digital [21]

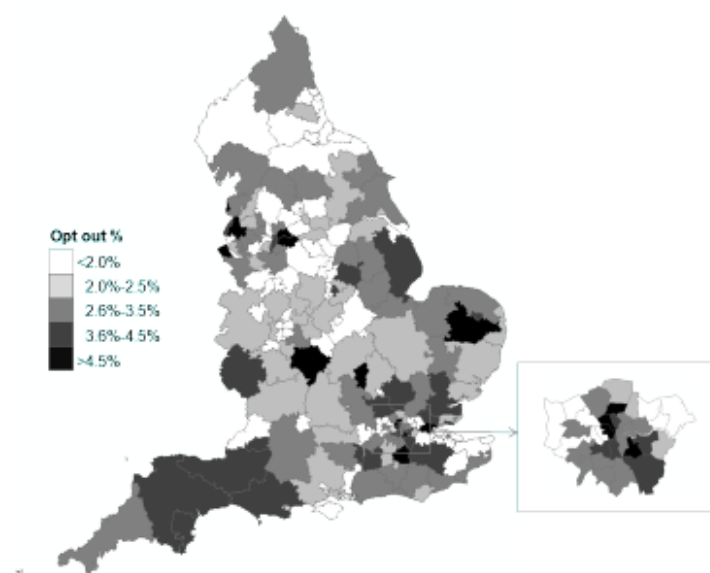


Table 1: Published and simulated number of events, patients and event rates, by indicator and CCG. Data from the National Child and Maternal Health Intelligence Network [25] and NHS Digital [26]

	Caesarean section			Very low birth weight		
	Bradford City CCG	Merton CCG	Oldham CCG	Bradford City CCG	Merton CCG	Oldham CCG
CCG opt out rate (%) [26]	0.3	2.74	10.08	0.3	2.74	10.08
Published maternity data [25]						
Events	478	948	792	36	39	39
Deliveries	2024	3265	2797	1657	3219	3195
Event rate	23.6	29	28.3	2.17	1.21	1.22
(95% CI)	(21.8,25.5)	(27.5,30.6)	(26.7,30.0)	(1.57,2.99)	(0.89,1.65)	(0.89,1.66)
New event rate (95% CI)*						
Scenario (1)	23.5 (21.8,25.4)	28.2 (26.7,30.0)	25.5 (24.0,27.0)	2.17 (1.57,2.99)	1.18 (0.86,1.61)	1.1 (0.80,1.50)
Scenario (2)	23.6 (21.8,25.5)	29 (27.5,30.6)	28.3 (26.8,30.0)	2.17 (1.57,2.99)	1.21 (0.89,1.64)	1.2 (0.90,1.63)
Scenario (3)	23.8 (22.0,25.7)	31 (29.4,32.6)	35.6 (34.0,37.3)	2.47 (1.82,3.33)	3.93 (3.32,4.64)	11.2 (10.2,12.3)

*Scenario 1 (no new events) = published events/(deliveries + extra deliveries) where extra deliveries= published deliveries/(1-opt out rate)-published deliveries; scenario 2 (average events) = ((extra deliveries*event rate)+published events)/(deliveries + extra deliveries); scenario 3 (max. events) = (events+ extra deliveries)/(deliveries + extra deliveries).

Lewis, KM et al. / International Journal of Population Data Science (2020) 5:1:03

Figure 2: Observed and simulated rates of deliveries by caesarean section: By CCG, 2016/17 [21, 25]

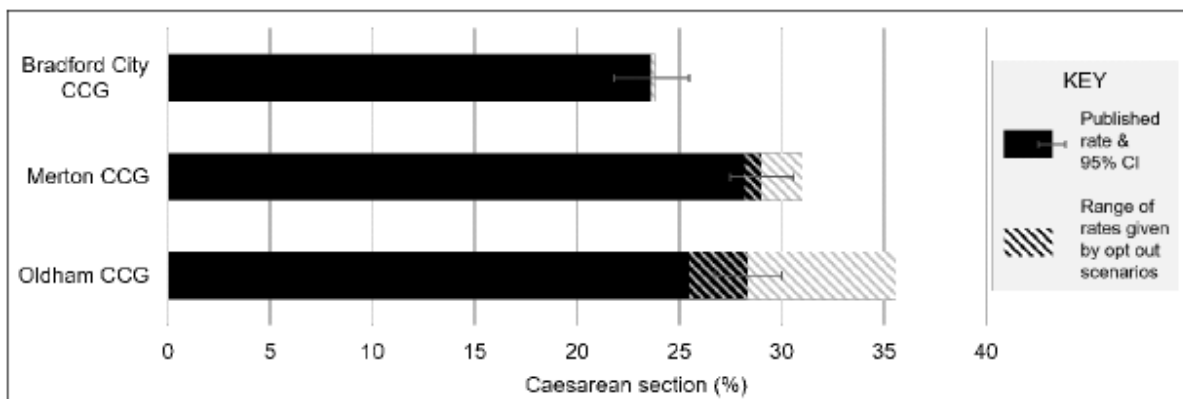
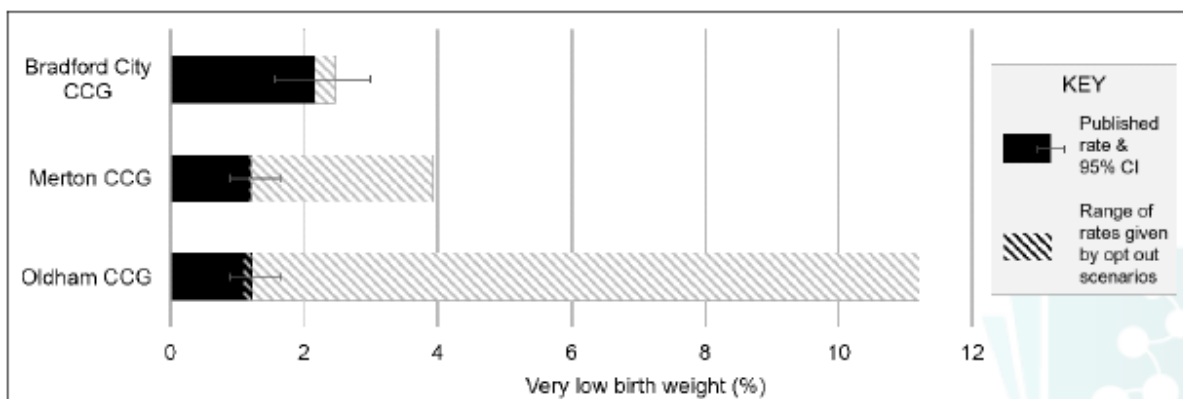


Figure 3: Observed and simulated rates of births with very low birth weight (<1500g) observed and simulated rates: by CCG, 2016 [21, 25]



such as expectant mothers, would help in applying such methods to tackle data missing due to opt outs.

An inevitable problem?

The experience of public health researchers in the Nordic countries demonstrate that data sharing can be achieved with buy-in from citizens and at great value to clinical research [30]. In these countries, residents are assigned a personal identity number from birth, which can be used to track individuals across time and generations. Unlike the UK equivalent (NHS number), which is used only in health and social care settings [31], personal identity numbers are used across a multitude of sectors. This means that individual data can be linked across health, education and social security datasets, for example, providing high quality comprehensive information on risk factors for ill-health and other outcomes. Research using these administrative datasets have contributed markedly to the evidence base regarding determinants of health and disease across the life-course. Valuable research outcomes include the long-term social and medical consequences of preterm birth and heritability of pre-eclampsia, amongst others [32, 33].

Before a data-based research project can begin, approval must be gained from a regional or national ethics committee. To protect confidentiality, personal identifiers are not shared with researchers and results cannot be published at an individual level [34]. In essence, the safeguards in place are very similar to those in England. In contrast, however, there is broad public general awareness and acceptance of the use of individual data in research and a long-standing culture of trust in public services and data donation for the good of the population [4, 30]. More research is needed on why this discrepancy in public perception of using administrative data for research between England and the Nordic countries has arisen, what can be done to improve public trust in England, and who is best placed to do it.

Evidence from England suggests that greater knowledge of research processes and safeguards improves the likelihood of acceptance of electronic health records being used without explicit consent [35]. An electronic real-time dataset integrating primary and secondary care was successfully implemented over a decade ago in Salford, Manchester. All patients were sent a letter with information and a query about opting out. Less than 0.2% of the nearly quarter of a million patients chose to opt out [36]. In contrast, information about the now withdrawn care.data scheme to integrate primary and secondary care records was disseminated by generic leaflets, which were reportedly not seen by the majority of the population [11]. However, given that some of the opt-outs may have been driven by GP practices rather than patient-level decisions [9], it is not clear to what extent whether it was the NHS information campaign directed at patients (or indeed lack of it) that led to the 2.8% opt-out rate, or a lack of buy-in from clinicians.

In terms of providing information to the general public, some lessons have been learnt since care.data. A national radio campaign ran for 6-weeks after the launch of the new opt out system and NHS Digital's website now links to Understanding Patient Data [15], an informative website run by Wellcome Trust. To further exhibit the benefits of sharing data for au-

ditions and research, NHS Digital could start by listing examples of how health care data have been used for research, as is available in other NHS held datasets (e.g. CPRD [37]). Other innovative examples of dissemination include the University of Manchester's citizen's jury on health records [38] and an animation created as part of the #datasaveslives campaign by the Farr Institute [39]. However, the impact and wider reach of these schemes are not clear.

Arguably, information about the benefit of data sharing can only go so far in raising public confidence. Evidence from the research literature and reflected in media coverage suggests that there is unease about the potential of commercial entities, such as pharmaceutical and insurance companies, to make profit from NHS data. This is in contrast to the largely positive view of university researchers or NHS staff making use of this data [22, 36]. Therefore, Wellcome Trust's call for clear examples of 'acceptable and unacceptable purposes' for which data can and cannot be used, amongst other steps, should be heeded [15]. NHS Digital is beginning to advertise the benefits of sharing NHS data for research and planning purposes, and it is vital these efforts are continued and extended. Clinicians play a pivotal role in the discourse of patient consent to use NHS data for research. Their concerns must be better understood and addressed in future consultations and information campaigns about using data for research.

Conclusion

When patients to choose to opt out of sharing data beyond their direct care, the reliability of service information and evaluation and wider research based on electronic health records is diminished. We call for more transparent, clear and detailed information on: who can apply to use NHS data and for what reasons; the safeguards in place to protect individual information; and, importantly, the wider consequences of opting out on population health research and public health service information. Only with this information can individuals be expected make an informed decision about opting out of sharing their data.

Acknowledgments

Patient involvement: We thank Maurice Hoffman, Shilpa Patel, Katherine Ruane and other representatives from the Farr Institute London Public Panel for their perspectives on the issues discussed, as well as specific comments on an earlier version of this paper.

Funding Statement

KML is funded by the Medical Research Council [grant number MR/N013867/1].

Statement on conflicts of Interest

The authors declare that they have no conflicts of interest.

Ethics Statement

This study used open source aggregate-level data and, therefore, ethical approval was not required.

References

- Häyrynen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: A review of the research literature. *International Journal of Medical Informatics*. 2008;77(5):291-304. <https://doi.org/10.1016/j.ijmedinf.2007.09.001>
- Mold F, de Lusignan S, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic review in primary care. *British Journal of General Practice*. 2015;65(632):e141-e51. <https://doi.org/10.3399/bjgp15X683941>
- Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *International Journal of Epidemiology*. 2017;46(4):1093-i. <https://doi.org/10.1093/ije/dyx015>
- Webster PC. Sweden's health data goldmine. *CMAJ : Canadian Medical Association Journal*. 2014;186(9):E310-E. <https://doi.org/10.1503/cmaj.109-4713>
- Guroi-Urganci I, Cromwell DA, Edozien LC, Smith GC, Onwere C, Mahmood TA, et al. Risk of placenta previa in second birth after first birth cesarean section: a population-based study and meta-analysis. *BMC Pregnancy and Childbirth*. 2011;11(1):95. <https://doi.org/10.1186/1471-2393-11-95>
- Love ER, Bhattacharya S, Smith NC, Bhattacharya S. Effect of interpregnancy interval on outcomes of pregnancy after miscarriage: retrospective analysis of hospital episode statistics in Scotland. *BMJ*. 2010;341:c3967. <https://doi.org/10.1136/bmj.c3967>
- Meeraus WH, Petersen I, Gilbert R. Association between Antibiotic Prescribing in Pregnancy and Cerebral Palsy or Epilepsy in Children Born at Term: A Cohort Study Using The Health Improvement Network. *PLOS ONE*. 2015;10(3):e0122034. <https://doi.org/10.1371/journal.pone.0122034>
- NHS England. The NHS Long Term Plan. London: NHS England; 2019.
- Piel FB, Parkes BL, Daby H, Hansell AL, Elliott P. The challenge of opt-outs from NHS data: a small-area perspective. *Journal of Public Health*. 2018. <https://doi.org/10.1093/pubmed/fty059>
- National Data Guardian. Information: To share or Not to Share? The Information Governance Review. Department of Health; 2013.
- McCartney M. Care.data doesn't care enough about consent. *BMJ*. 2014;348. <https://doi.org/10.1136/bmj.g2831>
- Armstrong S. Data deadlines loom large for the NHS. *BMJ*. 2018;360.
- Heather B. Poor uptake of new scheme to give patients control of confidential records: *Health Service Journal*; 2018 [Available from: <https://www.hsj.co.uk/technology-and-innovation/poor-uptake-of-new-scheme-to-give-patients-control-of-confidential-records/7023553.article>].
- Caldicott F. Review of Data Security, Consent and Opt-Outs. 2016.
- Wellcome Trust. Understanding Patient Data 2018 [Available from: <https://understandingpatientdata.org.uk/>].
- Culshaw N, Pasupathy D, Kyle P. The value of obstetric surveillance systems within the National Health Service. *The Obstetrician & Gynaecologist*. 2013;15(2):85-9. <https://doi.org/10.1111/tog.12005>
- Which? Where to give birth 2018 [Available from: <https://www.which.co.uk/birth-choice/where-to-give-birth>].
- National Maternity and Perinatal Audit Royal College of Obstetricians and Gynaecologist 2017 [Available from: <http://www.maternityaudit.org.uk/>].
- Aylin P, Bennett P, Bottle A, et al. e. Estimating the risk of adverse birth outcomes in pregnant women undergoing non-obstetric surgery using routinely collected NHS data: an observational study NIHR Journals Library. 2016;4:29. <https://doi.org/10.3310/hsdr04290>
- Zylbersztejn A, Gilbert R, Hjern A, Wijlaars L, Hardelid P. Child mortality in England compared with Sweden: a birth cohort study. *The Lancet*. [https://doi.org/10.1016/S0140-6736\(18\)30670-6](https://doi.org/10.1016/S0140-6736(18)30670-6)
- NHS Digital. [MI] National Data Opt-Out, December 2018 2018 [Available from: <https://digital.nhs.uk/data-and-information/publications/statistical/national-data-opt-out/december-2018/december-2018>].
- Hill EM, Turner EL, Martin RM, Donovan JL. "Let's get the best quality research we can": public awareness and acceptance of consent to use existing data in health research: a systematic review and qualitative study. *BMC Medical Research Methodology*. 2013;13:72. <https://doi.org/10.1186/1471-2288-13-72>
- Sakshaug JW, Couper MP, Ofstedal MB, Weir DR. Linking survey and administrative records: mechanisms of consent. *Social Methods Res*. 2012;41(4):535-69. <https://doi.org/10.1177/0049124112460381>




APPENDIX - CHAPTER 4

Appendix 4.1. Thorax publication: Geospatial and seasonal variation of bronchiolitis in England

Respiratory infection

ORIGINAL RESEARCH

Geospatial and seasonal variation of bronchiolitis in England: a cohort study using hospital episode statistics

Kate Marie Lewis , Bianca De Stavola , Pia Hardelid 

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/thoraxjnl-2019-213764>).

Population, Policy and Practice Research and Teaching Department, University College London Great Ormond Street Institute of Child Health, London, UK

Correspondence to

Kate Marie Lewis, Population, Policy and Practice, University College London Institute of Child Health, London, UK; kate.lewis.14@ucl.ac.uk

Received 27 June 2019
Revised 10 December 2019
Accepted 16 December 2019
Published Online First
20 January 2020

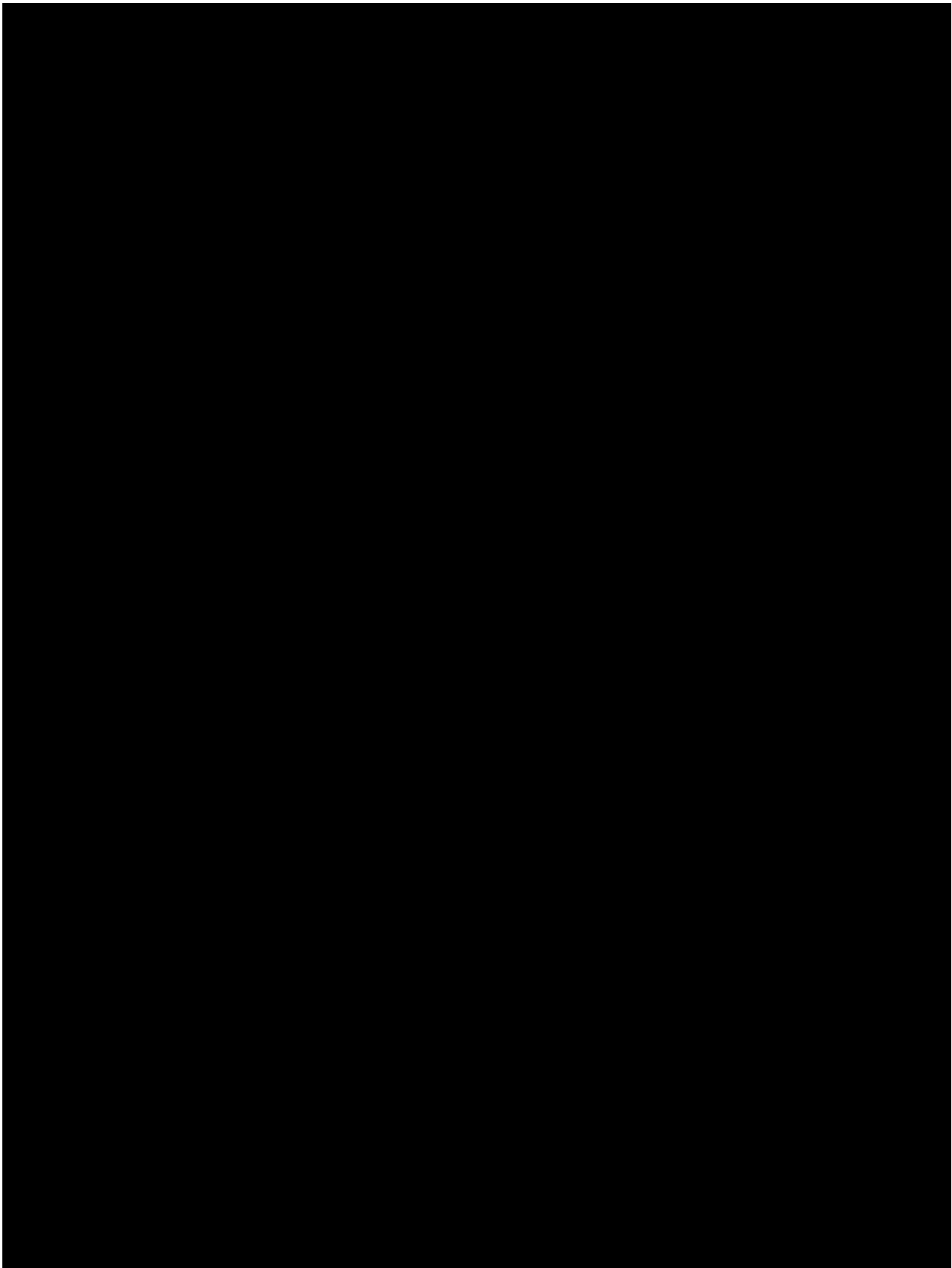
 Check for updates

© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

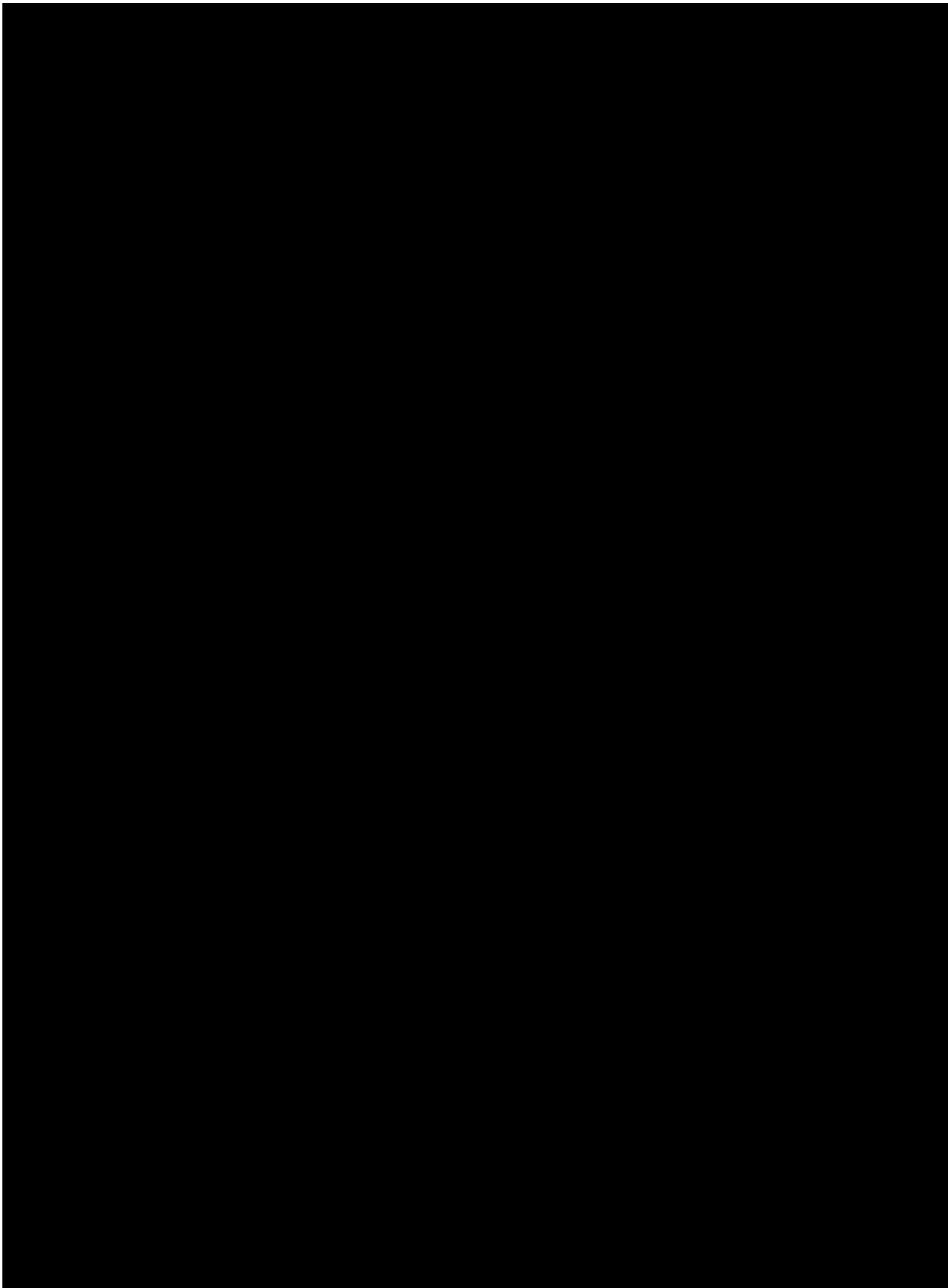
To cite: Lewis KM, De Stavola B, Hardelid P. *Thorax* 2020;**75**:262–268.



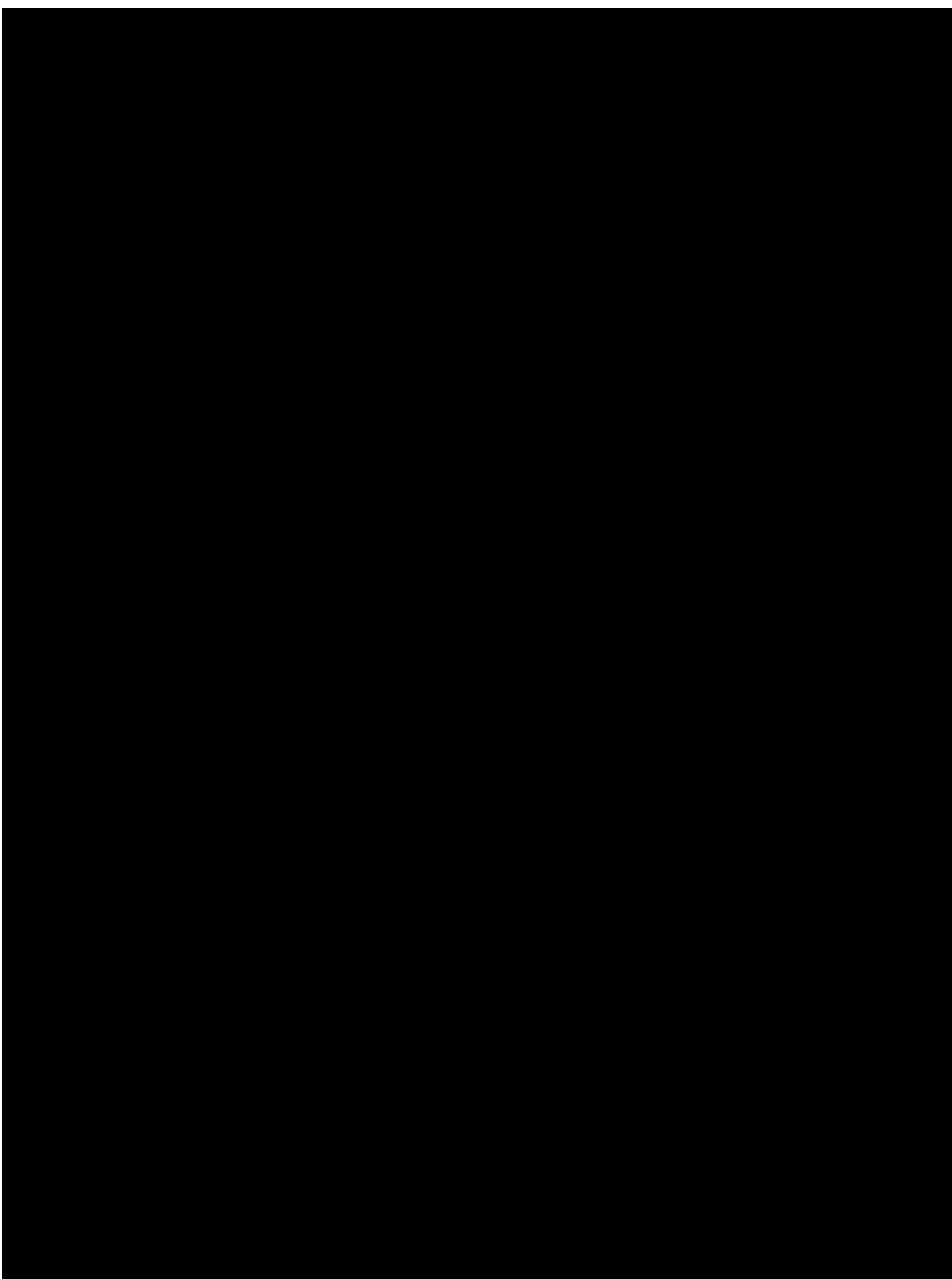
Respiratory infection

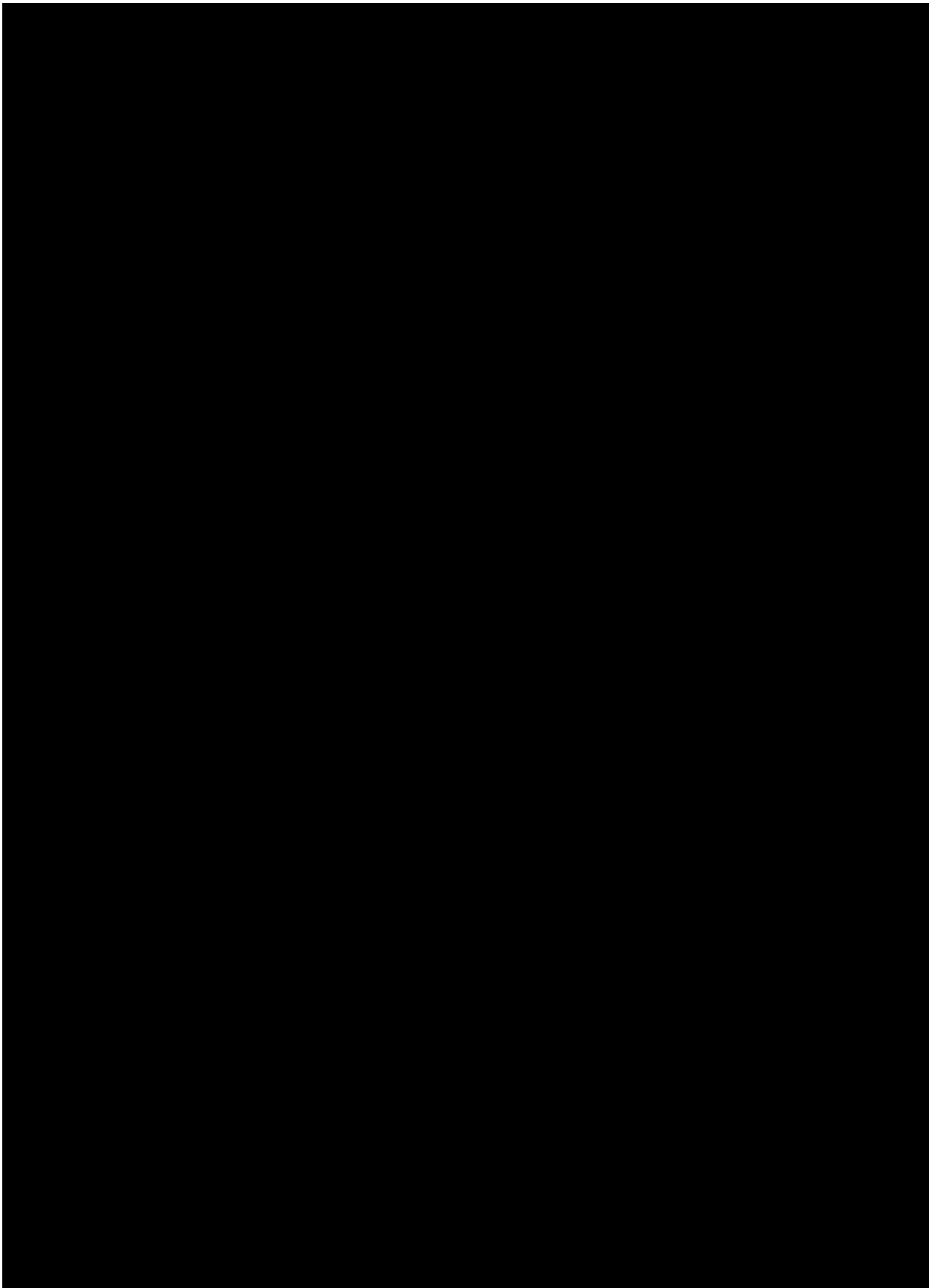


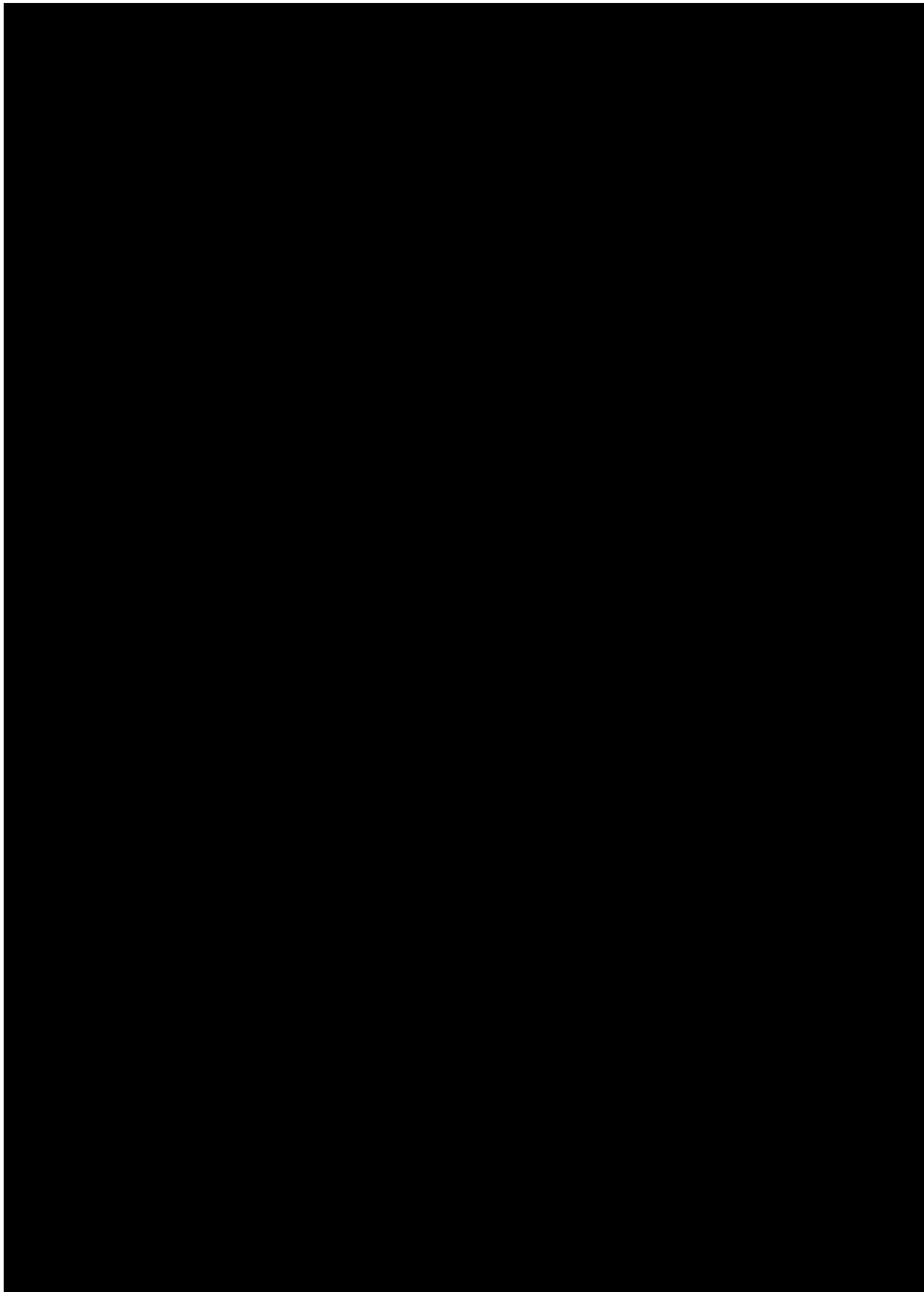
Respiratory infection

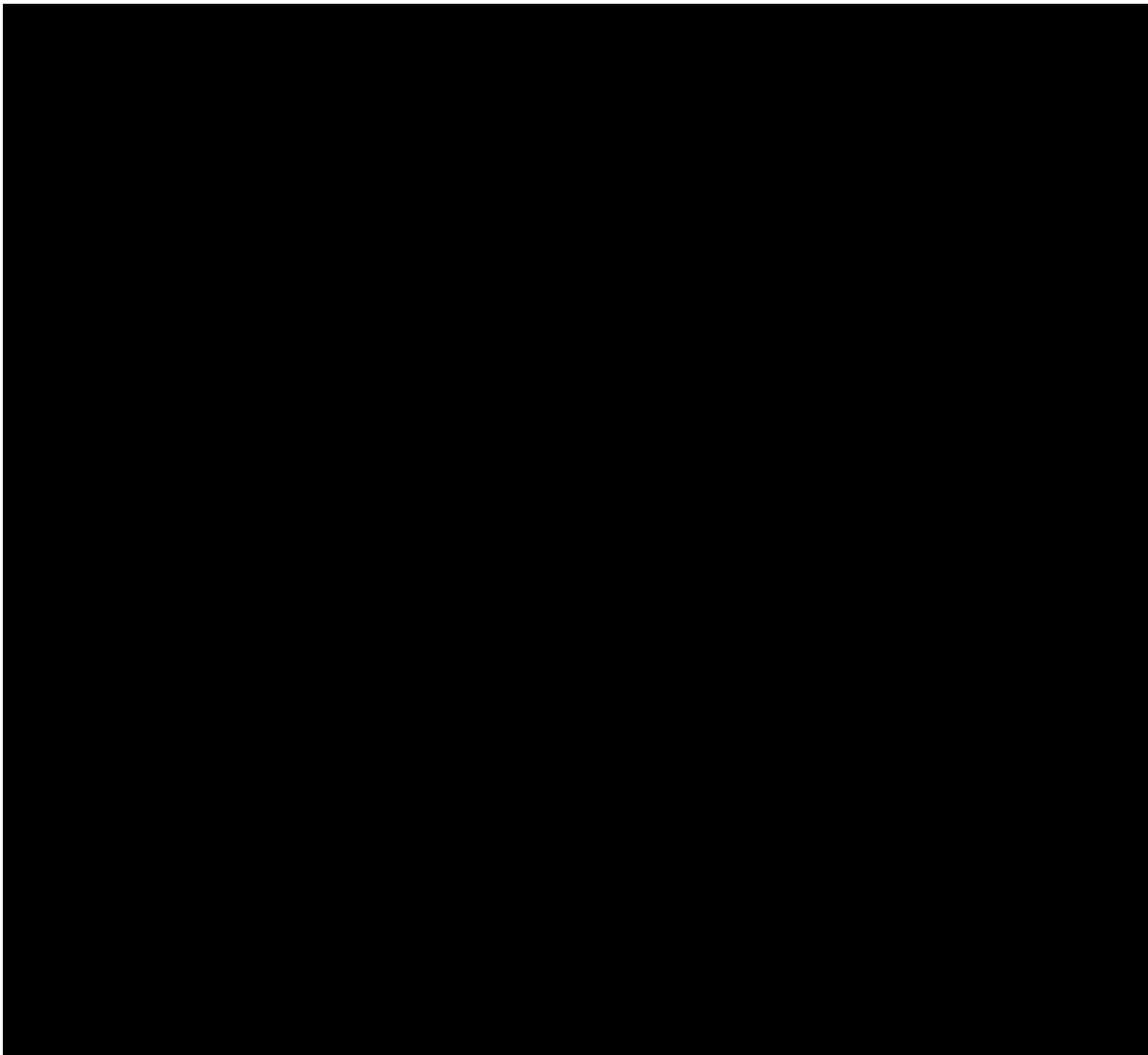


Respiratory infection









Appendix 4.2. Stata code for Chapter 4 harmonic Poisson regression analyses

```

use "dataset\stsplitt_2012to2016.dta", clear // survival set dataset

*create person-time
stsplitt      week_pyr, at(0(7.024038)365.25)
gen          pyr=_t-_t0

*createtimeseries: from dates
gen         date=_origin+_t
format     date %td
gen         year_event=year(date)
gen         month_event=month(date)
gen         week_event=week(date)
drop       date

*invalid rows
drop       if _st==0
drop       if year_event==2011 | year_event==2017

save       "geo\geo_stsplitt.dta", replace

*+++++++ HARMONIC REGRESSION MODEL: by REGION ++++++++

*convert data into easier to manage file
use       "geo\geo_stsplitt.dta",clear
keep     week_event region year_event hesid d_entry end_fup end ///
        start_bronchiolitis_any_use end_bronchiolitis_any_use ///
        event_st_d _origin _t _t0 year_event
qui strate region week_event year_event, output ///
("geo\geo_stsplitt_strate region.dta", replace)

use       "geo\geo_stsplitt_strate region.dta",clear

*generate cos sine terms
gen       sine=sin(2*_pi*1/52*(week_event-1))
gen       cos=cos(2*_pi*1/52*(week_event-1))

*final model
poisson  _D c.sine#ib7.region c.cos#ib7.region sine cos ib2013.year_event ///
        ib7.region, vce(robust) exp(_Y)

*test interaction
testparm  c.sine#ib7.region c.cos#ib7.region

*event prediction for duration of epidemic

```

```

predict      event_region
preserve
collapse    (sum) event_region _Y, by (region week_event)
gen          rate=(event_region/_Y)*365250
list        region week_event rate, divider clean
restore

*lincom for week 51
forvalues   k=1(1)9{
            lincom `k'.region+(.9709418*cos#`k'.region)+(-.2393157*sine#`k'.region),irr
            }

*SEASON PARAMETERS
*!calculations for seasonality parameters following Barnett AG, Dobson AJ. Analysing
seasonal health data. Berlin: Springer-Verlag, 2010.
gen         b0=_b[_cons]
gen         b1=_b[sine]
gen         b2=_b[cos]
forvalues   k=1(1)9{
            gen alpha`k'=_b[`k'.region]
            gen delta1`k'=_b[c.sine#`k'.region]
            gen delta2`k'=_b[c.cos#`k'.region]
            }
forvalues   k=1(1)9{
            gen b0`k'=b0+alpha`k'
            gen b1`k'=b1+delta1`k'
            gen b2`k'=b2+delta2`k'
            }
forvalues   k=1(1)9{
            label var b0`k' "intercept for region_`k'"
            label var b1`k' "sine coef for region_`k'"
            label var b2`k' "cosine coef for region_`k'"
            }
*amplitude, gamma: a*sqrt(b1^2+b2^2), where a=-1 if b2<0
gen         a=1
forvalues   k=1(1)9{
            replace a=-1 if b2`k'<0
            gen gamma`k'=a*(sqrt(b1`k'^2+b2`k'^2))
            }
forvalues   k=1(1)9{
            nlcom sqrt((_b[sine]+_b[sine#`k'.region])^2+(_b[cos]+_b[cos#`k'.region])^2)
            }

**exp(amplitude) in poisson models
forvalues   k=1(1)9{
            gen gamma`k'_rate=exp(gamma`k')
            }
forvalues   k=1(1)9{

```

```

        nlcom
        exp(sqrt((_b[sine]+_b[sine#`k'.region])^2+(_b[cos]+_b[cos#`k'.region])^2))
        }

*phase (in radians), psi:arctan(b1/b2)
forvalues k=1(1)9{
gen      psi`k'=atan(b1`k'/b2`k')
}

forvalues k=1(1)9{
nlcom atan((_b[sine]+_b[sine#`k'.region])/(_b[cos]+_b[cos#`k'.region]))
}

*estimates
forvalues k=1(1)9{
gen      peak_week`k'=52.14*(psi`k'/(2*_pi))+1
}

forvalues k=1(1)9{
nlcom 52.14* ((atan((_b[sine]+_b[sine#`k'.region]) / (_b[cos]+_b[cos#`k'.region])))/(2*_pi))+1
}

*+++++++ HARMONIC REGRESSION MODEL: by CCG ++++++++

*convert data into easier to manage file
use      "geo\geo_stsplit.dta",clear
keep     week_event region CCG_number year_event hesid d_entry end_fup ///
        end_start_bronchiolitis_any_use end_bronchiolitis_any_use ///
        event_st_d_origin_t_t0 year_event
qui strate CCG_number week_event year_event, output ("geo\geo_stsplit_strate.dta",
replace)
use      "geo\geo_stsplit_strate.dta", clear

*generate cos sine terms
gen      sin=sin(2*_pi*1/52.14*(week_event-1))
gen      cos=cos(2*_pi*1/52.14*(week_event-1))

mepoisson _D sin cos ib2013.year_event, exp(_Y) || CCG_number: sin cos

*random effects intercept:
predict      rem_sin rem_cos rem_cons, remeans
predict      re_sin re_cos re_cons,reflects
gen          b1=_b[sine]+re_sin
gen          b2=_b[cos]+re_cos
gen          b0=_b[_cons]+re_cons
gen          IR_CCG=rem_cons

*SEASON PARAMETERS

```

*!calculations for seasonality parameters following Barnett AG, Dobson AJ. Analysing seasonal health data. Berlin: Springer-Verlag, 2010.

*amplitude, gamma: $a \cdot \sqrt{b_1^2 + b_2^2}$, where $a = -1$ if $b_2 < 0$

gen $a = 1$

replace $a = -1$ if $b_2 < 0$

gen $\text{gamma} = a \cdot (\sqrt{b_1^2 + b_2^2})$

**exp(amplitude) in poisson models

gen $\text{gamma_rate} = \exp(\text{gamma})$

*phase (in radians), psi: $\arctan(b_1/b_2)$

gen $\text{psi} = \text{atan}(b_1/b_2)$

*estimates

gen $\text{peak_week} = 52.14 \cdot (\text{psi} / (2 \cdot \pi)) + 1$

gen $\text{mean_rate} = \exp(b_0) \cdot 365250$

Appendix 4.3. Harmonic Poisson regression models

Table A4.1. Output from the harmonic Poisson regression model for rates of bronchiolitis, by region and year

		Coefficient	<i>p</i>	(95% CI)
Region	North East	0.51	<0.001	(0.47, 0.56)
	North West	0.68	<0.001	(0.65, 0.71)
	Yorkshire & the Humber	0.49	<0.001	(0.45, 0.52)
	East Midlands	0.31	<0.001	(0.27, 0.35)
	West Midlands	0.50	<0.001	(0.47, 0.54)
	East of England	0.16	<0.001	(0.12, 0.19)
	London	Reference		
	South East	0.26	<0.001	(0.23, 0.29)
	South West	0.36	<0.001	(0.32, 0.39)
Sine		-0.41	<0.001	(-0.43, -0.39)
Cosine		1.16	<0.001	(1.13, 1.18)
Sine* Region	North East	0.07	<0.001	(0.03, 0.11)
	North West	-0.07	<0.001	(-0.10, -0.04)
	Yorkshire & the Humber	-0.01	0.46	(-0.05, 0.02)
	East Midlands	0.03	0.15	(-0.01, 0.07)
	West Midlands	-0.04	0.03	(-0.07, 0.00)
	East of England	-0.03	0.16	(-0.06, 0.01)
	London	Reference		
	South East	-0.06	<0.001	(-0.09, -0.02)
	South West	0.06	<0.001	(0.02, 0.09)
Cosine* Region	North East	0.41	<0.001	(0.36, 0.46)
	North West	0.21	<0.001	(0.18, 0.24)
	Yorkshire & the Humber	0.25	<0.001	(0.21, 0.29)
	East Midlands	0.32	<0.001	(0.28, 0.37)
	West Midlands	0.23	<0.001	(0.19, 0.27)
	East of England	0.37	<0.001	(0.32, 0.41)
	London	Reference		
	South East	0.28	<0.001	(0.25, 0.32)
	South West	0.32	<0.001	(0.28, 0.37)
Year of event	2012	0.06	<0.001	(0.05, 0.08)
	2013	Reference		
	2014	0.06	<0.001	(0.04, 0.07)
	2015	0.22	<0.001	(0.20, 0.23)
	2016	0.27	<0.001	(0.25, 0.30)
Constant		-9.85	<0.001	(-9.88, -9.83)

Wald test testing interaction between region and sine and cosine functions, $p < 0.0001$; likelihood ratio test, $p < 0.0001$

Table A8.2. Output of the harmonic multilevel mixed-effect Poisson regression analysis for rates of bronchiolitis

		Coefficient	(95% CI)
Individual-level			
Sine		- 0.44	(-0.46, -0.42)
Cos		1.42	(1.39, 1.45)
Year of event	2012	0.06	(0.04, 0.07)
	2013	Ref.	
	2014	0.06	(0.04, 0.08)
	2015	0.22	(0.20, 0.23)
	2016	0.27	(0.25, 0.28)
Constant		- 9.54	(-9.58, -9.49)
CCG-level			
var(sine)		0.01	(0.01, 0.02)
var(cos)		0.05	(0.04, 0.06)
var(constant)		0.11	(0.10, 0.13)

Excluding London-based CCGs**Table A4.3.** Harmonic multilevel mixed-effect Poisson regression analysis for rates of RSV bronchiolitis: Excluding London-based CCGs

		Coefficient	<i>p</i>	(95% CI)
Individual-level				
Sine		- 0.44	<0.001	(-0.46, -0.42)
Cos		1.48	<0.001	(1.45, 1.50)
Year of event	2012	0.07	<0.001	(0.05, 0.09)
	2013	Ref.		
	2014	0.06	<0.001	(0.05, 0.08)
	2015	0.22	<0.001	(0.20, 0.24)
	2016	0.27	<0.001	(0.25, 0.28)
Constant		- 9.44	<0.001	(-9.49, -9.40)
CCG-level				
var(sine)		0.01		(0.01, 0.02)
var(cos)		0.03		(0.03, 0.04)
var(constant)		0.10		(0.08, 0.13)

Appendix 4.4: Sensitivity analyses

Bootstrapped standard errors (SEs)

Table A4.4. Multivariable linear regression analyses assessing the effect of population density and IMD score on IRR and peak timing at the CCG-level following two stage analysis with 1,000 bootstraps on a 20% sample of model

	<u>No bootstrapped SEs</u>			<u>1000 bootstrapped SEs</u>		
	Coefficient	Lower CI	Upper CI	Coefficient	Lower CI	Upper CI
IRR						
Population density*	-0.1046	-0.1375	-0.0718	-0.1046	-0.1410	-0.0683
IMD	0.0214	0.0160	0.0268	0.0214	0.0159	0.0269
Peak week						
Population density*	-0.1471	-0.1532	-0.1411	-0.1471	-0.2288	-0.0655
IMD	0.0496	0.0457	0.0534	0.0496	-0.0079	0.1070
IMD ²	-0.0007	-0.0007	-0.0006	-0.0007	-0.0018	0.0004

*Log-scale

Note: Year of admission was excluded from model due to convergence problem

One admission per child

Table A4.5. Derived average annual seasonal estimates following harmonic Poisson regression (not shown)*, by region: England, 2012 to 2016 (one admission per child)

Region*	N	<u>Admissions</u>		Amplitude $\hat{\gamma}^{**}$ (95% CI)	Peak week \hat{P} (95%CI)
		Rate per 1000 infant- years (95% CI)			
North East	8,471	57.8	(56.6, 59.1)	5.66 (4.91, 6.52)	51.3 (50.8, 51.7)
North West	20,017	61.2	(60.4, 61.9)	4.86 (4.30, 5.49)	50.1 (49.7, 50.6)
Yorkshire & the Humber	15,163	51.6	(50.7, 52.4)	4.88 (4.28, 5.56)	50.5 (50.0, 51.0)
East Midlands	10,939	45.4	(44.6, 46.3)	5.23 (4.57, 5.98)	50.9 (50.4, 51.4)
West Midlands	17,155	51.4	(50.6, 52.2)	4.97 (4.36, 5.67)	50.3 (49.9, 50.8)
East of England	13,543	40.3	(39.6, 41.0)	5.56 (4.83, 6.40)	50.6 (50.2, 51.1)
London	16,144	27.5	(27.0, 27.9)	3.92 (3.47, 4.42)	50.0 (49.5, 50.6)
South East	20,232	42.2	(41.6, 42.8)	5.24 (4.55, 6.04)	50.3 (49.8, 50.8)
South West	12,868	46.8	(46.0, 47.6)	5.43 (4.73, 6.23)	50.9 (50.5, 51.4)
		Phase shift $\hat{\psi}$ (95%CI)		Duration (weeks)	IRR at week 51 (95% CI)
North East		-0.23	(-0.29, -0.17)	26 (26-52)	2.46 (2.11, 2.87)
North West		-0.36	(-0.42, -0.31)	26 (25-51)	2.43 (2.10, 2.82)
Yorkshire & the Humber		-0.32	(-0.38, -0.26)	26 (25-51)	2.06 (1.78, 2.40)
East Midlands		-0.27	(-0.33, -0.21)	26 (26-52)	1.88 (1.61, 2.18)
West Midlands		-0.34	(-0.40, -0.29)	26 (25-51)	2.07 (1.78, 2.40)
East of England		-0.30	(-0.36, -0.25)	26 (26-52)	1.69 (1.46, 1.97)
London		-0.37	(-0.43, -0.31)	26 (25-51)	Reference
South East		-0.34	(-0.40, -0.28)	26 (25-51)	1.73 (1.49, 2.02)
South West		-0.26	(-0.32, -0.21)	26 (26-52)	1.96 (1.68, 2.29)

*Adjusted for year of admission, **amplitude exponentiated

Table A4.6. Harmonic multilevel mixed-effect Poisson regression analysis for rates of RSV bronchiolitis (one admission per child)

	Coefficient	<i>p</i>	(95% CI)
Individual-level			
Sine	-0.53	<0.001	(-0.55, -0.51)
Cos	1.54	<0.001	(1.51, 1.57)
Constant	-9.60	<0.001	(-9.64, -9.56)
variance			
var(sine)	0.02		(0.01, 0.02)
var(cos)	0.05		(0.04, 0.06)
var(constant)	0.10		(0.09, 0.12)

Note: Year of admission was excluded from model due to convergence problems

Table A4.7. Multivariable linear regression analyses assessing the effect of population density and IMD on IRR compared the national average and peak timing of CCG seasonal epidemic (one admission per child)

	All CCGs		
	Coefficient (95% CI)		Partial η^2 (95% CI)
IRR			
Population density*	-0.09	(-0.12, -0.06)	0.13 (0.06, 0.22)
IMD	0.02	(0.02, 0.03)	0.23 (0.14, 0.32)
Constant	1.22	(1.02, 1.42)	
	Total η^2		0.24 (0.14, 0.34)
Peak week			
Population density*	-0.26	(-0.31, -0.21)	0.32 (0.22, 0.41)
IMD	0.11	(0.08, 0.15)	0.19 (0.10, 0.28)
	-	-	
IMD ²	0.002	(-0.002, 0.001)	0.12 (0.05, 0.21)
Constant	50.59	(50.11, 51.06)	
	Total η^2		0.40 (0.29, 0.48)


*Log-scale

APPENDIX - CHAPTER 5

Appendix 5.1. JECH publication: Is socioeconomic position associated with bronchiolitis seasonality?

Original research

Is socioeconomic position associated with bronchiolitis seasonality? A cohort study

Kate Lewis , Bianca De Stavola, Pia Hardelid

► Supplemental material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jech-2019-213056>).

Population, Policy and Practice Research and Teaching Department, UCL Great Ormond Street Institute of Child Health, London, UK

Correspondence to

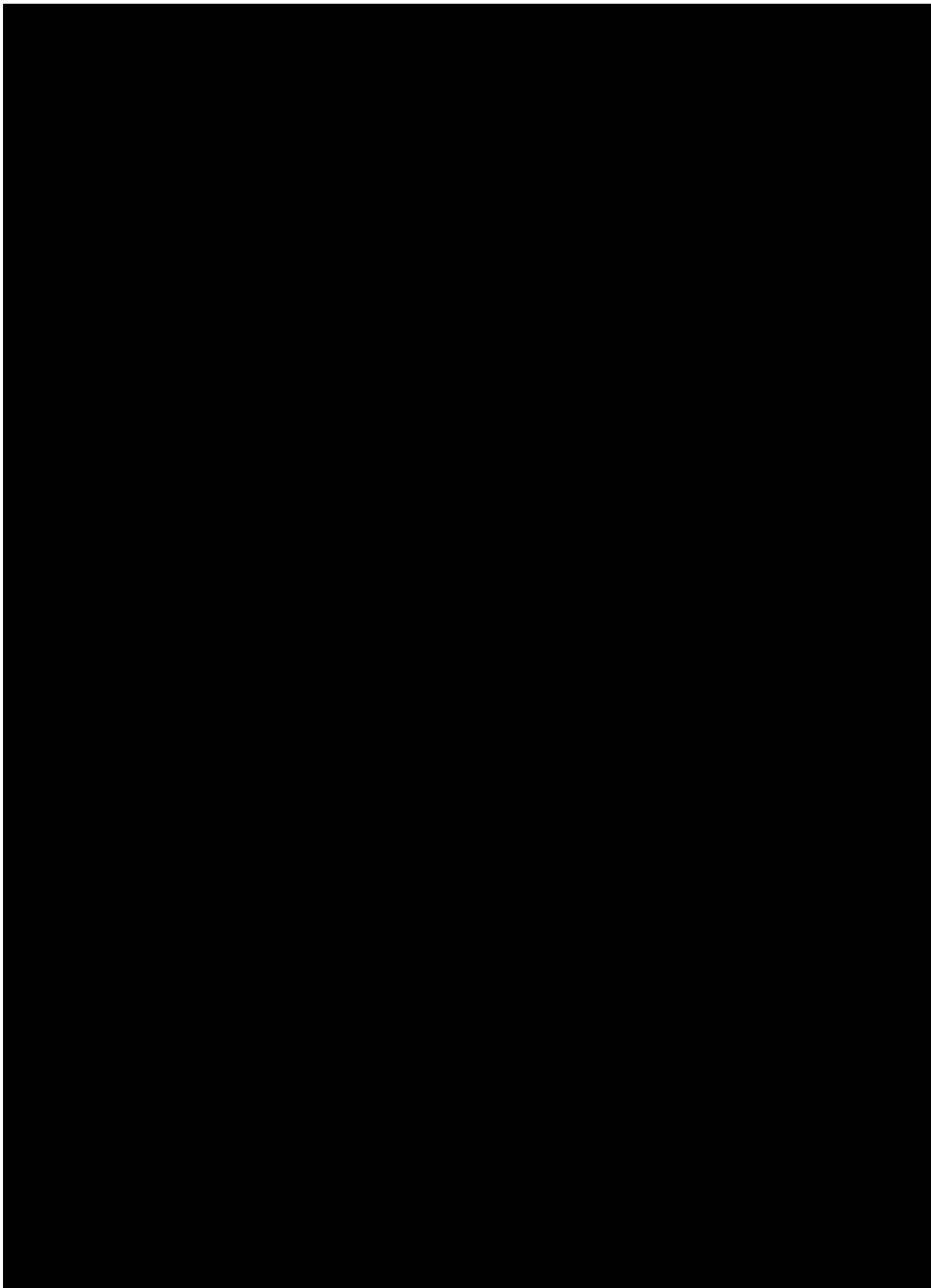
Kate Lewis, Department of Population, Policy and Practice, UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK; kate.lewis.14@ud.ac.uk

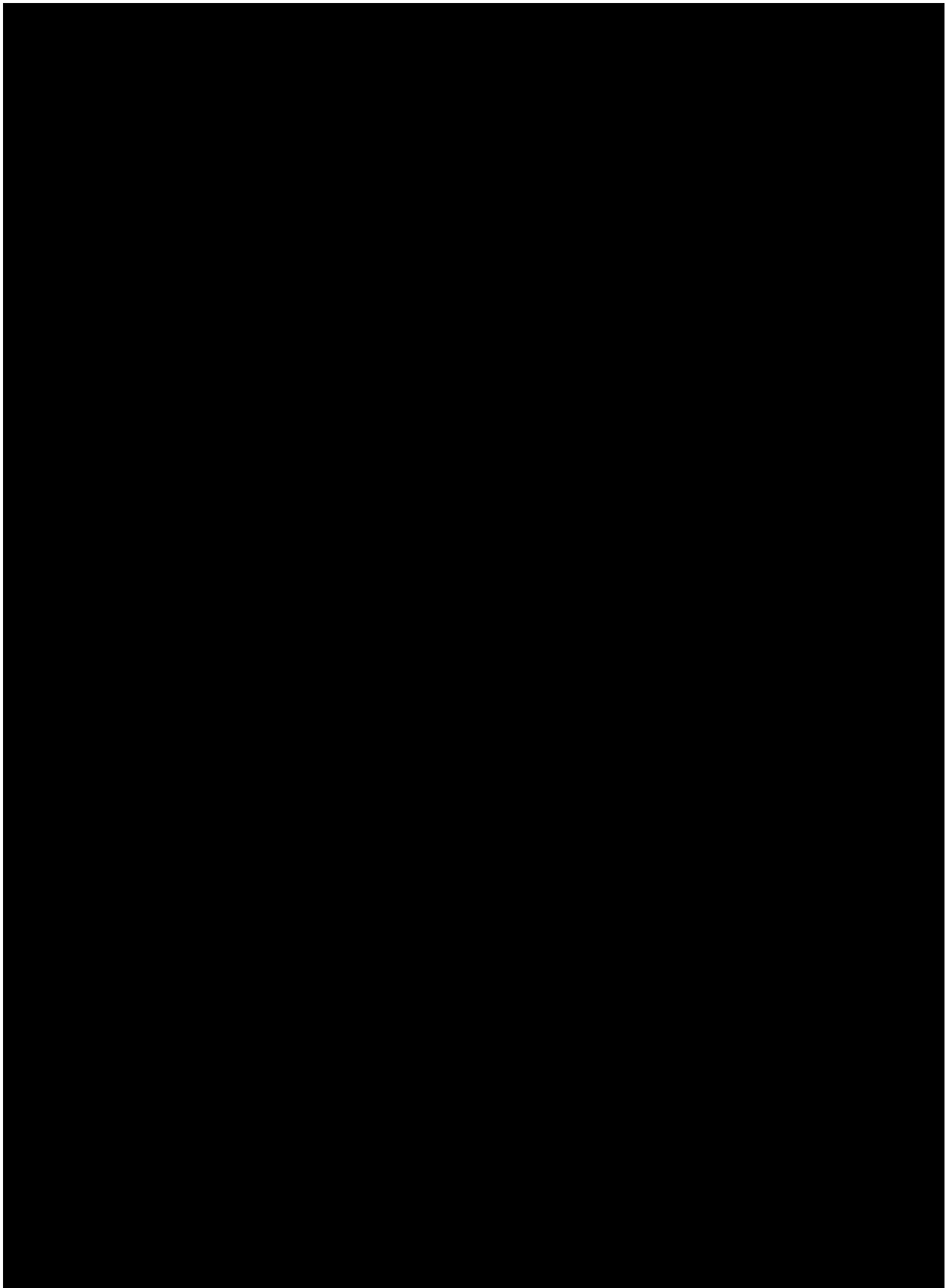
Received 9 August 2019
Revised 22 June 2020
Accepted 12 August 2020

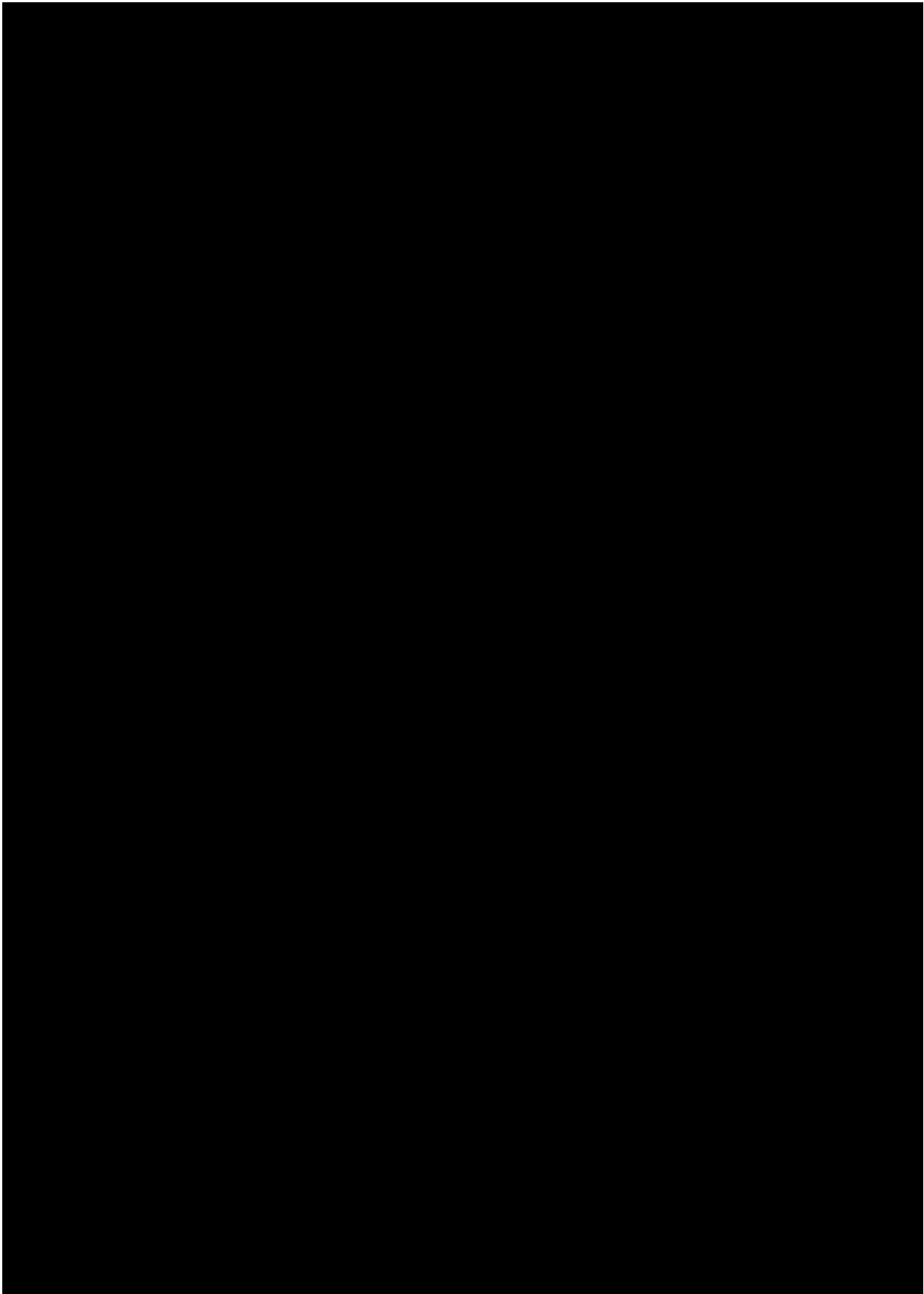


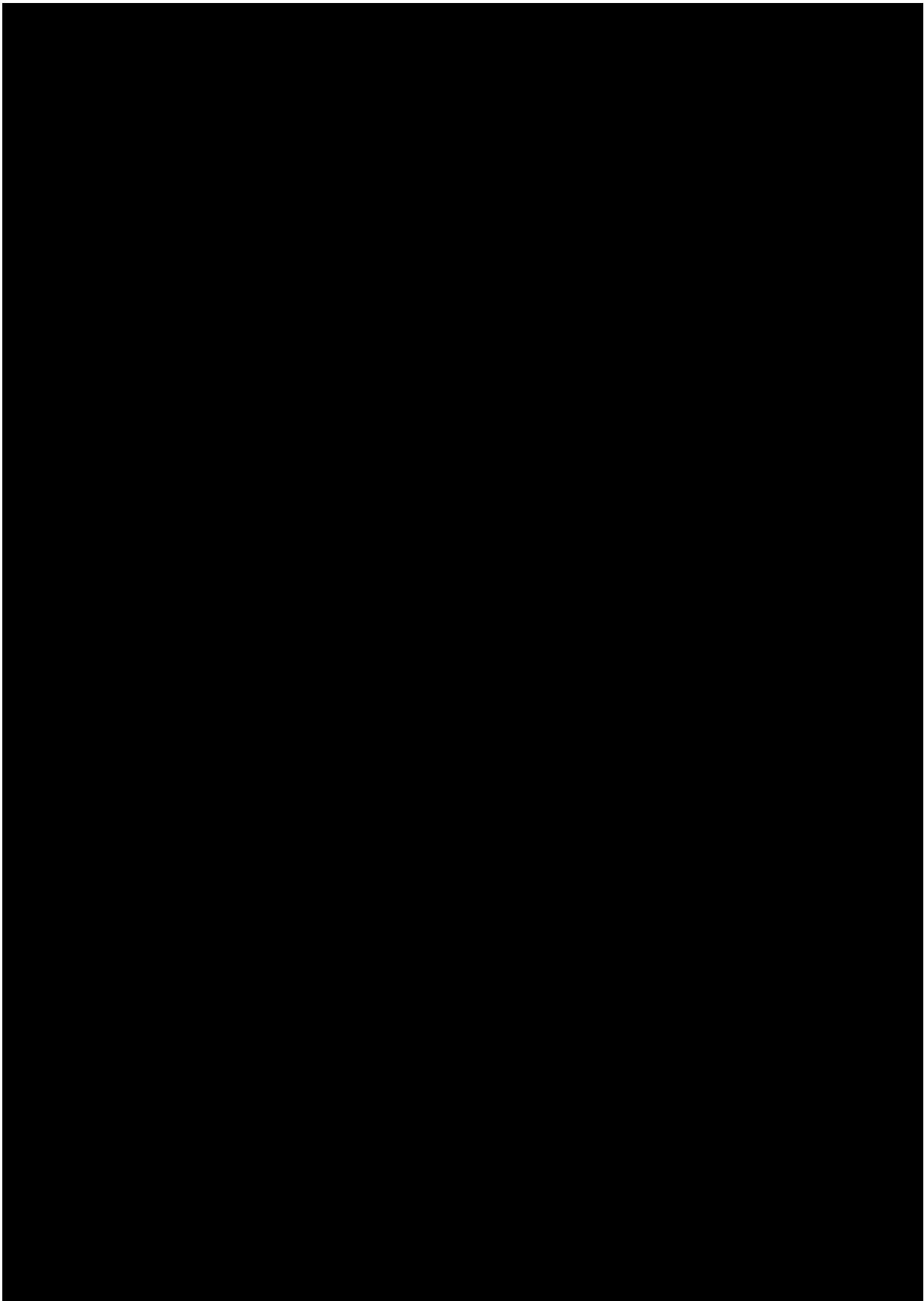
© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

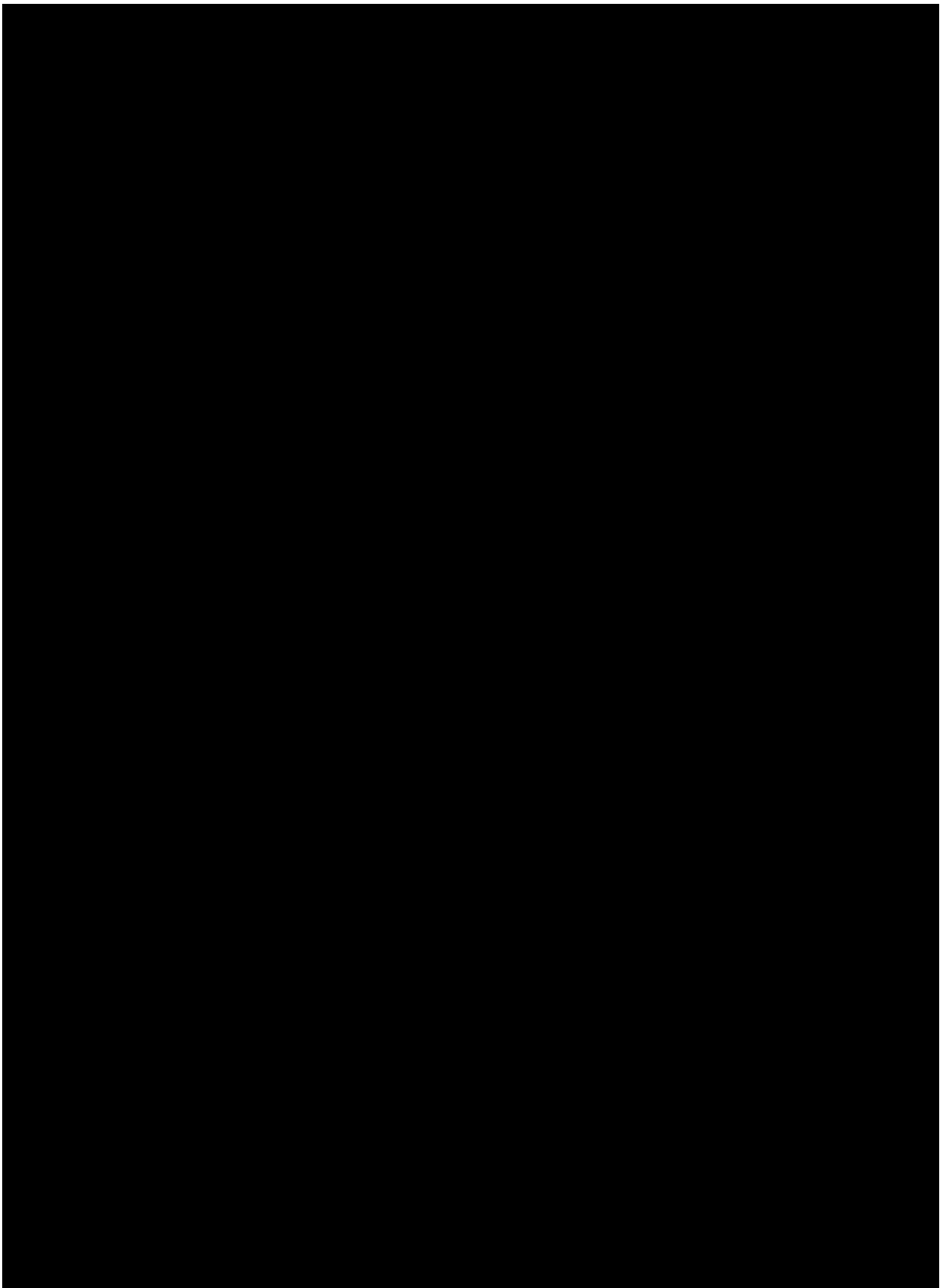
To cite: Lewis K, De Stavola B, Hardelid P. *J Epidemiol Community Health* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jech-2019-213056

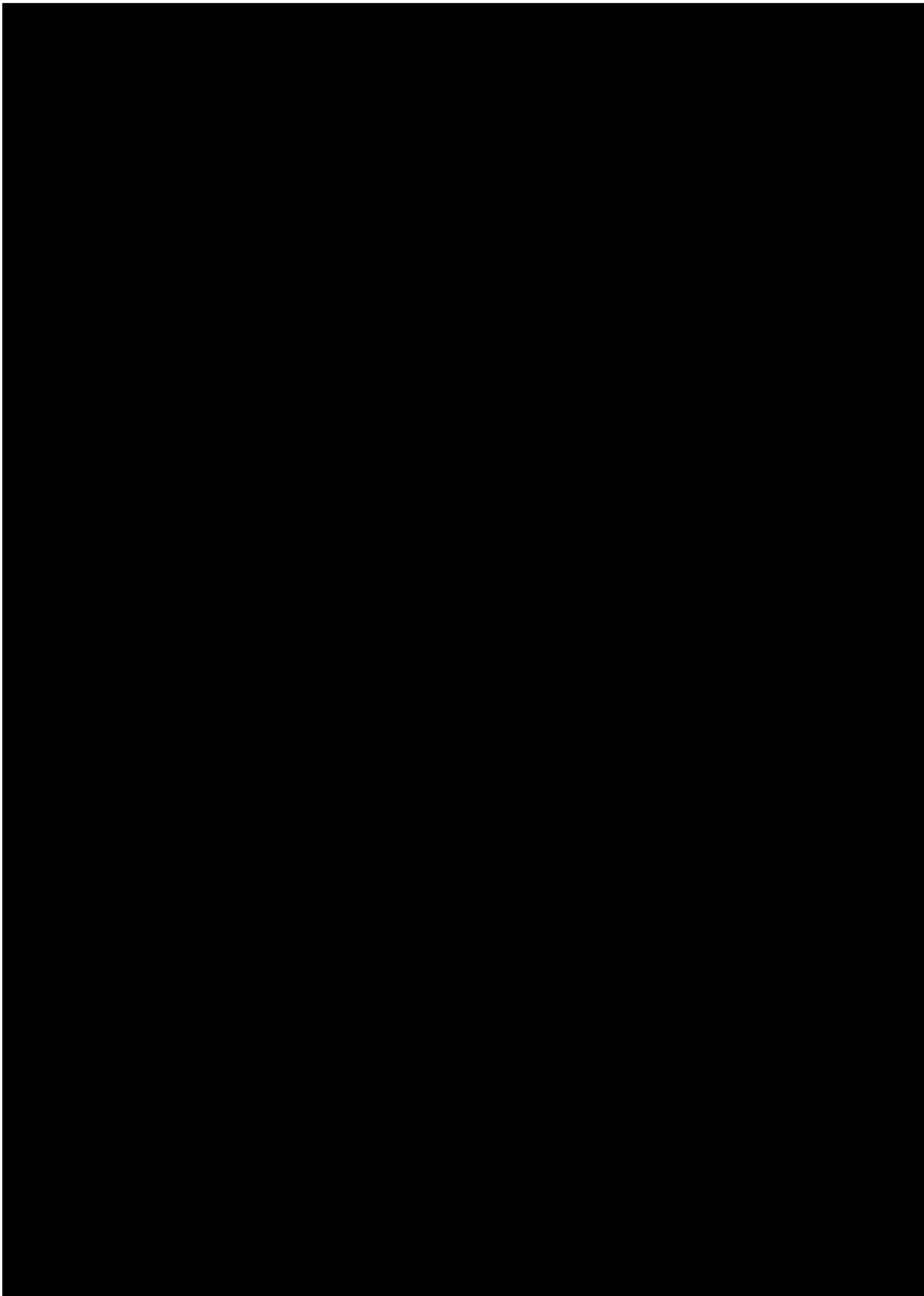


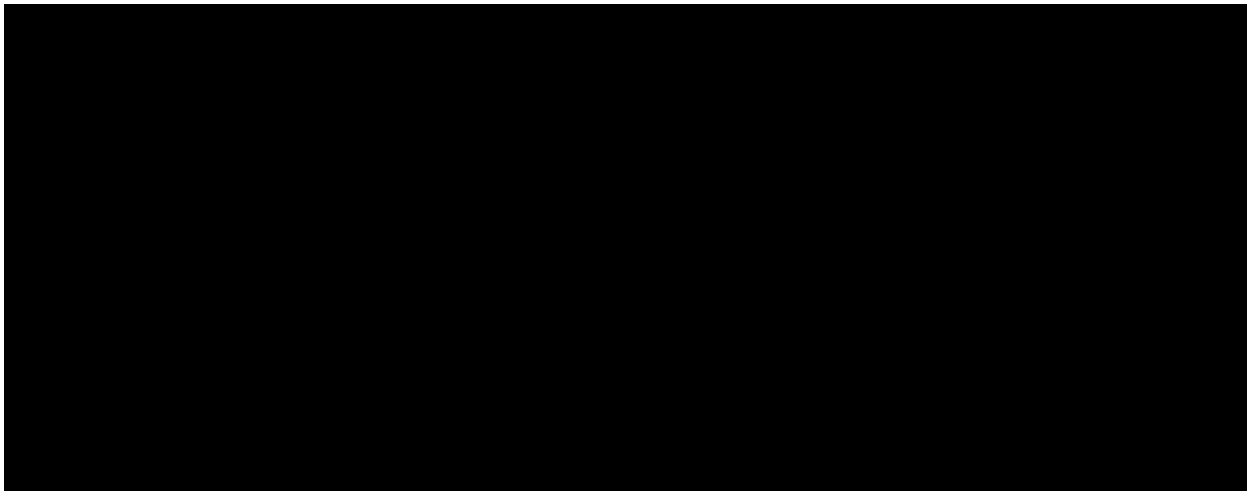












Appendix 5.2. Stata code for Chapter 5 harmonic regression analyses

```

use "dataset\stsplit_2012to2016_2.dta", clear // survival set dataset

*create person-time
stsplit      week_pyr, at(0(7.024038)365.25)
stsplit      ageband, at(0, 91.3125, 182.625, 365.25)

*create age var
gen          age=2 if ageband!=.
replace      age=1 if ageband<180
replace      age=0 if ageband==0
label        variable ageband "<3/4-<6/6-<12 months"

*createtimeseries: from dates
gen          date=_origin+_t
format       date %td
gen          year_event=year(date)
gen          month_event=month(date)
gen          week_event=week(date)
drop         date

*invalid rows
drop         if _st==0
drop         if year_event==2011 | year_event==2017
gen          pyrs=_t-_t0

*+++++++ HARMONIC REGRESSION MODEL ++++++++

*generate harmonics
gen          sine=sin(2*_pi*1/52.14*(week_event-1))
gen          cosine=cos(2*_pi*1/52.14*(week_event-1))

*final model
poisson      _d ib1.mdob ib1.sex ib2013.year_event ib7.region ib2.age c.sine c.cosine ///
             c.sine#i.IMD c.cosine#i.IMD c.sine#i.age c.cosine#i.age i.IMD, ///
             exp(pyrs) vce(cluster hesid) irr

testparm     c.sine#i.IMD c.cosine#i.IMD
testparm     c.sine#i.age c.cosine#i.age

predict      event_IMD
estat        ic

*****
*duration of epidemic by IMD
preserve

```

```
collapse      (sum) event_IMD _d pyrs, by (IMD week_event)
gen           rate_obs=(_d/pyrs)*365250
gen           rate_pre=(event_IMD/pyrs)*365250
list          IMD week_event rate_obs rate_pre, divider clean
restore
```

```
*lincom for week 50
forvalues k=1(1)5{
lincom `k'.IMD+(.929261*cosine#`k'.IMD)+(-.3694239*sine#`k'.IMD),irr
}
```

```
*lincom for week 51
forvalues k=1(1)5{
lincom `k'.IMD+(.9669321*cosine#`k'.IMD)+(-.2550341*sine#`k'.IMD),irr
}
```

*SEASON PARAMETERS

*!calculations for seasonality parameters following Barnett AG, Dobson AJ. Analysing seasonal health data. Berlin: Springer-Verlag, 2010.

```
gen b01=_b[_cons]
gen b11=_b[sine]
gen b21=_b[cosine]
```

```
forvalues k=2(1)5{
gen alpha`k'=_b[`k'.IMD]
gen delta1`k'=_b[`k'.IMD#c.sine]
gen delta2`k'=_b[`k'.IMD#c.cosine]
}
```

```
forvalues k=2(1)5{
gen b0`k'=b01+alpha`k'
gen b1`k'=b11+delta1`k'
gen b2`k'=b21+delta2`k'
}
```

```
forvalues k=1(1)5{
label var b0`k' "intercept for IMD_`k'"
label var b1`k' "sine coef for IMD_`k'"
label var b2`k' "cosine coef for IMD_`k'"
}
```

amplitude, gamma: $a\sqrt{b1^2+b2^2}$, where $a=-1$ if $b2<0$

```
gen a=1
forvalues k=1(1)5{
replace a=-1 if b2`k'<0
gen gamma`k'=a*(sqrt(b1`k'^2+b2`k'^2))
}
forvalues k=1(1)5{
```

```
nlcom sqrt((_b[sine]+_b[sine#`k'.IMD])^2+(_b[cosine]+_b[cosine#`k'.IMD])^2)
}
```

****exp(amplitude) in poisson models**

```
forvalues k=1(1)5{
gen gamma_rate`k'=exp(gamma`k')
}
forvalues k=1(1)5{
nlcom exp(sqrt((_b[sine]+_b[sine#`k'.IMD])^2+(_b[cosine]+_b[cosine#`k'.IMD])^2))
}
```

***phase (in radians), psi: -arctan(b1/b2)**

```
forvalues k=1(1)5{
gen psi`k'=-atan(b1`k'/b2`k')
}
forvalues k=1(1)5{
nlcom atan((_b[sine]+_b[sine#`k'.IMD])/(_b[cosine]+_b[cosine#`k'.IMD]))
}
```

***estimates**

```
forvalues k=1(1)5{
gen peak_week`k'=52.14*(psi`k'/(2*_pi))+1
}
forvalues k=1(1)5{
nlcom 52.14*((atan((_b[sine]+_b[sine#`k'.IMD])/(_b[cosine]+_b[cosine#`k'.IMD]))/(2*_pi))+1)
}
```

Appendix 5.3. Harmonic Poisson regression models

Table A5.1. IRRs (95% CI) from multivariable harmonic Poisson regression models using IMD quintiles including and excluding the health deprivation and disability domain

Variable		Full IMD† IRR (95% CI)	IMD excluding health domain‡ IRR (95% CI)
IMD quintile	1 - Least deprived	Reference	Reference
	2	1.10 (1.06, 1.14)	1.10 (1.06, 1.14)
	3	1.21 (1.17, 1.25)	1.21 (1.17, 1.26)
	4	1.37 (1.32, 1.42)	1.34 (1.30, 1.39)
	5 - Most deprived	1.62 (1.57, 1.67)	1.62 (1.57, 1.67)
Sine		0.55 (0.53, 0.56)	0.55 (0.53, 0.56)
Cosine		3.45 (3.33, 3.57)	3.45 (3.33, 3.57)
Age group	< 3 months	1.85 (1.81, 1.89)	1.85 (0.00, 0.00)
	3 to < 6 months	1.68 (1.65, 1.72)	1.68 (1.81, 1.89)
	6 to < 12 months	Reference	Reference
Sex	Male	1.46 (1.44, 1.47)	1.46 (1.44, 1.47)
Year of event	2012	1.06 (1.04, 1.08)	1.06 (1.04, 1.08)
	2013	Reference	Reference
	2014	1.06 (1.04, 1.08)	1.06 (1.04, 1.08)
	2015	1.24 (1.22, 1.26)	1.24 (1.22, 1.26)
	2016	1.31 (1.29, 1.33)	1.31 (1.29, 1.33)
Region	North East	2.04 (1.98, 2.09)	2.08 (2.03, 2.14)
	North West	2.17 (2.13, 2.22)	2.22 (2.17, 2.27)
	Yorkshire and the Humber	1.82 (1.78, 1.87)	1.85 (1.80, 1.89)
	East Midlands	1.66 (1.62, 1.70)	1.68 (1.64, 1.72)
	West Midlands	1.84 (1.80, 1.88)	1.86 (1.82, 1.90)
	East of England	1.53 (1.49, 1.57)	1.54 (1.50, 1.57)
	London	Reference	Reference
	South East	1.66 (1.63, 1.70)	1.67 (1.63, 1.71)
South West	1.79 (1.74, 1.83)	1.80 (1.76, 1.84)	
Month of birth	January	Reference	Reference
	February	1.12 (1.08, 1.16)	1.12 (1.08, 1.16)
	March	1.22 (1.18, 1.26)	1.22 (1.18, 1.26)
	April	1.30 (1.26, 1.35)	1.30 (1.26, 1.35)
	May	1.35 (1.31, 1.40)	1.35 (1.31, 1.40)
	June	1.37 (1.33, 1.42)	1.37 (1.33, 1.42)
	July	1.42 (1.37, 1.47)	1.42 (1.37, 1.47)
	August	1.48 (1.43, 1.53)	1.48 (1.43, 1.53)
	September	1.52 (1.47, 1.57)	1.52 (1.47, 1.57)
	October	1.54 (1.49, 1.58)	1.54 (1.49, 1.58)
	November	1.34 (1.30, 1.38)	1.34 (1.30, 1.38)
	December	1.03 (1.00, 1.06)	1.03 (1.00, 1.06)
Sine*age	< 3 months	1.15 (1.12, 1.18)	1.15 (1.12, 1.18)
	3 to < 6 months	1.12 (1.09, 1.16)	1.12 (1.09, 1.16)
	6 to < 12 months	Reference	Reference
Cosine*age	< 3 months	1.65 (1.61, 1.70)	1.65 (1.61, 1.70)
	3 to < 6 months	1.04 (1.00, 1.07)	1.04 (1.00, 1.07)
	6 to < 12 months	Reference	Reference
Sine*IMD quintile	1 - Least deprived	Reference	Reference
	2	1.07 (1.03, 1.11)	1.06 (1.02, 1.10)
	3	1.14 (1.10, 1.18)	1.14 (1.11, 1.18)
	4	1.18 (1.14, 1.22)	1.17 (1.13, 1.21)
	5 - Most deprived	1.21 (1.18, 1.25)	1.21 (1.18, 1.25)
Cosine*IMD quintile	1 - Least deprived	Reference	Reference

2	0.99 (0.95, 1.04)	1.00 (0.95, 1.04)
3	0.99 (0.95, 1.04)	0.99 (0.95, 1.03)
4	0.95 (0.91, 0.98)	0.96 (0.93, 1.00)
5 - Most deprived	0.90 (0.87, 0.94)	0.90 (0.87, 0.93)

† Wald test of interaction between age and sine and cosine functions, $X^2(8)=239.2$, $p<0.0001$, Wald test of interaction between IMD quintiles and sine and cosine functions, $X^2(4)=1835.9$, $p<0.0001$; ‡Wald test of interaction between age and sine and cosine functions, $X^2(8)=253.8$, $p<0.0001$, Wald test of interaction between IMD quintiles and sine and cosine functions, $X^2(4)=1836.0$, $p<0.0001$

APPENDIX - CHAPTER 6

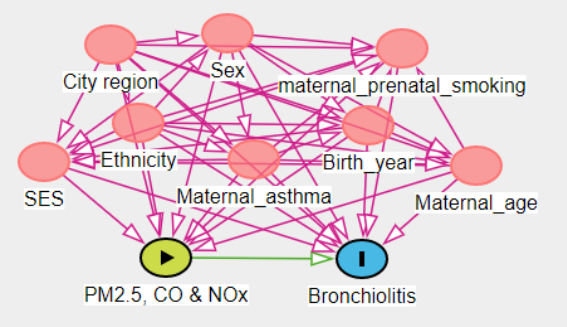
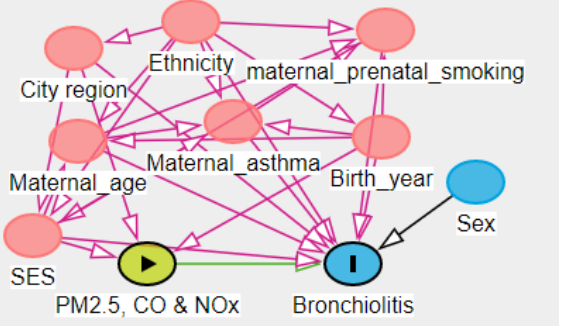
Appendix 6.1. ESC-DAG output

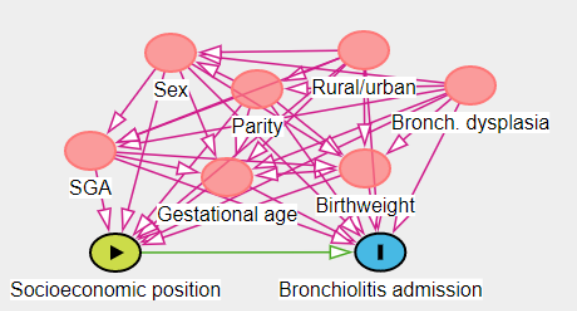
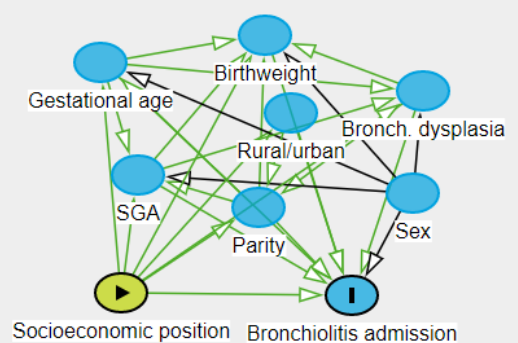
Table A6.1. Study decision logs

Publication details #1 ¹⁰³	
Author	Kengo Inagaki, Chad Blackshear, Paul A Burns, Charlotte Hobbs
Year	2020
Study title	Racial/Ethnic Disparity in the Incidence of Bronchiolitis Requiring Hospitalization
Design	Population-based longitudinal observational study using the State Inpatient Database from New York state in the United States.
Methods	Logistic regression
Sample	Infants born between 2009 and 2013 at term without comorbidities, N = 877,465
Variables	
Outcome	Rates of bronchiolitis admissions (ICD-9 codes)
Study exposure	Race/ethnicity of participants
My exposure of interest	Median household income quartile
Control(s)	Sex, delivery method, birth season, birth year
Mediator(s)	-
Moderator(s)	-
Excluded	Primary insurance (private, medicaid and other). Reason: not applicable in UK emergency hospital admission setting
Diagrams	
Implied graph*	
DAG*	

21 edges assessed; 3 edges reversed; 8 edges retained; 10 edges removed	
Publication details #2⁸²	
Author	Hannah C Moore, Nicholas de Klerk, Christopher C Blyth, et al.
Year	2019
Study title	Temporal trends and socioeconomic differences in acute respiratory infection hospitalisations in children: an intercountry comparison of birth cohort studies in Western Australia, England and Scotland
Design	Retrospective population-based cohort studies: Linked birth, death and hospitalisation data
Methods	Negative binomial regression models
Sample	6,976,508 children 0-4 years
Variables	
Outcome	Bronchiolitis hospitalisations
Exposure(s)	Socioeconomic deprivation; year
Control(s)	sex
Mediator(s)	-
Moderator(s)	-
Excluded	-
ALL EDGES HAVE ALREADY BEEN ASSESSED	
Publication details #3¹³⁸	
Author	Lawder et al.
Year	2019
Study title	Impact of maternal smoking on early childhood health: a retrospective cohort linked dataset analysis of 697 003 children born in Scotland 1997–2009
Design	Retrospective population-based cohort study; Linked birth, death, maternity, infant health, child health surveillance and admission records
Methods	Cox and logistic regression. Population attributable fraction (PAF)
Sample	Singleton births between 1997 and 2009 (n=697 003) followed to March 2012.
Variables	
Outcome	Hospital admissions with a primary diagnosis of bronchiolitis
Exposure(s)	Smoking during pregnancy
Control(s)	Scottish Index of Multiple Deprivation (SIMD), maternal age, infant gender, parity, delivery method, infant breastfed at 6–8 weeks, mother's country of birth
Mediator(s)	-
Moderator(s)	-
Excluded/changed	Mother's socioeconomic status and father's socioeconomic status. Reason: high correlation with SIMD.
Diagrams	

Implied graph*	
DAG*	
<p>30 edges assessed; 6 edges reversed; 19 edges retained; 5 edges removed</p>	
<p align="center">Publication details #4⁴⁸⁷</p>	
Author	Kennedy et al.
Year	2018
Study title	Associations of mobile source air pollution during the first year of life with childhood pneumonia, bronchiolitis, and otitis media
Design	Birth cohort; Kaiser Air Pollution and Pediatric Asthma Study, a retrospective birth cohort of children born during 2000–2010 and insured by Kaiser Permanente Georgia (USA)
Methods	Cox proportional hazards models
Sample	22,441 children followed until 2 years
Variables	
Outcome	Diagnosis of bronchiolitis (ICD-9 466.XX) before 2 nd birthday
Exposure	Residential exposure to primary fine particulate matter (PM2.5), nitrogen oxides (NOx), and carbon monoxide (CO) from traffic
Control(s)	Neighborhood socioeconomic status (SES), city region, child ethnicity, child sex, maternal asthma, Maternal smoking, birth year, and maternal age.
Mediator(s)	-
Moderator(s)	-
Excluded/changed	
Diagrams	

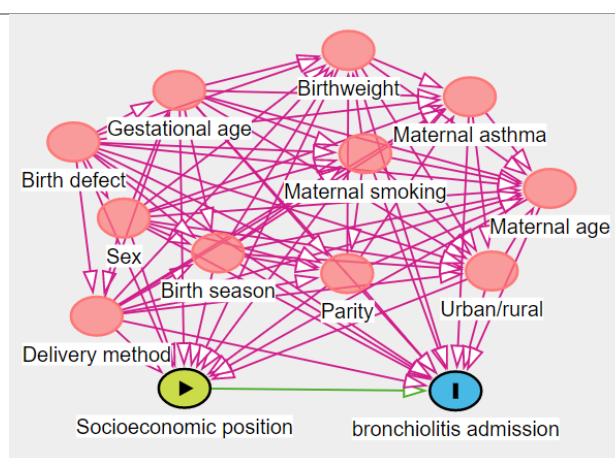
Implied graph*	
DAG*	
28 edges assessed; 5 edges reversed; 12 edges retained; 11 edges removed	
Publication details #5⁴⁸⁸	
Author	Olivia Laugier, Patricia Garcia, Mohamed Boucékine et al.
Year	2017
Study title	Influence of Socioeconomic Context on the Rehospitalization Rates of Infants Born Preterm
Design	Cohort study
Methods	Multivariable logistic regression analysis
Sample	3149 preterm infants with GA of <32 + 6 weeks with residency in the West- Provence-Alpes-Côte-d'Azur region (France) and survived to the neonatal period
Variables	
Outcome	Area-level deprivation
Exposure(s)	Emergency hospitalisation in first year of life (including bronchiolitis)
Control(s)	Gestational age, birth weight, SGA, sex, parity, bronchopulmonary dysplasia, urbanicity
Mediator(s)	-
Moderator(s)	-
Excluded	Necrotizing enterocolitis (inflammation of the bowel) and cerebral injury - related to other outcomes used in the this study
Diagrams	

Implied graph*	
DAG*	
29 edges assessed; 13 edges reversed; 13 edges retained; 3 edges removed	
Publication details #6³⁹⁷	
Author	Cintia Muñoz-Quiles, Mónica López-Lacort, Isabel Úbeda-Sansano et al.
Year	2016
Study title	Population-based Analysis of Bronchiolitis Epidemiology in Valencia, Spain
Design	Retrospective cohort; population and health databases in Valencia
Methods	Bayesian mixed Poisson regressions adjusted by fixed effects
Sample	198,223 children
Variables	
Outcome	Bronchiolitis hospitalisation <2 years
Exposure(s)	Social exclusion risk: obtained from electronic database (SIP), and its classification was based on multiple aspects such as unemployment, foreigner in irregular situation or without resources.
Control(s)	age, sex, year, prematurity, urban/rural residence (random effect: the month of the year (autocorrelated) and the health department)
Mediator(s)	-
Moderator(s)	-
Excluded	-
ALL EDGES HAVE ALREADY BEEN ASSESSED	
Publication details #7²⁶³	
Author	Khadra A Jama-Alol, Hannah C Moore, Peter Jacoby, Carol Bower and Deborah Lehmann
Year	2014
Study title	Morbidity due to acute lower respiratory infection in children with birth defects: a total population-based linked data study
Design	Retrospective cohort study

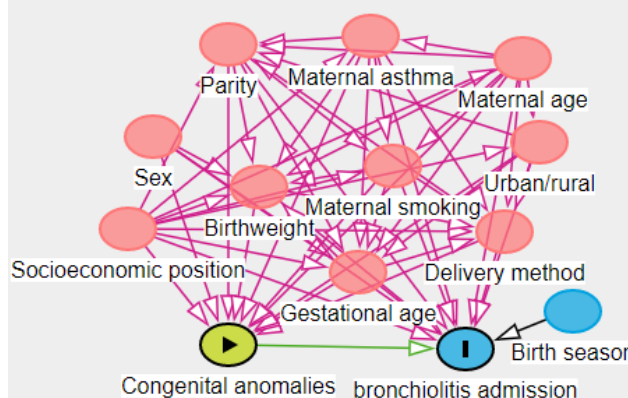
Methods	Negative binomial regression
Sample	245,249 singleton births in Western Australia (1996-2005)
Variables	
Outcome	Number of ALRI hospitalisation before age 2 years (including Bronchiolitis <12 months)
Exposure(s)	Congenital anomalies
Control(s)	SEIFA index of disadvantage, sex, gestational age, birthweight, parity, season of birth, delivery method, Smoking during pregnancy, Asthma during pregnancy, maternal age, urban/rural
Mediator(s)	-
Moderator(s)	-
Excluded/changes	Birth defect changed to congenital anomalies

Diagrams

Implied graph*



DAG*



38 edges assessed; 18 edges reversed; 7 edges retained; 13 edges removed

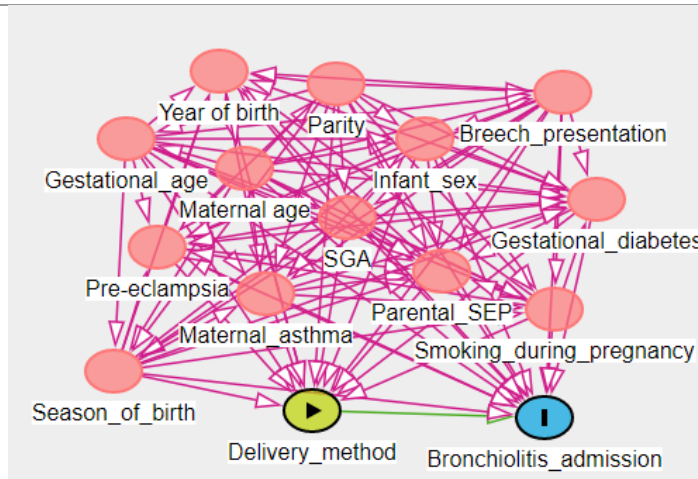
Publication details #8¹⁸²

Author	Hannah C Moore, Nicholas de Klerk, Patrick Holt, et al
Year	2011
Study title	Hospitalisation for bronchiolitis in infants is more common after elective caesarean delivery
Design	Retrospective population-based data linkage cohort study, Western Australia
Methods	Negative binomial regression
Sample	212 068 non-Aboriginal singleton births of 37–42 weeks gestation.

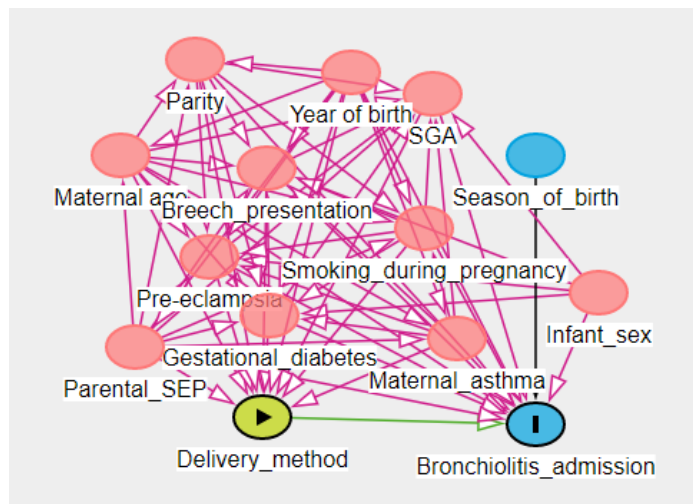
Variables	
Outcome	Hospitalisations for bronchiolitis in children aged <12 months
Exposure(s)	Delivery method
Control(s)	Socioeconomic index, pre-eclampsia, gestational diabetes, breech presentation, gestational age, smoking during pregnancy, maternal asthma, infant sex, season of birth, SGA, number of previous pregnancies, maternal age (years), year of birth,
Mediator(s)	-
Moderator(s)	-
Excluded/changed	Excluded number of admissions aged <12 months. Reason: not applicable to bronchiolitis in infancy outcome. Proportion of optimal birth weight (POBW), a measure which takes into account gestational duration, fetal sex, maternal age, maternal height and parity: changed to SGA

Diagrams

Implied graph*



DAG*



51 edges assessed; 11 edges reversed; 35 edges retained; 5 edges removed

Publication details #9⁴⁸⁹

Author	Carme Puig, Jordi Sunyer, Oscar Garcia-Algar, et al.
Year	2008

Study title	Incidence and risk factors of lower respiratory tract illnesses during infancy in a Mediterranean birth cohort
Design	Birth cohort recruited in Barcelona, Spain: home visit and follow-up questionnaire
Methods	Multivariate regression
Sample	A total of 487 infants were recruited at birth and followed up for 1 year
Variables	
Outcome	LRTI in infancy including bronchiolitis
Exposure(s)	Paternal occupation using the UK Registrar General's 1990 classification
Control(s)	Maternal asthma, number of siblings, breastfed duration
Mediator(s)	-
Moderator(s)	-
Excluded	-
ALL EDGES HAVE ALREADY BEEN ASSESSED	

*In the diagrams, the yellow node is the exposure, the red nodes are ancestors of the exposure and outcome, the blue node with a black outline is the outcome and all other blue nodes are ancestors of the outcome. Red edges represent biasing paths and green edges represent causal paths.

Appendix 6.2. Lancet abstract: Could socioeconomic inequity in bronchiolitis admissions be reduced through intervening on gestational age?

Meeting Abstracts

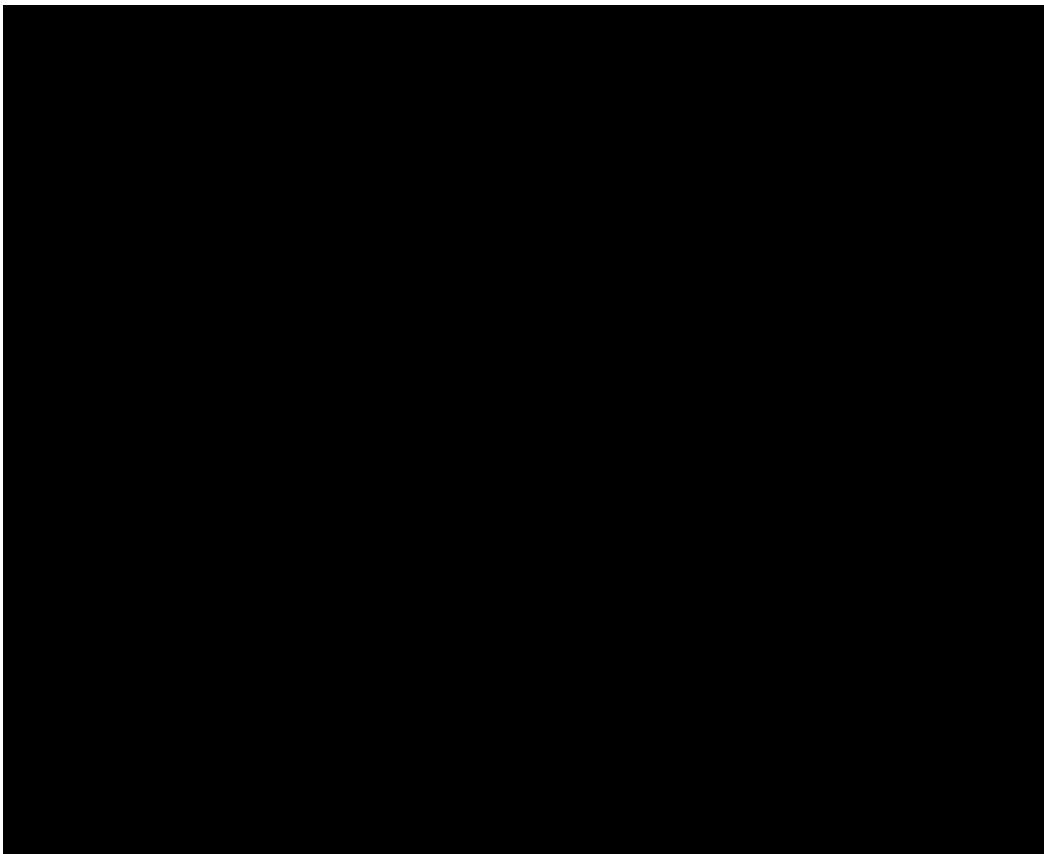
Could socioeconomic inequity in bronchiolitis admissions be reduced through intervening on gestational age? A principled analysis of cohort data

Kate Marie Lewis, Bianca De Stavola, Pia Hardelid

Published Online
November 22, 2018

UCL Great Ormond Street
Institute of Child Health,
London, UK (K M Lewis MSc,
Prof B De Stavola PhD,
P Hardelid PhD)

Correspondence to:
Ms Kate M Lewis, Population,
Policy and Practice, UCL Great
Ormond Street Institute of Child
Health, London, WC1N 1EH, UK
kate.lewis.14@ucl.ac.uk



Appendix 6.3. Stata code for causal mediation analyses

Set up file

```
*****
* Sets up file for:
* 1) IPW of MSMs – for categorical X
* 2) G-estimation of structural nested model - for categorical X
* M-Y confounders c1 c2 c3 c4 c5 c6
* X-Y counfounders c1 c2 c6
*****
use      "$temp\cohort_main.dta", clear
gen      id= childid
gen      x= deprivation3 /*where 1 low 2 medium 3 high*/
gen      y= bronch_indicator

gen      m = smoke_preg
gen      c1 = birthyear_child
gen      c2 = mother_born_UK
gen      c3 = parity
gen      c4 = mat_age_cat
gen      c5 = mat_asthma
gen      c6 = area_groups

misstable summarize
misstable pattern id x y m c1 c2 c3 c4 c5 c6

*analyses based on complete records
gen      to_keep= x<. & y<.& m<. & c1<. & c2<. & c3<. & c4<. & c5<. & c6<.
ta      to_keep
keep     if to_keep==1
keep     id y x m c*
su      m

label    define x 1 "Low depriv" 2 "Medium depriv" 3 "High depriv"
label    value x x

```

```
*****
```

1) IPW of MSMs method

```
***firstly calculate baseline risk using IPW without BSs ***
**Step 1. model X
mlogit      x, rrr
predict     num_1 num_2 num_3
```

```

mlogit      x i.c1 c2 i.c6, rrr
predict     den_1 den_2 den_3

gen         ipw1 = num_1/den_1 if x==1
replace    ipw1 = num_2/den_2 if x==2
replace    ipw1 = num_3/den_3 if x==3

su         ipw1
cap       drop num* den*

*model for Y (weighted) giving risk differences
binreg    y i.x [pw=ipw],rd

// _b[_cons] = baseline risk

***** run program for other estimates *****

cap       program drop CDM_ipw
program   CDM_ipw, rclass
          version 15
          args y x m c1 c2 c3 c4 c5 c6

cap       drop num* den*
cap       drop ipw*
cap       drop X_M*

**Step 1. model X
mlogit    `x', rrr
predict   num_1 num_2 num_3

mlogit    `x' i.`c1' `c2' i.`c6', rrr
predict   den_1 den_2 den_3

su        num* den*

gen       ipw1 = num_1/den_1 if `x'==1
replace   ipw1 = num_2/den_2 if `x'==2
replace   ipw1 = num_3/den_3 if `x'==3

su        ipw1
cap       drop num* den*

*****

* Marginal risk difference using IPW
*****

*model for Y (weighted) giving risk differences
binreg    `y' i.`x' [pw=ipw],rd

```

*** results

```
local      mrd_2=_b[2.`x']
scalar    mrd_2=`mrd_2'
return    scalar mrd_2=`mrd_2'
```

```
local      mrd_3=_b[3.`x']
scalar    mrd_3=`mrd_3'
return    scalar mrd_3=`mrd_3'
```

*** back to CDM ***

*Step 2. Model M

```
logit      `m', or
predict    num
logit      `m' `x' i.`c1' `c2' `c3' i.`c4' `c5' i.`c6', or
predict    den
```

```
su         num den
gen        ipw2 = cond(`m', num/den, (1-num)/(1-den))
su         ipw2
cap        drop num den
```

```
gen        ipw=ipw1*ipw2
su         ipw
```

*Step 3. model for Y (weighted) giving risk differences

```
gen        X_M_2=(`x'==2)*`m'
gen        X_M_3=(`x'==3)*`m'
binreg     `y' i.`x' `m' X_M_2 X_M_3 [pw=ipw],rd
```

*** results CDM

```
local      cdm_2=_b[2.`x']
scalar    cdm_2=`cdm_2'
return    scalar cdm_2=`cdm_2'
```

```
local      cdm_3=_b[3.`x']
scalar    cdm_3=`cdm_3'
return    scalar cdm_3=`cdm_3'
```

*** results %

```
local      per_2=(`mrd_2'-`cdm_2')/`mrd_2'*100
scalar    per_2=`per_2'
return    scalar per_2=`per_2'
```

```
local      per_3=(`mrd_3'-`cdm_3')/`mrd_3'*100
scalar    per_3=`per_3'
return    scalar per_3=`per_3'
```

end

*check that it works:

CDM_ipw y x m i.c1 c2 c3 i.c4 c5 i.c6

*run with 1000 bootstraps

```
bootstrap MRD_2=r(mrd_2) MRD_3=r(mrd_3) ///
           CDM_2=r(cdm_2) CDM_3=r(cdm_3) ///
           PER_2=r(per_2) PER_3= r(per_3), reps(1000) seed(1234): ///
           CDM_ipw y x m i.c1 c2 c3 i.c4 c5 i.c6
```

ex

2) G-estimation of structural nested models method

```
cap program drop CDM_snm
program CDM_snm, rclass
    version 15
    args y x m c1 c2 c3 c4 c5 c6
```

```
cap drop m_hat
cap drop r_M
cap drop X_r_M
cap drop X2_r_M X3_r_M
cap drop Y_tilda2
cap drop x_hat
cap drop mrd2 mrd3
cap drop r_X
cap drop r_X2 r_X3
cap drop y1 y2 y3
cap drop x1_hat x2_hat x3_hat
```

* Marginal risk difference

```
logit `y' i.`c1' `c2' i.`c6' if `x'==1
predict y1
```

```
logit `y' i.`c1' `c2' i.`c6' if `x'==2
predict y2
```

```
logit `y' i.`c1' `c2' i.`c6' if `x'==3
predict y3
```

```
sum y1, meanonly
local meany1=r(mean)
```

```

sum          y2, meanonly
local        meany2=r(mean)
sum          y3, meanonly
local        meany3=r(mean)

local        risk1=`meany1'
local        mrd2=`meany2'-`meany1'
local        mrd3=`meany3'-`meany1'
scalar       risk1=`risk1'
scalar       mrd2=`mrd2'
scalar       mrd3=`mrd3'
return       scalar risk1=`risk1'
return       scalar mrd2=`mrd2'
return       scalar mrd3=`mrd3'
scalar       list risk1 mrd2 mrd3

*** back to CDM ***

**Step 1a. predict M probabilities
logit        `m' i.`x' i.`c1' `c2' `c3' i.`c4' `c5' i.`c6', or
predict      m_hat

*Step 1b. Create M residuals
gen          r_M=`m'-m_hat

**Step 2. model for Y
gen          X2_r_M=(`x'==2)*r_M
gen          X3_r_M=(`x'==3)*r_M
regress      `y' r_M X2_r_M X3_r_M i.`c1' `c2' `c3' i.`c4' `c5' i.`c6'
scalar       gamma2=_b[r_M]
scalar       gamma3_2=_b[X2_r_M]
scalar       gamma3_3=_b[X3_r_M]

**Step 3.intermediate transformation
gen          Y_tilda2=`y'-gamma2*`m'-gamma3_2*(`x'==2)*`m'-gamma3_3*(`x'==3)*`m'

**Step 4a. predict X probabilities
mlogit       `x' i.`c1' `c2' i.`c6', rrr
predict      x1_hat x2_hat x3_hat

*Step 4b. Create X residuals
gen          r_X2=(`x'==2)-x2_hat
gen          r_X3=(`x'==3)-x3_hat

**Step 5. Regress transformed outcome
regress      Y_tilda2 r_X2 r_X3 i.`c1' `c2' i.`c6'

*** results CDM

```

```
local      cdm2=_b[r_X2]
local      cdm3=_b[r_X3]
scalar     cdm2=`cdm2'
scalar     cdm3=`cdm3'
return     scalar cdm2=`cdm2'
return     scalar cdm3=`cdm3'
scalar     list cdm2 cdm3

*** results %
local      per2=(`mrd2'-`cdm2')^mrd2*100
scalar     per2=`per2'
return     scalar per2=`per2'

local      per3=(`mrd3'-`cdm3')^mrd3*100
scalar     per3=`per3'
return     scalar per3=`per3'

end

*****
*check that it works:
CDM_snm    y x m i.c1 c2 c3 i.c4 c5 i.c6

*run with 1000 bootstraps
bootstrap  RISK_1= r(risk1) MRD_2=r(mrd2) MRD_3=r(mrd3) ///
           CDM_2=r(cdm2) CDM_3=r(cdm3) ///
           PER_2=r(per2) PER_3=r(per3), reps(1000) seed(1234): ///
           CDM_snm y x m i.c1 c2 c3 i.c4 c5 i.c6

ex
*****
```

APPENDIX - CHAPTER 7

Appendix 7.1. Defining chronic conditions

Table A7.1. ICD-10 codes defining chronic conditions (from Kristensen)¹⁶³

Group	ICD-10 code	Condition/s
Malformations of the respiratory system		
	Q30.0–Q30.9	Other malformations of the respiratory system
	Q31.0–Q31.9	Malformations of the larynx; trachea and bronchi
	Q32.0–Q32.4	
	Q33.0–Q33.9	Malformations of the lungs
	Q34.0–Q34.9	
	Q35.0–Q37.9	Cleft lip and palate
Other conditions associated with respiratory symptom		
	E84.0–E84.9	Cystic fibrosis
	P27.0–P27.9	Bronchopulmonary dysplasia
	Q39.0–Q39.1	Esophageal atresia
	Q39.2–Q39.9	Other malformations of the esophagus
	Q79.0	Congenital diaphragmatic hernia
Neuromuscular disease		
	G60.0–G60.9	
	G70.2	
	G80.0–G80.9	Cerebral palsy
	P94.1–P94.9	Congenital disturbances of muscle tonus, peripheral nerve disease, congenital myasthenia
	Q01.0–Q01.9	Encephalocele
	Q02.9	Microcephalus
	Q03.0–Q03.9	Congenital hydrocephalus
	Q04.0–Q04.9	Other cerebral malformations
	Q05.0–Q05.9	Spina bifida and malformations of the spinal cord
	Q06.0–Q06.9	
	Q07.0–Q07.9	Other malformations in the nervous system
	G12.0–G12.9	Spinal muscular atrophy
	G71.0–G71.3	Muscular dystrophy
Congenital diseases of the heart and the urinary system, chromosomal abnormalities, and others		
	D80.0–D82.9	Congenital immunodeficiencies
	E70.0–E73.0	Inborn errors of metabolism
	E74.0–E83.9	
	I27.0	
	N07.0–N07.9	Conditions affecting renal function
	N13.0–N13.9	Malformations of the urinary system
	N25.0–N25.9	
	P35.0–P35.9	Congenital viral infection
	Q20.0–Q26.9	Congenital heart disease

Q40.0–Q43.9	Malformations of the gastrointestinal tract, liver, biliary system, pancreas, and the abdominal wall
Q44.0–Q44.7	
Q45.0–Q45.3	
Q60.0–Q64.9	
Q79.2–Q79.3	
Q86.0	Fetal alcohol syndrome
Q90.0–Q90.9	Down syndrome
Q91.0–Q99.9	Other chromosomal abnormalities

Appendix 7.2. Example Mplus code

Deriving asthma/wheeze trajectory groups

```
! CODE FOR LCGA 4 CLASSES CUBIC MODEL
! TERMS TO INCLUDE FOR GMM ALSO SHOWN
! MANUAL 3 STEP APPROACH
```

```
TITLE: Y on LCGA 4 CLASS ASTHMA/WHEEZE: AW2-AW9
```

```
!AWx = asthma/wheeze at age x
```

```
DATA: FILE IS "file_name.dat";
      Format is free;
```

```
VARIABLES:
```

```
  NAMES = AW2-AW9 ID ;
  USEV = AW2-AW9;
  MISSING are all (-9999);
  IDVARIABLE = CHILDDID;
  CLASSES = C(4)
  CATEGORICAL = AW2-AW9;
```

```
ANALYSIS:
```

```
  TYPE = MIXTURE;
  ALGORITHM = INTEGRATION;
  STARTS = 500 100;
  STITERATIONS = 10;
  PROCESS = 10(STARTS);
```

```
MODEL:
```

```
! This describes the model to be estimated
```

```
! i s q cu are names for intercept, slope, quadratic and cubic terms, respectively
```

```
  %overall%
  i s q cu | aw2@0 aw3@1 aw4@2 aw5@3 aw6@4
            aw7@5 aw8@6 aw9@7;
  i-cu @ 0;
```

```
! Delete "@ 0" to allow within class variation (GMM model)
```

```
! This allows for the additional terms needed for GMM models
```

```
OUTPUT      : TECH1 TECH8 TECH11 TECH 14 CINTERVAL;
```

```
! TECH1 produces starting values for all free parameters in the model
```

Add details from best fitting model into the 11 imputed datasets

```
! CODE FOR adding LCGA details into imputation datasets
! Repeat on each imputed dataset
! use parameters from output of above code to replicate the same LCGA output in each
dataset
```

```
TITLE: Adding LCGA 4 CLASS ASTHMA/WHEEZE
```

```
DATA: FILE IS
```

```
"\imputed_dataset_1.dat";
    Format is free;
```

```
VARIABLES:
```

```
    NAMES = AW2-AW9 ID SGA area_groups asthma_mother season
           birthyear bronch chronic_cond delivery_meth deprivation3 mat_age
           mother_born _uk preterm sex siblings smoke_preg;
    USEV = AW2-AW9;
    MISSING are all (-9999);
    IDVARIABLE = ID;
    CLASSES = C(4)
    CATEGORICAL = AW2-AW9;
    AUXILIARY = SGA area_groups asthma_mother season
               birthyear bronch chronic_cond delivery_meth deprivation3 mat_age
               mother_born _uk preterm sex siblings smoke_preg;
```

```
ANALYSIS:
```

```
    TYPE = MIXTURE;
    ALGORITHM = INTEGRATION;
    STARTS = 0;
    !OPTSEED = 2341;
```

```
MODEL:
```

```
    %overall%
    i s q cu | aw2@0 aw3@1 aw4@2 aw5@3 aw6@4
              aw7@5 aw8@6 aw9@7;
```

```
! Class-specific means
```

```
    [c#1*-3.931];
    [c#2*-3.745];
    [c#3*-3.584];
```

```
! Class-specific growth factors from the final LCGA model
```

```
    %C#1%
    [i*4.983];
    [s*1.635];
    [q*-0.269];
    [cu*0.010];
    i-cu@0;
```

```
%C#2%
[i*1.380];
[s*0.855];
[q*0.079];
[cu*-0.016];
i-cu@0;
```

```
%C#3%
[i*4.935];
[s*1.047];
[q*-0.446];
[cu*0.036];
i-cu@0;
```

```
%C#4%
[i*0];
[s*-0.210];
[q*-0.058];
[cu*0.013];
i-cu@0;
```

```
OUPUT      : FILE is \LCGA_step1_imp1.dat;
             save is Cprob;
```

Multinomial logistic regression

```
! CODE FOR multinomial logistic regression looking at risk factors by trajectory group
! example of SIMD group as exposure
! uses all imputed datasets and Rubin's rules applied
```

```
TITLE: Multinomial logistic regression
```

```
DATA: FILE IS
```

```
"\imputed_dataset_list.dat";
      Format is free;
      Type = imputation;
```

```
VARIABLES:
```

```
NAMES = AW2-AW9 SGA area_groups asthma_mother season
         birthyear bronch chronic_cond delivery_meth deprivation3 mat_age
         mother_born _uk preterm sex siblings smoke_preg Cprobs1-Cprobs4 N ID;
```

```
USEV = N SIMD1 SIMD2 SIMD3;
```

```
IDVARIABLE = ID;
CLASSES = C(4)
NOMINAL = N;
```

DEFINE:

```
SIMD1=0;
SIMD2=0;
SIMD3=0;
If (deprivation3 eq 3) then SIMD1=1;
If (deprivation3 eq 2) then SIMD2=1
If (deprivation3 eq 1) then SIMD3=1
```

ANALYSIS:

```
TYPE = MIXTURE;
STARTS = 0;
```

MODEL:

```
%overall%
C ON SIMD1 SIMD2 SIMD3;
```

```
! including logit classification probabilities from LCGA model
! read table output by row
```

```
%C#1%
[N#1@5.186];
[N#2@2.229];
[N#3@1.917];
```

```
%C#2%
[N#1@-1.113];
[N#2@2.068];
[N#3@-0.542];
```

```
%C#3%
[N#1@-1.520];
[N#2@-1.546];
[N#3@1.525];
```

```
%C#4%
[N#1@-13.815];
[N#2@8.62];
[N#3@-7.754];
```

```
OUTPUT : CINTERVAL;
```

Appendix 7.3. Stata code for mediation analysis

Set up file and run analyses in one

```

*****
* File for:
* 1) overall % chronic asthma attributable to bronchiolitis
* 2) IPW of MSMs – for categorical X
* begins with imputation
*X: deprivation3
*M: bronch
*Y: binary_asthma
*M-Y counfounders: area_groups, birthyear_child, mother_born_UK
*X-Y confounders: area_groups, birthyear_child, mother_born_UK, birth_season, sex,
mat_age_cat, preterm, smoke_preg, chronic_conditions, SGA, asthma_mother, siblings
*keep birthweight, gestation, firstborn to calculate SGA post imputation
*also keep lbw, mat_age, socio_economic_group, intensive_flag, preterm_conditions,
hospital_code
*****
use          "$temp\CDMasthma_dataset.dta", clear

rename      deprivation3 x
rename      bronch m
generate    y=0
replace     y=1 if class==1 | class==2

rename      area_groups c1
gen         c2=0
replace     c2=1 if birthyear_child==2008
rename      mother_born_UK c3
rename      birth_season c4
rename      sex c5
rename      mat_age_cat c6 // not in imputation model
rename      preterm c7 // not in imputation model
rename      smoke_preg c8
rename      chronic_condition c9
rename      SGA c10 // not in imputation model
rename      asthma_mother c11
rename      siblings_cat c12 // not in imputation model
rename      delivery_method_3cat c13
replace     c13=c13-1 //so categorised 0 1 2

rename      preterm_conditions iv1
rename      birthweight iv2
rename      gestation iv3
rename      mat_age iv4
rename      intensive_flag iv5

```

```

rename          socio_economic_group iv6
rename          hosp_G iv7
rename         siblings iv8

```

```
*****
```

```

cap program drop CDM_asthma
program define CDM_asthma, rclass

```

```
*****
```

```
*SINGLE IMPUTATION
```

```
*for missing data
```

```
*FOR iv2 = birthweight (to create SGA c10) regress
```

```
*FOR iv3 = gestation (to create preterm c7) regress
```

```
*FOR iv4 = mat_age (to create mat age cats c13) regress
```

```
*FOR iv8 = siblings (to create categorical siblings c12 and SGA c10 too) mlogit
```

```
*FOR c13 = delivery_method mlogit
```

```
*FOR c8 = smoke_preg logit
```

```
*FOR c11 = asthma_mother logit
```

```
*****
```

```
***generate imputation vars
```

```

qui foreach var in x m y c1 c2 c3 c4 c5 c6 c7 ///
    c8 c9 c10 c11 c12 c13 ///
    iv1 iv2 iv3 iv4 iv5 iv6 iv7 iv8 {
    gen imp_`var'=`var'
    }

```

```
***replace missingness in vars with average values
```

```
*binary vars with missing values
```

```

qui foreach var in c8 c11 {
    summ `var'
    replace imp_`var'=runiform() $<$ r(mean) if imp_`var'==.
}

```

```
*continuous vars with missing values
```

```

qui foreach var in iv2 iv3 iv4 iv8 {
    summ `var'
    replace imp_`var'=r(mean) if imp_`var'==.
}

```

```
*categorical var with missing values
```

```

qui ta          c13,m
qui gen         imp_c13_0=(c13==0)
qui gen         imp_c13_2=(c13==2)
qui replace    imp_c13_0=. if c13==.
qui replace    imp_c13_2=. if c13==.

```

```

qui replace    imp_c13_0=runiform()<0.05 if imp_c13_0==.
qui replace    imp_c13_2=0 if imp_c13_0==1
qui replace    imp_c13_2=runiform()>0.81 if imp_c13_2==.
qui replace    imp_c13=imp_c13_0+imp_c13_2*2 if imp_c13==.

```

```

qui          su imp_*

```

```

*Need to do a lead in to imputation to stop convergence
forvalues cycle=1(1)10 {

```

```

***impute binary vars first

```

```

*for smoke_preg c8

```

```

qui logit      c8 i.imp_x imp_m imp_y i.imp_c1 imp_c3 i.imp_c4 ///
               imp_c5 imp_c9 imp_c11 i.imp_c13 ///
               imp_iv1 imp_iv2 imp_iv3 imp_iv4 imp_iv5 imp_iv6 imp_iv7 imp_iv8

```

```

qui predict    ec8

```

```

qui replace    imp_c8=runiform()<ec8 if c8==.

```

```

qui tab        imp_c8

```

```

qui drop       ec8

```

```

*for asthma_mother c11

```

```

qui logit      c11 i.imp_x imp_m imp_y i.imp_c1 imp_c3 i.imp_c4 imp_c5 ///
               imp_c8 imp_c9 i.imp_c13 ///
               imp_iv1 imp_iv2 imp_iv3 imp_iv4 imp_iv5 imp_iv6 imp_iv7 imp_iv8

```

```

qui predict    ec11

```

```

qui replace    imp_c11=runiform()<ec11 if c11==.

```

```

qui tab        imp_c11

```

```

qui drop       ec11

```

```

***impute continuous vars

```

```

*for birthweight iv2

```

```

qui regress    iv2 i.imp_x imp_m imp_y i.imp_c1 imp_c3 i.imp_c4 imp_c5 ///
               imp_c8 imp_c9 imp_c11 i.imp_c13 ///
               imp_iv1 imp_iv3 imp_iv4 imp_iv5 imp_iv6 imp_iv7 imp_iv8

```

```

qui predict    eiv2

```

```

qui replace    imp_iv2=eiv2+e(rmse)*rnormal() if iv2==.

```

```

qui drop       eiv2

```

```

*for gestation iv3

```

```

qui regress    iv3 i.imp_x imp_m imp_y i.imp_c1 imp_c3 i.imp_c4 imp_c5 ///
               imp_c8 imp_c9 imp_c11 i.imp_c13 ///
               imp_iv1 imp_iv2 imp_iv4 imp_iv5 imp_iv6 imp_iv7 imp_iv8

```

```

qui predict    eiv3

```

```

qui replace    imp_iv3=eiv3+e(rmse)*rnormal() if iv2==.

```



```

qui drop      eiv3

*for mat_age iv4
qui regress   iv4 i.imp_x imp_m imp_y i.imp_c1 imp_c3 i.imp_c4 imp_c5 ///
              imp_c8 imp_c9 imp_c11 i.imp_c13 ///
              imp_iv1 imp_iv2 imp_iv3 imp_iv5 imp_iv6 imp_iv7 imp_iv8

qui predict   eiv4
qui replace   imp_iv4=eiv4+e(rmse)*rnormal() if iv4==.
qui drop      eiv4

*for siblings iv8
qui regress   iv8 i.imp_x imp_m imp_y i.imp_c1 imp_c3 i.imp_c4 imp_c5 ///
              imp_c8 imp_c9 imp_c11 i.imp_c13 ///
              imp_iv1 imp_iv2 imp_iv3 imp_iv4 imp_iv5 imp_iv6 imp_iv7

qui predict   eiv8
qui replace   imp_iv8=eiv8+e(rmse)*rnormal() if iv8==.
qui drop      eiv8

***impute categorical var
*for delivery_method c13
qui mlogit    c13 i.imp_x imp_m imp_y i.imp_c1 imp_c3 i.imp_c4 imp_c5 ///
              imp_c8 imp_c9 imp_c11 ///
              imp_iv1 imp_iv2 imp_iv3 imp_iv4 imp_iv5 imp_iv6 imp_iv7 imp_iv8

qui predict   ec13_0 ec13_1 ec13_2
qui replace   imp_c13_0=runiform()<ec13_0 if c13==.
qui replace   imp_c13_2=0 if imp_c13_0==1
qui replace   imp_c13_2=runiform()>(1-ec13_2) if c13==.
qui replace   imp_c13=(imp_c13_0==1)*0+(imp_c13_0==0)*(imp_c13_2==0)*1 ///
              + (imp_c13_2==1)*2 if c13==.

qui tab       imp_c13
qui drop      ec13*
              }

*keep original vars only
qui foreach var of varlist iv2 iv3 iv4 iv8 c8 c11 c13 {
    replace   `var'=imp_`var' if `var'==.
}

qui drop      imp_*

***create the new vars needed for analysis
*preterm c7 from gestation iv3
qui replace   c7=0 if iv3<37 & c7==.
qui replace   c7=1 if iv3>=37 & c7==.

*mat age cats c6 from mat_age iv4

```

```

qui replace    c6=0 if iv4<=19 & c6==.
qui replace    c6=1 if (iv4>19 & iv4<30) & c6==.
qui replace    c6=2 if (iv4>29 & iv4<40) & c6==.
qui replace    c6=3 if iv4>=40 & c6==.

*categorical siblings c12 from siblings iv8
qui replace    c12=0 if (round(iv8)==0 | iv8<0) & c12==.
qui replace    c12=1 if round(iv8)==1 & c12==.
qui replace    c12=2 if round(iv8)>1 & c12==.

*SGA c10 from birthweight iv2 gestation iv3 and siblings c12
qui generate   firstborn=1 if c12==0
qui replace    firstborn=0 if c12==1 | c12==2
qui generate   gestation=round(iv3)
qui replace    gestation=24 if gestation==23
qui replace    gestation=43 if gestation>43
qui generate   sex=c5
qui merge      m:m sex gestation using "$temp\SGA.dta"
qui replace    c10=0 if iv2>tenth_centile & c10==.
qui replace    c10=0 if iv2<=tenth_centile & c10==.
qui drop       _merge tenth_centile gestation firstborn sex

*****
* CDM analysis
*****

cap            drop num*
cap            drop den*
cap            drop ipw*
cap            drop X_M*

*1. Model x
qui mlogit     x, base(1) rrr
qui predict    num_1 num_2 num_3

qui mlogit     x i.c1 c2 c3, base(1) rrr
qui predict    den_1 den_2 den_3

qui gen        ipw1 = num_1/den_1 if x==1
               qui replace ipw1 = num_2/den_2 if x==2
               qui replace ipw1 = num_3/den_3 if x==3

*model for asthma_binary (weighted) giving risk differences for TCE
qui binreg     y i.x [pw=ipw],rd

local          tce_2=_b[2.x]
scalar         tce_2=`tce_2'
return         scalar tce_2=`tce_2'

```

```

local          tce_3=_b[3.x]
scalar        tce_3=`tce_3'
return        scalar tce_3=`tce_3'

qui cap              drop num* den*

*Step 2. Model M
qui logit           m, or
qui predict         num

qui logit           m i.x i.c1 c2 c3 i.c4 c5 i.c6 c7 c8 c9 c10 c11 i.c12 i.c13

qui predict         den

qui gen             ipw2 = cond(m, num/den, (1-num)/(1-den))

qui cap            drop num* den*
qui cap            drop ipw
qui gen            ipw=ipw1*ipw2

*Step 3. model for Y (weighted) giving risk differences
qui gen            X_M_2=(x==2)*m
qui gen            X_M_3=(x==3)*m
qui binreg         y i.x X_M_2 X_M_3 [pw=ipw],rd

*** Counterfactual Disparity Measures
local            cdm_2=_b[2.x]
scalar          cdm_2=`cdm_2'
return          scalar cdm_2=`cdm_2'

local            cdm_3=_b[3.x]
scalar          cdm_3=`cdm_3'
return          scalar cdm_3=`cdm_3'

local            risk_1=_b[_cons]
scalar          risk_1=`risk_1'
return          scalar risk_1=`risk_1'

*** results %
local            per_2=(`tce_2'-`cdm_2')/tce_2*100
scalar          per_2=`per_2'
return          scalar per_2=`per_2'

local            per_3=(`tce_3'-`cdm_3')/tce_3*100
scalar          per_3=`per_3'
return          scalar per_3=`per_3'

end

```

*check that it works:

CDM_asthma y x m i.c1 c2 c3 i.c4 c5 i.c6 c7 c8 c9 c10 c11 i.c12 i.c13

*Bootstrap

bootstrap TCE_2=r(tce_2) TCE_3=r(tce_3) RISK_1=r(risk_1) ///
CDM_2=r(cdm_2) CDM_3=r(cdm_3) ///
PER_2=r(per_2) PER_3= r(per_3) ///
, reps(1000) seed(1234): CDM_asthma

Appendix 7.4. Risk factor distribution by asthma/wheeze trajectories

Table A7.2. Risk factor distribution by asthma/wheeze trajectories (groups based on most likely class membership)

		Never or infrequent (N=78,803)	Early- Transient (N=1,853)	Early- Persistent (N=1,547)	Intermediate- onset (N=1,645)
SIMD	Low	26.6%	20.5%	19.8%	21.4%
	Medium	37.9%	36.8%	37.0%	37.9%
	High	35.5%	42.7%	43.2%	40.7%
≥1 bronchiolitis admission	No	97.2%	89.2%	91.0%	94.5%
	Yes	2.8%	10.8%	9.0%	5.5%
Birth season	Spring	33.4%	30.5%	34.1%	35.3%
	Summer	22.8%	23.8%	23.7%	22.7%
	Autumn	16.8%	17.1%	15.4%	15.9%
	Winter	27.0%	28.7%	26.8%	26.1%
Sex	Female	49.5%	39.2%	37.4%	40.2%
	Male	50.5%	60.8%	62.6%	59.8%
Area	Northern	13.0%	12.0%	12.7%	12.6%
	Central	28.5%	28.4%	26.0%	25.5%
	Glasgow	23.1%	25.4%	24.9%	26.1%
	Edinburgh	17.1%	13.6%	14.1%	16.5%
	Southern	18.2%	20.6%	22.4%	19.2%
Maternal birth country	Non-UK	10.9%	9.1%	8.7%	8.9%
	UK	89.1%	90.9%	91.3%	91.1%
Maternal age	<20	7.1%	8.0%	7.6%	9.6%
	20-29	42.8%	49.4%	47.1%	45.5%
	30-39	39.7%	34.2%	35.9%	35.2%
	40+	5.2%	4.1%	4.5%	4.9%
	Missing	5.2%	4.3%	5.0%	4.8%
Asthma	No	92.6%	91.3%	88.2%	89.5%
	Yes	2.2%	4.4%	6.9%	5.7%
	Missing	5.2%	4.3%	5.0%	4.8%
Smoking	No	70.7%	67.4%	64.6%	67.2%
	Yes	19.6%	23.1%	24.8%	22.6%
	Missing	9.7%	9.5%	10.5%	10.2%
Siblings	0 siblings	44.2%	47.1%	44.5%	46.6%
	1 sibling	31.8%	32.1%	31.9%	31.4%
	2+ siblings	18.4%	15.9%	18.4%	16.5%
	Missing	5.6%	4.8%	5.2%	5.4%
Chronic condition	No	96.2%	93.5%	92.9%	94.4%
	Yes	3.9%	6.5%	7.1%	5.6%
Delivery method	Vaginal	80.0%	69.8%	67.6%	69.8%
	Elective	9.8%	9.6%	10.2%	10.2%
	Emergency	14.1%	16.3%	17.3%	15.3%
	Missing	5.2%	4.3%	5.0%	4.8%
Preterm	No	88.8%	85.2%	83.4%	87.6%

	Yes	5.9%	10.5%	11.6%	7.5%
	Missing	5.3%	4.3%	5.0%	4.9%
SGA	No	85.4%	85.5%	83.2%	84.7%
	Yes	8.8%	9.7%	11.2%	9.8%
	Missing	5.8%	4.8%	5.6%	5.5%

BIBLIOGRAPHY

1. Smyth, R. L. & Openshaw, P. J. M. Bronchiolitis. *The Lancet* **368**, 312–322 (2006).
2. Borchers, A. T., Chang, C., Gershwin, M. E. & Gershwin, L. J. Respiratory syncytial virus--a comprehensive review. *Clin Rev Allergy Immunol* **45**, 331–379 (2013).
3. NICE. Bronchiolitis in children: diagnosis and management. <https://www.nice.org.uk/guidance/ng9/chapter/1-Recommendations#assessment-and-diagnosis> (2015).
4. Green, C. A. *et al.* Admission to hospital for bronchiolitis in England: trends over five decades, geographical variation and association with perinatal characteristics and subsequent asthma. *Archives of Disease in Childhood* **101**, 140–146 (2016).
5. Ricci, V., Delgado Nunes, V., Murphy, M. S. & Cunningham, S. Bronchiolitis in children: summary of NICE guidance. *BMJ : British Medical Journal* **350**, (2015).
6. Reeves, R. M. *et al.* Estimating the burden of respiratory syncytial virus (RSV) on respiratory hospital admissions in children less than five years of age in England, 2007-2012. *Influenza and Other Respiratory Viruses* **11**, 122–129 (2017).
7. Kutter, J. S., Spronken, M. I., Fraaij, P. L., Fouchier, R. A. M. & Herfst, S. Transmission routes of respiratory viruses among humans. *Current Opinion in Virology* **28**, 142–151 (2018).
8. Kulkarni, H. *et al.* Evidence of Respiratory Syncytial Virus Spread by Aerosol. Time to Revisit Infection Control Strategies? *Am J Respir Crit Care Med* **194**, 308–316 (2016).
9. ONS. *Mortality statistics - underlying cause, sex and age.* <https://www.nomisweb.co.uk/datasets/mortsa> (2020).
10. National Records of Scotland. *Vital Event Reference Tables 2019.* <https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/general-publications/vital-events-reference-tables/2019/list-of-data-tables#section7> (2020).
11. NHS Digital. *Hospital Admitted Patient Care Activity, 2018-19: Diagnosis.* <https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/2018-19> (2019).
12. Piché-Renaud, P.-P. *et al.* Parents' perspectives, information needs and healthcare preferences when consulting for their children with bronchiolitis: A qualitative study. *Acta Paediatrica* **110**, 944–951 (2021).
13. Campbell, A., Hartling, L., Louie-Poon, S. & Scott, S. D. Parent Experiences Caring for a Child With Bronchiolitis: A Qualitative Study. *Journal of Patient Experience* **7**, 1362–1368 (2020).
14. Lapillonne, A. *et al.* Impact on parents of bronchiolitis hospitalization of full-term, preterm and congenital heart disease infants. *BMC Pediatr* **12**, 171–171 (2012).
15. Cheung, C. R. & Semple, M. G. Stemming the tide of hospital admissions for bronchiolitis. *Arch Dis Child* **101**, 118 (2016).
16. Taylor, S. *et al.* Modelling estimates of the burden of respiratory syncytial virus infection in children in the UK. *BMJ Open* **6**, (2016).

17. NICE. Asthma: diagnosis, monitoring and chronic asthma management. <https://www.nice.org.uk/guidance/ng80/informationforpublic> (2017).
18. Vos, T. *et al.* Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* **396**, 1204–1222 (2020).
19. Institute for Health Metrics and Evaluation. *GBD Results Tool*. <http://ghdx.healthdata.org/gbd-results-tool> (2020).
20. Gibson GJ, Loddenkemper R, Sibille Y & Lundbäck B. *The European Lung White Book*. (European Respiratory Society, 2013).
21. Panico, L., Stuart, B., Bartley, M. & Kelly, Y. Asthma Trajectories in Early Childhood: Identifying Modifiable Factors. *PLOS ONE* **9**, e111922 (2014).
22. Trivedi, M. & Denton, E. Asthma in Children and Adults-What Are the Differences and What Can They Tell us About Asthma? *Front Pediatr* **7**, 256–256 (2019).
23. Dharmage, S. C., Perret, J. L. & Custovic, A. Epidemiology of Asthma in Children and Adults. *Front Pediatr* **7**, 246–246 (2019).
24. Sbihi, H., Koehoorn, M., Tamburic, L. & Brauer, M. Asthma Trajectories in a Population-based Birth Cohort. Impacts of Air Pollution and Greenness. *American Journal of Respiratory and Critical Care Medicine* **195**, 607–613 (2017).
25. Arathimos, R., Granel, R., Henderson, J., Relton, C. L. & Tilling, K. Sex discordance in asthma and wheeze prevalence in two longitudinal cohorts. *PLOS ONE* **12**, e0176293 (2017).
26. Wu, T.-J., Pan, S.-C., Chen, B.-Y., Chin, W.-S. & Guo, Y. L. Different seasonal effect on asthma trajectories: A population-based birth cohort study. *Pediatric Allergy and Immunology* **29**, 873–877 (2018).
27. De Meer, G., Reijneveld, S. A. & Brunekreef, B. Wheeze in children: the impact of parental education on atopic and non-atopic symptoms. *Pediatric Allergy and Immunology* **21**, 823–830 (2010).
28. Garcia-Marcos, L. *et al.* A different pattern of risk factors for atopic and non-atopic wheezing in 9-12-year-old children. *Pediatr Allergy Immunol* **16**, (2005).
29. Galobardes, B. *et al.* Childhood wheezing, asthma, allergy, atopy, and lung function: different socioeconomic patterns for different phenotypes. *Am J Epidemiol* **182**, 763–774 (2015).
30. Cosio, B. G., Rosado, J. R. & Rossi, F. F. Chapter 38 - Asthma: Epidemiology, Pathophysiology, and Risk Factors. in *Clinical Respiratory Medicine (Fourth Edition)* (eds. Spiro, S. G., Silvestri, G. A. & Agustí, A.) 487–500 (W.B. Saunders, 2012).
31. Moraes, T. J., Sears, M. R. & Subbarao, P. Epidemiology of Asthma and Influence of Ethnicity. *Seminars in respiratory and critical care medicine* **39**, 3–11 (2018).
32. Mukherjee, M. *et al.* The epidemiology, healthcare and societal burden and costs of asthma in the UK and its member nations: analyses of standalone and linked national databases. *BMC Medicine* **14**, 113 (2016).
33. Royal College of Physicians. *Why asthma still kills: The National Review of Asthma Deaths (NRAD)*. <https://www.asthma.org.uk/293597ee/globalassets/campaigns/nrad-full-report.pdf> (2014).

34. Shah, R., Hagell, A. & Cheung, R. *International comparisons of health and wellbeing in adolescence and early adulthood*. https://www.nuffieldtrust.org.uk/files/2019-02/1550657729_nt-ayph-adolescent-health-report-web.pdf (2019).
35. Fleming, M. *et al.* Educational and health outcomes of children treated for asthma: Scotland-wide record linkage study of 683716 children. *Eur Respir J* **54**, 1802309 (2019).
36. Dudeney, J., Sharpe, L., Jaffe, A., Jones, E. B. & Hunt, C. Anxiety in youth with asthma: A meta-analysis. *Pediatric Pulmonology* **52**, 1121–1129 (2017).
37. Calam, R., Gregg, L. & Goodman, R. Psychological Adjustment and Asthma in Children and Adolescents: The UK Nationwide Mental Health Survey. *Psychosomatic Medicine* **67**, (2005).
38. Silva, N., Carona, C., Crespo, C. & Canavarro, M. C. Quality of life in pediatric asthma patients and their parents: a meta-analysis on 20 years of research. *null* **15**, 499–519 (2015).
39. Landau, L. I. Current pharmacological treatments for bronchiolitis are useless: The case for the con's. *Paediatric Respiratory Reviews* **7**, S101–S103 (2006).
40. NICE. Palivizumab. <https://bnfc.nice.org.uk/drug/palivizumab.html> (2019).
41. Public Health England. *Respiratory syncytial virus: the green book, chapter 27a*. <https://www.gov.uk/government/publications/respiratory-syncytial-virus-the-green-book-chapter-27a> (2015).
42. GlaxoSmithKline. GSK starts phase 3 study of RSV maternal candidate vaccine. <https://www.gsk.com/en-gb/media/press-releases/gsk-starts-phase-3-study-of-rsv-maternal-candidate-vaccine/> (2020).
43. Rossey, I. & Saelens, X. Vaccines against human respiratory syncytial virus in clinical trials, where are we now? *null* **18**, 1053–1067 (2019).
44. Hardelid, P., Verfuerden, M., McMenamin, J., Smyth, R. L. & Gilbert, R. The contribution of child, family and health service factors to respiratory syncytial virus (RSV) hospital admissions in the first 3 years of life: birth cohort study in Scotland, 2009 to 2015. *Euro Surveill* **24**, 1800046 (2019).
45. Murray, J. *et al.* Risk Factors for Hospital Admission with RSV Bronchiolitis in England: A Population-Based Birth Cohort Study. *PLOS ONE* **9**, e89186 (2014).
46. Ebmeier, S. *et al.* Trends in international asthma mortality: analysis of data from the WHO Mortality Database from 46 countries (1993–2012). *The Lancet* **390**, 935–945 (2017).
47. Pavord, I. D. *et al.* After asthma: redefining airways diseases. *The Lancet* **391**, 350–400 (2018).
48. Healthcare Improvement Scotland. *SIGN158 British guideline on the management of asthma: A national clinical guideline*. <https://www.sign.ac.uk/sign-158-british-guideline-on-the-management-of-asthma> (2019).
49. Bønnelykke, K., Vissing, N. H., Sevelsted, A., Johnston, S. L. & Bisgaard, H. Association between respiratory infections in early life and later asthma is independent of virus type. *Journal of Allergy and Clinical Immunology* **136**, 81-86.e4 (2015).
50. Moore, H. C., Hall, G. L. & de Klerk, N. Infant respiratory infections and later respiratory hospitalisation in childhood. *European Respiratory Journal* (2015) doi:10.1183/13993003.00587-2015.

51. Carroll, K. N. *et al.* The Severity-Dependent Relationship of Infant Bronchiolitis on the Risk and Morbidity of Early Childhood Asthma. *The Journal of allergy and clinical immunology* **123**, 1055-1061.e1 (2009).
52. Escobar, G. J., Masaquel, A. S., Li, S. X., Walsh, E. M. & Kipnis, P. Persistent recurring wheezing in the fifth year of life after laboratory-confirmed, medically attended respiratory syncytial virus infection in infancy. *BMC pediatrics* **13**, 97–97 (2013).
53. Wu, P. *et al.* Evidence of a Causal Role of Winter Virus Infection during Infancy in Early Childhood Asthma. *American Journal of Respiratory and Critical Care Medicine* **178**, 1123–1129 (2008).
54. Skirrow, H. *et al.* Preschool respiratory hospital admissions following infant bronchiolitis: a birth cohort study. *Archives of Disease in Childhood* archdischild-2018-316317 (2019) doi:10.1136/archdischild-2018-316317.
55. Kuehni, C. E., Spycher, B. D. & Silverman, M. Causal links between RSV infection and asthma – no clear answers to an old question. *Editorial for the American Journal of Respiratory and Critical Care Medicine* (2009).
56. Dupont, W., Wu, P., Gebretsadik, T. & Hartert, T. RSV prevention in infancy and asthma in later life. *The Lancet Respiratory Medicine* **6**, e32 (2018).
57. Carbonell-Estrany, X. RSV prevention in infancy and asthma in later life. *The Lancet Respiratory Medicine* **6**, e31 (2018).
58. Scheltema, N. M. *et al.* Respiratory syncytial virus prevention and asthma in healthy preterm infants: a randomised controlled trial. *The Lancet Respiratory Medicine* **6**, 257–264 (2018).
59. Embrett, M. G. & Randall, G. E. Social determinants of health and health equity policy research: Exploring the use, misuse, and nonuse of policy analysis theory. *Social Science & Medicine* **108**, 147–155 (2014).
60. WHO. *A conceptual framework for action on the social determinants of health.* (2010).
61. Kawachi, I., Subramanian, S. V. & Almeida-Filho, N. A glossary for health inequalities. *Journal of Epidemiology and Community Health* **56**, 647–652 (2002).
62. Galobardes, B., Shaw, M., Lawlor, D. A., Lynch, J. W. & Davey Smith, G. Indicators of socioeconomic position (part 1). *Journal of Epidemiology and Community Health* **60**, 7–12 (2006).
63. GOV.UK. English indices of deprivation 2019. <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019>.
64. Scottish Government. The Scottish Index of Multiple Deprivation. <https://www2.gov.scot/Topics/Statistics/SIMD> (2019).
65. Ministry of Housing, Communities & Local Government. The English Indices of Deprivation 2015: Research report. <https://www.gov.uk/government/publications/english-indices-of-deprivation-2015-research-report> (2015).
66. Yousaf, S. & Bonsall, A. *UK Townsend Deprivation Scores from 2011 census data.* <https://www.statistics.digitalresources.jisc.ac.uk/dataset/2011-uk-townsend-deprivation-scores/resource/f500b512-2f49-4e6f-8403-9417b9133f64> (2017).
67. Galobardes, B., Lynch, J. & Smith, G. D. Measuring socioeconomic position in health research. *British Medical Bulletin* **81–82**, 21–37 (2007).

68. Tagiyeva, N., Devereux, G., Fielding, S., Turner, S. & Douglas, G. Outcomes of Childhood Asthma and Wheezy Bronchitis. A 50-Year Cohort Study. *Am J Respir Crit Care Med* **193**, 23–30 (2015).
69. Gupta, R. P., Mukherjee, M., Sheikh, A. & Strachan, D. P. Persistent variations in national asthma mortality, hospital admissions and prevalence by socioeconomic status and region in England. *Thorax* **73**, 706 (2018).
70. Goodman, A., Wilkinson, P., Stafford, M. & Tonne, C. Characterising socio-economic inequalities in exposure to air pollution: A comparison of socio-economic markers and scales of measurement. *Health & Place* **17**, 767–774 (2011).
71. Deas, I., Robson, B., Wong, C. & Bradford, M. Measuring Neighbourhood Deprivation: A Critique of the Index of Multiple Deprivation. *Environ Plann C Gov Policy* **21**, 883–903 (2003).
72. Heymann, S. Health and Social Policy. in *Social Epidemiology* 368–382 (Oxford University Press, 2000).
73. Ruckert, A., Huynh, C. & Labonté, R. Reducing health inequities: is universal basic income the way forward? *Journal of Public Health* **40**, 3–7 (2018).
74. The Scottish Government. *Tackling inequalities in the early years: key messages from 10 years of the growing up in scotland study*. <https://www.gov.scot/publications/tackling-inequalities-early-years-key-messages-10-years-growing-up-scotland-study/pages/1/> (2015).
75. Marmot, M. *Fair society, healthy lives : the Marmot Review : strategic review of health inequalities in England post-2010*. <https://www.gov.uk/research-for-development-outputs/fair-society-healthy-lives-the-marmot-review-strategic-review-of-health-inequalities-in-england-post-2010> (2010).
76. Adams, J. & White, M. Removing the health domain from the Index of Multiple Deprivation 2004—effect on measured inequalities in census measure of health. *Journal of Public Health* **28**, 379–383 (2006).
77. Bloom-Feshbach, K. *et al.* Latitudinal Variations in Seasonal Activity of Influenza and Respiratory Syncytial Virus (RSV): A Global Comparative Review. *PLOS ONE* **8**, (2013).
78. Bartley, M. *Health inequality*. (Polity Press, 2004).
79. Lynch, J. & Davey Smith, G. A life course approach to chronic disease epidemiology. *Annual Review of Public Health* **26**, 1–35 (2005).
80. Hans Bisgaard, Signe M. Jensen & Klaus Bønnelykke. Interaction between Asthma and Lung Function Growth in Early Life. *American Journal of Respiratory and Critical Care Medicine* **185**, 1183–1189 (2012).
81. Stocks, J., Hislop, A. & Sonnappa, S. Early lung development: lifelong effect on respiratory health and disease. *The Lancet Respiratory Medicine* **1**, 728–742 (2013).
82. Moore, H. C. *et al.* Temporal trends and socioeconomic differences in acute respiratory infection hospitalisations in children: an intercountry comparison of birth cohort studies in Western Australia, England and Scotland. *BMJ Open* **9**, e028710 (2019).
83. University of Glasgow. CARSTAIRS SCORES. <https://www.gla.ac.uk/researchinstitutes/healthwellbeing/research/mrccsocialandpublichealthsciencesunit/programmes/inequalities/healthinequalities/determinantsofhealthandhealthinequalitiesnscotland/carstairsscores/>.
84. O'Donnell, D. R., Parslow, R. C. & Draper, E. S. Deprivation, ethnicity and prematurity in infant respiratory failure in PICU in the UK. *Acta Pædiatrica* **99**, 1186–1191 (2010).

85. Zheng, D. X. *et al.* Socioeconomic Status and Bronchiolitis Severity Among Hospitalized Infants. *Academic Pediatrics* (2019) doi:10.1016/j.acap.2019.06.003.
86. US Census Bureau. US Population By Zip Code: For both 2000 and 2010. https://www.kaggle.com/census/us-population-by-zip-code?select=population_by_zip_2010.csv (2017).
87. *National Health Service Act 2006*. vol. 41.
88. Lewis, K. M. *et al.* Mother's education and offspring asthma risk in 10 European cohort studies. *European Journal of Epidemiology* **32**, 797–805 (2017).
89. ScotCen Social Research. *Scottish Health Survey 2018: main report - revised 2020*. <https://www.gov.scot/publications/scottish-health-survey-2018-volume-1-main-report/pages/70/> (2020).
90. Ardura-Garcia, C., Stolbrink, M., Zaidi, S., Cooper, P. J. & Blakey, J. D. Predictors of repeated acute hospital attendance for asthma in children: A systematic review and meta-analysis. *Pediatr Pulmonol* **53**, 1179–1192 (2018).
91. Centre for Longitudinal Studies. Millennium Cohort Study. *Centre for Longitudinal Studies* <https://cls.ucl.ac.uk/cls-studies/millennium-cohort-study/>.
92. Uphoff, E. *et al.* A systematic review of socioeconomic position in relation to asthma and allergic diseases. *Eur Respir J* **46**, 364 (2015).
93. Taylor-Robinson, D. C., Pearce, A., Whitehead, M., Smyth, R. & Law, C. Social inequalities in wheezing in children: findings from the UK Millennium Cohort Study. *European Respiratory Journal* (2015) doi:10.1183/13993003.01117-2015.
94. Strachan, D. P. Hay fever, hygiene, and household size. *BMJ* **299**, 1259 (1989).
95. Litonjua, A. A. & Weiss, S. T. Is vitamin D deficiency to blame for the asthma epidemic? *Journal of Allergy and Clinical Immunology* **120**, 1031–1035 (2007).
96. Sbihi, H. *et al.* Thinking bigger: How early life environmental exposures shape the gut microbiome and influence the development of asthma and allergic disease. *Allergy* **0**, (2019).
97. Bhopal, R. Glossary of terms relating to ethnicity and race: for reflection and debate. *J Epidemiol Community Health* **58**, 441 (2004).
98. WHO Europe. *How health systems can address health inequities linked to migration and ethnicity*. (2010).
99. ONS. *Socioeconomic status*. <https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/socioeconomic-status/latest> (2018).
100. Krishna, M. T., Hackett, S., Bethune, C. & Fox, A. T. Achieving equitable management of allergic disorders and primary immunodeficiency in a Black, Asian and Minority Ethnic population. *Clinical & Experimental Allergy* **50**, 880–883 (2020).
101. Coleman, D. *Immigration, Population and Ethnicity: The UK in International Perspective*. <https://migrationobservatory.ox.ac.uk/resources/briefings/immigration-population-and-ethnicity-the-uk-in-international-perspective/> (2013).
102. Bjur, K. A. *et al.* Socioeconomic Status, Race/Ethnicity, and Health Disparities in Children and Adolescents in a Mixed Rural-Urban Community—Olmsted County, Minnesota. *Mayo Clinic Proceedings* **94**, 44–53 (2019).

103. Inagaki, K., Blackshear, C., Burns, P. A. & Hobbs, C. V. Racial/Ethnic Disparity in the Incidence of Bronchiolitis Requiring Hospitalization. *Clinical Infectious Diseases* (2020) doi:10.1093/cid/ciaa113.
104. Sheikh, A. *et al.* Ethnic variations in asthma hospital admission, readmission and death: a retrospective, national cohort study of 4.62 million people in Scotland. *BMC Medicine* **14**, 3 (2016).
105. Petherick, E. S., Pearce, N., Sunyer, J. & Wright, J. Ethnic and socio-economic differences in the prevalence of wheeze, severe wheeze, asthma, eczema and medication usage at 4 years of age: Findings from the Born in Bradford birth cohort. *Respiratory Medicine* **119**, 122–129.
106. Kuehni, C. E., Strippoli, M.-P. F., Low, N. & Silverman, M. Asthma in young south Asian women living in the United Kingdom: the importance of early life. *Clin Exp Allergy* **37**, 47–53 (2007).
107. Bråbäck, L., Vogt, H. & Hjern, A. Migration and asthma medication in international adoptees and immigrant families in Sweden. *Clinical & Experimental Allergy* **41**, 1108–1115 (2011).
108. Li, Y. *et al.* Global patterns in monthly activity of influenza virus, respiratory syncytial virus, parainfluenza virus, and metapneumovirus: a systematic analysis. *The Lancet Global Health* **7**, e1031–e1045 (2019).
109. Collins, P. L. & Graham, B. S. Viral and Host Factors in Human Respiratory Syncytial Virus Pathogenesis. *Journal of Virology* **82**, 2040–55 (2008).
110. Radhakrishnan, D. *et al.* The association between climate, geography and respiratory syncytial virus hospitalizations among children in Ontario, Canada: a population-based study. *BMC Infectious Diseases* **20**, 157 (2020).
111. Pitzer, V. E. *et al.* Environmental Drivers of the Spatiotemporal Dynamics of Respiratory Syncytial Virus in the United States. *PLOS Pathogens* **11**, e1004591 (2015).
112. Cherrie, M. P. C. *et al.* Pathogen seasonality and links with weather in England and Wales: a big data time series analysis. *BMC Public Health* **18**, 1067 (2018).
113. Donaldson, G. C. Climate Change and the End of the Respiratory Syncytial Virus Season. *Clinical Infectious Diseases* **42**, 677–679 (2006).
114. D'Amato, G. & Akdis, C. A. Global warming, climate change, air pollution and allergies. *Allergy* **75**, 2158–2160 (2020).
115. Cecchi, L., D'Amato, G. & Annesi-Maesano, I. Climate change and outdoor aeroallergens related to allergy and asthma: Taking the exposome into account. *Allergy* **75**, 2361–2363 (2020).
116. DEFRA. *Air Quality and Social Deprivation in the UK: an environmental inequalities analysis*. https://uk-air.defra.gov.uk/assets/documents/reports/cat09/0701110944_AQinequalitiesFNL_AEAT_0506.pdf (2006).
117. Krstić, G. Asthma prevalence associated with geographical latitude and regional insolation in the United States of America and Australia. *PLoS One* **6**, e18492–e18492 (2011).
118. Martineau, A. *et al.* Vitamin D for the management of asthma. *Cochrane Database of Systematic Reviews* (2016) doi:10.1002/14651858.CD011511.pub2.
119. Majeed, H. & Moore, G. W. K. Influence of the Scandinavian climate pattern on the UK asthma mortality: a time series and geospatial study. *BMJ Open* **8**, e020822 (2018).

120. Weiland, S. K. *et al.* Climate and the prevalence of symptoms of asthma, allergic rhinitis, and atopic eczema in children. *Occup Environ Med* **61**, 609–615 (2004).
121. Gelormino, E., Melis, G., Marietta, C. & Costa, G. From built environment to health inequalities: An explanatory framework based on evidence. *Preventive Medicine Reports* **2**, 737–745 (2015).
122. Perdue, W. C., Stone, L. A. & Gostin, L. O. The built environment and its relationship to the public's health: the legal framework. *Am J Public Health* **93**, 1390–1394 (2003).
123. Rao, M., Prasad, S., Adshead, F. & Tissera, H. The built environment and health. *The Lancet* **370**, 1111–1113 (2007).
124. Noveroske, D. B., Warren, J. L., Pitzer, V. E. & Weinberger, D. M. Local variations in the timing of RSV epidemics. *BMC Infectious Diseases* **16**, 674 (2016).
125. MacIntyre, E. A. *et al.* Air pollution and respiratory infections during early childhood: an analysis of 10 European birth cohorts within the ESCAPE Project. *Environ Health Perspect* **122**, 107–113 (2014).
126. N. Krieger, D. R. Williams & N. E. Moss. Measuring Social Class in US Public Health Research: Concepts, Methodologies, and Guidelines. *Annual Review of Public Health* **18**, 341–378 (1997).
127. Gjelsvik, A. *et al.* Neighborhood Risk and Hospital Use for Pediatric Asthma, Rhode Island, 2005–2014. *Prev Chronic Dis* **16**, E68 (2019).
128. Mayne, S. L., Pellissier, B. F. & Kershaw, K. N. Neighborhood Physical Disorder and Adverse Pregnancy Outcomes among Women in Chicago: a Cross-Sectional Analysis of Electronic Health Record Data. *J Urban Health* **96**, 823–834 (2019).
129. Teixeira, S. & Zuberi, A. Neighborhood Social and Environmental Factors and Asthma Among Children Living in Low-Income Neighborhoods: The Importance of Informal Social Control. *Family & Community Health* **41**, (2018).
130. Cantu, P. *et al.* Downward Neighborhood Poverty Mobility during Childhood Is Associated with Child Asthma: Evidence from the Geographic Research on Wellbeing (GROW) Survey. (2019) doi:10.1007/s11524-019-00356-2.
131. NHS Digital. Infant Feeding Survey - UK, 2010. <https://digital.nhs.uk/data-and-information/publications/statistical/infant-feeding-survey/infant-feeding-survey-uk-2010> (2012).
132. Hiscock, R., Bauld, L., Amos, A. & Platt, S. Smoking and socioeconomic status in England: the rise of the never smoker and the disadvantaged smoker. *Journal of Public Health* **34**, 390–396 (2012).
133. Scottish Public Health Observatory. *Tobacco use: maternal smoking*. <https://www.scotpho.org.uk/behaviour/tobacco-use/data/maternal-smoking> (2020).
134. NHS Digital. *Statistics on Women's Smoking Status at Time of Delivery: England - Quarter 4, 2018-19*. <http://digital.nhs.uk/pubs/wsstdq418> (2019).
135. Public Health England. *Public Health Profiles*. <https://fingertips.phe.org.uk>.
136. Carroll, K. N. *et al.* Maternal Asthma and Maternal Smoking Are Associated With Increased Risk of Bronchiolitis During Infancy. *Pediatrics* **119**, 1104 (2007).
137. Koehoorn, M. *et al.* Descriptive epidemiological features of bronchiolitis in a population-based cohort. *Pediatrics* **122**, 1196–1203 (2008).

138. Lawder, R., Whyte, B., Wood, R., Fischbacher, C. & Tappin, D. M. Impact of maternal smoking on early childhood health: a retrospective cohort linked dataset analysis of 697 003 children born in Scotland 1997–2009. *BMJ Open* **9**, e023213 (2019).
139. Jones, L. L. *et al.* Parental and household smoking and the increased risk of bronchitis, bronchiolitis and other lower respiratory infections in infancy: systematic review and meta-analysis. *Respiratory Research* **12**, 5 (2011).
140. Strina, A., Barreto, M. L., Cooper, P. J. & Rodrigues, L. C. Risk factors for non-atopic asthma/wheeze in children and adolescents: a systematic review. *Emerging Themes in Epidemiology* **11**, 5 (2014).
141. Scottish Government. The Scottish Health Survey: Volume 1: Main Report. <http://www.gov.scot/Publications/2011/09/27084018/0>.
142. Davidson, R., Roberts, S. E., Wotton, C. J. & Goldacre, M. J. Influence of maternal and perinatal factors on subsequent hospitalisation for asthma in children: evidence from the Oxford record linkage study. *BMC Pulmonary Medicine* **10**, 14 (2010).
143. Firoozi, F. *et al.* Impact of maternal asthma on perinatal outcomes: a two-stage sampling cohort study. *European Journal of Epidemiology* **27**, 205–214 (2012).
144. Schoos, A.-M. M. *et al.* Parent-specific effects on risk of developing allergic sensitization and asthma in childhood. *Clinical & Experimental Allergy* **50**, 915–921 (2020).
145. Fecht, D. *et al.* Associations between air pollution and socioeconomic characteristics, ethnicity and age profile of neighbourhoods in England and the Netherlands. *Environmental Pollution* **198**, 201–210 (2015).
146. Milojevic, A. *et al.* Socioeconomic and urban-rural differentials in exposure to air pollution and mortality burden in England. *Environmental Health* **16**, 104 (2017).
147. Guarnieri, M. & Balmes, J. R. Outdoor air pollution and asthma. *The Lancet* **383**, 1581–1592 (2014).
148. Thurston, G. D. *et al.* A joint ERS/ATS policy statement: what constitutes an adverse health effect of air pollution? An analytical framework. *Eur Respir J* **49**, 1600419 (2017).
149. Brugha, R. & Grigg, J. Urban Air Pollution and Respiratory Infections. *Paediatric Respiratory Reviews* **15**, 194–199 (2014).
150. Khreis, H., Alotaibi, R., Horney, J. & McConnell, R. The impact of baseline incidence rates on burden of disease assessment of air pollution and onset childhood asthma: analysis of data from the contiguous United States. *Annals of Epidemiology* **53**, 76-88.e10 (2021).
151. Buteau, S. *et al.* A population-based birth cohort study of the association between childhood-onset asthma and exposure to industrial air pollutant emissions. *Environment International* **121**, 23–30 (2018).
152. Holst, G. J. *et al.* Air pollution and family related determinants of asthma onset and persistent wheezing in children: nationwide case-control study. *BMJ* **370**, m2791 (2020).
153. Brauer, M. *et al.* Air pollution and development of asthma, allergy and infections in a birth cohort. *Eur Respir J* **29**, 879 (2007).
154. Hsiao-Hsien Leon Hsu *et al.* Prenatal Particulate Air Pollution and Asthma Onset in Urban Children. Identifying Sensitive Windows and Sex Differences. *American Journal of Respiratory and Critical Care Medicine* **192**, 1052–1059 (2015).

155. Wright, R. J. & Brunst, K. J. Programming of respiratory health in childhood: influence of outdoor air pollution. *Current Opinion in Pediatrics* **25**, 232–239 (2013).
156. King, C., Kirkham, J., Hawcutt, D. & Sinha, I. The effect of outdoor air pollution on the risk of hospitalisation for bronchiolitis in infants: a systematic review. *PeerJ* **6**, e5352–e5352 (2018).
157. Mehta, S., Shin, H., Burnett, R., North, T. & Cohen, A. J. Ambient particulate air pollution and acute lower respiratory infections: a systematic review and implications for estimating the global burden of disease. *Air Quality, Atmosphere & Health* **6**, 69–83 (2013).
158. ONS. *Dataset: Births by parents' characteristics*. <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/datasets/birthsbyparentscharacteristics> (2020).
159. Public Health Scotland. *Births in Scottish hospitals*. <https://beta.isdscotland.org/find-publications-and-data/population-health/births-and-maternity/births-in-scottish-hospitals/> (2020).
160. NHS Inform. *Congenital heart disease*. <https://www.nhsinform.scot/illnesses-and-conditions/heart-and-blood-vessels/conditions/congenital-heart-disease> (2021).
161. Public Health England. *National Congenital Anomaly and Rare Disease Registration Service: Congenital anomaly statistics 2018*. <https://www.gov.uk/government/publications/ncardsr-congenital-anomaly-annual-data> (2020).
162. Thébaud, B. *et al.* Bronchopulmonary dysplasia. *Nat Rev Dis Primers* **5**, 78 (2019).
163. Kristensen, K., Hjuler, T., Ravn, H., Simões, E. A. F. & Stensballe, L. G. Chronic Diseases, Chromosomal Abnormalities, and Congenital Malformations as Risk Factors for Respiratory Syncytial Virus Hospitalization: A Population-Based Cohort Study. *Clinical Infectious Diseases* **54**, 810–817 (2012).
164. Fauroux, B. *et al.* Risk factors for bronchiolitis hospitalization in infants: A French nationwide retrospective cohort study over four consecutive seasons (2009-2013). *PLOS ONE* **15**, e0229766 (2020).
165. Knowles, R. L. *et al.* Ethnic and socioeconomic variation in incidence of congenital heart defects. *Arch Dis Child* **102**, 496 (2017).
166. Vrijheid, M. *et al.* Socioeconomic inequalities in risk of congenital anomaly. *Arch Dis Child* **82**, 349–352 (2000).
167. WHO. *Congenital anomalies*. <https://www.who.int/news-room/fact-sheets/detail/congenital-anomalies> (2016).
168. Smith, L. K., Budd, J. L. S., Field, D. J. & Draper, E. S. Socioeconomic inequalities in outcome of pregnancy and neonatal mortality associated with congenital anomalies: population based study. *BMJ* **343**, d4306–d4306 (2011).
169. Tulloh, R. M. & Bury, S. Prevention and prophylaxis of respiratory syncytial virus in pediatric cardiology: a UK perspective. *Future Cardiology* **10**, 235–242 (2014).
170. Dik, N., Tate, R. B., Manfreda, J. & Anthonisen, N. R. Risk of Physician-Diagnosed Asthma in the First 6 Years of Life. *Chest* **126**, 1147–1153 (2004).
171. Lefebvre, G. & Samoilenko, M. On the use of the outcome variable “small for gestational age” when gestational age is a potential mediator: a maternal asthma perspective. *BMC Medical Research Methodology* **17**, 165 (2017).

172. Martinson, M. L. & Reichman, N. E. Socioeconomic Inequalities in Low Birth Weight in the United States, the United Kingdom, Canada, and Australia. *Am J Public Health* **106**, 748–754 (2016).
173. Opondo, C. *et al.* Joint contribution of socioeconomic circumstances and ethnic group to variations in preterm birth, neonatal mortality and infant mortality in England and Wales: a population-based retrospective cohort study using routine data from 2006 to 2012. *BMJ Open* **9**, e028227 (2019).
174. Bonellie, S. *et al.* Centile charts for birthweight for gestational age for Scottish singleton births. *BMC Pregnancy and Childbirth* **8**, 5 (2008).
175. Xu, X.-F. *et al.* Effect of low birth weight on childhood asthma: a meta-analysis. *BMC Pediatr* **14**, 275 (2014).
176. Mebrahtu, T. F., Feltbower, R. G. & Parslow, R. C. Effects of birth weight and growth on childhood wheezing disorders: findings from the Born in Bradford Cohort. *BMJ Open* **5**, (2015).
177. Liu, X. *et al.* Birth weight, gestational age, fetal growth and childhood asthma hospitalization. *Allergy, Asthma & Clinical Immunology* **10**, 13 (2014).
178. Sandall, J. *et al.* Short-term and long-term effects of caesarean section on the health of women and children. *The Lancet* **392**, 1349–1357 (2018).
179. Lindquist, A., Kurinczuk, J., Redshaw, M. & Knight, M. Experiences, utilisation and outcomes of maternity care in England among women from different socio-economic groups: findings from the 2010 National Maternity Survey. *BJOG: An International Journal of Obstetrics & Gynaecology* **122**, 1610–1617 (2015).
180. Bragg, F. *et al.* Variation in rates of caesarean section among English NHS trusts after accounting for maternal and clinical risk: cross sectional study. *BMJ* **341**, c5065 (2010).
181. Essex, H., Green, J., Baston, H. & Pickett, K. Which women are at an increased risk of a caesarean section or an instrumental vaginal birth in the UK: an exploration within the Millennium Cohort Study. *BJOG: An International Journal of Obstetrics & Gynaecology* **120**, 732–743 (2013).
182. Moore, H. C., de Klerk, N., Holt, P., Richmond, P. C. & Lehmann, D. Hospitalisation for bronchiolitis in infants is more common after elective caesarean delivery. *Archives of Disease in Childhood* (2011) doi:10.1136/archdischild-2011-300607.
183. Huang, L. *et al.* Is elective cesarean section associated with a higher risk of asthma? A meta-analysis. *Journal of Asthma* **52**, 16–25 (2015).
184. Biasucci, G., Benenati, B., Morelli, L., Bessi, E. & Boehm, G. Cesarean Delivery May Affect the Early Biodiversity of Intestinal Bacteria. *The Journal of Nutrition* **138**, 1796S-1800S (2008).
185. Almqvist, C., Cnattingius, S., Lichtenstein, P. & Lundholm, C. The impact of birth mode of delivery on childhood asthma and allergic diseases--a sibling study. *Clin Exp Allergy* **42**, 1369–1376 (2012).
186. Bråbäck, L., Ekéus, C., Lowe, A. J. & Hjertqvist, A. Confounding with familial determinants affects the association between mode of delivery and childhood asthma medication – a national cohort study. *Allergy, Asthma & Clinical Immunology* **9**, 14 (2013).
187. Almqvist, C. & Öberg, A. S. The association between caesarean section and asthma or allergic disease continues to challenge. *Acta Paediatrica* **103**, 349–351 (2014).
188. Berrington, A., Stone, J. & Beaujouan, E. Educational differences in timing and quantum of childbearing in Britain: A study of cohorts born 1940–1969. *Demographic Research* **33**, 733–764 (2015).

189. Hardelid, P., Verfuenden, M., McMenamin, J., Smyth, R. & Gilbert, R. The contribution of child, family and health service factors to respiratory syncytial virus hospital admissions in the first three years of life: birth cohort study in Scotland. (*under review in Pediatric Infectious Disease Journal*) (2018).
190. Haerskjold, A. *et al.* Risk Factors for Hospitalization for Respiratory Syncytial Virus Infection: A Population-based Cohort Study of Danish Children. *Pediatr Infect Dis J* **35**, 61–65 (2016).
191. Caudri, D. *et al.* Perinatal risk factors for wheezing phenotypes in the first 8 years of life. *Clinical & Experimental Allergy* **43**, 1395–1405 (2013).
192. Strachan, D. P. *et al.* Siblings, asthma, rhinoconjunctivitis and eczema: a worldwide perspective from the International Study of Asthma and Allergies in Childhood. *Clinical & Experimental Allergy* **45**, 126–136 (2015).
193. WHO. *Nutrition: breastfeeding.* https://apps.who.int/nutrition/topics/exclusive_breastfeeding/en/index.html (2021).
194. Public Health Scotland. *Infant Feeding Statistics Scotland: Financial Year 2019/20.* <https://beta.isdscotland.org/find-publications-and-data/population-health/child-health/infant-feeding-statistics/> (2020).
195. Shi, T. *et al.* Risk factors for respiratory syncytial virus associated with acute lower respiratory infection in children under five years: Systematic review and meta-analysis. *J Glob Health* **5**, 020416–020416 (2015).
196. Tromp, I. *et al.* Breastfeeding and the risk of respiratory tract infections after infancy: The Generation R Study. *PLOS ONE* **12**, e0172763 (2017).
197. Owora, A. H. & Zhang, Y. Childhood wheeze trajectory-specific risk factors: a systematic review and meta-analysis. *Pediatric Allergy and Immunology* *n/a*, (2020).
198. Tin Tin, S. *et al.* Internal living environment and respiratory disease in children: findings from the Growing Up in New Zealand longitudinal child cohort study. *Environmental Health* **15**, 120 (2016).
199. Fisk, W. J., Lei-Gomez, Q. & Mendell, M. J. Meta-analyses of the associations of respiratory health effects with dampness and mold in homes. *Indoor Air* **17**, 284–296 (2007).
200. Ministry of Housing, Communities & Local Government. *English Housing Survey.* <https://www.gov.uk/government/collections/english-housing-survey> (2020).
201. Ministry of Housing, Communities & Local Government. *Statistical data set: English Housing Survey data on dwelling condition and safety.* <https://www.gov.uk/government/statistical-data-sets/dwelling-condition-and-safety> (2020).
202. WHO. *WHO guidelines for indoor air quality: dampness and mould.* https://www.euro.who.int/__data/assets/pdf_file/0017/43325/E92645.pdf (2009).
203. Hersoug, L. -G. Viruses as the causative agent related to ‘dampness’ and the missing link between allergen exposure and onset of allergic disease. *Indoor Air* **15**, 363–366 (2005).
204. Ministry of Housing, Communities & Local Government. *English Housing Survey 2019 to 2020: headline report.* <https://www.gov.uk/government/statistics/english-housing-survey-2019-to-2020-headline-report> (2020).
205. Ministry of Housing, Communities and Local Government. *Overcrowded households.* <https://www.ethnicity-facts-figures.service.gov.uk/housing/housing-conditions/overcrowded-households/latest> (2020).

206. Ingham, T. *et al.* Damp mouldy housing and early childhood hospital admissions for acute respiratory infection: a case control study. *Thorax* **74**, 849 (2019).
207. Tischer, C., Chen, C.-M. & Heinrich, J. Association between domestic mould and mould components, and asthma and allergy in children: a systematic review. *Eur Respir J* **38**, 812 (2011).
208. Colosia, A. D. *et al.* Residential crowding and severe respiratory syncytial virus disease among infants and young children: a systematic literature review. *BMC Infect Dis* **12**, 95–95 (2012).
209. NHS. When do I have to pay for NHS treatment? <https://www.nhs.uk/common-health-questions/nhs-services-and-treatments/when-do-i-have-to-pay-for-nhs-treatment/> (2018).
210. Kelly, E., Stoye, G. & Vera-Hernández, M. Public Hospital Spending in England: Evidence from National Health Service Administrative Records. *Fiscal Studies* **37**, 433–459 (2016).
211. Cookson, R., Propper, C., Asaria, M. & Raine, R. Socio-Economic Inequalities in Health Care in England. *Fiscal Studies* **37**, 371–403 (2016).
212. Tudor Hart, J. THE INVERSE CARE LAW. *The Lancet* **297**, 405–412 (1971).
213. Marmot, M. An inverse care law for our time. *BMJ* **362**, k3216 (2018).
214. Purdy, S., Griffin, T., Salisbury, C. & Sharp, D. Ambulatory care sensitive conditions: terminology and disease coding need to be more specific to aid policy makers and clinicians. *Public Health* **123**, 169–173 (2009).
215. Carande, E. J., Cheung, C. R., Pollard, A. J. & Drysdale, S. B. Change in viral bronchiolitis management in primary care in the UK after the publication of NICE guideline. *Thorax* **73**, 674 (2018).
216. Cecil, E. *et al.* Primary Care Access, Emergency Department Visits, and Unplanned Short Hospitalizations in the UK. *Pediatrics* **137**, (2016).
217. Fleetcroft, R. *et al.* Emergency hospital admissions for asthma and access to primary care: cross-sectional analysis. *Br J Gen Pract* **66**, e640 (2016).
218. Meissner, H. C. Viral Bronchiolitis in Children. *New England Journal of Medicine* **374**, 62–72 (2016).
219. Betts, K. S., Soares Magalhães, R. J. & Alati, R. The role of neonatal pulmonary morbidity in the longitudinal patterns of hospitalisation for respiratory infection during the first year of life. *Epidemiology and Infection* 1–8 (2018) doi:10.1017/S0950268818001103.
220. Papenburg, J. *et al.* Comparison of Risk Factors for Human Metapneumovirus and Respiratory Syncytial Virus Disease Severity in Young Children. *The Journal of Infectious Diseases* **206**, 178–189 (2012).
221. NHS. *High temperature (fever) in children*. <https://www.nhs.uk/conditions/fever-in-children/> (2020).
222. Becklake, M. R. & Kauffmann, F. Gender differences in airway behaviour over the human life span. *Thorax* **54**, 1119–1138 (1999).
223. Verhein, K. C., Vellers, H. L. & Kleeberger, S. R. Inter-individual variation in health and disease associated with pulmonary infectious agents. *Mamm Genome* **29**, 38–47 (2018).
224. Alvarez, A. E., Marson, F. A. de L., Bertuzzo, C. S., Arns, C. W. & Ribeiro, J. D. Epidemiological and genetic characteristics associated with the severity of acute viral bronchiolitis by respiratory syncytial virus. *Jornal de Pediatria* **89**, 531–543 (2013).

225. Meyers, D. A., Bleecker, E. R., Holloway, J. W. & Holgate, S. T. Asthma genetics and personalised medicine. *Lancet Respir Med* **2**, 405–415 (2014).
226. Feldman, A. S., He, Y., Moore, M. L., Hershenson, M. B. & Hartert, T. V. Toward Primary Prevention of Asthma. Reviewing the Evidence for Early-Life Respiratory Viral Infections as Modifiable Risk Factors to Prevent Childhood Asthma. *American Journal of Respiratory and Critical Care Medicine* **191**, 34–44 (2015).
227. Thomsen, S. F. *et al.* Exploring the Association between Severe Respiratory Syncytial Virus Infection and Asthma. *American Journal of Respiratory and Critical Care Medicine* **179**, 1091–1097 (2009).
228. Husby, A. *et al.* CDHR3 gene variation and childhood bronchiolitis. *Journal of Allergy and Clinical Immunology* **140**, 1469–1471.e7 (2017).
229. Larkin, E. K. & Hartert, T. V. Genes associated with RSV lower respiratory tract infection and asthma: the application of genetic epidemiological methods to understand causality. *Future Virology* **10**, 883–897 (2015).
230. Singh, A. M., Moore, P. E., Gern, J. E., Lemanske, R. F. & Hartert, T. V. Bronchiolitis to Asthma. *American Journal of Respiratory and Critical Care Medicine* **175**, 108–119 (2007).
231. The Scottish Parliament. Devolved and Reserved Matters. <https://www.parliament.scot/visitandlearn/Education/18642.aspx>.
232. *Scotland Act 1998 c. 46*.
233. GOV.UK. Public Health England: About us. <https://www.gov.uk/government/organisations/public-health-england/about>.
234. NHS Health Scotland. A Fairer Healthier Scotland - a strategic framework for action 2017-2022. <http://www.healthscotland.scot/publications/a-fairer-healthier-scotland-a-strategic-framework-for-action-2017-2022> (2017).
235. Public Health England. *Respiratory syncytial virus (RSV): symptoms, transmission, prevention, treatment*. <https://www.gov.uk/government/publications/respiratory-syncytial-virus-rsv-symptoms-transmission-prevention-treatment> (2008).
236. Jefferson, T. *et al.* Physical interventions to interrupt or reduce the spread of respiratory viruses. *Cochrane Database of Systematic Reviews* (2011) doi:10.1002/14651858.CD006207.pub4.
237. Head, M. G., Fitchett, J. R., Holmes, A. H. & Atun, R. Funding healthcare-associated infection research: a systematic analysis of UK research investments, 1997–2010. *Journal of Hospital Infection* **87**, 84–91 (2014).
238. Stone, S. P. *et al.* Evaluation of the national Cleanyourhands campaign to reduce *Staphylococcus aureus* bacteraemia and *Clostridium difficile* infection in hospitals in England and Wales by improved hand hygiene: four year, prospective, ecological, interrupted time series study. *BMJ* **344**, e3005 (2012).
239. NHS NSS. Seasonal respiratory pathogens. *Health Protection Scotland* <https://www.hps.scot.nhs.uk/a-to-z-of-topics/seasonal-respiratory-pathogens/>.
240. Zylbersztejn, A. & Hardelid, P. G492 Access to palivizumab against respiratory syncytial virus in high-risk children in England. *Arch Dis Child* **104**, A198 (2019).

241. World Health Organisation. Immunization, Vaccines and Biologicals: RSV vaccine research and development. https://www.who.int/immunization/research/development/ppc_rsv_vaccines/en/ (2021).
242. Dunn, A. Novavax shares slump as RSV vaccine fails in Phase 3 trial. *BioPharma Dive*.
243. Madhi, S. A. *et al.* Respiratory Syncytial Virus Vaccination during Pregnancy and Effects in Infants. *N Engl J Med* **383**, 426–439 (2020).
244. Public Health England. *Pertussis immunisation in pregnancy: vaccine coverage (England)*. <https://www.gov.uk/government/publications/pertussis-immunisation-in-pregnancy-vaccine-coverage-estimates-in-england-october-2013-to-march-2014>.
245. Public Health England. *Seasonal influenza vaccine uptake in GP patients: winter season 2018 to 2019*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/804889/Seasonal_influenza_vaccine_uptake_in_GP_patients_1819.pdf (2019).
246. NHS Digital. Childhood Vaccination Coverage Statistics - England 2018-19. <https://digital.nhs.uk/data-and-information/publications/statistical/nhs-immunisation-statistics/england-2018-19> (2019).
247. The Scottish Public Health Observatory. Immunisations: inequalities in uptake. <https://www.scotpho.org.uk/health-wellbeing-and-disease/immunisations/data/inequalities-in-uptake/>.
248. Wilcox, C. R. *et al.* Attitudes of Pregnant Women and Healthcare Professionals Toward Clinical Trials and Routine Implementation of Antenatal Vaccination Against Respiratory Syncytial Virus: A Multicenter Questionnaire Study. *The Pediatric Infectious Disease Journal* **38**, (2019).
249. AbbVie. More than a Cold. <https://www.morethanacold.co.uk/> (2018).
250. NHS England. Childhood asthma. <https://www.england.nhs.uk/childhood-asthma/>.
251. Department of Health. *An Outcomes Strategy for Chronic Obstructive Pulmonary Disease (COPD) and Asthma in England*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/216139/dh_128428.pdf (2011).
252. NICE. *Asthma: QOF indicators*. <https://cks.nice.org.uk/topics/asthma/goals-outcome-measures/qof-indicators/> (2020).
253. Public Health England. All Our Health: personalised care and population health. *GOV.UK* <https://www.gov.uk/government/collections/all-our-health-personalised-care-and-population-health#resources> (2019).
254. NHS. Welcome. *Respiratory Futures* <https://www.respiratoryfutures.org.uk/> (2019).
255. The Scottish Government. *Good Places Better Health for Scotland's Children*. <http://www.spokes.org.uk/wp-content/uploads/2010/07/11112-Good-Places-better-health-for-Scotlands-children-0124383.pdf> (2011).
256. The Scottish Government. *Good Places Better Health for Scotland's Children: Childhood Asthma Evidence Assessment*. <https://www2.gov.scot/Resource/0039/00391409.pdf> (2011).
257. NHS England. NHS Long Term Plan. (2019).

258. Trueman, D., Woodcock, F. & Hancock, E. *The battle for breath - the economic burden of lung disease*. <https://www.blf.org.uk/policy/economic-burden> (2017).
259. Public Health England. *PHE Strategy 2020-25*. <https://www.gov.uk/government/publications/phe-strategy-2020-to-2025> (2019).
260. NHS Health Scotland. *NHS Health Scotland Delivery plan 2019–20*. <http://www.healthscotland.scot/media/2549/delivery-plan-2019-20.pdf> (2019).
261. NHS Digital. *Hospital Admitted Patient Care Activity, 2016-17; Hospital Admitted Patient Care Activity, 2016-17: Diagnosis*. <https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/2016-17> (2017).
262. Lai, C. K. W. *et al.* Global variation in the prevalence and severity of asthma symptoms: Phase Three of the International Study of Asthma and Allergies in Childhood (ISAAC). *Thorax* **64**, 476 (2009).
263. Jama-Alol, K. A., Moore, H. C., Jacoby, P., Bower, C. & Lehmann, D. Morbidity due to acute lower respiratory infection in children with birth defects: a total population-based linked data study. *BMC Pediatr* **14**, 80–80 (2014).
264. Wu, P. & Hartert, T. V. Evidence for a causal relationship between respiratory syncytial virus infection and asthma. *Expert review of anti-infective therapy* **9**, 731–745 (2011).
265. Pavis, S. & D Morris, A. Unleashing the power of administrative health data: the Scottish model. *Public Health Research & Practice* (2015).
266. Connelly, R., Playford, C. J., Gayle, V. & Dibben, C. The role of administrative data in the big data revolution in social science research. *Social Science Research* **59**, 1–12 (2016).
267. Harron, K. *et al.* Challenges in administrative data linkage for research. *Big Data & Society* **4**, 2053951717745678 (2017).
268. Dattani, N., Datta-Nemdharry, P. & Macfarlane, A. Linking maternity data for England, 2005-06: methods and data quality. *Health Statistics Quarterly* **49**, 53–79 (2011).
269. Wijlaars, L. P. *et al.* Contribution of recurrent admissions in children and young people to emergency hospital admissions: retrospective cohort analysis of hospital episode statistics. *Arch Dis Child* *archdischild-2014-307771* (2015) doi:10.1136/archdischild-2014-307771.
270. NHS Digital. Hospital Episode Statistics Data Dictionary. (2018).
271. NHS Digital. Hospital Episode Statistics (HES). (2019).
272. NHS Digital. Understanding the health and care information we collect: Your health and care information - what we collect, and why. <https://digital.nhs.uk/about-nhs-digital/our-work/keeping-patient-data-safe/how-we-look-after-your-health-and-care-information/understanding-the-health-and-care-information-we-collect#section-child-pages> (2019).
273. Harron, K., Gilbert, R., Cromwell, D. & van der Meulen, J. Linking Data for Mothers and Babies in De-Identified Electronic Health Data. *PLOS ONE* **11**, e0164667 (2016).
274. Herbert, A., Wijlaars, L., Zylbersztejn, A., Cromwell, D. & Hardelid, P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *International Journal of Epidemiology* **46**, 1093–1093i (2017).

275. Méray, N., Reitsma, J. B., Ravelli, A. C. J. & Bonsel, G. J. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. *Journal of Clinical Epidemiology* **60**, 883.e1-883.e11 (2007).
276. Doidge, J. C. & Harron, K. Demystifying probabilistic linkage: Common myths and misconceptions. *Int J Popul Data Sci* **3**, 410–410 (2018).
277. Burns, E. M. *et al.* Systematic review of discharge coding accuracy. *Journal of Public Health* **34**, 138–148 (2011).
278. NHS Digital. Hospital Episode Statistics (HES) Analysis Guide. (2019).
279. Murray, J. *et al.* Quality of routine hospital birth records and the feasibility of their use for creating birth cohorts. *Journal of Public Health* **35**, 298–307 (2012).
280. Knight, H. E. *et al.* Evaluating maternity care using national administrative health datasets: How are statistics affected by the quality of data on method of delivery? *BMC Health Services Research* **13**, 200 (2013).
281. Lewis, K. M. & Hardelid, P. National data opt out programme: consequences for maternity services in England. *International Journal of Population Data Science* **5**, (2020).
282. The King's Fund. *The UK private health market*. <https://www.kingsfund.org.uk/sites/default/files/media/commission-appendix-uk-private-health-market.pdf>.
283. Medical Research Council and NHS Research Authority. Do I need NHS REC approval? <http://www.hra-decisiontools.org.uk/ethics/> (2020).
284. Zylbersztejn, A., Gilbert, R., Hjern, A., Wijlaars, L. & Hardelid, P. Child mortality in England compared with Sweden: a birth cohort study. *The Lancet* **391**, 2008–2018 (2018).
285. Hagger-Johnson, G. *et al.* Identifying Possible False Matches in Anonymized Hospital Administrative Data without Patient Identifiers. *Health Services Research* **50**, 1162–1178 (2015).
286. Office for National Statistics. *User guide to mortality statistics*. <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/methodologies/userguidetomortalitystatisticsjuly2017> (2019).
287. NHS Digital. [MI] Care Information Choices, England - September 2017. <https://digital.nhs.uk/data-and-information/publications/statistical/care-information-choices/mi-care-information-choices-england-september-2017> (2017).
288. ISD Scotland. Data Dictionary A-Z: CHI Number. *ISD Scotland Home* <https://www.ndc.scot.nhs.uk/Dictionary-A-Z/Definitions/index.asp?ID=128&Title=CHI%20Number> (2020).
289. NHS National Services Scotland. Patient data uses/disclosure opt out. *NHSNSS* <https://nhsnss.org/foi-disclosure/patient-data-usesdisclosure-opt-out/> (2016).
290. Alvarez-Madrado, S., McTaggart, S., Nangle, C., Nicholson, E. & Bennie, M. Data Resource Profile: The Scottish National Prescribing Information System (PIS). *Int J Epidemiol* **45**, 714–715f (2016).
291. GOV.UK. National Records of Scotland. <https://www.gov.uk/national-records-of-scotland>.
292. HM Passport Office. General Register Office. (2019).

- 293.Scottish Public Health Observatory. *Using the Linked Scottish Health Survey to Predict Hospitalisation & Death*. <https://www.scotpho.org.uk/publications/reports-and-papers/using-the-linked-scottish-health-survey-to-predict-hospitalisation-and-death/> (2007).
- 294.ISD Scotland. SBR & UNHS: Scottish Birth Record & Universal Newborn Hearing Screening. (2010).
- 295.ISD Scotland. SMR02 - Maternity Inpatient and Day Case. <http://www.ndc.scot.nhs.uk/Dictionary-A-Z/Definitions> (2018).
- 296.Hardelid, P., Verfuenden, M., McMenamin, J., Smyth, R. & Gilbert, R. The contribution of child, family and health service factors to respiratory syncytial virus hospital admissions in the first three years of life: birth cohort study in Scotland. (*under review in Pediatric Infectious Disease Journal*) (2018).
- 297.Hardelid, P., Verfuenden, M., McMenamin, J. & Gilbert, R. Risk factors for admission to hospital with laboratory-confirmed influenza in young children: birth cohort study. *Eur Respir J* **50**, 1700489 (2017).
- 298.National Records of Scotland. Vital Events Reference Tables 2018 (Section 3: Births). (2019).
- 299.Sinha, S., Peach, G., Poloniecki, J. D., Thompson, M. M. & Holt, P. J. Studies using English administrative data (Hospital Episode Statistics) to assess health-care outcomes—systematic review and recommendations for reporting. *European Journal of Public Health* **23**, 86–92 (2013).
- 300.National Data Guardian. *Information: To share or Not to Share? The Information Governance Review*. <https://www.gov.uk/government/publications/the-information-governance-review> (2013).
- 301.Piel, F. B., Parkes, B. L., Daby, H., Hansell, A. L. & Elliott, P. The challenge of opt-outs from NHS data: a small-area perspective. *Journal of Public Health* **40**, e594–e600 (2018).
- 302.Heather, B. *Poor uptake of new scheme to give patients control of confidential records*. <https://www.hsj.co.uk/technology-and-innovation/poor-uptake-of-new-scheme-to-give-patients-control-of-confidential-records/7023553.article> (2018).
- 303.NHS Digital. National Data Opt-Out, December 2018. <https://digital.nhs.uk/data-and-information/publications/statistical/national-data-opt-out/december-2018/december-2018> (2018).
- 304.National Data Guardian. *Review of data security, consent and opt-outs*. <https://www.gov.uk/government/publications/review-of-data-security-consent-and-opt-outs> (2016).
- 305.McCartney, M. Care.data doesn't care enough about consent. *BMJ* **348**, g2831 (2014).
- 306.Armstrong, S. Data deadlines loom large for the NHS. *BMJ* **360**, k1215 (2018).
- 307.Culshaw, N., Pasupathy, D. & Kyle, P. The value of obstetric surveillance systems within the National Health Service. *The Obstetrician & Gynaecologist* **15**, 85–89 (2013).
- 308.Which? Where to give birth? <https://www.which.co.uk/birth-choice/where-to-give-birth> (2018).
- 309.Royal College of Obstetricians and Gynaecologists. National Maternity and Perinatal Audit. <http://www.maternityaudit.org.uk> (2017).
- 310.Aylin, P. *et al. Estimating the risk of adverse birth outcomes in pregnant women undergoing non-obstetric surgery using routinely collected NHS data: an observational study*. (NIHR Journals Library, 2016).
- 311.Public Health England. Child and Maternal Health. <https://fingertips.phe.org.uk/profile/child-health-profiles/> (2018).

- 312.NHS Digital. [MI] National Data Opt-Out, December 2018. <https://digital.nhs.uk/data-and-information/publications/statistical/national-data-opt-out/december-2018/december-2018> (2018).
- 313.Public Health England. Technical guide - confidence intervals. <https://fingertips.phe.org.uk/profile/guidance> (2018).
- 314.StataCorp. *Stata Statistical Software: Release 15*. (StataCorp LP, 2015).
- 315.Sterne, J. A. C. *et al.* Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* **338**, b2393 (2009).
- 316.Langhoff-Roos, J. *et al.* The Nordic medical birth registers--a potential goldmine for clinical research. *Acta Obstet Gynecol Scand* **93**, 132–137 (2014).
- 317.NHS Digital. NHS Number. <https://digital.nhs.uk/services/nhs-number#benefits-of-using-the-nhs-number>.
- 318.Webster, P. C. Sweden's health data goldmine. *CMAJ* **186**, E310–E310 (2014).
- 319.Riordan, F. *et al.* Patient and public attitudes towards informed consent models and levels of awareness of Electronic Health Records in the UK. *Int J Med Inform* **84**, 237–247 (2015).
- 320.New, J. P., Leather, D., Bakerly, N. D., McCrae, J. & Gibson, J. M. Putting patients in control of data from electronic health records. *BMJ* **360**, j5554 (2018).
- 321.Wellcome Trust. Understanding Patient Data. <https://understandingpatientdata.org.uk/> (2020).
- 322.CPRD. Bibliography. <https://www.cprd.com/Bibliography> (2020).
- 323.NHS Digital. Register of approved data releases. <https://digital.nhs.uk/services/data-access-request-service-dars/register-of-approved-data-releases> (2021).
- 324.Health eResearch Centre. Health Data on Trial: The Citizens' Juries. https://www.herc.ac.uk/case_studies/health-data-trial-citizens-juries/.
- 325.The Farr Institute. #datasaveslives. <https://farrinstitute.org/public-engagement-involvement/datasaveslives> (2018).
- 326.Hill, E. M., Turner, E. L., Martin, R. M. & Donovan, J. L. "Let's get the best quality research we can": public awareness and acceptance of consent to use existing data in health research: a systematic review and qualitative study. *BMC Medical Research Methodology* **13**, 72 (2013).
- 327.Cheung, C. R., Smith, H., Thurland, K., Duncan, H. & Semple, M. G. Population variation in admission rates and duration of inpatient stay for bronchiolitis in England. *Archives of Disease in Childhood* **98**, 57–59 (2013).
- 328.Christiansen, C. F., Pedersen, L., Sørensen, H. T. & Rothman, K. J. Methods to assess seasonal effects in epidemiological studies of infectious diseases—exemplified by application to the occurrence of meningococcal disease. *Clinical Microbiology and Infection* **18**, 963–969 (2012).
- 329.Lofgren, E., Fefferman, N., Doshi, M. & Naumova, E. N. Assessing Seasonal Variation in Multisource Surveillance Data: Annual Harmonic Regression. in *Intelligence and Security Informatics: Biosurveillance* (eds. Zeng, D. *et al.*) 114–123 (Springer Berlin Heidelberg, 2007).
- 330.Ramanathan, K. *et al.* Assessing Seasonality Variation with Harmonic Regression: Accommodations for Sharp Peaks. *Int J Environ Res Public Health* **17**, 1318 (2020).

331. Panozzo, C. A., Fowlkes, A. L. & Anderson, L. J. Variation in Timing of Respiratory Syncytial Virus Outbreaks: Lessons From National Surveillance. *The Pediatric Infectious Disease Journal* **26**, (2007).
332. Moore, H. C. *et al.* Seasonality of Respiratory Viral Identification Varies With Age and Aboriginality in Metropolitan Western Australia. *The Pediatric Infectious Disease Journal* **28**, (2009).
333. Park, N. Small area population estimates in England and Wales: mid-2016. <https://www.ons.gov.uk/releases/smallareapopulationestimatesinenglandandwalesmid2016> (2017).
334. Department of Housing, Communities & Local Government. English indices of deprivation 2015: File 13: clinical commissioning group summaries. GOV.UK <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015> (2015).
335. Reeves, R. M. Respiratory Syncytial Virus in Young Children in England: Burden and Risk Factors for Severe Disease. (UCL, 2017).
336. Ralston, S. L. *et al.* Clinical practice guideline: the diagnosis, management, and prevention of bronchiolitis. *Pediatrics* **134**, e1474-1502 (2014).
337. ONS. *Regions (Former GORs), Guidance and Methodology*. <http://webarchive.nationalarchives.gov.uk/20160128190831/http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/administrative/england/government-office-regions/index.html> [accessed 9 July 2019].
338. Ministry of Housing, Communities & Local Government. *English indices of deprivation 2015: technical report*. <https://www.gov.uk/government/publications/english-indices-of-deprivation-2015-technical-report> (2015).
339. QGIS Development Team. *QGIS Geographic Information System*. <http://qgis.osgeo.org> (2018).
340. Lofgren, E. T. *et al.* Disproportional effects in populations of concern for pandemic influenza: insights from seasonal epidemics in Wisconsin, 1967-2004. *Influenza Other Respir Viruses* **4**, 205–212 (2010).
341. Naumova E.N., MacNeill I.B. Seasonality Assessment for Biosurveillance Systems. in *In: Auget JL., Balakrishnan N., Mesbah M., Molenberghs G. (eds) Advances in Statistical Methods for the Health Sciences. Statistics for Industry and Technology*. (Birkhäuser, 2007).
342. Barnett, A. G. & Dobson, A. J. *Analysing Seasonal Health Data*. (Springer-Verlag, 2010).
343. Tsai, P. J., Lin, M. L. & Chu, C. M. Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006. *BMC Public Health*. **9**, 464 (2009).
344. Semple, M. G., Taylor-Robinson, D. C., Lane, S. & Smyth, R. L. Household Tobacco Smoke and Admission Weight Predict Severe Bronchiolitis in Infants Independent of Deprivation: Prospective Cohort Study. *PLOS ONE* **6**, e22425 (2011).
345. Sheikh, A. & Alves, B. Age, sex, geographical and socio-economic variations in admissions for anaphylaxis: analysis of four years of English hospital data. *Clin Exp Allergy* **31**, 1571–1576 (2001).
346. Tennant, A. Admission to hospital following head injury in England: incidence and socio-economic associations. *BMC Public Health* **5**, 21–21 (2005).
347. National Child and Maternal Health Intelligence Network. Child and Maternal Health. <https://fingertips.phe.org.uk/profile/child-health-profiles/> (2018).

348. Nath, S., Viner, R., Lewis, K. & Hardelid, P. Geographical variations in pathways to emergency hospital admissions in infants in England. in (HSR UK, 2020).
349. NHS. *Paediatric Intensive Care Winter Surge Standard Operating Procedure*. <https://www.england.nhs.uk/publication/paediatric-intensive-care-surge-sop/> (2020).
350. British Lung Foundation. *Out in the cold: lung disease, the hidden driver of NHS winter pressure*. <https://www.blf.org.uk/policy/out-in-the-cold> (2017).
351. Broberg, E. K. *et al.* Seasonality and geographical spread of respiratory syncytial virus epidemics in 15 European countries, 2010 to 2016. *Eurosurveillance* **23**, 17–00284 (2018).
352. Tang, J. W. & Loh, T. P. Correlations between climate factors and incidence—a contributor to RSV seasonality. *Reviews in Medical Virology* **24**, 15–34 (2014).
353. Eccles, R. An explanation for the seasonality of acute upper respiratory tract viral infections. *Acta Otolaryngol* **122**, 183–191 (2002).
354. Al-Mahtot, M., Barwise-Munro, R., Wilson, P. & Turner, S. Changing characteristics of hospital admissions but not the children admitted—a whole population study between 2000 and 2013. *Eur J Pediatr* **177**, 381–388 (2018).
355. Anderson, A. J., Snelling, T. L., Moore, H. C. & Blyth, C. C. Advances in Vaccines to Prevent Viral Respiratory Illnesses in Children. *Pediatric Drugs* **19**, 523–531 (2017).
356. Cai, W. *et al.* Evaluation of using ICD-10 code data for respiratory syncytial virus surveillance. *Influenza Other Respir Viruses* **14**, 630–637 (2020).
357. Ministry of Housing, Communities & Local Government. *English Indices of Deprivation 2015 - LSOA Level*. <https://opendatacommunities.org/data/societal-wellbeing/imd/indices> (2015).
358. Ministry of Housing, Communities & Local Government. *English indices of deprivation 2015*. <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015> (2015).
359. Lewis, K. M., De Stavola, B. & Hardelid, P. Could socioeconomic inequity in bronchiolitis admissions be reduced through intervening on gestational age? A principled analysis of cohort data. *The Lancet* **392**, S52 (2018).
360. Schisterman, E. F., Cole, S. R. & Platt, R. W. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* **20**, 488–495 (2009).
361. Rogers, W. Regression standard errors in clustered samples. *Stata Technical Bulletin* **13**, 19–2 (1993).
362. Zhu, Y., Matsuyama, Y., Ohashi, Y. & Setoguchi, S. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. *Journal of Biomedical Informatics* **56**, 80–86 (2015).
363. Stolwijk, A., Straatman, H. & Zielhuis, G. Studying seasonality by using sine and cosine functions in regression analysis. *Journal of epidemiology and community health* **53**, 235–238 (1999).
364. Foley, D., Best, E., Reid, N. & Berry, M. (Mary) J. Respiratory health inequality starts early: The impact of social determinants on the aetiology and severity of bronchiolitis in infancy. *Journal of Paediatrics and Child Health* **55**, 528–532 (2019).
365. Public Health England. *Laboratory reporting to Public Health England: A guide for diagnostic laboratories*. (2016).

366. Zhao, H. *et al.* A new laboratory-based surveillance system (Respiratory DataMart System) for influenza and other respiratory viruses in England: results and experience from 2009 to 2012. *Eurosurveillance* **19**, 20680 (2014).
367. Colosia, A. D. *et al.* Residential crowding and severe respiratory syncytial virus disease among infants and young children: A systematic literature review. *BMC Infectious Diseases* **12**, 95 (2012).
368. Daniel, R. M., De Stavola, B. L., Cousens, S. N. & Vansteelandt, S. Causal mediation analysis with multiple mediators. *Biometrics* **71**, 1–14 (2015).
369. Rehkopf, D. H., Glymour, M. M. & Osypuk, T. L. The Consistency Assumption for Causal Inference in Social Epidemiology: When a Rose Is Not a Rose. *Current Epidemiology Reports* **3**, 63–71 (2016).
370. Naimi, A. I., Schnitzer, M. E., Moodie, E. E. M. & Bodnar, L. M. Mediation Analysis for Health Disparities Research. *Am J Epidemiol* **184**, 315–324 (2016).
371. VanderWeele, T. J. Mediation Analysis: A Practitioner's Guide. *Annu. Rev. Public Health* **37**, 17–32 (2016).
372. Lange, T., Vansteelandt, S. & Bekaert, M. A Simple Unified Approach for Estimating Natural Direct and Indirect Effects. *American Journal of Epidemiology* **176**, 190–195 (2012).
373. Daniel, R. M. & De Stavola, B. L. Mediation Analysis for Life Course Studies. in *Pathways to Health* (Springer, 2019).
374. Steen, J., Loeys, T., Moerkerke, B. & Vansteelandt, S. medflex: An R Package for Flexible Mediation Analysis using Natural Effect Models. *Journal of Statistical Software; Vol 1, Issue 11* (2017) (2017).
375. Steen, J., Loeys, T., Moerkerke, B. & Vansteelandt, S. Flexible Mediation Analysis With Multiple Mediators. *American Journal of Epidemiology* **186**, 184–193 (2017).
376. Rudolph, K. E. *et al.* Causal Mediation Analysis With Observational Data: Considerations and Illustration Examining Mechanisms Linking Neighborhood Poverty to Adolescent Substance Use. *American Journal of Epidemiology* **188**, 598–608 (2019).
377. Richiardi, L., Bellocco, R. & Zugna, D. Mediation analysis in epidemiology: methods, interpretation and bias. *International Journal of Epidemiology* **42**, 1511–1519 (2013).
378. Mann, C. J. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emerg Med J* **20**, 54 (2003).
379. Pearl, J. An introduction to causal inference. *Int J Biostat* **6**, 7–7 (2010).
380. Hernán, M. A. & Robins, J. *Causal Inference: What If*. (Chapman & Hall/CRC, 2020).
381. Hernán, M. A. & Robins, J. M. Estimating causal effects from epidemiological data. *J Epidemiol Community Health* **60**, 578 (2006).
382. Vanderweele, T. J., Vansteelandt, S. & Robins, J. M. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology* **25**, 300–306 (2014).
383. Pearl, J. Direct and Indirect Effects. in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence* 411–420 (Morgan Kaufmann Publishers Inc., 2001).
384. VanderWeele, T. J. & Tchetgen Tchetgen, E. J. Mediation analysis with time varying exposures and mediators. *J R Stat Soc Series B Stat Methodol* **79**, 917–938 (2017).

385. Pearl, J. *The Book of Why*.
386. Kaplan, G. A. What's Wrong with Social Epidemiology, and How Can We Make It Better? *Epidemiologic Reviews* **26**, 124–135 (2004).
387. Tchetgen Tchetgen, E. J. & VanderWeele, T. J. On causal inference in the presence of interference. *Stat Methods Med Res* **21**, 55–75 (2012).
388. Berkman, L. & Kawachi, I. *Social Epidemiology*. (Oxford University Press, 2000).
389. Shrier, I. & Platt, R. W. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol* **8**, 70–70 (2008).
390. Pearce, N. & Lawlor, D. A. Causal inference—so much more than statistics. *International Journal of Epidemiology* **45**, 1895–1903 (2017).
391. Ferguson, K. D. *et al.* Evidence synthesis for constructing directed acyclic graphs (ESC-DAGs): a novel and systematic method for building directed acyclic graphs. *International Journal of Epidemiology* (2019) doi:10.1093/ije/dyz150.
392. Pearl, J. *Causal inference in statistics: An overview*. (Department of Statistics, UCLA, 2009).
393. Elwert, F. Graphical Causal Models. in *Handbook of Causal Analysis for Social Research* 245–273 (2013). doi:10.1007/978-94-007-6094-3_13.
394. Tennant, P., Textor, J., Gilthorpe, M. & Ellison, G. OP87 Dagitty and directed acyclic graphs in observational research: a critical review. *J Epidemiol Community Health* **71**, A43 (2017).
395. Tennant, P. W. *et al.* Use of directed acyclic graphs (DAGs) in applied health research: review and recommendations. *medRxiv* 2019.12.20.19015511 (2019) doi:10.1101/2019.12.20.19015511.
396. Kennedy, C. M. *et al.* Associations of mobile source air pollution during the first year of life with childhood pneumonia, bronchiolitis, and otitis media. *Environ Epidemiol* **2**, e007 (2018).
397. Muñoz-Quiles, C. *et al.* Population-based Analysis of Bronchiolitis Epidemiology in Valencia, Spain. *Pediatr Infect Dis J* **35**, 275–280 (2016).
398. Textor, J., van der Zander, B., Gilthorpe, M., Liskiewicz, M. & Ellison, G. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology* **45**, 1887–1894 (2016).
399. Naimi, A. I., Kaufman, J. S. & MacLehose, R. F. Mediation misgivings: ambiguous clinical and public health interpretations of natural direct and indirect effects. *International Journal of Epidemiology* **43**, 1656–1661 (2014).
400. VanderWeele, T. J. Controlled direct and mediated effects: definition, identification and bounds. *Scand Stat Theory Appl* **38**, 551–563 (2011).
401. Naimi, A. I., Kaufman, J. S. & MacLehose, R. F. Mediation misgivings: ambiguous clinical and public health interpretations of natural direct and indirect effects. *International Journal of Epidemiology* **43**, 1656–1661 (2014).
402. Silverwood, R. J., Williamson, L., Grundy, E. M. & De Stavola, B. L. Pathways between Socioeconomic Disadvantage and Childhood Growth in the Scottish Longitudinal Study, 1991–2001. *PLOS ONE* **11**, e0164853 (2016).
403. VanderWeele, T. J., Vansteelandt, S. & Robins, J. M. Effect Decomposition in the Presence of an Exposure-Induced Mediator-Outcome Confounder. *Epidemiology* **25**, (2014).

404. Tchetgen Tchetgen, E. J. & Vanderweele, T. J. Identification of natural direct effects when a confounder of the mediator is directly affected by exposure. *Epidemiology* **25**, 282–291 (2014).
405. Andrews, R. & Didelez, V. Insights into the ‘cross-world’ independence assumption of causal mediation analysis. *arXiv* (2020) doi:arXiv:2003.10341.
406. Vansteelandt, S. & Daniel, R. M. Interventional Effects for Mediation Analysis with Multiple Mediators. *Epidemiology* **28**, 258–265 (2017).
407. Bellavia, A., Zota, A. R., Valeri, L. & James-Todd, T. Multiple mediators approach to study environmental chemicals as determinants of health disparities. *Environmental Epidemiology* **2**, (2018).
408. Oude Groeniger, J., de Koster, W. & van der Waal, J. Time-varying Effects of Screen Media Exposure in the Relationship Between Socioeconomic Background and Childhood Obesity. *Epidemiology* **31**, (2020).
409. ISD Scotland. SMR datasets. <https://www.ndc.scot.nhs.uk/Data-Dictionary/SMR-Datasets/> (2019).
410. BNF. *British National Formulary*. (2007).
411. Naimi, A. I., Cole, S. R. & Kennedy, E. H. An introduction to g methods. *Int J Epidemiol* **46**, 756–762 (2017).
412. Bang, H. & Robins, J. M. Doubly robust estimation in missing data and causal inference models. *Biometrics* **61**, 962–973 (2005).
413. White, I. R. & Carlin, J. B. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in Medicine* **29**, 2920–2931 (2010).
414. Knox, S. *et al.* The challenge of using routinely collected data to compare hospital admission rates by ethnic group: a demonstration project in Scotland. *Journal of Public Health* **42**, 748–755 (2020).
415. Bansal, N. *et al.* Myocardial infarction incidence and survival by ethnic group: Scottish Health and Ethnicity Linkage retrospective cohort study. *BMJ Open* **3**, e003415 (2013).
416. Katikireddi, S. V. *et al.* Assessment of health care, hospital admissions, and mortality by ethnicity: population-based cohort study of health-system performance in Scotland. *The Lancet Public Health* **3**, e226–e236 (2018).
417. Gray, L. *et al.* Cohort profile: the Scottish health surveys cohort: linkage of study participants to routinely collected records for mortality, hospital discharge, cancer and offspring birth characteristics in three nationwide studies. *Int J Epidemiol* **39**, 345–350 (2010).
418. Dietz, P. M. *et al.* Estimates of Nondisclosure of Cigarette Smoking Among Pregnant and Nonpregnant Women of Reproductive Age in the United States. *American Journal of Epidemiology* **173**, 355–359 (2011).
419. Shipton, D. *et al.* Reliability of self reported smoking status by pregnant women for estimating smoking prevalence: a retrospective, cross sectional study. *BMJ* **339**, b4347 (2009).
420. NHS Scotland. *Child health surveillance programme pre-school*. https://www.isdscotland.org/Health-Topics/Child-Health/Child-Health-Programme/_docs/CHSP-PS-Clinical-Guidelines-2018-06-26-FINAL.pdf.
421. SMITH, D. H. *et al.* A National Estimate of the Economic Costs of Asthma. *Am J Respir Crit Care Med* **156**, 787–793 (1997).

422. Raheison, C. & Girodet, P.-O. Epidemiology of COPD. *EUROPEAN RESPIRATORY REVIEW* **18**, 213 (2009).
423. Textor, J., Hardt, J. & Knüppel, S. DAGitty: A Graphical Tool for Analyzing Causal Diagrams. *Epidemiology* **22**, (2011).
424. ONS. *Families and households in the UK: 2019*. <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/bulletins/familiesandhouseholds/2019> (2019).
425. McGowan, A. *et al.* 'Breathe': the stop smoking service for pregnant women in Glasgow. *Midwifery* **26**, e1–e13 (2010).
426. Daniel, R. M., Cousens, S. N., De Stavola, B. L., Kenward, M. G. & Sterne, J. A. C. Methods for dealing with time-dependent confounding. *Statistics in Medicine* **32**, 1584–1618 (2013).
427. Henderson, J. *et al.* Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood. *Thorax* **63**, 974 (2008).
428. Savenije, O. E. *et al.* Comparison of childhood wheezing phenotypes in 2 birth cohorts: ALSPAC and PIAMA. *Journal of Allergy and Clinical Immunology* **127**, 1505-1512.e14 (2011).
429. Martinez, F. D. *et al.* Asthma and Wheezing in the First Six Years of Life. *N Engl J Med* **332**, 133–138 (1995).
430. Depner, M. *et al.* Clinical and Epidemiologic Phenotypes of Childhood Asthma. *Am J Respir Crit Care Med* **189**, 129–138 (2013).
431. Koppelman, G. H. & Gehring, U. Data-driven Asthma Phenotypes in Childhood. Does the Environment Hold the Clue? *Am J Respir Crit Care Med* **195**, 545–546 (2017).
432. Paediatric Formulary Committee. *BNF for Children 2012-2013*. (Pharmaceutical Press, 2012).
433. Elphick, H. E. *et al.* Survey of respiratory sounds in infants. *Arch Dis Child* **84**, 35 (2001).
434. Ng, M. C. W. & How, C. H. Recurrent wheeze and cough in young children: is it asthma? *Singapore Med J* **55**, 236–241 (2014).
435. Scott, M., Kurukulaaratchy, R. J. & Arshad, S. H. Definitions are important and not all wheeze is asthma. *Thorax* **66**, 633 (2011).
436. Cole, T. J., Williams, A. F. & Wright, C. M. Revised birth centiles for weight, length and head circumference in the UK-WHO growth charts. *null* **38**, 7–11 (2011).
437. National Records of Scotland. Vital Events - General Background Information. <https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/general-background-information> (2019).
438. Jones, B. L. & Nagin, D. S. A Note on a Stata Plugin for Estimating Group-based Trajectory Models. *Sociological Methods & Research* **42**, 608–613 (2013).
439. Jung, T. & Wickrama, K. A. S. An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass* **2**, 302–317 (2008).
440. Frankfurt, S., Frazier, P., Syed, M. & Jung, K. R. Using Group-Based Trajectory and Growth Mixture Modeling to Identify Classes of Change Trajectories. *The Counseling Psychologist* **44**, 622–660 (2016).

441. Reinecke, J. & Seddig, D. Growth mixture models in longitudinal research. *AStA Advances in Statistical Analysis* **95**, 415–434 (2011).
442. Muthén, L. K. & Muthén, B. O. Chapter 8: Mixture modeling with longitudinal data. in *Mplus User's Guide*. 221–260 (Muthén & Muthén, 1998).
443. Herle, M. *et al.* Identifying typical trajectories in longitudinal data: modelling strategies and interpretations. *European Journal of Epidemiology* **35**, 205–222 (2020).
444. Muthén, B. The potential of growth mixture modelling. *Infant and Child Development* **15**, 623–625 (2006).
445. *Mplus Discussion: Missing Data Modeling*.
<http://www.statmodel.com/discussion/messages/22/22.html>.
446. Muthén, L. K. & Muthén, B. O. Chapter 11: Special modeling issues. in *Mplus User's Guide*. 515–562 (Muthén & Muthén, 1998).
447. Asparouhov, T. & Muthén, B. Using Mplus TECH11 and TECH14 to test the number of latent classes. in *Mplus Web Notes: No. 14* (2012).
448. Nylund, K. L., Asparouhov, T. & Muthén, B. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal* **14**, 535–569 (2007).
449. Nagin, D. S. & Odgers, C. L. Group-Based Trajectory Modeling in Clinical Research. *Annu. Rev. Clin. Psychol.* **6**, 109–138 (2010).
450. Wheeler, A. P., Worden, R. E. & McLean, S. J. Replicating Group-Based Trajectory Models of Crime at Micro-Places in Albany, NY. *Journal of Quantitative Criminology* **32**, 589–612 (2016).
451. Feldman, B. J., Masyn, K. E. & Conger, R. D. New approaches to studying problem behaviors: a comparison of methods for modeling longitudinal, categorical adolescent drinking data. *Dev Psychol* **45**, 652–676 (2009).
452. White, I. R., Royston, P. & Wood, A. M. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine* **30**, 377–399 (2011).
453. Azur, M. J., Stuart, E. A., Frangakis, C. & Leaf, P. J. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res* **20**, 40–49 (2011).
454. Asparouhov, T. & Muthén, B. Auxiliary Variables in Mixture Modeling: Three-Step Approaches Using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal* **21**, 329–341 (2014).
455. Heron, J. E., Croudace, T. J., Barker, E. D. & Tilling, K. A comparison of approaches for assessing covariate effects in latent class analysis. *Longitudinal and Life Course Studies; Vol 6, No 4 (2015): Longitudinal and Life Course Studies* (2015) doi:10.14301/llcs.v6i4.322.
456. Vermunt, J. K. Latent Class Modeling with Covariates: Two Improved Three-Step Approaches. *Political Analysis* **18**, 450–469 (2010).
457. Micali, N., Daniel, R. M., Ploubidis, G. B. & De Stavola, B. L. Maternal Prepregnancy Weight Status and Adolescent Eating Disorder Behaviors: A Longitudinal Study of Risk Pathways. *Epidemiology* **29**, (2018).
458. Bodner, T. E. What Improves with Increased Missing Data Imputations? *Structural Equation Modeling: A Multidisciplinary Journal* **15**, 651–675 (2008).

459. Chawla, J., Seear, M., Zhang, T., Smith, A. & Carleton, B. Fifty years of pediatric asthma in developed countries: how reliable are the basic data sources? *Pediatr Pulmonol* **47**, 211–219 (2012).
460. Granell, R., Henderson, A. J. & Sterne, J. A. Associations of wheezing phenotypes with late asthma outcomes in the Avon Longitudinal Study of Parents and Children: A population-based birth cohort. *J Allergy Clin Immunol* **138**, 1060–1070.e11 (2016).
461. Berlin, K. S., Parra, G. R. & Williams, N. A. An Introduction to Latent Variable Mixture Modeling (Part 2): Longitudinal Latent Class Growth Analysis and Growth Mixture Models. *Journal of Pediatric Psychology* **39**, 188–203 (2014).
462. Lanari, M. *et al.* Risk factors for bronchiolitis hospitalization during the first year of life in a multicenter Italian birth cohort. *Italian Journal of Pediatrics* **41**, 40 (2015).
463. DEFRA. *Modelled background pollution data*. <https://uk-air.defra.gov.uk/data/pcm-data> (2021).
464. Straatmann, V. S. *et al.* How do early-life factors explain social inequalities in adolescent mental health? Findings from the UK Millennium Cohort Study. *J Epidemiol Community Health* **73**, 1049 (2019).
465. Inoue, Y. & Shimojo, N. Epidemiology of virus-induced wheezing/asthma in children. *Front Microbiol* **4**, 391–391 (2013).
466. Collin, S. M. *et al.* Pet ownership is associated with increased risk of non-atopic asthma and reduced risk of atopy in childhood: findings from a UK birth cohort. *Clinical & Experimental Allergy* **45**, 200–210 (2015).
467. Man, W. H. *et al.* Infant respiratory syncytial virus prophylaxis and nasopharyngeal microbiota until 6 years of life: a subanalysis of the MAKI randomised controlled trial. *The Lancet Respiratory Medicine* **8**, 1022–1031 (2020).
468. Brunwasser, S. M. *et al.* Assessing the strength of evidence for a causal effect of respiratory syncytial virus lower respiratory tract infections on subsequent wheezing illness: a systematic review and meta-analysis. *The Lancet Respiratory Medicine* **8**, 795–806 (2020).
469. Verwey, C., Nunes, M. C., Dangor, Z. & Madhi, S. A. Pulmonary function sequelae after respiratory syncytial virus lower respiratory tract infection in children: A systematic review. *Pediatric Pulmonology* **55**, 1567–1583 (2020).
470. Thompson, S. S. & Sloan, C. Seasonality and Virus Transmission. in *C57. VIRAL LUNG INFECTION CLINICAL STUDIES AND CASE REPORTS* A5216–A5216 (American Thoracic Society, 2019). doi:10.1164/ajrccm-conference.2019.199.1_MeetingAbstracts.A5216.
471. Dales, R. E., White, J., Bhumgara, C. & McMullen, E. Parental reporting of childrens' coughing is biased. *European Journal of Epidemiology* **13**, 541–545 (1997).
472. Government Equalities Office. Government updates on identifying and tackling COVID-19 disparities. <https://www.gov.uk/government/news/government-updates-on-identifying-and-tackling-covid-19-disparities> (2021).
473. Gill, P. J. *et al.* Increase in emergency admissions to hospital for children aged under 15 in England, 1999–2010: national database analysis. *Arch Dis Child* **98**, 328 (2013).
474. Public Health England. Research and analysis: Sources of COVID-19 surveillance systems. <https://www.gov.uk/government/publications/national-covid-19-surveillance-reports/sources-of-covid-19-systems> (2020).

- 475.UCL. PICNIC Study. <https://www.ucl.ac.uk/child-health/research/population-policy-and-practice-research-and-teaching-department/cenb-clinical-12> (2021).
- 476.Royal College of Paediatrics and Child Health. *The inside story: Health effects of indoor air quality on children and young people*. <https://www.rcpch.ac.uk/resources/inside-story-health-effects-indoor-air-quality-children-young-people> (2020).
- 477.To, T. *et al.* Early life exposure to air pollution and incidence of childhood asthma, allergic rhinitis and eczema. *Eur Respir J* **55**, 1900913 (2020).
- 478.Goldizen, F. C., Sly, P. D. & Knibbs, L. D. Respiratory effects of air pollution on children. *Pediatric Pulmonology* **51**, 94–108 (2016).
- 479.WHO. WHO Coronavirus (COVID-19) Dashboard. <https://covid19.who.int/> (2021).
- 480.Public Health England. *Official Statistics Weekly national flu reports: 2019 to 2020 season*. 2020 <https://www.gov.uk/government/statistics/weekly-national-flu-reports-2019-to-2020-season>.
- 481.Yeoh, D. K. *et al.* The impact of COVID-19 public health measures on detections of influenza and respiratory syncytial virus in children during the 2020 Australian winter. *Clin Infect Dis* ciaa1475 (2020) doi:10.1093/cid/ciaa1475.
- 482.NHS Digital. *Provisional Monthly Hospital Episode Statistics for Admitted Patient Care, Outpatient and Accident and Emergency data April 2020 - December 2020*. <https://digital.nhs.uk/data-and-information/publications/statistical/hospital-episode-statistics-for-admitted-patient-care-outpatient-and-accident-and-emergency-data/april-2020---december-2020> (2021).
- 483.Department of Health and Social Care & The Scottish Government. *Palivizumab passive immunisation against respiratory syncytial virus (RSV) in at-risk pre-term infants*. <https://www.sehd.scot.nhs.uk/publications/DC20201111RSV.pdf> (2020).
- 484.Jones, N. How Covid-19 is changing the cold and flu season. *Nature* **588**, 388–390 (2020).
- 485.Harron, K., Goldstein, H. & Dibben, C. Introduction. in *Introduction. In Methodological Developments in Data Linkage* (2016).
- 486.Winkler, W. E. Probabilistic linkage. in *Methodological Developments in Data Linkage* (2016).
- 487.Kennedy, C. M. *et al.* Associations of mobile source air pollution during the first year of life with childhood pneumonia, bronchiolitis, and otitis media. *Environmental Epidemiology* **2**, (2018).
- 488.Laugier, O. *et al.* Influence of Socioeconomic Context on the Rehospitalization Rates of Infants Born Preterm. *The Journal of Pediatrics* **190**, 174-179.e1 (2017).
- 489.Puig, C. *et al.* Incidence and risk factors of lower respiratory tract illnesses during infancy in a Mediterranean birth cohort. *Acta Paediatrica* **97**, 1406–1411 (2008).