Using administrative linked datasets to explain differences in child mortality between England and Sweden

Anna Maria Zylbersztejn

A thesis submitted for the degree of
Doctor of Philosophy

UCL Great Ormond Street Institute of Child Health,
University College London (UCL)

The Farr Institute of Health Informatics Research

March 2018

Declaration

I, Anna Maria Zylbersztejn 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

Background: Child mortality (under-5 years old) is almost twice as high in England as in Sweden. Policy makers need to know whether preventive strategies should address adverse birth characteristics (e.g., preterm birth, low birth weight), or focus on care after birth. This PhD used administrative linked datasets in England and Sweden to determine the contribution of birth characteristics and socio-economic factors to inter-country differences in child mortality.

Methods: I developed nationally-representative birth cohorts using an administrative hospital database in England, and a medical birth register in Sweden for births in 2003-2012, with longitudinal follow-up from linked hospitalisation and mortality records. I compared all-cause mortality, and mortality from potentially preventable causes in England relative to Sweden using Cox proportional hazards regression models. The models were adjusted for birth characteristics (gestational age, birth weight, sex, congenital anomalies), and socio-economic factors (maternal age and socio-economic status).

Results: Birth characteristics accounted for 77% and 68% of excess risk of death in England at 2-27 days and 28-364 days, respectively. Socio-economic factors contributed a further 3% and 11%, respectively. After adjustment for all risk factors, small but statistically significant differences in mortality remained in infancy; the differences were negligible, however, at 1-4 years.

The risk of respiratory tract infection-related mortality at 31-364 days in England relative to Sweden decreased from 50% to 16% after adjusting for birth characteristics, and from 58% to 32% at 1-4 years. A third of the excess mortality from sudden unexpected infant deaths in England was explained by each birth characteristics and socio-economic factors.

Conclusions: The biggest reductions in child mortality in England relative to Sweden could be achieved by reducing the prevalence of adverse birth characteristics. Policies to reduce child mortality in England should focus on improving the health of women and reducing socio-economic disadvantage before and during pregnancy.

Streszczenie

Wprowadzenie: Umieralność dzieci poniżej piątego roku życia w Anglii jest niemal dwa razy wyższa niż w Szwecji. Aby zmniejszyć liczbę dziecięcych zgonów w Anglii, należy ocenić, czy działania prewencyjne podejmowane w tym celu powinny skupić się na poprawie stanu zdrowia noworodków (np. poprzez zmniejszenie liczby przedwczesnych porodów, porodów z niską masą urodzeniową), czy też na poprawie jakości opieki nad dziećmi po narodzinach.

Cel: Celem tego doktoratu było ustalenie w jakim stopniu wysoka śmiertelność dzieci w Anglii w stosunku do Szwecji wynika z różnic w stanie zdrowia noworodków i czynnikach społeczno-ekonomicznych w obu krajach. Analiz dokonano korzystając z połączonych administracyjnych baz danych w Anglii i Szwecji.

Metody: Stworzyłam krajowe kohorty narodzin w Anglii (korzystając z bazy danych o przyjęciach do szpitala) i w Szwecji (korzystając z narodowego rejestru narodzin) w latach 2003-2012. Kohorty były połączone z historią hospitalizacji dzieci i informacją o zgonach do piątego roku życia.

Śmiertelność dzieci w Anglii i Szwecji (uwzględniając wszystkie przyczyny śmierci oraz oddzielnie przyczyny którym można potencjalnie zapobiec) była porównana modelem proporcjonalnego hazardu Coxa, z uwzględnieniem wpływu na śmiertelność wyznaczników zdrowia noworodków (masa ciała, długość ciąży, płeć, wykryte wady wrodzone) i czynników społeczno-ekonomicznych (wiek matki, status społeczno-ekonomiczny).

Wyniki: 77% i 68% podwyższonego ryzyka śmierci w wieku 2-27 dni i 28-364 dni w Anglii w stosunku do Szwecji wynikało z różnic w stanie zdrowia noworodków, dodatkowe 3% i 11% było wytłumaczone przez różnice w czynnikach społeczno-ekonomicznych między tymi dwoma krajami. Małe, ale znamienne statystycznie różnice w umieralności niemowląt pozostały po uwzględnieniu wszystkich czynników ryzyka w końcowym modelu. Umieralność w wieku 1-4 lat była porównywalna miedzy Anglią i Szwecją w końcowym modelu.

Ryzyko śmierci powiązanej z infekcjami układu oddechowego w wieku 31-364 dni i 1-4 lat w Anglii w stosunku do Szwecji obniżyło się odpowiednio z 50% do 16% i z 58% do 32% po uwzględnieniu różnic w zdrowiu noworodków. Różnice w zdrowiu noworodków i czynniki społeczno-ekonomiczne wytłumaczyły dwie trzecie podwyższonego ryzyka nagłej i niespodziewanej śmierci wśród niemowląt.

Wnioski: Największe obniżenie śmiertelności dzieci w Anglii w porównaniu ze Szwecją można osiągnąć poprzez poprawę stanu zdrowia noworodków. W tym celu należy

podjąć działania w zakresie ochrony i promocji zdrowia matek i zmniejszenia nierówności społeczno-ekonomicznych przed i w trakcie ciąży.

Acknowledgements

Firstly, I would like to thank my supervisors Pia Hardelid, Ruth Gilbert and Anders Hjern, for their encouragement, support and enthusiasm for my work over past three years. I have learnt so much from working with you and it has been a wonderful adventure. I would particularly like to thank my primary supervisor, Pia, who came up with the initial idea to compare child mortality in England and Sweden using administrative datasets, which led to my PhD.

Thank you to everyone I meet in CHESS during my visits to Stockholm, for making me feel welcome and introducing me to Friday "fika" (coffee and cake break). A particularly big thank you to Anders Hjern and Can Liu for their invaluable help with organising my visits to Sweden and for answering my limitless questions about the Swedish National Registers.

I would also like to acknowledge my colleagues from London, especially Linda Wijlaars and Katie Harron for their help with mother-baby linkage in HES, which turned to be invaluable for developing the English birth cohort; and fellow PhD students Rachel Reeves and Louise McGrath-Lone – we shared many milestones throughout our PhDs and you were the best PhD support team.

I am grateful to the Farr Institute of Health Informatics Research and Administrative Data Research Centre who financially supported my PhD. I would also like to thank Louise, Greg and Calum, who helped with proof reading and editing.

Last but not least, I would like to thank my family and friends for their huge support throughout. A special thank you to my parents, Maria and Michał, for encouraging me early on to study abroad and eventually to pursue a PhD, and for their continued interest and enthusiasm for my work – I couldn't have done it without you. Many thanks to my family in Lyon, Adam, Alexandra, Vincent and Bianka; and to Jess. Lastly, special thanks to Calum, for your patience, good humour and support through ups and downs – you're my rock.

Table of contents

List of Figures

List of Tables

List of Equations

List of Boxes

Abbreviations

BMI	Body Mass Index
CA	Congenital Anomaly
CHESS	Centre for Health Equity Studies
CI	Confidence Interval
DRG	Diagnosis-Related Group
EU	European Union
EUROCAT	European Surveillance of Congenital Anomalies network
GDP	Gross Domestic Product
GP	General Practitioner
HES	Hospital Episode Statistics
HESID	Hospital Episode Statistics Identifier
HiB	Haemophilus Influenzae type B
HR	Hazard Ratio
HRG	Healthcare Resource Groups
ICD-10	International Statistical Classification of Diseases and Related Health Problems version 10
IMD	Index of Multiple Deprivation
INRICH	International Network for Research on Inequalities in Child Health
IUGR	Intrauterine Growth Restriction
IVF	*In Vitro* Fertilisation
LISA	Longitudinal Integration Database for Health Insurance and Labour Market Studies
LMP	Last Menstrual Period
LSOA	Lower Layer Super Output Area
MCDS	Maternity and Children's Data Set
MI	Multiple Imputation
MNAR	Missing Not at Random
NBHW	National Board of Health and Welfare
NCSP	Nordic Medico-Statistical Committee Classification of Surgical Procedures
NHS	National Health Service
NN4B	NHS Number for Babies

NNRD	National Neonatal Research Database
NNU	Neonatal Unit
NOMESCO	Nordic Medico-Statistical Committee
NorCAS	Northern Congenital Abnormality Survey
OECD	Organisation for Economic Co-operation and Development
ONS	Office of National Statistics
OPCS	Office of Population Censuses and Surveys Classification of Interventions and Procedures
PbR	Payment by Results
PDS	Personal Demographics Service
PERM	Percentage of Excess Risk Mediated
PH	Proportional Hazards
PICANet	Paediatric Intensive Care Audit Network
PIN	Personal Identity Numbers
PPP	Purchasing Power Parity
RSV	Respiratory Syncytial Virus
RTI	Respiratory Tract Infections
SCDR	Swedish Cause of Death Register
SD	Standard Deviation
SES	Socio-economic Status
SHDR	Swedish Hospital Discharge Register
SIDS	Sudden Infant Death Syndrome
SMBR	Swedish Medical Birth Register
SUDI	Sudden Unexpected Death in Infancy
TOP	Termination of Pregnancy
UK	United Kingdom
USA	United States of America
WHO	World Health Organization

Chapter 1. Thesis background and rationale

1.1 Chapter overview

The aim of this chapter is to describe the rationale for comparing mortality in children aged less than five years old in England and Sweden using administrative linked datasets. Much of the evidence to date focused on comparing child or infant mortality in the United Kingdom (UK) relative to Sweden. This PhD, however, largely focuses on England due to data availability. Because England is the biggest and the most diverse of the four UK countries (it covers 85% of births and child deaths in the UK),[1] evidence from comparisons of the UK and Sweden is highly relevant for this thesis. Where possible I report information for England only.

I first present an overview of child mortality in the UK and Sweden, and discuss hypothesised origins of the differences in mortality between the two countries (Section 1.2). Next, I summarise birth characteristics, maternal risk factors operating during pregnancy and risk factors operating after birth which contribute to an increased risk of child death, and compare their prevalence in England and Sweden. I also discuss differences in upstream determinants of child health such as social determinants, welfare policies or organisation of healthcare (Section 1.3). I then describe the rationale for this thesis (Section 1.4), and I set out the overall aim and specific objectives of this thesis (Section 1.5). Finally, in Section 1.6, I describe the structure of the thesis.

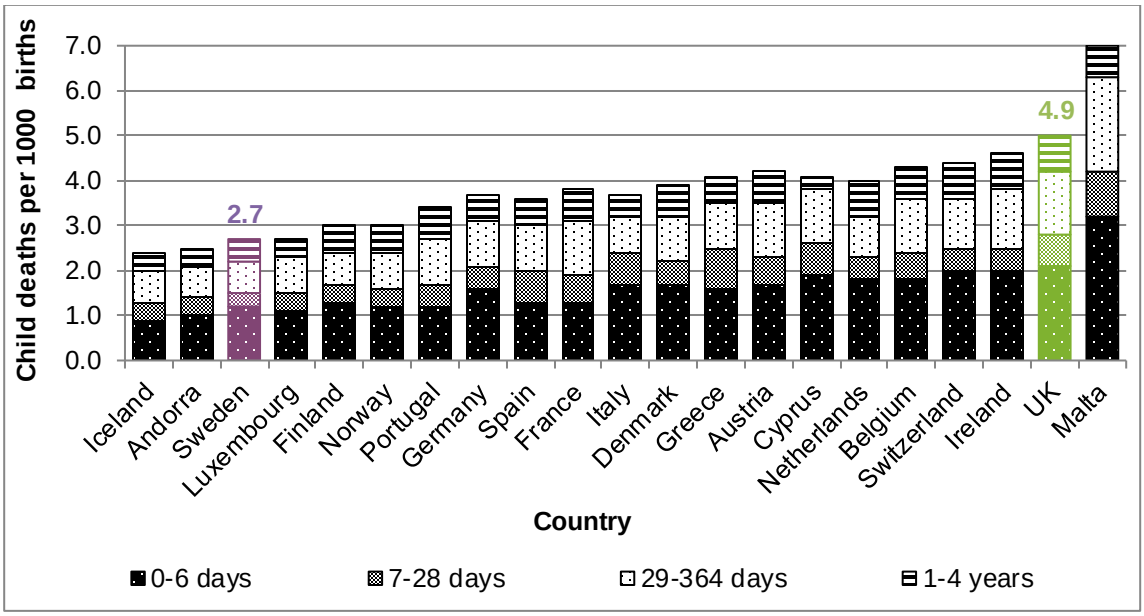
Throughout this thesis, *child mortality* refers to deaths in children aged less than 5 years old (i.e. before their fifth birthday). Unless stated otherwise, *neonatal mortality* refers to deaths in children aged 0-27 days, *post-neonatal mortality* refers to deaths at 28-364 days and *early childhood mortality* relates to deaths at ages 1-4 years.

1.2 Child mortality in the UK compared to Sweden

1.2.1 Overview

The UK has one of the highest child mortality rates in Western Europe (Figure 1.1).[2–4] In 2013, child mortality in the UK was almost twice as high as in Sweden (4.9 deaths/1000 births compared to 2.7/1000 births, respectively).[3] The UK and Sweden are both high-income economies with similar levels of gross domestic product (GDP) per capita and comparable government spending on health (see Table 1.1), and universally accessible healthcare.[5] Thus, Sweden is often viewed as a benchmark for the levels of reductions in child mortality which should be achievable in the UK.[6–10]

Figure 1.1 – Child mortality per 1000 births in Western Europe in 2013



UK=United Kingdom. Data are number of deaths per 1000 live births, tabulated by age at death. Deaths in the first year of life accounted for 86% of all child deaths in the UK and 81% in Sweden. This figure is based on data from Wang et al.[3] Note that Wang et al. used different cut-offs for neonatal and post-neonatal deaths than those used in this PhD (0-28 days and 29-364 days vs 0-27 days and 28-364 days, respectively).

Table 1.1 – Selected indicators of economic development in the UK and Sweden, compared to average for OECD countries

Indicator	UK	Sweden	OECD average
GDP per capita in 2013 based on current prices and current PPP *(in USA dollars)*	38,743	44,586	37,815
Household gross adjusted disposable income per capita in 2013 *(in USA dollars at current PPP)*	28,669	30,124	N/A
Total government expenditure of health in 2012 *(as % of GDP)*	7.5%	6.9%	N/A

GDP=gross domestic product; N/A=not available; OECD=Organisation for Economic Co-operation and Development; PPP=purchasing power parity; UK=United Kingdom; USA=United States of America. Data comes from OECD "National Accounts at a Glance" publication.[5]

Researchers commonly report the number of excess child deaths in the UK compared to Sweden.[7–10] This number refers to deaths that could have been prevented if the UK had the same child mortality rates as Sweden.[7–10] In 2013, the differences in child mortality between Sweden and the UK accounted for 1,713 excess deaths in the UK (out of 3,816 child deaths in total).[3,11] A key question for the policy makers is: where do these differences originate?

1.2.2 Hypothesised origins of higher child mortality rates in the UK relative to Sweden

There is limited evidence about the origins of differences in child mortality in the UK and Sweden. I identified 14 international comparisons of child mortality published since 1st January 2000, which included England or the UK and Sweden in their analyses (details of search strategy are presented in Appendix A). Of these studies, five described hypothesised origins of increased child mortality in the UK relative to Sweden (Table 1.2). Three of these studies compared cause-specific mortality (based on the underlying cause of death recorded at death registration), and two were descriptive.

Table 1.2 – Summary of studies which compared child mortality in the UK and in Sweden published in 2000-2017

Study (main author)	Study description	Study type	Compared statistics	Risk factors adjusted for
Wolfe[8]	Overview of the differences in organisation and provision of child healthcare services in the UK relative to selected European countries	Descriptive study	- Number of excess deaths in the UK relative to Sweden at 0-14 years - Mortality rates at 0-14 years for deaths from meningococcal disease, pneumonia and asthma	Age at death Underlying cause of death
Wolfe[7]	Overview of the differences in organisation and provision of child healthcare services in Europe	Descriptive study	- Number of excess deaths in the UK relative to Sweden at 0-14 years - Levels of spending on social protection for families and child death rates - Mortality rates at 0-14 years for deaths from pneumonia and asthma	Age at death Underlying cause of death
Wolfe[10]	Overview of child mortality in the UK	Descriptive study	- Number of excess deaths in the UK relative to Sweden at 0-14 years - Prevalence of preterm birth, low birth weight, teenage pregnancy, maternal smoking and child poverty	None
Wolfe[9]	Overview of child mortality in the UK	Descriptive study	- Number of excess deaths in the UK relative to Sweden at 0-14 years	None
Tambe[6]	Comparison of cause-specific child mortality rates in the UK and in Sweden	Observa-tional study	- cause-specific mortality rates at 0-4 years (deaths were grouped based on the underlying cause of death)	Age at death Underlying cause of death

UK=United Kingdom

It has been suggested that the differences in child mortality between the UK and Sweden reflect wider socio-economic inequalities in the UK, leading to higher rates of adverse birth characteristics (such as preterm birth or presence of congenital anomalies).[6–10] Preterm birth is the leading cause of child death in the UK, accounting for approximately 7% of all births but one fifth of all child deaths.[12] In 2006-08, the consequences of preterm birth accounted for 138.5 deaths/100,000 births in the UK.[6] This rate was almost 14 times higher than in Sweden (10.1/100,000 births).[6] Congenital anomalies are the second most common cause of death in the UK, and the leading cause of death in Sweden according to death registration data.[6] In 2006-8, 112.1 children died from a congenital anomaly per 100,000 births in the UK, compared to 88.6/100,000 births in Sweden.[6] Since preterm birth and congenital anomalies are more common among the most deprived mothers,[13,14] the differences in mortality from these two causes have been attributed to wider socio-economic inequalities in the UK, leading to an increased proportion of babies born preterm or with a congenital anomaly.[6,9,10] For example, in 2008-2010, the least deprived 20% of the UK's population had seven times higher income than the most deprived 20% of the population, compared with an approximately four-fold difference in Sweden.[15]

Wide differences in mortality rates were also observed for deaths due to disorders that could be amenable to healthcare. For example, infections are the third most common cause of death in both countries.[6] In 2006-08, infection-related mortality was almost twice as high in the UK as in Sweden (63.9 deaths/100,000 births compared to 34.8/100,000 births).[6] Infection-related deaths are considered to be healthcare amenable, as they can be prevented through vaccination programs, and timely antibiotic treatment.[16–18] The differences in infection-related mortality were, therefore, attributed to delays in the diagnosis of acute life threatening infections.[6] Respiratory disorders are also a more common cause of death in the UK than in Sweden. In 2006-8, the UK had almost seven times higher mortality from paediatric respiratory disorders than Sweden (5.9/100,000 births vs 0.9/100,000 births), and four times higher mortality from neonatal respiratory disorders (34.2/100,000 births vs 8.9/100,000 births).[6] Because respiratory conditions are often managed in the primary care setting, increased mortality due to these conditions has been attributed to differences in organisation and provision of child health services. Child health professionals have, therefore, called for a better integration of primary care and paediatric services and additional paediatric training for general practitioners (GP) beyond that received during undergraduate studies, to reduce mortality from conditions seen as amenable to healthcare.[6,8,9]

1.2.3 Gaps in current research

Previous comparisons of child mortality rates in the UK and Sweden were based either on unadjusted all-cause mortality rates (used to calculate the number of excess deaths in the UK) or on cause-specific mortality rates, calculated using data aggregated by the underlying cause of death (see fourth column of Table 1.2). Such data are routinely collected by international agencies such as the World Health Organization (WHO), and readily available for studies.[19] However, comparisons of cause-specific mortality based on such aggregated data provide limited evidence about the origins of differences in child mortality between countries.

First, comparisons based on cause-specific mortality do not account for inter-country differences in the distribution of birth characteristics such as gestational age, birth weight or the presence of congenital anomalies. Without adjustment for birth characteristics, it is not possible to determine whether the differences in mortality due to preterm birth or congenital anomalies reflect an increased prevalence of these risk factors in the UK relative to Sweden, or differences in the quality of care that these vulnerable babies receive after birth. Adverse birth characteristics, such as preterm birth or low birth weight, can also make babies more susceptible to infections and respiratory illness.[20,21] Therefore, some of the excess mortality from infections or respiratory disorders in the UK relative to Sweden could be in part explained by the increased prevalence of adverse birth characteristics, rather than by poor performance of the healthcare system.

Second, comparisons of child mortality based on the death registration data are prone to bias due to inter-country differences in death certification practices. For example, in England neonatal deaths are certified using a neonatal death certificate, which gives equal weighting to health conditions of the mothers and the babies that contributed to death.[22] In Sweden, a standard death certificate is used for all deaths (regardless of the age at death), which details the sequence of health conditions that led directly to death, and any additional conditions that contributed to death, but were not part of the causal sequence ending in death.[23] Different coding rules apply to the selection of the underlying cause of death for the two types of death certificates.[24] Furthermore, international comparisons of crude child mortality rates are susceptible to bias due to differences in the reporting of live births, stillbirths and deaths occurring around the time of birth.[25,26] Some of the differences in cause-specific mortality could, therefore, be due to data artefacts.

For a fair comparison, we need to account for differences in the prevalence of key risk factors at birth such as preterm birth or the presence of congenital anomalies. An adjusted comparison of child mortality rates can inform policy makers as to whether the

excess child mortality in the UK relative to Sweden can be attributed to an increased prevalence of risk factors operating before and during pregnancy, affecting the healthy development of a foetus, or to differences in the care of babies after birth, given their birth characteristics. Such a distinction is not obvious when comparing unadjusted all-cause or cause-specific child mortality rates. Therefore, any suggested explanations for increased child mortality in the UK relative to Sweden remain speculative.

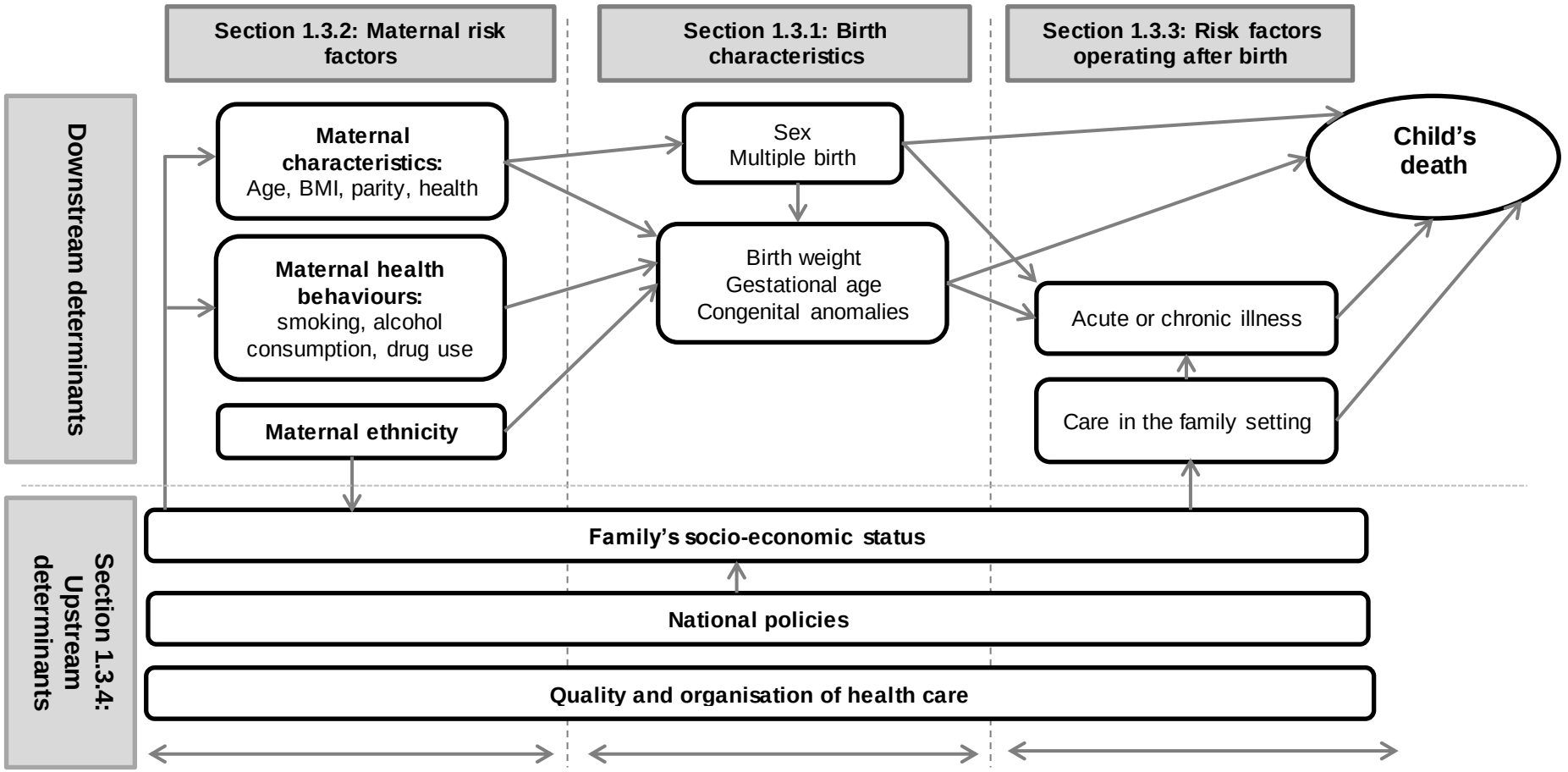
1.3 Risk factors associated with child mortality in high-income countries

Child mortality is associated with a range of risk factors operating before and during pregnancy, as well as after birth, which are described in this section. Figure 1.2 presents an overview of the key risk factors associated with the risk of child death, and maps them out to subsections in this chapter.

Characteristics of a child at birth are key determinants of child mortality.[27] For example, boys, babies born prematurely, with a low birth weight or with a congenital anomaly, have an increased risk of death.[27] In Section 1.3.1, I describe the key birth characteristics associated with the risk of child death and compare their prevalence in England and Sweden. The healthy development of a foetus during pregnancy is in turn associated with maternal health during pregnancy. The differences in the characteristics of mothers in England and Sweden are described in Section 1.3.2. After birth, the risk of child death is associated with acute and chronic illness, and with the care received in a home setting. These factors are described in Section 1.3.3. Finally, in Section 1.3.4 I discuss the upstream determinants of child mortality, which operate throughout the life course. These include a family's socio-economic circumstance, national welfare policies and the healthcare system.

Previous comparisons of child mortality described in Section 1.2 focussed on the UK and Sweden. This thesis, however, focuses on England. Therefore, the evidence presented in the next section is largely based on data for England, or England and Wales where more detailed data were not available (Wales contributes to approximately 4% of births and child deaths in the UK).[1]

Figure 1.2 – Overview of risk factors associated with the risk of child death in high-income countries, which could contribute to the differences in child mortality between UK and Sweden. Each sub-section from Section 1.3 is mapped onto the figure.



BMI=Body Mass Index; UK=United Kingdom.

1.3.1 Birth characteristics

The key birth characteristics associated with an increased risk of child death described in this section are: gestational age, birth weight, presence of congenital anomalies, sex and multiple birth.[12,27] A table summarising all presented evidence is available at the end of Section 1.3.1 (Table 1.5).

1.3.1.1 Gestational age

Gestational age is a key determinant of the risk of a child's death, especially in infancy. The risk of death is inversely related with gestational age, with the highest mortality rates associated with preterm birth (that is, birth before 37 weeks of gestation, see Box 1.1 for classification of gestational age used in this thesis), as indicated by gestation-specific mortality rates presented in Table 1.3.[12] In 2010, preterm births accounted for 57% of all infant deaths in Sweden and 61% of all infant deaths in England and Wales, despite affecting only 5.9% and 7.1% of live births, respectively.[12] Preterm birth can be spontaneous, or induced due to maternal or foetal complications; induced preterm birth accounts for approximately 30-35% of all preterm births in high-income countries.[28]

Table 1.3 – Gestation-specific infant mortality rates per 1000 live births in England and Wales, and in Sweden in 2010

	England and Wales			Sweden		
Gestational age (weeks)	**% of live births**	**% of infant deaths**	**Infant mortality rate**	**% of live births**	**% of infant deaths**	**Infant mortality rate**
<28	0.4%	36.7%	353.2	0.3%	32.0%	238.6
28-31	0.8%	10.3%	48.4	0.5%	7.9%	35.1
32-36	6.3%	13.6%	8.8	5.0%	17.3%	8.3
≥37	92.5%	39.3%	1.6	94.1%	42.8%	1.1
TOTAL (n)	711,365	2,686		114,706	278	

Data are % of all live births, % of infant deaths, and infant deaths per 1000 live births. Information comes from the EURO-PERISTAT project,[12] and is based on gestational age recoded at birth.

Box 1.1 – Gestational age classification (based on the number of completed weeks) used throughout this thesis

- <37 weeks: preterm birth
  - <28 weeks: extremely preterm birth
  - 28-31 weeks: very preterm birth
  - 32-36 weeks: moderate to late preterm birth
- 37-41 weeks: term birth
- ≥42 weeks: post-term birth

Extremely and very preterm birth

Babies born extremely prematurely (at <28 weeks' gestation) and very prematurely (at 28-31 weeks) have the highest risk of death. In 2010, babies born at <28 weeks and at 28-31 weeks were over 200 times and 30 times more likely to die in infancy than babies born at ≥37 weeks, respectively, both in England, Wales, and Sweden (see Table 1.3).[12] A study from Western Australia showed that babies born at 24-31 weeks also had 40% higher risk of death at 1-5 years of life compared to term babies (risk ratio: 1.4, 95% confidence interval (CI): 0.7-3.0); however, the differences were not statistically significant (likely due to insufficient sample size).[29] In 2010, mortality in infants born at <28 weeks and 28-31 weeks was 46% and 37% higher in England and Wales, than in Sweden (Table 1.3).[12]

Babies born at <32 weeks' gestation are also the most likely to suffer from long term morbidity due to prematurity, including increased susceptibility to infections and lung disease, and neurological impairments such as cerebral palsy and visual and auditory deficits.[14,20,28] The increased risk of mortality and morbidity in children born at <32 week's gestation is a result of the immaturity of baby's organs (in particular, brain and lungs), which are not sufficiently developed to support baby's growth outside the womb.

Moderate and late preterm birth

Moderate and late preterm births (at 32-36 weeks' gestation) carry much lower risks compared to births at <32 weeks. Compared to term babies, however, they have an increased risk of infant death (five times higher in England and Wales, and seven times higher in Sweden in 2010, Table 1.3),[12] and of other neonatal morbidity, such as jaundice, temperature instability, respiratory distress and feeding difficulties.[20,21,30] Two studies from Australia suggested that babies born at 32-36 weeks' gestation also have an increased risk of death beyond infancy compared to term babies; however, the differences were not statistically significant (20% higher risk according to study of births in 1980-2010 in Western Australia,[29] and 47-48% higher according to a study of singleton live births in 2001-2010 in New South Wales).[31] In 2010, mortality in moderate and late preterm births was only 5% higher in England and Wales, than in Sweden (Table 1.3).[12]

Term births

In term babies (born at 37-41 weeks), the risks of neonatal and infant mortality and other neonatal morbidity are higher for early term babies (at 37-38 weeks' gestation) than for full term babies (born at 39-41 weeks' gestation).[32,33] In 2010, infant mortality in early term babies was twice as high as for full term babies in England and Wales (2.8 deaths/1000 births vs. 1.3/1000 births).[34] This has important health implications as in

many Western countries there has been a shift towards early term elective delivery, instead of postponing the delivery until full term.[32,35]

Post-term births

Post-term babies (born at ≥42 weeks' gestation) have an increased risk of stillbirth and neonatal death relative to full term babies, but the differences diminish in the post-neonatal period.[32] In England and Wales, post-term births had 27% higher neonatal mortality rate than full term births in 2010 (0.8/1000 births compared to 0.6/1000 births in full term babies).[34] Babies born post-term are also more likely to experience obstetric complications such as birth asphyxia, peripheral nerve damage, umbilical cord complications, bone fracture or aspiration (with odds ratios for these conditions in the range of 1.75-2.13, according to a study of singleton births in 1978-1993 in Denmark).[36]

The rates of post-term birth are higher in Sweden than in England and Wales (6.6% of live births compared to 4.1% of live births in 2010).[12] This is likely to reflect differences in management of overdue pregnancies: in England, overdue births are induced at 41 weeks' gestation,[37] while in Sweden they are induced at the minimum of 42 weeks.[38]

1.3.1.2 Birth weight

Birth weight is another key determinant of infant and child mortality.[27] Box 1.2 summarises the classification of birth weight used throughout this thesis.

Box 1.2 – Classification of birth weight used throughout this thesis

- <1000g: extremely low birth weight
- <2500g: low birth weight
- ≥2500g: normal birth weight

Low birth weight

Like preterm birth, low birth weight (i.e. <2500g) accounts for a small proportion of live births (4.2% in Sweden and 7.0% in England and Wales in 2010), but it is one of the key determinants of child mortality, accounting for approximately 60% of all infant deaths (see Table 1.4).[12] Low birth weight can be a result of preterm birth, whereby the child has a low birth weight which is appropriate for their gestational age but they are born too early, or intrauterine growth restriction (IUGR), where children are too light for their gestational age.[14,39]

Table 1.4 – Birth weight-specific infant mortality rates per 1000 live births in England and Wales, and in Sweden in 2010

Birth weight category (g)	England and Wales			Sweden		
	% of live births	% of infant deaths	Infant mortality rate	% of live births	% of infant deaths	Infant mortality rate
<500	0.2%	6.4%	120.6	0.03%	8.0%	552.6
500-1499	1.0%	39.7%	141.5	0.7%	31.0%	97.0
1500-2499	5.8%	16.3%	10.0	3.4%	19.5%	13.1
≥2500	93.0%	37.5%	1.4	95.8%	41.4%	1.0
TOTAL (n)	712,938	2,548		114,498	261	

Data are % of all live births, % of infant deaths, and infant deaths per 1000 live births. Data comes from the EURO-PERISTAT project,[12] and is based on birth weight recoded at birth.

The risk of infant and child death decreases as birth weight increases. In England and Wales, babies born in 1993-2011 weighing 500-1499g at birth had a 145 times higher risk of death in infancy, and nearly seven times higher risk of death at 1-18 years compared to babies with a birth weight of ≥3500g (after accounting for sex, maternal age, multiple birth, and an area level indicator of deprivation); babies with a birth weight of 1500-2499g had 9.8 and 2.9 times higher risks of death in infancy and at 1-18 years, respectively.[40]

Normal birth weight

For babies with normal birth weight (≥2500g), the risk of mortality and morbidity also decreases as birth weight increases, and a birth weight of 3500-4499g is seen as 'optimal' for long term health outcomes.[41] In 2010, infant mortality in England and Wales was 3.3/1000 births for a birth weight of 2500-2999g, 1.4/1000 births for a birth weight of 3000-3499g, 0.9/1000 births for a birth weight of 3500g-3999g and 1.0 for a birth weight of ≥4000g.[42] Therefore, the differences in the distribution of normal birth weight could also contribute to the increased child mortality rate in England and Wales, relative to Sweden: in 2010, 49.2% of all births in Sweden weighed 3500-4499g,[43] compared to 38.7% in England and Wales.[44]

The risk of neonatal mortality and stillbirth is higher for babies weighing ≥4500g than for those with birth weight of 3500-4499g, due to an increased risk of birth trauma and neonatal morbidity.[41,45] However, such births are rare – in 2010, they accounted for 3.6% of births in Sweden and 1.7% of births in England and Wales.[12]

1.3.1.3 Congenital anomalies

Prevalence of congenital anomalies in the UK, England, and Sweden

Congenital anomalies are one of the three leading causes of deaths in infancy and early childhood in the UK, England and Sweden (according to death registration data). In 2011, infant mortality due to congenital anomalies was 103.5 deaths/100,000 births in the UK, compared to 45.6/100,000 births in Sweden; the corresponding figures for deaths at 1-4 years were 11.1/100,000 births and 8.1/100,000, respectively.[19] According to the EUROCAT, a European network of population-based congenital anomaly registries, the prevalence of congenital anomalies in all live births is higher in England than in Sweden (2.0% of live births in England in 2010, based on data from six registers covering 31.9% of all births, compared to 1.7% of live births in Sweden, based on data from a whole-country register of congenital anomalies).[46]

The risk of death and the prevalence of congenital anomalies in live born babies vary between different types of anomalies. According to the UK Northern Congenital Abnormality Survey register (NorCAS, covering a region with approximately 35,000 deliveries a year in the North England), three most common groups of anomalies recorded for babies born in 1985-2003 were: anomalies of cardiovascular system (accounting for 39.1% of all live births with any anomaly in the register), chromosomal abnormalities (12.2%), and anomalies of urinary system (11.5%).[47] Children with these anomalies had high survival rates: 91.1%, 81.1% and 93.5% of these children lived until the fifth birthday, respectively.[47] Five-year survival was lowest for anomalies of respiratory system (64.3%), skeletal dysplasia (65.3%) and anomalies of the nervous system (71.7%), but the prevalence of these anomalies was low in live births (0.8%, 0.5% and 5.0%, respectively).[47] The most common groups of congenital anomalies in Sweden were comparable with those reported by NorCAS for England: congenital heart defects accounted for 39% of live births with an anomaly in 2010, genital anomalies contributed 16% and chromosomal abnormalities a further 11%, according to the EUROCAT.[46]

Terminations of pregnancy for foetal anomaly

The prevalence of congenital anomalies is linked to rates of terminations of pregnancy (TOP) due to foetal anomalies detected during pregnancy. The overall rate of TOP for foetal anomaly was comparable between England and Sweden (accounting for 24.4% of all pregnancies with a diagnosed anomaly in England, and 22.7% in Sweden in 2007-2012, according to the EUROCAT).[46] In Sweden, however, a higher proportion of pregnancies with a chromosomal abnormality lead to a TOP (66.4% of all pregnancies diagnosed with a chromosomal anomaly, compared to 53.8% in England).[46]

Differences in the uptake of TOP could reflect differences in regulations for TOP and antenatal screening practices. In Sweden, TOP is free and available on women's request with no legal indication required until the 18th week of pregnancy.[48] After that, TOPs are only permitted following a review from a multidisciplinary committee at the Swedish government agency, the National Board of Health and Welfare (NBHW), and very few TOPs are allowed after 22 weeks' gestation.[49] In England, TOPs are permitted for indications related to physical and/or mental health or social reasons until 24 weeks of gestation (conditional on a formal confirmation from two doctors) and are free (only 2% were funded privately in 2016).[48,50] In both countries women are offered an early dating scan in the first trimester, and a later anomaly ultrasound scan (at 15-18 weeks in Sweden and at 18-21 weeks in England, as of 2010).[49] While in both England and Sweden over 90% of women attend antenatal care before 20 weeks of gestation, the proportion of women who receive a dating scan is higher in Sweden (estimated >95% in Sweden and 75-95% in England and Wales in 2016).[51] I could not find information about the uptake of the second anomaly scan in the two countries. Some of the differences in rates of TOP could reflect differences in the timing of the detection of foetal anomalies.

1.3.1.4 Sex

Mortality rates are higher in boys than in girls, both in infancy and in early childhood.[27] In 1993-2010, infant mortality in girls was 21% lower than for boys in England and Wales, and 18% lower at 1-18 years.[40] These differences could reflect increased risk of congenital anomalies and pregnancy complications in boys than girls.[39,52] Boys also have poorer lung function, which could contribute to an increased risk of death from respiratory conditions.[53] For example, boys have approximately 50% higher infant mortality from respiratory infections and sudden infant death syndrome (SIDS, these are unexpected and unexplained deaths in infants).[53] The mechanisms behind these differences between sexes are insufficiently studied and could include a combination of biological factors (e.g., related to the absence of second X chromosome in boys), as well as psychological factors.[27]

1.3.1.5 Multiple birth

Compared with singletons, multiple births carry an increased risk of stillbirth, neonatal and infant mortality.[12] The risk of stillbirth is higher at all gestational ages while the increased risk of infant mortality is primarily driven by an increased risk of preterm delivery.[54] In 2010, multiple births accounted for only 2.8% of all births in Sweden and 3.1% of live births in England and Wales, but 22% and 23% of all preterm births, respectively.[12]

Multiple births are more common in older mothers, both due to an increased prevalence of a spontaneous multiple birth and a more frequent use of assisted reproduction techniques.[54,55] Due to the association with advanced maternal age, multiple births also carry an increased risk of stillbirth, low birth weight and congenital anomalies (risks associated with maternal age are described in Section 1.3.2.1).[12]

Over the past decade, increased use of assisted reproduction techniques, such as *in vitro* fertilisation (IVF), has resulted in a rise in the rates of twins.[54,55] However, currently both countries have a single embryo transfer policy, applied in over 75% of IVF cases, which reduces the odds of having a multiple pregnancy.[51]

Table 1.5 - Summary of the prevalence of key risk factors at birth in live births in the UK, England and Wales, and in Sweden in 2010

Risk factors at birth	UK	England and Wales	Sweden
Live births*	801,003	718,266	114,706
Preterm birth (at <37 weeks)*	7.0%	7.0%	5.9%
Low birth weight (<2500g)*	6.9%	7.0%	4.2%
Prevalence of congenital anomalies (in live births)**	2.0%	2.0%	1.7%
Multiple births (as % of all live births)*	3.1%	3.1%	2.8%

*UK=United Kingdom. *Information comes from the EURO-PERISTAT project.[12] **Information comes from the EUROCAT network.[46]*

1.3.2 Maternal risk factors associated with birth characteristics

Key maternal risk factors operating before and during pregnancy, which contribute to the increased risk of adverse birth characteristics presented in this section are: maternal age, body mass index (BMI), parity, maternal health, smoking, alcohol consumption and drug use, and ethnicity. A table summarising the presented evidence is available at the end of Section 1.3.2 (Table 1.6).

1.3.2.1 Maternal age at birth

Maternal age at birth shows a U-shaped association with child mortality. The risks are increased for teenage mothers (<20 years old) as well as for older mothers (>35 years old).

Teenage mothers

The children of teenage mothers (<20 years old) have an increased risk of adverse birth characteristics such as preterm birth, or low birth weight.[56,57] These increased risks are often attributed to socio-economic deprivation, inadequate prenatal care, or inadequate weight gain due to continued growth of the expectant mother (especially for mothers aged <17 years old); however, these factors do not explain all of the increased

risk of adverse birth characteristics.[56,57] A study of almost four million mothers from the United States of America (USA) in 1995-2000 showed that even after accounting for these factors, the risks of preterm birth and low birth weight remained 20% and 15% higher, respectively, in babies of teenage mothers compared to those of mothers aged 20-24 years old.[56] The relative risks were higher for mothers aged <17 than for those aged 18-19 years old, and could be associated with the continued growth of the expectant mother (e.g., foetal growth could be hindered if mother's pelvis is not fully developed).[56,57]

Babies of teenage mothers also have an increased risk of child mortality, independent of the effect of birth characteristics.[58] In England and Wales, the babies of teenage mothers were approximately 1.5 times more likely to die in the neonatal period, 2.75 times more likely to die in the post-neonatal period, and twice as likely to die at 1-4 years relative to children of mothers aged 30-34, after accounting for birth weight (according to data from 1993-2010).[58] The increased child mortality rates could be attributed to socio-economic factors – teenage childbearing can limit mothers educational and employment opportunities. A 1991 survey of mothers in the UK found that teenage mothers had a 12–24% lower probability of returning to education, and had 5-22% lower pay.[59]

In 2010, the proportion of teenage mothers was much higher in England and Wales than in Sweden (5.7% vs 1.6%).[12] Increased rates of teenage pregnancy were not explained by differences in rates of TOP, as the rates were comparable. In 2008, 24.4/1000 women aged 15–19 years terminated their pregnancy in Sweden compared to 23.8/1000 in England, Wales and Scotland, accounting for 19.9% and 22.1% of all TOPs in the two countries respectively.[48]

Older mothers (aged >35 years old)

Older mothers (aged >35 years old) have an increased risk of adverse birth characteristics such as chromosomal abnormalities,[60] preterm birth,[12,27] and pregnancy complications such as hypertension or diabetes,[12] which are associated with an increased risk of infant death. In England and Wales, babies of older mothers had approximately 20% higher risk of death in the first year of life in 1993-2010 (after adjustment for birth weight), but the differences diminished in early childhood.[58]

The proportion of older mothers was comparable in England and Wales (19.7%) and in Sweden (22.5%).[12] The rates of TOP in women aged >35 years, however, were higher in Sweden (8.2 terminations/1000 women, covering 20.1% of all TOPs) than in England, Wales and Scotland (4.3/1000 women, covering 14.0% of all TOPs).[48] These differences are likely to reflect differences in decisions about TOP following the

detection of chromosomal anomalies. Older mothers have a higher risk of a chromosomal anomaly in their offspring, and TOPs due to chromosomal anomalies are more common in Sweden (as explained in Section 1.3.1.3).

1.3.2.2 Maternal Body Mass Index (BMI)

Maternal obesity (defined as BMI≥30) is associated with an increased prevalence of some congenital malformations,[61,62] and spontaneous extreme preterm birth.[63] Obese women also have a higher risk of pregnancy complications, such pre-eclampsia or gestational diabetes (which can lead to an induced preterm labour on medical grounds)[12] and of mortality in term infants (due to birth asphyxia, or other neonatal morbidity).[64,65] In 2010, 12.6% of pregnant women in Sweden were obese,[12] compared to approximately 20% in England (based on obesity in all females aged 16-44 years in 2010).[66]

Mothers with low pre-pregnancy BMI and short stature, on the other hand, have an increased risk of delivering growth-restricted babies.[12,14] The prevalence of underweight mothers (defined as BMI<18.5) was 2.5% in Sweden in 2010 (I could not identify comparable number for England).[12]

1.3.2.3 Parity

Women who give birth for the first time (primiparous women) and women who have had five or more pregnancies (grand multiparous women, with a parity of four) have an increased risk of pregnancy complications, neonatal morbidity, stillbirth or neonatal death.[12,67] In 2010, the proportion of primiparous women was comparable between England (42.9%) and Sweden (46.3%), but the proportion of grand multiparous women was higher in England (5.4% in England vs 2.1% in Sweden).[12]

1.3.2.4 Maternal health state: infections and chronic illness

Some maternal infections could contribute to the risk of adverse birth characteristics. For example, bacterial vaginosis and other vaginal infections are one of the key risk factors for spontaneous preterm birth.[14] Rubella, varicella, toxoplasmosis and cytomegalovirus infections during pregnancy could increase the risk of congenital anomalies.[60,61,68] I could not identify figures on the whole-country prevalence of these infections in England and Sweden.

Adverse birth characteristics and neonatal morbidity are also more common in mothers with chronic health conditions. For example, babies of diabetic mothers have an increased risk of IUGR, congenital anomalies and stillbirths.[69] In Sweden, 1.3% of women who gave birth in 1997-2006 had diabetes, including 0.9% who had gestational diabetes.[70] In the UK, the estimates varied between 1-3% (based on cohort studies

from before 2010, using consistent criteria recommended by the WHO to define gestational diabetes).[71] Chronic conditions such as thyroid disease, hypertension, diabetes or asthma could lead to an early induction of birth to prevent maternal or foetal complications, contributing to the burden of preterm birth.[14,28]

Maternal hypertensive disorders (both chronic and pregnancy-induced) are also associated with an increased risk of adverse birth characteristics. Preterm birth and low birth weight are three times more common in mothers with chronic hypertension, than in the general population, while the risk of stillbirth or neonatal death is four times higher (based on a population of women in the USA).[72] Pregnancy-induced hypertension could lead to pregnancy complications such as pre-eclampsia or eclampsia, which contribute to the increased risks of IUGR and preterm birth.[73] It is estimated that pre-eclampsia affects approximately 3% of pregnancies, while all hypertensive disorders affect 5-10% of pregnancies in high-income countries.[73] In 1997-2006, 0.5% of Swedish mothers in 1997-2006 had chronic hypertension, and 3.9% had pregnancy induced hypertension.[70]

1.3.2.5 Smoking, alcohol consumption and drug use during pregnancy

The exposure of a foetus to toxic substances such as alcohol, tobacco, or drugs, can impair healthy development *in utero*. For example, maternal smoking it is one of the key risk factors for low birth weight and IUGR.[14] It is also associated with an increased risk of a preterm birth, and some congenital anomalies.[12,74] The prevalence of maternal smoking is higher in England than in Sweden: in 2010, 12% of mothers in England smoked during pregnancy (according to survey data), compared to 6.5% of mothers who smoked in the 1st trimester, and 4.9% who continued to smoke in the 3rd trimester in Sweden (according to data from antenatal care clinics).[12] Heavy alcohol consumption and drug use are also associated with an increased risk of preterm birth and IUGR, congenital anomalies and neurodevelopmental and growth problems (due to foetal alcohol and neonatal withdrawal syndromes).[14,28,75] However, the prevalence of these behaviours is relatively low amongst expectant mothers, for example prevalence of neonatal withdrawal syndrome was estimated to be 0.3% in England between 1997 and 2011.[75]

1.3.2.6 Ethnicity

Infant mortality and the prevalence of adverse birth characteristics vary between ethnic groups in England and Wales (see Appendix A, Table A.1 for a detailed comparison). For example, infant mortality rates are highest for Caribbean and Pakistani babies (7.8 and 8.8/1000 births, respectively, compared to 3.6/1000 births for White babies in 2010).[34] The prevalence of low birth weight is highest in Asian babies (10.0%, 10.5%

and 9.8% in Bangladeshi, Indian and Pakistani babies, respectively compared to 6.0% for White babies in 2005).[76] Infants of Caribbean origin have the highest rates of preterm birth (9.5% compared to 6.9% in White babies in 2010).[34] Infant mortality due to congenital anomalies is nearly five times higher in Pakistani babies than in White babies (4.8/1000 births vs. 1.0/1000 births in 2005, based on the underlying cause of death).[76] A cohort study from Bradford in England ("Born in Bradford" study, including almost 14,000 births in 2007-2011) found that the prevalence of congenital anomalies in live born babies is also higher in babies of Pakistani origin (5% compared to 3% in all study participants).[77]

Ethnic variation in infant mortality rates and the prevalence of adverse birth characteristics in England and Wales reflects the complex interplay between socio-economic disadvantage, cultural factors and some biological factors. For example, a lower proportion of Pakistani and Black babies have fathers in managerial and professional occupations, indicating lower socio-economic status (SES; 21.6% and 25.2% respectively, vs. 37.1% for all babies in England and Wales in 2005).[76] The increased infant mortality from congenital anomalies in babies of Pakistani origin can be partially attributed to higher rates of consanguineous marriages amongst couples of Pakistani origin.[60] According to the "Born in Bradford" study, consanguinity accounted for 31% of the anomalies among Pakistani babies.[77] The increased prevalence of and mortality from congenital anomalies in Pakistani babies is also associated with lower rates of TOP for foetal anomaly than for White British or Indian women (46% vs 71%, according to a cohort study from East Midlands and South Yorkshire regions of England in 1998-2007).[13] The increased prevalence of low birth weight in Asian babies can, in part, be explained by shorter parental statures, which are associated with an offspring's birth weight.[78] According to the Millennium Cohort Study, Asian mothers in England are shorter and weigh on average 7kg less than White mothers.[79] However, mothers from ethnic minority groups are less likely than White mothers to smoke or consume alcohol during pregnancy, and more likely to breastfeed.[80]

Information on ethnicity is not routinely collected in any population register in Sweden. Instead, information on mother's country of birth is available. Births to foreign-born mothers accounted for 24.4% of all pregnancies in Sweden in 2010 (compared to 26.5% in England and Wales).[12] Birth outcomes vary between women born in Sweden and abroad. For example, in 1995-2005 stillbirth rates were higher for immigrant mothers from Africa, the Middle East, and recently settled immigrants; but comparable for women from USA, Canada and Western Europe.[81] Mothers from East Asia, South Asia, and Sub-Saharan Africa were more likely to give birth prematurely.[82] Infants of foreign-born parents are also more likely to be born small for gestational age.[82]

The differences in ethnic make-up of the populations in England and Sweden could contribute to the differences in child mortality between the two countries. However, I do not investigate this in this PhD as ethnic composition of the population is not a modifiable factor.

Table 1.6 - Summary of distribution of selected maternal risk factors at birth in the UK, England and Wales and in Sweden in 2010

Maternal risk factors	UK	England	Wales	Sweden
Number of pregnant women	781,000	662,913	36,199	113,488
Teenage mothers (<20 years old)	5.7%	5.7%		1.6%
Older mothers (>35 years old)	19.7%	19.7%		22.5%
Maternal obesity (BMI≥30)*	N/A	20%	N/A	12.6%
% of women born outside of country or of foreign origin using another definition	26.1%	26.5%		24.4%
Primiparity (1st pregnancy)	43.6%	42.9%	52.9%	46.3%
Grand multiparty (5th pregnancy)	4.9%	5.4%	2.6%	2.1%
Smoking during pregnancy	12.0%	12.0%	16.0%	6.5% (1st trimester)

*BMI=body mass index. N/A=not available, UK=United Kingdom. Data comes from the EURO-PERISTAT project.[12] *Data from Public Health England, based on obesity in all females aged 16-44 years.[66]*

1.3.3 Risk factors operating after birth

1.3.3.1 Acute and chronic illness

Given a child's characteristics at birth, acute and chronic illness can contribute to the risk of death. According to the WHO, non-communicable diseases are the leading cause of death in the UK and Sweden beyond the first month of life, accounting for 34.3% and 39.8% of deaths at age 28 days-4 years, respectively.[19] In 2010, mortality from non-communicable diseases was 40% higher in the UK than in Sweden (79.9/100,000 births vs. 57.5/100,000 births).[19] Deaths due to lower respiratory tract infection (an example of acute illness) accounted for approximately 7% of deaths in the UK and in Sweden and mortality was 70% higher in the UK (16.6/100,000 compared to 9.6/100,000 in 2010).[19]

The differences in the prevalence of chronic conditions could contribute to the increased child mortality in the UK relative to Sweden. However, I could not identify representative figures on the prevalence of chronic conditions, measured in a comparable way between England, or the UK and Sweden.

Available data on mortality due to non-communicable diseases reported by the WHO are likely to undercount the true number of deaths due to non-communicable disease in the UK and Sweden, because it is based on the underlying cause of death (rather than all causes of death). A study based on information from child's hospital records and all causes of death, estimated that over 70% of all children who died in 2001-2010 at age 1-4 years in England had at least one chronic condition, with neurological or sensory conditions accounting for approximately 40% of all deaths.[58] Furthermore, children with at least one chronic condition accounted for almost 90% of all deaths due to respiratory tract infections in England.[18] Therefore, differences in the prevalence of chronic conditions may be wider than suggested by the differences in mortality from non-communicable diseases.

1.3.3.2 Care in the family setting

Child's physical environment

A child's physical environment could contribute to the risk of death from accidents, injury and poisoning.[27] According to the WHO Mortality Database, injuries accounted for 8.3% of all deaths at 28 days-4 years in the UK, compared to only 4.8% in Sweden in 2010; mortality due to injury was three times higher in the UK (19.4/100,000 vs. 7.0/100,000).[19]

Housing conditions could also influence a child's health. For example, mould and damp could lead to worse respiratory health in the child.[83] However, the prevalence of these risk factors is difficult to measure between countries.

Exposure to environmental tobacco smoke at home can also affect infant's respiratory function and is associated with an increased risk of respiratory infections and deaths from SIDS.[84] In 2009, 6.5% of mothers and 11% of fathers smoked when the baby was 8 months old in Sweden.[39] I could not identify corresponding figures for England. However, the proportion of smokers in the population in the UK is higher than in Sweden (19% as of 2014 in the UK, compared to 13% as of 2011 in Sweden).[85,86]

Breastfeeding

Breastfeeding is known to be beneficial for a child's health outcomes, and exclusive breastfeeding is recommended by the WHO in the first 6 months of baby's life.[87] In high-income countries, breastfeeding is associated with a reduced risk of hospitalisation due to respiratory tract infections (72% reduction in children exclusively breastfed for a minimum of four months).[88] Children with any history of breastfeeding have a third lower risk of SIDS (compared to never breastfed children).[88] Breastfeeding is also associated with a reduced risk of diabetes in childhood.[88] Therefore, differences

in rates of breastfeeding could contribute to differences in child mortality in England and Sweden.

Rates of breastfeeding are higher in Sweden than in the UK. According to national survey data, 98% of infants were ever breastfed in Sweden, 52% were breastfed for at least 6 months and 16% were breastfed for at least one year in 2010.[87] In the UK, 81% of infants were ever breastfed, 34% were breastfed for at least 6 months and only 0.5% were breastfed for at least a year.[87]

Sleeping practices

SIDS accounts for a high proportion of deaths in infancy, especially in the post-neonatal period (13% of post-neonatal deaths in England and Wales were from SIDS in 2010).[44] SIDS refers to an infant death which occurred suddenly and unexpectedly, for which no cause of death could be identified.[89,90] Sleeping practices are an important risk factor for SIDS.

In most countries, the rates of SIDS have declined since the introduction of public health campaigns in the 1990s, which recommended that infants be put to sleep on their backs ('Back to sleep' campaigns). In England, the proportion of SIDS deaths due to prone position reduced from 89% in 1984-88 to 24% in 1999-2003.[91] However, unsafe sleeping practices remained an important risk factor for SIDS. For example, an increased risk of SIDS is associated with co-sleeping on the sofa or armchair, and in hazardous environment (e.g., when parents used alcohol, drugs or smoked cigarettes).[89] A case-control study of SIDS in Bristol and surrounding regions found that maternal alcohol consumption (of more than 2 units within 24h of death) increased the risk of SIDS 40 times, while co-sleeping on a sofa lead to a 20-fold increase in the risk of SIDS.[92]

Differences in unsafe sleeping practices could contribute to differences in infant mortality in England and Sweden. However, I could not identify information about prevalence of unsafe sleeping practices in the two countries.

1.3.4 Upstream determinants of child mortality

1.3.4.1 Maternal socio-economic status (SES)

SES of the mother has a substantial impact on the healthy development of a baby in the womb. Babies of more disadvantaged mothers are more likely to be born preterm,[14,93] with low birth weight,[14,93] or a congenital anomaly[13]. Socio-economic inequality in preterm birth and congenital anomalies explained almost 80% of the increased neonatal mortality rates in the most deprived 10% of the population in England compared to the least deprived 10% in 1997-2007.[94]

Socio-economic disadvantage determines the risk of adverse birth characteristics upstream, through the responses and behaviours of mothers exposed to poverty or financial hardship.[14] Many of the maternal characteristics described in Section 1.3.2 are known to show socio-economic gradients. For example, maternal obesity,[66] young maternal age,[56] and low pregnancy weight gain[14] are more common in the most deprived mothers. Socio-economic gradients are also observed for bacterial vaginosis, alcohol and drug use, and smoking.[14] Almost one third of the excess infant mortality in the most deprived 20% of the Scottish population compared to the least deprived 20% in 1994-2003 was attributed to variation in smoking during pregnancy. Socio-economic inequalities were also observed in rates of TOP in England in 1998-2007 (based on data from a register of congenital anomalies), and contributed to increased neonatal mortality from serious congenital anomalies in the most deprived 10% of women relative to the least deprived 10% in England (as antenatal detection rates were comparable).[13] Wider socio-economic inequality (in terms of income) could, therefore, explain some of the increased prevalence of adverse birth characteristics and associated maternal characteristics in the UK relative to Sweden.

A family's SES determines not only the prevalence of adverse birth outcomes, but also the types of risks that the child is exposed to after birth.[27,39] For example, causes of death like SIDS and infections are more prevalent amongst the more deprived groups in the UK.[95,96] In Sweden, mothers from low-income households are less likely to breastfeed.[39]

As detailed in Section 1.2.2, relative poverty (defined as the ratio of average incomes of the most deprived and the least deprived 20% of the population) is twice as high in the UK as in Sweden.[97] Therefore, the differences in socio-economic factors are likely to contribute to the increased child mortality in the UK and England, relative to Sweden.

1.3.4.2 Family policies

Public policies that impact on family income levels are another upstream determinant of child mortality.[98,99] Evidence from ecological studies has shown that public spending on social protection for families is inversely associated with infant mortality rates: mortality tends to be lower in countries with higher spending.[7] The effect of spending, however, depended on the design of family policies – data from 18 OECD countries showed that the benefits from higher spending were limited to countries where family policies supported families with two earning parents (these policies included paid parental leave, universal child benefits, and childcare support). One percentage point increase in spending on family policies was associated with a reduction in infant mortality by 4/100,000 births.[100]

Family policies differ between the UK and Sweden (Table 1.7). While the proportion of GDP spent on family benefits is lower in Sweden than in the UK (3.6% compared to 4.2% in 2012), a higher proportion is spent on in-kind benefits, which, in particular, enable women to resume work after having children (2.1% compared to 1.4%, Table 1.7).[101] These benefits in Sweden include, for example, affordable day care, which is heavily subsidised and costs approximately £70 a month.[85] In contrast, day care is mainly privately owned and operated in the UK (with average costs estimated around £900 per month) and 15-hours of free child care is only available for children aged 3-4 years old.[102] These differences likely explain the higher proportion of children aged 0-2 enrolled in formal childcare and the higher proportion of mothers who were employed (both with partners and as single parents) in Sweden than in the UK (Table 1.7).[101] Paid maternity leave, available to both mothers and fathers, is also longer in Sweden than in the UK (combined 70 weeks compared to 41 weeks).[101]

Table 1.7 – Summary of differences in family policies in the UK and in Sweden in 2012

Family policy	UK	Sweden
Length of paid maternity and parental leave available to mothers (weeks)	39	60
Length of paid paternity and parental leave reserved for fathers (weeks)	2	10
Proportion of children aged 0-2 enrolled in formal childcare and pre-school	31.0%	48.2%
Total public expenditure on families (% of GDP)	4.2%	3.6%
Public expenditure on cash benefits for families (% of GDP)	2.6%	1.4%
Public expenditure on services and in-kind benefits for families (% of GDP)	1.4%	2.1%
Proportion of all mothers (15-64 years old) with at least one child under 15 in employment	65.5%	82.7%
Proportion of partnered mothers (15-64 years olds) with at least one child under 15 in employment	69.5%	83.9%
Proportion of sole-parent mothers (15-64 years old) with at least one child under 15 in employment	54.5%	76.0%

GDP=Gross Domestic Product; OECD=Organisation for Economic Co-operation and Development; UK=United Kingdom. Data comes from OECD Family Database.[101]

Both countries have a universal child allowance. In the UK, the allowance is approximately £80 for one child and £55 for any additional children per month, until the child's 16th birthday.[102] There is also an additional means-tested child benefit ('Child tax credit') of £315-8,800 a year depending on family income and the number of children.[102] In Sweden, all parents receive £80 per child per month for children <16 years old, with an additional allowance for families with two or more children.[85]

There are further differences in the support for parents of sick and disabled children. In Sweden, parents can take up to 120 days of paid leave a year to take care of their sick child (given child is aged 0-11 years of age). Children with disabilities can also receive an extra personal assistance from an external carer.[85] In England, parents of children with disabilities can receive additional financial support,[102] but there is no similar legal entitlement to compassionate leave to care for a sick child.

1.3.4.3 Provision of healthcare

Some of the differences in child mortality in the UK, England and Sweden have been hypothesised to reflect differences in organisation and provision of healthcare.[27] This section provides an overview of known differences.

Overview

The UK and Sweden have comparable levels of public spending on healthcare, which is universally accessible in both countries (7.5% of GDP in the UK compared to 6.9% in Sweden in 2012).[5] However, the two countries differ in the organisation of healthcare.

In the UK, the National Health Service (NHS) is free at point-of-use and is publicly funded. The care provided through the NHS comes at no direct charge for the UK residents, with the exception of dental and optical care, and prescriptions for adults. The NHS is managed independently in the four UK countries – this section focuses on NHS in England.[102]

In Sweden, healthcare is largely publicly funded; however, approximately 17% of Swedish healthcare is privately funded. This is primarily through user charges – patients are charged a flat-rate for appointments in the primary care or with a specialist, with the total cost per year capped at approximately £110 per year for appointments and £220 per year for prescriptions.[85] Therefore in total the spending on healthcare is higher in Sweden than in the UK.

Obstetric and neonatal care for high-risk mothers and babies

Differences in neonatal mortality could reflect differences in the quality and organisation of obstetric and neonatal care in England and Sweden (Table 1.8).[51] Obstetric practices for uncomplicated pregnancies are comparable in the two countries – uncomplicated, low-risk pregnancies are delivered by midwives, with support from an obstetrician if needed.[85,102]

Neonatal intensive care is also organised in a similar manner in the two countries. In England, neonatal intensive care is managed within 24 networks, each covering between 4 and 16 maternity departments. Each network has one leading neonatal

intensive care unit, which provides the full range of specialist care (including surgery or cardiology) and is responsible for the transport of high-risk babies.[102] In Sweden, neonatal intensive care is organised nationally (even though health services are generally provided and managed regionally) and services are centralised due to relatively few children requiring neonatal care.[103] This is reflected by a lower number of tertiary neonatal units per live births <32 weeks (Table 1.8).

Table 1.8 – Summary of differences in organisation of obstetric and neonatal care for high-risk mothers and babies

	UK	Sweden
National recommendation for transfer of pregnant women to tertiary neonatal units (NNU)*	<28 weeks or clinical need	<26 weeks or clinical need
Number of tertiary NNUs* in the country	179 (approximately 1 per 50 live births <32 weeks)	7 (approximately 1 per 190 live births <32 weeks)
Designated neonatal transport teams	Yes	Yes
Proportion of acute transfers carried out by designated transport teams	>95%	<50%
% of infants <1500g managed at level-2 NNUs**	10-50	10-50
Proportion of babies retro-transferred to level-2 units** from tertiary units* before discharging home (%)	10-50%	>75%
Percent of level-2 units** offering respiratory support:		
A) Short-term (⩽ 2 days) mechanical ventilation (%)	>75%	>75%
B) Long-term (>2 days) mechanical ventilation (%)	>75%	<25%
C) Continuous positive airway pressure/high flow (%)	>75%	>75%

*NNU=Neonatal Unit; UK=United Kingdom. Data were obtained from a recent survey, which compared organisation of NNU in selected European countries.[51] *Tertiary units were defined as units which provide highly-specialised care for sickest children (e.g., born extremely prematurely).**Level two units ('Step down' units) were defined as units which provide care for preterm babies prior to discharge home or for sick babies born at higher gestational ages.*

Paediatric and primary care services

Primary and paediatric care services are organised differently in England and Sweden. In England, primary and paediatric care are funded and managed independently.[7] GPs are the first point of access to healthcare, and can be seen as "gatekeepers" for referrals to specialists.[102] In Sweden, primary care for children is provided by GPs, but patients can also get appointments with specialists directly.[85] Unlike in England, primary care services are often co-located in paediatric centres, enabling better

coordination between the two services.[7] GPs in Sweden receive at least 3 months training in paediatrics or gynaecology and obstetrics.[7] In England, paediatric training is not mandatory.[7] The ratio of primary care doctors to children is comparable between the two countries (1 GP to 266 children in England compared to 1 GP to 286 children in Sweden in 2006-08), Sweden, however, has a higher ratio of paediatricians to children (1:1,215 children compared to 1:3,928 children in England in 2008), possibly addressing increased demand due to self-referrals.

It has been previously argued that differences in the provision of primary care in the UK (and England) and Sweden (in particular, the lack of integration of primary care and paediatric services and no mandatory paediatric training for GPs in the UK – although in Sweden GPs could choose obstetric training instead) could cause delays in diagnosis and treatment for acute illness, such as infections or respiratory conditions, leading to increased child mortality rates.[6,8,9] This hypothesis, however, has not been formally tested so it cannot be confirmed.

1.4 Thesis rationale

As outlined above, there are a range of risk factors operating before and during pregnancy, as well as after birth, which could contribute to the differences in child mortality in England, relative to Sweden. Policy makers need to know which preventive strategies are likely to have the biggest impact on reducing child mortality in England. Should they invest in improving women's health before and during pregnancy to reduce the prevalence of adverse birth characteristics, or are improvements needed in the care received after birth (through changes in policy or provision of healthcare), given the underlying health conditions which a child is born with? Or should they invest equally in both? This is not obvious when looking at crude mortality rates or data aggregated by the underlying cause of death, which were used to compare child mortality between the UK and Sweden previously.[6–9]

As discussed in section 1.3, a child's characteristics at birth is a key determinant of the risk of death throughout childhood, and it is strongly influenced by maternal risk factors operating before and during pregnancy. To inform policies aimed at reducing child mortality in the UK, or England relative to Sweden, we need a comparison accounting for the differences in risk factor exposures during pregnancy, as indicated by birth characteristics such as preterm birth, low birth weight or presence of congenital anomalies. Such comparison can indicate the contribution of risk factors operating before and during pregnancy to the excess risk of child death in the UK relative to Sweden. I assumed that any differences in child mortality remaining after adjustment

for these birth characteristics would indicate excess mortality attributable to risk factors in the care received after birth.

In this PhD, I use individual-level data from administrative linked datasets in England (covering 85% of births in the UK[1]) and in Sweden to develop comparable national birth cohorts, with information about key birth characteristics (gestational age, birth weight, sex and presence of congenital anomalies) and socio-economic circumstances at birth (maternal age and SES). I combine the cohorts to compare adjusted all-cause mortality in England and Sweden. This enables me to quantify the relative contribution of birth characteristics and socio-economic factors to the excess mortality in England relative to Sweden. I also compare adjusted mortality from two potentially preventable causes, associated with the quality of care and health advice received after birth: deaths related to respiratory tract infections (RTI), which are amenable to healthcare through vaccination and antibiotics treatment,[18] and sudden unexpected deaths in infancy (SUDI), which are amenable to public health interventions, such as advice on safe sleeping practices or smoking cessation programs.[89] SUDIs cover causes of all unexpected infant deaths, including deaths from unexplained causes (e.g., SIDS)[24] and from explained causes (e.g., accidental suffocation),[90] minimising bias due to inter-country differences in death certification practices. The results from this thesis can be used to guide policy decisions to reduce child mortality in England relative to Sweden.

This thesis focuses on England, as England is the biggest and the most diverse of the four UK countries. Furthermore, data for England were available from the start of my PhD (a data sharing agreement to use a de-identified extract of linked hospitalisation and mortality records for a programme of research on child mortality was in place from the start of my PhD).

1.5 PhD aims and objectives

The overall aim of this thesis is to quantify the contribution of birth characteristics and socio-economic factors to higher child mortality rates in England relative to Sweden.

The specific objectives are to:

1. Determine whether aggregate data tabulated by a key risk factor at birth (such as gestational age or birth weight) can be used to inform policy about the origins of differences in infant mortality rates.
2. Develop comparable national birth cohorts using administrative linked datasets in England and in Sweden, with information on birth characteristics (birth weight, gestational age, sex and presence of congenital anomalies) and socio-economic factors (maternal age and quintile of SES).

3. Compare the risk of child mortality in England and Sweden using individual-level data and determine to what extent the differences can be explained by birth characteristics and socio-economic factors (after accounting for birth characteristics).
4. Compare the risk of child mortality from causes which could be potentially preventable by improving the quality of care received after birth: RTI-related deaths, which are amenable to healthcare, and SUDI deaths, which are amenable to public health interventions.

1.6 Thesis structure

In **Chapter 2**, I present two metrics which can be used to conduct more policy-relevant inter-country comparisons of infant mortality using aggregated data tabulated by a risk factor at birth to address objective 1. The two metrics describe the contribution of exposures during pregnancy to inter-country differences in infant mortality rates. I discuss the limitations of inter-country comparisons of childhood mortality based on aggregate data and the need for analyses based on individual-level data.

In **Chapter 3,** I present work towards objective 2. I describe administrative linked datasets used to develop a birth cohort in England. I present methods for identifying births, enhancing information on birth characteristics (birth weight, gestational age and sex) and socio-economic factors (maternal age and quintile of socio-economic status) by linking mothers and babies, validating the cohort and dealing with missing data.

Further work towards objective 2 is described in **Chapter 4**. I present the linked Swedish national registers used for this thesis and describe methods for developing a Swedish birth cohort. I compare datasets available in England and in Sweden, in terms of data collection process, recorded variables, and diagnostic practices to determine whether there are likely biases that need to be addressed in the analyses.

In **Chapter 5,** I compare child mortality in England and Sweden using comparable birth cohorts from Chapters 3 and 4 to address objective 3. I determine to what extent the differences in the risks of child mortality between the two countries can be explained by birth characteristics (birth weight, gestational age, sex and presence of congenital anomalies) and socio-economic factors (maternal age and quintile of SES). These results can be used to inform policies to reduce child mortality in England.

Chapter 6 presents work to address objective 4. I compare mortality from causes which could be potentially prevented through either public health interventions (SUDI) or health care interventions (RTI-related deaths).

In **Chapter 7** I summarise the key findings from the thesis, describe the limitations of using aggregate and administrative data for inter-country comparisons of child mortality, and discuss the implications of the presented results for policy and future research.

Chapter 2. Policy-relevant comparisons of infant mortality in Europe using aggregate data

What is already known:

- Child and infant mortality rates vary between European countries.
- Some of these differences reflect variation in the prevalence of key risk factors at birth such as preterm birth or low birth weight.

What this chapter adds:

- I present a simple method for decomposing the differences in crude infant mortality into two policy-relevant metrics.
    - Metric 1 (within-country difference in crude and standardised mortality) shows excess mortality attributable to differences in prevalence of preterm birth, reflecting influence of prenatal risk factors.
    - Metric 2 (between-country difference in gestation-standardised mortality) reflects excess mortality due to differences in quality of

2.1 Chapter overview

As discussed in Chapter 1, the risk of child death is associated with maternal risk factors operating before and during pregnancy, which determine a child's health outcomes at birth, and with the care that the child receives after birth, in a healthcare setting or at home. To reduce child mortality, policy makers need to know whether preventive strategies should focus on maternal health, improve the care received after birth, or address both. Such distinction requires analyses of child mortality rates according to birth characteristics.

This chapter presents work towards objective 1: "*to determine whether aggregate data tabulated by a key risk factor at birth (such as gestational age or birth weight) can be used to inform policy about the origins of differences in infant mortality rates*". I compare infant mortality in eleven European countries and I present two metrics based on counts of live births and deaths tabulated by gestational age category to estimate the contribution of risk factors operating before and after birth to inter-country differences in infant mortality.

A paper on the two metrics described in this chapter has been accepted for publication in BMC Pregnancy and Childbirth. Some of the work comparing infant and child mortality in England and Sweden using aggregated data, which contributed to development of the two presented metrics, was published as a letter in Archives of Disease in Childhood and presented at the 2015 International Network for Research on Inequalities in Child Health (INRICH) Workshop (Montreal, Canada), the 2015 Farr Institute International Conference (St. Andrews, United Kingdom (UK)) and the 2015 Public Health Science Conference (London, UK).

2.2 Background

Infant mortality is often used to compare health profiles of different populations. Global data on infant mortality are routinely collected and collated by the World Health Organisation (WHO),[19] UNICEF reports it as one of their indicators of child wellbeing,[98] while the Organisation for Economic Co-operation and Development (OECD) uses it as one of its health indicators.[99] International rankings of infant mortality are an important tool for policy makers, as they illustrate potential improvements in infant survival which should be achievable, relative to countries with similar levels of economic development but lower infant mortality rates. For example, the Nordic countries have some of the lowest infant mortality rates in the world and are often used in the literature as benchmarks for achievable reductions in infant mortality among the high-income OECD countries such as the United States of America (USA) and in the UK, where infant mortality rates are among the highest.[8,9,104–106]

As discussed in Chapter 1, infant mortality is associated with a range of risk factors operating before and/or after birth. Between-country differences in infant mortality can be explained at least partly by variation in the prevalence of adverse birth characteristics including prematurity, low birth weight and the presence of congenital anomalies. A child's health at birth is in turn associated with the health, wellbeing and socio-economic circumstances of mothers before and during pregnancy. This implies that infant mortality could reflect welfare policies that impact on levels of poverty and distribution of wealth in a society, and, specifically, how these welfare policies impact mothers and families.[15,97] Infant mortality rates can also reflect the effectiveness of public health preventive strategies targeting modifiable risk factors after birth in the home setting, such as reducing parental smoking or advice about sleeping practices.[89,107] Infant mortality rates also reflect the quality of healthcare, especially obstetric and neonatal care for high-risk babies.[12] Lastly, some of the between-country differences in infant mortality are likely to be artefactual due to differences in definitions and registration practices for live births, stillbirths and deaths occurring shortly after birth.[12,25,26]

To reduce infant mortality, policy makers need to know when and how to target interventions to prevent the largest number of deaths in early life. Should they focus primarily on maternal health and the wellbeing of women before or during pregnancy, or on improving the care that children receive at or after birth, or focus on both women and children? Answering this question requires establishing the contribution of birth characteristics such as birth weight or gestational age (reflecting maternal health during pregnancy) to the overall rates of infant mortality.

2.2.1 Chapter aims

In this chapter, I used country-level aggregate data on stillbirths and infant mortality broken down by gestational age, a key risk factor for infant mortality. I developed two metrics which estimate the contribution of birth characteristics and risk factors operating after birth to inter-country differences in infant mortality rates. The first metric (the within-country difference between crude and gestation-standardised mortality rates) is associated with maternal health and wellbeing before and during pregnancy. The second metric (the between-country difference in gestation-standardised rates) reflects the quality of care for infants after birth, given their birth characteristics. I demonstrate how this simple decomposition of international differences in crude infant mortality rates could be used to guide policies to reduce infant deaths.

2.3 Methods

2.3.1 Data Sources

2.3.1.1 EURO-PERISTAT project

I used data from the EURO-PERISTAT project, a collaboration between countries in the European Union (EU), which aimed to design and collect internationally comparable indicators of maternal and perinatal health.[12] The project collected national-level aggregate data on 30 indicators from routinely-collected sources such as administrative datasets, health registers or routine surveys in 2010.[12] The indicators published by the EURO-PERISTAT that I used for these analyses included counts of total and live births, neonatal deaths (at 0-27 days), and all infant deaths (at 0-364 days) tabulated by gestational age and birth weight. Using these indicators, I also derived tabulations for stillbirths (as the difference in total and live births per country) and post-neonatal deaths (defined as deaths at 28-364 days, calculated as the difference in number of deaths in infancy and in the neonatal period).

2.3.1.2 Country selection

The EURO-PERISTAT project collected data for 31 European countries (there are 28 EU member states, but data for the UK was provided separately for England & Wales (combined), Scotland and Northern Ireland). Of the 31 countries, a complete set of aggregate data for total births, live births, neonatal and infant deaths tabulated by both birth weight and gestational age were available for 18 countries. Sweden did not provide tabulations for neonatal deaths to the EURO-PERISTAT project. To allow comparisons with Sweden, I generated tables of neonatal deaths by gestational age and birth weight using the Swedish Medical Birth Register, a database covering all births in Sweden to resident mothers, described in detail in Chapter 4, and used for analyses in Chapters 5 and 6. Thus, complete data were initially available for 19 countries before further exclusions.

I excluded one country which provided only data for selected regions (Belgium) and five countries with <20,000 births per year (Estonia, Iceland, Latvia, Luxembourg, Malta), as their counts of neonatal and infant deaths per birth weight and gestational age categories were prone to chance variations. Finally, two countries were excluded due to inconsistencies in recorded data (Northern Ireland, Slovenia; see Table 2.1 for details of exclusion criteria). Thus, 11 countries were included in the analyses: Austria, Czech Republic, Denmark, England & Wales, Finland, Norway, Poland, Romania, Scotland, Sweden, Switzerland.

Table 2.1 – Details of exclusion criteria for the study

Exclusion criteria	Excluded countries
Not all required tables with aggregate data were provided to the EURO-PERISTAT	Cyprus, France, Germany, Greece, Hungary, Ireland, Italy, Lithuania, Netherlands, Portugal, Slovakia, Spain
Regional data only	Belgium
<20,000 births per year	Estonia (15,884 births in 2010), Iceland (4,903 births in 2010), Latvia (19,248 births in 2010), Luxembourg (6,560 births in 2010), Malta (4,036 births in 2010)
Recording errors	Northern Ireland: the number of stillbirths (from 2 different tables) and neonatal deaths per birth weight category was identical, therefore it is likely that neonatal data were not correct. Slovenia: there were fewer infant deaths at 28-31 weeks than neonatal deaths with the same gestational age, leading to a negative number of post-neonatal deaths.

All data came from the EURO-PERISTAT project[12]

2.3.1.3 Allowing for inter-country differences in registration practices

International comparisons of early life mortality are prone to bias due to differences in definitions for registration of live and stillbirths.[25,26] In the majority of the included countries (9 out of 11), registration of live and stillbirths was limited to births with gestational age ≥22 weeks or birth weight ≥500g (Table 2.2). Foetal losses before 22 weeks were not recorded in any vital registration system. In England, Wales and Scotland, a higher cut off value of ≥24 weeks was used to distinguish between late foetal losses and stillbirths. Therefore, I excluded all births (live or still) with gestational age <24 weeks and birth weight <500g to minimise bias from inter-country differences in definitions and registration requirements.[25,26]

Some between-country differences in definitions of stillbirths remained, including four countries (Czech Republic, England & Wales, Scotland and Switzerland) that included terminations of pregnancy (TOPs) in the stillbirth category (Table 2.2). The gestational age limit for late TOPs was <24 weeks or lower (except for when mother's life is in danger) in all countries apart from Switzerland, where there was no limit for carrying out TOP. Therefore, excluding births at <24 weeks' gestation helped to minimise the contribution of TOPs to stillbirth counts.[108]

Table 2.2 – Differences in registration practices for still- and live births in the included eleven European countries

Country	Definition of live birth*	Definition of stillbirth*	TOPs included in stillbirth category*	Gestational age limit for carrying out TOP**
Austria	≥500g	≥500g	No	Late TOP rare
Czech Republic	≥22 weeks	≥22 weeks	Yes	<24 weeks
Denmark	≥22 weeks	≥22 weeks	No	<22 weeks
England & Wales	≥22 weeks	≥24 weeks	TOP registered as stillbirths from ≥24 weeks gestation	<24 weeks
Finland	≥22 weeks	≥22 weeks, if missing ≥500g	No	<24 weeks
Norway	≥22 weeks	≥22 weeks	No	<22 weeks
Poland	≥22 weeks	≥500g	No	Access to late TOP restricted
Romania	≥22 weeks	≥22 weeks	No	<24 weeks
Scotland	≥22 weeks	≥22 weeks; not complete at 22-23 weeks	Yes	<24 weeks
Sweden	≥22 weeks	≥22 weeks	No	<22 weeks
Switzerland	≥22 weeks, if missing ≥500g	≥22 weeks, if missing ≥500g	Yes	No limit

*TOP=terminations of pregnancy. *Information came from the EURO-PERISTAT project*[12] ***Not including TOP when mother's life is in danger. Information came from Blondel et al.*[108]

2.3.2 Outcomes

The primary outcome in this study was extended infant mortality rate per 1000 total births defined as per Equation 2.1:

Equation 2.1 – Definition of extended infant mortality rate used throughout this chapter

$$Infant\,mortality\,rate = \frac{Infant\,deaths\,and\,stillbirths}{Live\,births\,and\,stillbirths} \times 1000$$

Traditionally, WHO defines infant mortality as the number of infant deaths (at age 0-364 days) divided by the number of all live births.[24] Inclusion of stillbirths in the rate helped to account for possible inter-country differences in definitions of "signs of life" used to distinguish between still- and live births.[25,26] Furthermore, since many risk

factors for stillbirth and neonatal deaths are similar, including maternal obesity, smoking, socio-economic deprivation or high or low maternal age,[12,109] extended infant mortality which combines stillbirth and infant deaths better reflects the full potential benefits from reducing the prevalence of such risk factors on early life survival.

The secondary outcomes were the separate components of extended infant mortality rate defined above, grouped by the age at death: stillbirth rate, neonatal mortality rate and post-neonatal mortality rate per 1000 total births defined as per Equation 2.2:

Equation 2.2 – Definition of three subcomponents of extended infant mortality rate (defined in Equation 2.1) used throughout this chapter: stillbirth rate, neonatal mortality rate and post-neonatal mortality rate

$$Stillbirth\,rate = \frac{Stillbirths}{Live\,births + stillbirths} \times 1000$$

$$Neonatal\,mortality\,rate = \frac{Deaths\,at\,0-27\,days}{Live\,births + stillbirths} \times 1000$$

$$Post-neonatal\,mortality\,rate = \frac{Deaths\,at\,28-364\,days}{Live\,births + stillbirths} \times 1000$$

2.3.3 Risk factors

The EURO-PERISTAT project provided data tabulated by two important birth characteristics which are strongly associated with both the risk of stillbirth and infant mortality: birth weight and gestational age. The methods used to determine birth weight are considered to be more accurate and internationally standardised than methods used to calculate gestational age (e.g., using ultrasound scan or last menstrual period).[110] However, stillbirths and babies born at borderline viability are less likely to be systematically weighed at birth,[111] which is reflected by higher rates of missing data by birth weight (Table 2.3). Therefore, in this chapter I focused on gestational age, grouped as 24-27, 28-31, 32-36 and ≥37 weeks. All analyses were repeated using birth weight (categorised as 500-999g, 1000-1499g, 1500-2499g, ≥2500g) to check the robustness of my findings. The results of these sensitivity analyses are presented in Appendix B.

Table 2.3 – Proportion of births and deaths with missing birth weight and gestational age in each country

	Data tabulated by gestational age						Data tabulated by birth weight					
	% of missing data				Extended infant mortality rate (per 1000 total births)		% of missing data				Extended infant mortality rate (per 1000 total births)	
Country	Live births	Stillbirths	Neonatal deaths	Post-neonatal deaths	Known gestational age & ≥24 weeks	Gestation ≥24 weeks or unknown	Live births	Stillbirths	Neonatal deaths	Post-neonatal deaths	Known birth weight & ≥500g	Birth weight ≥500g or unknown
Austria	0.0%	0.0%	0.0%	0.0%	5.8	5.8	0.0%	0.0%	0.0%	0.0%	5.4	5.4
Czech Republic	0.0%	0.0%	3.1%	0.0%	4.9	4.9	0.0%	0.0%	2.1%	0.0%	5.3	5.3
Denmark	0.0%	0.0%	0.0%	0.0%	5.1	5.1	0.3%	34.1%	13.1%	6.5%	4.3	5.6
England & Wales	1.0%	3.0%	2.7%	1.1%	8.1	8.3	0.7%	5.2%	9.2%	3.8%	7.8	8.2
Finland	0.1%	3.2%	0.0%	0.0%	4.5	4.6	0.0%	24.2%	0.0%	0.0%	4.6	5.4
Norway	0.9%	2.4%	0.0%	5.3%	5.5	5.5	0.0%	13.8%	0.0%	0.0%	5.3	5.7
Poland	0.0%	0.0%	0.0%	0.0%	8.1	8.1	0.0%	0.0%	0.0%	0.0%	9.0	9.0
Romania	0.0%	0.0%	20.5%	35.7%	11.1	13.7	0.0%	0.4%	10.9%	22.9%	12.2	13.8
Scotland	0.0%	0.0%	2.4%	4.5%	8.3	8.4	0.0%	4.2%	6.5%	6.0%	8.1	8.5
Sweden	0.0%	0.0%	0.0%	0.0%	5.4	5.4	0.2%	8.5%	9.1%	1.0%	5.2	5.6
Switzerland	0.1%	0.0%	0.0%	0.0%	5.6	5.6	0.0%	0.8%	0.0%	0.0%	6.0	6.1

Extended infant mortality was defined as number of stillbirths and infant deaths per 1000 births (live or still). All calculations are based on births with gestational age ≥24 weeks (or missing) and birth weight ≥500g (or missing). Neonatal deaths were defined as deaths at 0-27 days; post-neonatal deaths were deaths at 28-364 days. Data came from the EURO-PERISTAT project,[12] except for Sweden, where data were obtained from the Swedish Medical Birth Register.[112]

2.3.4 Statistical analyses

2.3.4.1 Crude and standardised extended infant mortality rates

I calculated crude extended infant mortality rates (i.e. not adjusted for gestational age) and preterm birth rates (defined as the proportion of still and live births born at <37 weeks' gestation) for each country. I then calculated directly standardised rates to adjust for inter-country differences in the distribution of births by gestational age.

To calculate directly standardised rates for each country, I first calculated gestation-specific mortality rates (i.e. mortality rates within each gestational age category). I then multiplied the number of births per gestational age category in a chosen standard population (see below) by the gestation-specific mortality rate in a given country. This gave the expected number of stillbirths and infant deaths per gestational age category that would have occurred if the standard population had the same gestation-specific mortality as each of the compared countries. I then summed these "expected" stillbirths and deaths over all gestational age categories and I divided this number by the number of total births in the standard population to obtain the gestation-standardised rates.[113]

2.3.4.2 Choice of the standard population

Choosing a standard population required some consideration, since the directly standardised rates will vary depending on the chosen standard population. One option is to use a sum of all populations. Alternatively, one could select the country with the lowest prevalence of preterm birth to calculate the maximum possible reductions in early life mortality attainable by improving the distribution of gestational ages across countries in comparison with the standard population.

I chose Sweden as the standard population, since a comparison of child mortality between England and Sweden is the primary focus of my PhD study. Sweden had the second lowest prevalence of preterm birth (after Finland). Therefore, extended infant mortality rates standardised relative to Sweden reflect excess mortality attributable to an unfavourable distribution of gestational age for all countries apart from Finland.

2.3.4.3 Metrics

Given a set of crude and standardised mortality rates for each country, I decomposed the difference in crude extended infant mortality rates between country A and the standard population into two metrics as per Equation 2.3:

Equation 2.3 – Decomposition of between-country difference in crude extended infant mortality into two metrics using gestation-standardised infant mortality

$$Crude\ mortality\ rate_{Country\ A} - Crude\ mortality\ rate_{Standard\ Population} =$$

$$= Crude\ mortality\ rate_{Country\ A} - Standardised\ mortality\ rate_{Standard\ Population} =$$

$$= \left(Crude\ mortality\ rate_{Country\ A} - Standardised\ mortality\ rate_{Country\ A}\right) \quad (metric\ 1)$$

$$+ \left(Standardised\ mortality\ rate_{Country\ A} - Standardised\ mortality\ rate_{Standard\ Population}\right) \quad (metric\ 2)$$

For the standard population, the crude and directly standardised mortality rates are equal. Therefore, I simply added and subtracted the same term (standardised mortality rate for country A) to the difference in crude mortality rates between country A and the standard population. This simple but novel decomposition based on standard epidemiologic measures provided the two metrics.

Metric 1 is the within-country difference in crude and gestation-standardised mortality rates. It reflects the contribution of inter-country variation in gestational age distribution to the differences in infant mortality rates. Positive values indicate the number of stillbirths and infant deaths per 1000 births that could have been prevented if country A had the same distribution of total births by gestational age as in the standard population. If the distribution of gestational age is more favourable in country A than in the standard population (e.g., Finland had lower preterm birth rate than Sweden), metric 1 shows negative values. Metric 1 reflects the influence of prenatal risk factors on the infant mortality rate. Metric 1 can therefore be used as an indicator of maternal health, wellbeing and socio-economic circumstances before and during pregnancy.

Metric 2 is the difference in gestation-specific mortality between country A and the standard population. Positive values indicate higher gestation-specific mortality rates and negative values indicate lower gestation-specific mortality rates compared with the standard population. Metric 2 measures differences in extended infant mortality rates *given* the gestational age of the child. Metric 2 can therefore be seen as an indicator of quality of care the child received after birth, and the contribution of other risk factors such as congenital anomalies. Finally, a comparison of gestation-specific mortality rates between country A and the standard population can help to identify characteristics of births (by gestational age and age at death categories) with the largest differences relative to the standard population. Such a comparison can

therefore indicate characteristics of children that would benefit most from strategies to reduce deaths. Interpretation of Metrics 1 and 2 is summarised in Box 2.1.

Box 2.1 – Interpreting the two metrics

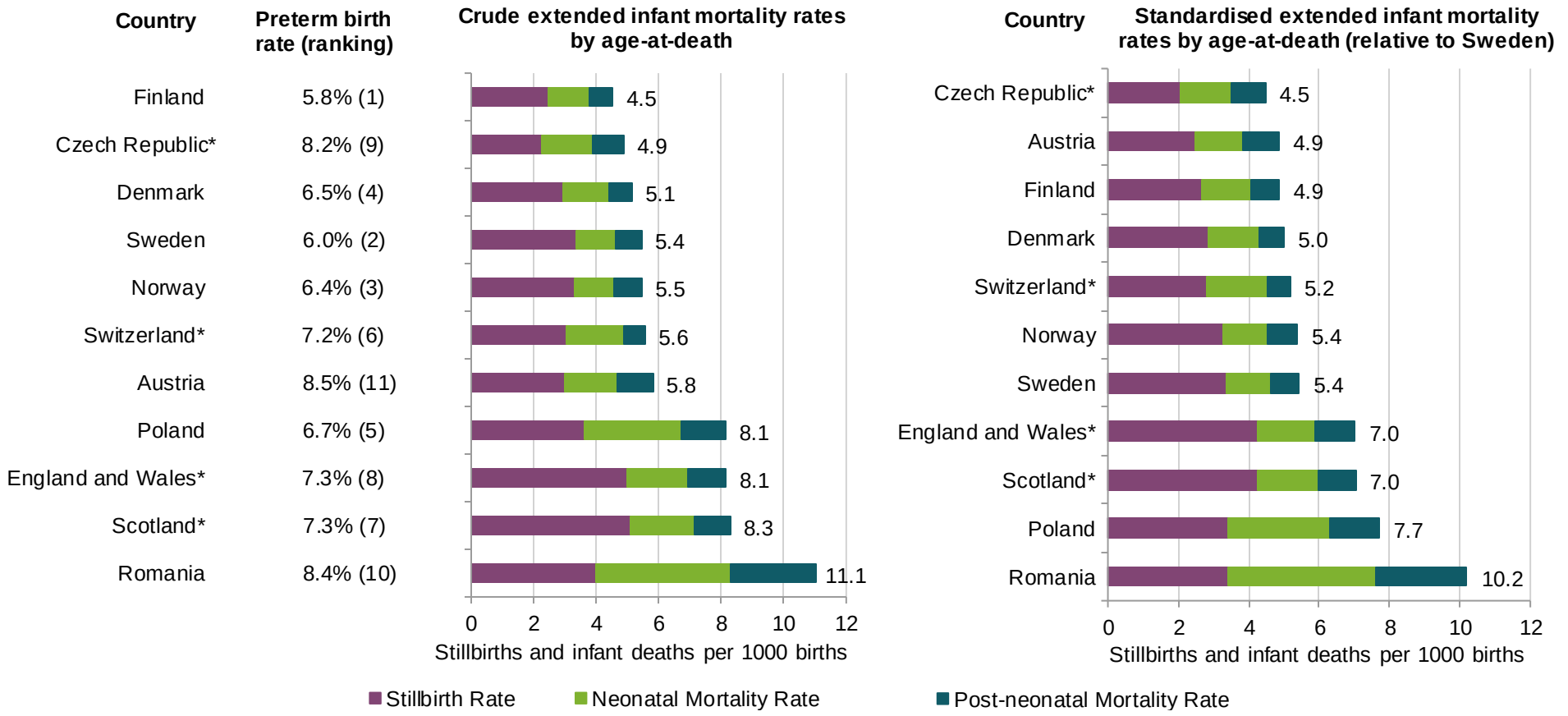
The difference in extended infant mortality between each country and Sweden can be decomposed into:

- Metric 1: within-country difference in crude and gestation-standardised mortality
  It indicates excess mortality attributable to prematurity, reflecting the influence of risk factors operating before and during pregnancy.
- Metric 2: between-country difference in gestation-standardised mortality
  This metric reflects excess mortality due to differences in the quality of infant care after birth.

2.4 Results

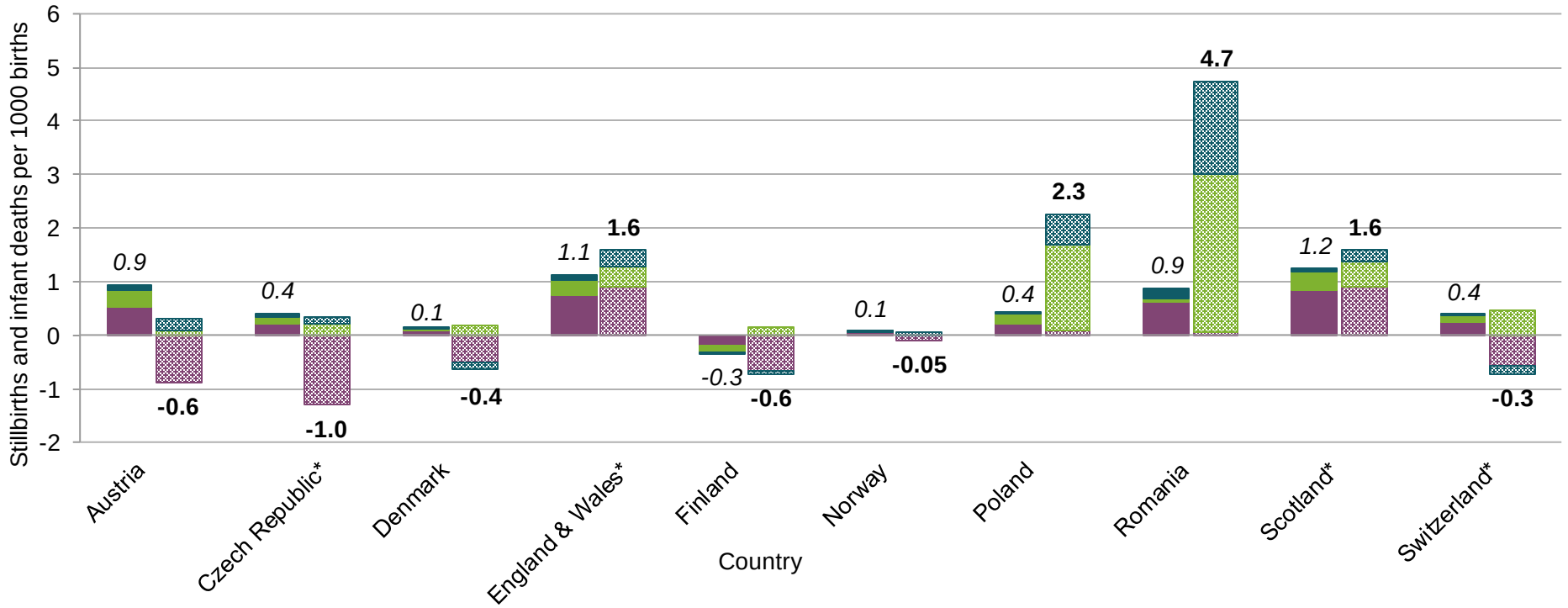
The study population comprised 1,977,051 births (1,969,173 live births and 7,878 stillbirths), 4,564 neonatal deaths and 2,625 post-neonatal deaths in eleven European countries. The key results for this chapter are presented in Figures 2.1-2.3. Figure 2.1 shows the proportion of preterm births and country rankings based on the crude extended infant mortality rates on the left-hand side, and the ranking based on gestation-standardised mortality rates on the right-hand side. Rates from Figure 2.1 were used to decompose inter-country differences in extended infant mortality rates into two metrics, presented in Figure 2.2. Bars on the left-hand side of Figure 2.2 show metric 1; bars on the right-hand side show metric 2. Figure 2.3 shows gestation-specific mortality rates, which are used to interpret high values of metric 2.

Figure 2.1 – Rankings of countries based on crude and gestation-standardised extended infant mortality rates by age at death (low to high mortality rates)



*Extended infant mortality was defined as the number of stillbirths and infant deaths per 1000 total births (live or still). The second column shows the proportion of total births born at <37 weeks' gestation. In Poland, access to terminations of pregnancy (TOP) was restricted. Countries with * included TOP in their counts of stillbirths. England & Wales and Scotland included terminations of pregnancy and stillbirths only after 24 weeks. All rates were calculated given gestational age was non-missing and ≥24 weeks.*

Figure 2.2 - Decomposition of the difference in crude extended infant mortality rates between each country and Sweden



Metric 1: Difference in crude and standardised mortality rates within the investigated country by age-at-death:

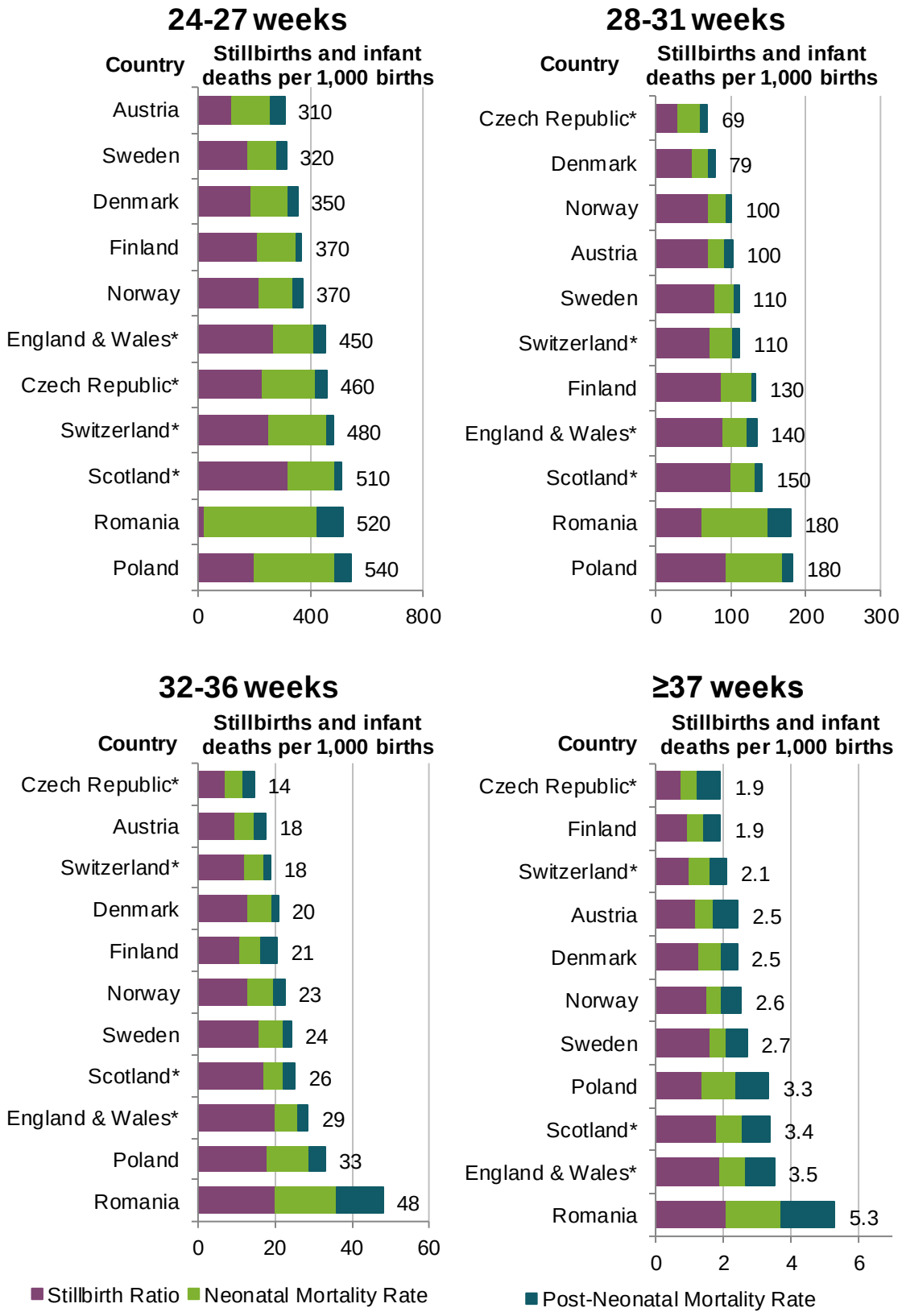
Stillbirth *Neonatal Period* *Post-neonatal period*

Metric 2: Difference in standardised mortality rates between the investigated country and Sweden by age-at-death:

Stillbirth **Neonatal Period** **Post-neonatal period**

*Extended infant mortality was defined as the number of stillbirths and infant deaths per 1000 total births (live or still). Solid bars (on the left-hand side) represent metric 1; patterned bars (on the right-hand side) represent metric 2. In Poland, access to terminations of pregnancy (TOP) was restricted. Countries with * included TOP in their counts of stillbirths. England & Wales and Scotland included terminations of pregnancy and stillbirths only after 24 weeks. Metrics were calculated given gestational age was non-missing and ≥24 weeks.*

Figure 2.3 – Gestation-specific extended infant mortality rates in each country by age at death.



*Extended infant mortality was defined as the number of stillbirths and infant deaths per 1000 total births (live or still). In Poland, access to terminations of pregnancy (TOP) was restricted. Countries with * included TOP in their counts of stillbirths. England & Wales and Scotland included terminations of pregnancy and stillbirths only after 24 weeks. All rates were calculated given gestational age was non-missing and ≥24 weeks.*

2.4.1 Countries with similar extended infant mortality rates as Sweden: Denmark, Finland and Norway

The four Nordic countries (Denmark, Finland, Norway and Sweden) had some of the lowest crude extended infant mortality rates (Figure 2.1, left-hand graph). This was driven by the low preterm birth rates observed in these countries (between 5.8% in Finland and 6.5% in Norway), and some of the lowest mortality rates for the high-risk babies born at 24-27 weeks (Figure 2.3).

Metrics 1 and 2 had values close to 0 for Denmark, Finland and Norway, indicating that both the gestational age distributions and the gestation-specific mortality rates were similar in these countries compared to Sweden (Figure 2.2). Small negative values of metric 2 for Denmark and Finland were primarily driven by lower gestation-specific stillbirth rates than in Sweden (by 0.5 stillbirths/1000 births in Denmark, and 0.7/1000 births in Finland, Figure 2.3), especially for stillbirths at ≥32 weeks' gestation (Figure 2.3).

2.4.2 Countries with significant contribution of both metrics: England & Wales and Scotland

The two metrics contributed almost equally to the differences between England, Wales, Scotland, and Sweden. England, Wales and Scotland had some of the highest extended infant mortality rates among the eleven countries studied (7.8 stillbirths and infant deaths/1000 births in England & Wales and 8.1/1000 births in Scotland, compared to 5.2/1000 births in Sweden). Metric 1 contributed to 41% of the difference in crude extended infant mortality between England & Wales relative to Sweden, and 44% of the difference between Scotland and Sweden (Figure 2.2). If England & Wales had the same prevalence of preterm birth as in Sweden, 1.1 fewer stillbirths and infant deaths per 1000 births would have occurred in 2010. Similarly, if Scotland had the same distribution of births by gestational age as in Sweden, 1.2 fewer stillbirths and infant deaths per 1000 births would have occurred in 2010. A slightly higher absolute reduction in extended infant mortality, 1.6/1000 births, could have been achieved if these countries had the same gestation-specific mortality rates as in Sweden. The differences in gestation-specific mortality relative to Sweden were largest for high-risk babies born <32 weeks (especially for stillbirths and neonatal deaths in babies born at 24-27 weeks and post-neonatal deaths in babies born at 28-31 weeks) and for neonatal and post-neonatal deaths in babies born at ≥37 weeks, that is, babies born at term, the vast majority of whom would be at low risk of death (Figure 2.3).

2.4.3 Countries where extended infant mortality was primarily driven by unfavourable gestational age distribution: Austria, Czech Republic and Switzerland

The absolute differences in extended infant mortality rates between Austria, Czech Republic, Switzerland and Sweden were small: the rate was lower by 0.5/1000 births in the Czech Republic, and higher by 0.4/1000 births in Austria and by 0.2/1000 births in Switzerland (Figure 2.1). Decomposing these differences into the two metrics illustrated that low infant mortality rates in Austria and Czech Republic were primarily driven by low gestation-specific stillbirth rates in these countries (metric 2, Figure 2.2). Reductions in extended infant mortality of 0.4/1000 births in Czech Republic and 1.0/1000 births in Austria relative to Sweden could be achieved by reducing the high prevalence of preterm birth, which was more than 2 percentage points higher in these two countries than in Sweden.

In Switzerland, approximately 0.4 stillbirths and infant deaths per 1000 births were attributable to unfavourable distribution of gestational age (metric 1, Figure 2.2). Metric 2 was overall close to 0, but decomposing by age-at-death showed that a further 0.4 excess stillbirths and infant deaths/1000 births were due to a higher gestation-specific neonatal mortality rate relative to Sweden, especially for babies born at <32 weeks' gestation. However, gestation-specific mortality was lower in Switzerland than in Sweden at 28-364 days of life for most gestational age categories (Figure 2.3).

2.4.4 Extended infant mortality rates primarily driven by high gestation-specific mortality: Poland and Romania

Poland and Romania had the highest rates of both crude and standardised extended infant mortality rates of the eleven countries studied (Figure 2.1). The differences between Poland, Romania and Sweden were primarily driven by the high gestation-specific mortality represented by metric 2 (Figure 2.2). If the two countries experienced Sweden's gestation-specific mortality rates, 2.3 fewer stillbirths and infant deaths per 1000 births would have occurred in Poland and 4.7/1000 births fewer in Romania. A further 0.4 stillbirths and infant deaths/1000 births in Poland and 0.9/1000 births in Romania were attributable to unfavourable gestational age distribution of births compared to Sweden. The differences in gestation-specific mortality rates between Poland and Sweden were largest for neonatal deaths for children born at <32 weeks, and for post-neonatal deaths for children born at 28-31 weeks. In Romania, the differences relative to Sweden were largest for neonatal and post-neonatal deaths across the distribution of gestational age.

2.4.5 Analyses based on birth weight

Analyses based on tabulations by birth weight showed similar results to analyses based on gestational age (Appendix B, Figures B.1-B.3). For Finland, Norway and Denmark, metric 1 based on standardisation by birth weight was close to 0. For all other investigated countries, metric 1 based on standardisation by birth weight was higher than when based on gestational age. An unfavourable distribution of birth weights compared to Sweden accounted for 1.1-1.7 excess stillbirths and infant deaths per 1000 births in these countries**.** Higher values of metric 1 reflected larger differences in the birth weight distributions between European countries than for gestational age. Babies born in Finland, Denmark, Sweden and Norway were heavier than infants born in other countries (Appendix B, Table B.1): a higher proportion weighed over 4.5kg (2.5-3.6% of births compared to 0.7-2.1% of births in other countries) and a lower proportion weighed <2.5kg (4.3-5.1% of births, compared to 5.9-8.2% of births in other countries).

2.5 Discussion

2.5.1 Key results

In this chapter, I presented two metrics for making international comparisons of early life mortality based on aggregated data more relevant for policy makers. In contrast to relying on crude extended infant mortality rates alone, these two metrics help to determine how much of inter-country variation in early life mortality can be explained by risk factors operating before and during pregnancy (leading to premature births), and what proportion reflects differences in care received after birth, given characteristics children are born with. These metrics can therefore be used to indicate the type and timing of interventions which would achieve the greatest reductions in stillbirth rates and infant mortality relative to Sweden.

2.5.1.1 England & Wales, and Scotland

In England & Wales and Scotland, the two metrics contributed almost equally to the difference in crude extended infant mortality rates relative to Sweden. This indicates that preventive strategies need to address both maternal health before and during pregnancy as well as the care of children after birth, given their gestational age. The prevalence of some of the risk factors operating during pregnancy which are associated with an increased risk of preterm birth was higher in the UK relative to Sweden and could be contributing to high values of metric 1: a higher proportion of mothers smoked during pregnancy (12.0% vs 4.9% in Sweden in 2010), were aged <20 years old (5.7% vs 1.6%) or were obese (20.7% vs 12.6%, based on information for Scotland).[12]

Some of the observed differences in gestation-specific mortality rates (illustrated by metric 2), especially in stillbirths and neonatal deaths at 24-27 weeks and neonatal deaths at ≥37 weeks, could reflect an increased prevalence of congenital anomalies in the UK compared to Sweden, which I could not adjust for (2.6% of total births in UK vs 2.3% in Sweden in 2010 according to the EUROCAT).[46] Many maternal risk factors associated with the risk of preterm birth, such as maternal smoking, obesity or age are also associated with increased risk of congenital anomalies.[62,63,74] Therefore, the benefits from reducing prenatal risk factors associated with preterm birth are likely to be higher than indicated by metric 1. Further comparison of cause-specific mortality (adjusted for gestational age) could identify risk factors operating after birth which contribute to high values of metric 2.

2.5.1.2 Austria, Czech Republic and Switzerland

While extended infant mortality rates were low in Austria, Czech Republic and Switzerland in relation to Sweden, metric 1 demonstrated that further reductions could be achieved by reducing the prevalence of preterm birth. Additional aggregated data on maternal risk factors during pregnancy including tabulations of births based on maternal smoking status or body mass index (BMI), which could have helped with interpretation of metric 1, were not available for these countries from the EURO-PERISTAT project.[12]

Czech Republic's high stillbirth rate at 24-27 weeks could reflect inclusion of TOPs in the count of stillbirths. Higher neonatal mortality at lower gestational ages relative to Sweden may indicate better neonatal care for high-risk babies in Sweden, since these high-risk babies are likely to be treated in neonatal intensive care units.

2.5.1.3 Poland and Romania

The differences in crude extended infant mortality rates between Poland, Romania and Sweden were primarily driven by differences in gestation-specific mortality (as indicated by metric 2). In Poland, the differences were largest for neonatal deaths at <32 weeks. This could reflect differences in obstetric and neonatal intensive care for high-risk babies and a higher prevalence of severe congenital anomalies, since access to TOPs for foetal abnormalities is restricted in Poland.[114]

In Romania, mortality differences relative to Sweden were observed across the gestational age distribution. This finding suggests that care needs to be improved both for high-risk babies (who are more likely to be cared for in hospital settings), as well as lower risk babies born at term who are more likely to be cared for at home. Some of the differences in gestation-specific mortality could reflect differences in the prevalence of congenital anomalies, however such additional data were not available from

EUROCAT for comparison.[46] Socio-economic factors could also be contributing to some of the differences in extended infant mortality between Romania and Sweden. Romania had the highest proportion of teenage mothers among the eleven countries in 2010 (10.6% compared to 1.6% in Sweden).[12] Babies born to teenage mothers are at an increased risk of adverse birth characteristics such as preterm birth, intrauterine growth restriction (IUGR), stillbirth or neonatal mortality.[12,27] Teenage motherhood is also strongly linked to lower socio-economic status (SES) (as discussed in Chapter 1, Section 1.3.2.1).[56] Furthermore, Roma people, who are more likely to be socially-deprived and have an increased prevalence of adverse birth characteristics and higher rates of child morbidity compared to non-Roma families,[115] constitute 8.3% of Romania's population.[116]

2.5.1.4 Standardisation by birth weight

Higher values of metric 1 based on standardisation by birth weight rather than gestational age indicated higher potential reductions in extended infant mortality rates relative to Sweden attributable to risk factors operating before and during pregnancy for all countries, apart from the three other Nordic countries. The distribution of birth weight in Norway, Denmark, Finland and Sweden was more favourable than in other countries included in the comparison, with a lower prevalence of low birth weight and a larger proportion of births weighing 3500-4499g (the birth weight category with highest infant survival).[41] This could at least partially reflect the fact that Scandinavian populations are taller,[117] as maternal and paternal heights are positively associated with birth weight.[78]

2.5.2 Strengths

The methods presented in this chapter provide important insights into the origin of differences in early life mortality between countries. The two metrics can approximate the contribution of exposures during pregnancy (as indicated by metric 1) and excess early life mortality due to the care after birth, given gestational age that a child is born with (as indicated by metric 2). These methods can be applied to aggregate data tables where stillbirth rates and infant mortality rates by age-at-death categories are broken down by one risk factor at a time, such as gestational age or birth weight. Detailed individual level data, which are difficult to access (due to privacy concerns) and time consuming to clean and analyse are not required.

2.5.3 Limitations

There is still substantial variation across Europe in the definitions used to define and report live and stillbirths by national statistics agencies. For example, Czech Republic, England & Wales, Scotland and Switzerland included terminations of pregnancy in the

counts of stillbirths, therefore leading to a higher stillbirth rate compared to other countries. To minimise the effect of these differences, the EURO-PERISTAT project recommended using a cut-off of 28 weeks of gestation for comparisons based on stillbirth rates. I included births at 24-27 weeks, since they are an important high-risk (but low prevalence) group for deaths in the first year of life, and the definitions were consistent in seven of the eleven included countries.

To enable fair inter-country comparisons, improvements in the completeness of the data are needed, especially for the recording of birth weight in stillbirths. Birth weight was more likely to be missing for stillbirths and infant deaths than for live births. Thus, comparisons based on aggregate data with known birth weight would underestimate extended infant mortality rates in countries with higher rates of missing data, biasing the results. Gestational age was more complete than birth weight; however, 20.5% of neonatal deaths and 35.7% of post-neonatal deaths in Romania had missing data on gestational age. As a result, extended infant mortality was underestimated by 2.6 stillbirths and infant deaths/1000 births and true values of metrics 1 and 2 are likely to be higher than these presented.

More detailed tabulations would provide a better understanding of the origins of inter-country differences in infant mortality. For example, I was not able to investigate mortality among post-term births separately to term births, as the EURO-PERISTAT report did not provide further breakdown of the ≥37 weeks' gestation category. Information about the timing of stillbirth (antepartum or intrapartum) could help distinguish between stillbirths due to prenatal risk factors and those related to the quality of obstetric care. However, more detailed tabulations would lead to small numbers per cell in countries with lower number of births. Thus, data for more than one birth year would be required to minimise the effect of chance variation.

While the metrics I presented are relatively simple, they are limited by investigating only one risk factor at a time. For example, some of the differences in gestation-specific mortality rates could reflect inter-country variation in the prevalence of congenital anomalies. For a fair comparison of extended infant mortality between countries, we need individual level data which account for multiple birth characteristics (such as birth weight, gestational age, and presence of congenital anomalies) and socio-economic factors. Such data would enable determining the relative contribution of gestational age, birth weight, congenital anomalies and other risk factors to the overall differences in extended infant mortality.

2.5.4 Implications of findings

Careful use of aggregate data tabulated by one key risk factor measured at birth could support the design of preventive strategies to reduce early life mortality. In order to

allow more informative comparisons of early life mortality between countries, national statistics agencies should routinely report counts of live births, stillbirths, neonatal and infant deaths tabulated by birth weight and/or gestational age categories to allow these metrics to be derived. The EURO-PERISTAT project has shown that many European countries have the capacity to collect and report such data.[12] Large perinatal datasets which could be used to derive these statistics are also available in regions of Australia,[118] Canada,[119] and the USA.[120] However, more funding is needed to ensure that complete data, based on standardised definitions of still- and live births, are collected on a regular basis in all high-income countries.

The conclusions reached by analysing aggregate data are limited by looking at only one risk factor at a time and some of the important risk factors determining a child's health at birth (such as congenital anomalies) remain unadjusted for. In order to carry out detailed analyses of origins of inter-country disparities in infant mortality, whole-country individual-level data with detailed information about characteristics at birth are needed. In the next two chapters, I describe administrative health databases in England and Sweden which can be used for such a detailed comparison, the results of which is presented in Chapter 5.

Chapter 3. Developing a national birth cohort using administrative linked datasets in England

What is already known:

- Aggregate data tabulated by one key risk factor at birth (such as gestational age) can provide important but limited insights into the origins of differences in infant mortality rates between countries.
- Inter-country comparisons of child mortality adjusted for multiple risk factors at birth (such as birth weight, gestational age, sex, presence of congenital anomalies) and maternal characteristics are needed to inform policies to reduce child mortality.

What this chapter adds:

- In this chapter, I develop a national birth cohort with information about characteristics of babies and mothers using administrative linked datasets in England.
- I evaluate whether the birth cohort is representative of the population of children in England and present approaches for dealing with missing data.

3.1 Chapter overview

In the previous chapter, I showed that country-level aggregate data tabulated by a key risk factor at birth (such as gestational age) provides important insights about the origins of inter-country differences in infant mortality. However, presented analyses were limited by focusing on only one risk factor at a time. International comparisons adjusted for multiple birth characteristics are needed to better inform policies to reduce infant and child deaths relative to a country with lower mortality rates.

This chapter presents work towards objective 2 of this thesis: "*to develop comparable national birth cohorts using administrative linked datasets in England and in Sweden with information on birth characteristics (birth weight, gestational age, sex and presence of congenital anomalies) and socio-economic factors (maternal age and quintile of socio-economic status (SES))*". I present methods for developing a birth cohort with longitudinal follow-up using an administrative hospital database linked to death registration data in England. I describe criteria for identifying births in the

administrative hospital database, methods for deriving information about birth characteristics and socio-economic factors of interest, and for validating the representativeness of the birth cohort against the population of children in England.

The methods described in this chapter were presented at the 2016 International Population Data Linkage Conference (Swansea, United Kingdom (UK)). The results are being prepared as a manuscript to submit for publication. I also intend to publish my Stata do-files for generating a birth cohort as a freely available resource for other researchers.

3.2 Background

3.2.1 National birth cohorts in high-income countries

Re-use of administrative linked datasets for research provides a rich source of data on health outcomes. Administrative data have the advantage of national coverage, which minimises selection bias due to loss to follow-up.[121,122] Linkage of administrative data from routinely collected maternity and child health records is time efficient and low cost compared to a birth cohort study involving de novo data collection. Large sample size and long follow-up times enable studying rare outcomes (such as child death) according to risk factors with low prevalence among children (such as congenital anomalies or extreme prematurity).[121,122]

Population-based birth cohorts from administrative linked datasets, covering key risk factors at birth, are increasingly being used in Australia,[118] Canada,[119] and the United States (USA).[120] The Nordic countries (Denmark, Finland, Iceland, Norway and Sweden) lead the way in this area, with a long tradition of collecting data from administrative sources in national registers covering information about all residents of the country.[122] A birth cohort in the Nordic countries can be defined using medical birth registers, containing details of antenatal, obstetric, and neonatal care and key characteristics of mothers and babies.[123] All residents in the Nordic countries are allocated a Personal Identity Number (PIN). PIN is a unique identifier which enables accurate linkage (with low error rates) between medical birth register and other health registers, such as hospital discharge registries (covering hospital admission trajectories), cause of death registries (covering information from death certificates) and non-health registers covering additional socio-demographic characteristics such as education, occupation or immigration status.[122]

3.2.2 National birth cohorts in England

In the UK, national birth cohorts based on birth registration datasets linked to longitudinal hospital admission data and death registration data are available and well

used for research in Scotland[124] and Wales,[125] though not in England. Since April 2015 maternity and child health services in England are required to contribute data collected in antenatal clinics (such as smoking status or body mass index (BMI) at first booking) and details of delivery and birth collected at the maternity ward (such as gestational age, delivery method, and diagnoses of the newborn baby) to the Maternity and Children's Data Set (MCDS).[126] Although this resource will be extremely valuable once established, it will take time to achieve whole-country coverage for all births (as of June 2017, only 88% of hospitals contribute data on births to MCDS).[127] Furthermore, the completeness of the key variables of interest requires improvements. For example, in June 2017, BMI was missing for 14% of women in the South of England and 23-24% elsewhere, while smoking status at booking was missing for 17% of women in the North of England and 8-9% elsewhere (rates of missing data were reported separately for London, the South of England, Midlands and East of England and the North of England).[127]

In the meantime, there are three existing administrative datasets covering births in England which could be used to create a whole country birth cohort. First, every birth in England and Wales is required by law to be registered within 42 days, thus birth registration data could be used to develop a national birth cohort.[128] The Office for National Statistics (ONS), the national statistics agency for England and Wales, collates official birth and death registration data from registry offices in each local authority. **ONS birth registration data** cover all births registered in England with near 100% completeness of birth weight and maternal age.[129] These data are routinely linked by ONS to death registration data to produce annual national statistics on child and infant mortality in England and Wales.[22] The second dataset is the **National Health Service (NHS) birth notification dataset** (formerly NHS number for babies (NN4B) dataset).[129] It was set up in October 2002 to issue NHS numbers (unique identifiers used in the healthcare setting in England) to all babies shortly after birth, rather than at birth registration, which could occur up to 6 weeks after birth.[130] The NHS birth notification dataset covers information on gestational age and ethnicity of the baby. It has been routinely linked by the ONS to birth and mortality registration data since 2005 for publication of annual national statistics on gestation-specific infant mortality.[131] These two datasets, however, lack information about clinical risk factors for child mortality, such as congenital anomalies.[129] Such information could be derived from diagnostic information recorded in **Hospital Episode Statistics (HES),** an administrative hospital dataset containing details of all inpatient admissions funded through the public health services in England (the NHS).[132]

Linkage of these three datasets (ONS birth registrations, NHS birth notifications, and HES) would provide a national birth cohort with high completeness of key risk factors at birth, whole-country coverage and several individual-level and area-level socio-economic indicators. However, linkage of the datasets is not straightforward and complex algorithms based on a number of identifying variables are required. Linkage for births in 2005-2014 was achieved by researchers at City University of London in 2016 (a year into my PhD). Accessing these data would involve seeking further permissions, including an application to the Confidentiality Advisory Group.[133] Instead, I looked for alternative solutions for my PhD which would be relevant and applicable to the substantial number of research groups in the UK who have access to de-identified extracts of HES data

3.2.3 Chapter aims

In this chapter, I present methods for developing a representative national birth cohort with information on birth characteristics (birth weight, gestational age, sex and congenital anomalies), socio-economic factors (maternal age and quintile of SES) and causes and timing of death using HES linked to ONS mortality data. An estimated 97% of all births in England occur in NHS hospitals and should therefore be recorded in HES since a hospital birth is considered an inpatient admission.[132] For each birth admission, HES contains additional details of the delivery and labour, which are comparable to the risk factor information recorded in the Nordic medical birth registers. Longitudinal linkage of a patient's hospital admissions trajectories and linkage to ONS mortality data are available for HES admissions from January 1998 onwards. The long period of data collection and national coverage indicate that HES can be a valuable source of information for studies of child health outcomes.

3.3 Datasets used to develop the English birth cohort

3.3.1 Hospital Episode Statistics (HES)

3.3.1.1 Overview

HES is the national administrative hospital database containing details of all admissions to NHS hospitals in England since 1989. It also contains information on all admissions to independent sector hospitals paid for by the English NHS.[134] Thus, HES covers an estimated 98-99% of all hospital activity in England.[132] HES is collated and maintained by NHS Digital, who provide extracts to researchers.[134]

Initially, HES was established to inform management and planning of healthcare services.[132] Since April 2004, data on all admissions is collected under Payment by Results (PbR), a pay for performance system of reimbursing hospitals based on the

interventions that the patients received and on the complexity of the conditions of the admitted patients.[135]

3.3.1.2 Structure

The basic analysis unit in HES is an episode of care (also known as a consultant episode), defined as the time during which a patient is under the care of one hospital consultant or other healthcare professional (e.g., uncomplicated pregnancies are fully managed by midwives). A hospital admission can consist of multiple episodes of care if a patient is seen by more than one consultant/healthcare professional.[132]

HES extracts are provided by financial years (which run from the 1st April to 31st March the following year in England).[132] An episode of care is marked as "finished" if it ended before the start of the new financial year.[20] If an episode started before the 31st March and finished after the 1st of April, it will be recorded twice: as an "unfinished" episode in financial year finishing on the 31st of March and as a "finished" episode in the following financial year.[136] Thus, unfinished episodes need to be removed to avoid duplication.[136]

3.3.1.3 Recorded information

Each episode of care recorded in HES includes a patient's details (e.g., age at the start and end of an episode, month and year of birth, sex, ethnicity, partial postcode), admission details (e.g., dates and methods of admission and discharge, episode start and end dates, discharge destination, hospital name) and clinical details (e.g., diagnoses, procedures and causes of injury).[134]

3.3.1.3.1 Birth and maternal characteristics

For every birth, at least two episodes of care get recorded in HES – a birth episode for each baby and a delivery episode for the mother. The maternal and birth episodes contain an additional 19 variables with the details of the delivery and labour, called the "baby tail". Information recorded in the "baby tail" includes gestational age, birth weight, sex and maternal age.[137,138] Maternal delivery records are often more complete than birth records.[129]

3.3.1.3.2 Birth dates

Accurate birth dates are crucial for precise calculation of age at death. This is estimated in HES as the admission date of the identified birth episode.

3.3.1.3.3 Clinical information

Diagnostic information recorded in HES can be used to identify comorbidities, such as congenital anomalies. For each episode of care, HES contains up to 20 diagnoses (up to 14 before April 2007, and up to 7 before April 2002).[136] Diagnoses are coded using

the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10), a medical classification list developed by the World Health Organization (WHO) and used in hospital databases in over 50 countries around the world (as of 2013).[132,139] Diagnostic information recorded at birth comes from maternity data systems and hospital notes. Therefore, congenital anomalies recorded in birth episodes may have been diagnosed through antenatal screening or during the postnatal stay in hospital, such as via the routine newborn physical examination.

Each episode of care can also contain up to 24 procedure codes (up to 12 before April 2007 and up to 4 before April 2002).[136] These include surgery, diagnostic imaging, ventilation and infusion/transfusion therapy.[132] Procedures are coded using the Office of Population Censuses and Surveys Classification of Interventions and Procedures (OPCS, currently version 4.7), a coding system unique to the UK.[132]

All diagnoses and procedures are entered by clinical coders who translate discharge notes into appropriate ICD-10 codes.[132] All coders follow a set of standardised national guidelines to ensure consistency in the recorded data across the country. However, coders rely solely on details recorded in hospital case notes and discharge notes (which they cannot interpret), so differences in diagnostic practices could remain between hospitals.[132]

Since the introduction of PbR in April 2004, diagnoses and procedures recorded in HES have been used to calculate the cost of each episode of care using Healthcare Resource Groups (HRG) – a grouping of patients' diagnoses and procedures which require use of common levels of healthcare resources.[140] This introduced a financial incentive for hospitals to improve the diagnostic and procedure coding depth and accuracy, as hospital reimbursement depends on the complexity of patient's conditions. For example, the number of diagnoses reported per episode has increased since 2004.[132,141] Thus, trends in admission rates for particular diagnoses, particularly those relating to chronic comorbidities, need to be interpreted with caution.

3.3.1.3.4 Socio-economic status indicator

Infant and child mortality rates are strongly associated with socio-economic status. In HES, SES is measured by the Index of Multiple Deprivation (IMD) score, a small area-level indicator of deprivation.[136] IMD scores are allocated at the Lower Layer Super Output Area (LSOA) level. Each LSOA covers between 200-1400 households.[132] Each patient in HES is assigned an IMD score based on their postcode at admission.

IMD scores are calculated based on indicators in seven domains: income, employment, health and disability, education, crime, barriers to housing and services, and living environment.[136] All domain-specific scores are also included in HES. These scores are

mainly based on indicators recorded in census data.[136] The scores reflect changes in deprivation of areas over time:

- Records up to financial year 2006/7 use IMD version 2004 (with scores based on data from 2001)
- Records in financial years 2007/08-2009/10 use IMD version 2007 (based on data from 2005)
- Records for financial years from 2009/10 onwards use IMD version 2010 (based on data from 2008).[136]

3.3.2 Office for National Statistics (ONS) mortality data

3.3.2.1 Overview

ONS mortality data cover all deaths registered in England in a given calendar year, both among residents and non-residents.[142] Deaths to English residents occurring abroad are not included (apart from members of Her Majesty's Armed Forces).[23]

3.3.2.2 Data collection process

All deaths in England are required by law to be registered within 5 days. In certain circumstances (e.g., if the cause of death is unknown or if the death was violent, unnatural or suspicious) the death might need to be referred to a coroner.[23] These deaths can only be registered once the coroner's investigation is closed and the causes of death are identified, and there are no time restrictions for the length of coroner's investigation.[23,143] Therefore, more recent deaths may be undercounted in ONS mortality data due to delayed registration. In 2011, 66.7% of neonatal deaths in England and Wales were registered within 5 days, and 21.2% were registered within 6-30 days.[144] Deaths from sudden infant death syndrome (SIDS) are more likely to require coroner's inquest, as certification is based on exclusion of other plausible causes of death.[89,90] In 2011, median registration delay for SIDS was 149 days in England and Wales (inter-quartile range: 97-220 days).[144]

3.3.2.3 Death certification in England

There are two types of death certificates used to register a death in England. For deaths at ≥28 days of life, information about the causes of death is recorded using a death certificate compatible with the international template recommended by the WHO, which consists of two parts.[23] Part I details the underlying condition and the sequence of conditions that lead directly to death.[23] Part II lists any additional conditions that contributed to death, but were not part of the causal sequence ending in death.[23] Since

1993, the selection and coding of the underlying cause of death is done using automated software based on WHO rules, comparable with that used in Sweden.[23]

England is one of the few countries in the world using the neonatal death certificate recommended by the WHO for stillbirths and deaths before 28 days of life.[22] The neonatal death certificate gives equal weighting to the main conditions in the foetus/child and the mother;[22] therefore, it is not possible to identify a single underlying cause of death for neonatal deaths using this death certificate.[22] For this study, information about stillbirths recorded in the ONS mortality data were not available.[145]

3.3.2.4 Recorded information

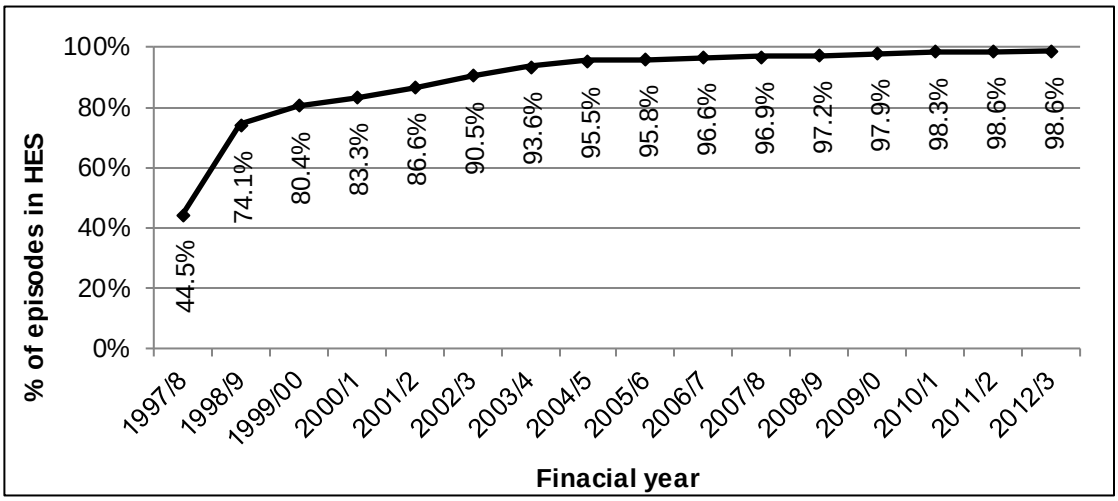
The ONS mortality data includes information on all causes of death, the place and date of death, and the date the death was registered.[23,143,145] Up to 15 causes of death can be reported, in addition to the underlying cause.[58] ICD-10 has been in use since January 2001.[58]

3.3.3 Following a patient across time in linked HES-ONS data

3.3.3.1 NHS number

Unlike in the Nordic countries, residents in England do not have a unique identifier like the PIN which is used extensively in all areas of society (such as healthcare and taxation). The NHS number is a unique identifier used in the publicly provided healthcare setting in England. However, it is not always recorded during hospital admissions (Figure 3.1). For example, the NHS number was likely to be missing for birth episodes before implementation of the NN4B service on 29th October 2002 (NN4B service is the basis of NHS Birth Notification dataset described in Section 3.2.2).[130] The NN4B system enabled midwives and other maternity unit staff to request an NHS number for newborns in hospital shortly after birth using an on-line system. Prior to 2002, babies had to wait until official birth registration at a local registrar's office to obtain their NHS number, which could take up to 6 weeks.[146] An NHS number allocated with a delay was unlikely to be updated in hospital birth record.

Figure 3.1 – Percentage of all episodes of care recorded in HES with a valid and complete NHS number by financial year



HES=Hospital Episode Statistics; NHS=National Health Service. Financial year is defined from the 1st April to 31st March the following year. Data from Hipisely-Cox et al.[147]

3.3.3.2 Linkage of a patient's records over time in HES

NHS Digital links patient's hospital admissions over time in HES using a unique pseudonymised patient identifier, called the HESID.[148] To ensure that a valid link is present even in the absence of the NHS number for any of the admissions, HESID is generated using three sets of rules, based on the NHS number, hospital code and local patient identifier in that hospital, date of birth, postcode and sex (see Box 3.1).[148] A set of records are allocated the same HESID if a match is found using any of these steps. If no match is found, a new HESID is generated.[148] A match cannot be made for records with conflicting NHS numbers. HESIDs are available for episodes of care which began after the 1st April 1997, when it became mandatory for hospitals to record a patient's NHS number.[132]

The linkage algorithm presented in Box 3.1 was designed to minimise the probability of false matches (that is, two different patients being assigned the same HESID). However, this led to an increased probability of missed matches (that is, the same person being assigned multiple HESIDs).[149] A study testing the HESID generating algorithm in a gold standard dataset with independently allocated patient identifier (the Paediatric Intensive Care Audit Network (PICANet) data, covering records from 33 paediatric intensive care units in England) estimated that the HESID algorithm resulted in false match rate of 0.2%, and missed match rate of 4.1% in children aged 0-19 years in 2004-2014.[149]

Box 3.1 – HESID generating linkage algorithm[148]

Step 1: Sex (exact match), date of birth (partial match), NHS number (exact match);

Step 2: If NHS number is not conflicting:
Sex (exact match), date of birth (partial match), postcode (exact match), local patient identifier within a hospital and hospital code (exact match);

Step 3: If NHS number is not conflicting:
Sex (exact match), date of birth (exact match), postcode (exact match)

3.3.3.3 Linkage of HES to ONS mortality data

Linkage between HES and ONS mortality data is available for deaths registered from the 1st January 1998 onwards and is carried out monthly by NHS Digital using date of birth, sex, NHS number and postcode.[145] There are eight hierarchal criteria, referred to as match ranks, which are used to match identifiers from ONS mortality data with those kept for each HESID in HES (listed in Box 3.2). If a death record is matched with more than one HESID then the best quality match is kept.[143]

Box 3.2 – Algorithm for linking HES with ONS mortality data based on agreement between NHS number, date of birth, sex and postcode[143]

Match rank	NHS number	Date of birth	Sex	Postcode
1 (best quality)	Exact match	Exact match	Exact match	Exact match
2	Exact match	Exact match	Exact match	
3	Exact match	Partial match	Exact match	Exact match
4	Exact match	Partial match	Exact match	
5	Exact match			Exact match
6	Not contradicting	Exact match & not 1st January	Exact match	Exact match & not a communal establishment (e.g., hospital, prison, army barracks, etc.)
7	Not contradicting	Exact match & not 1st January	Exact match	Exact match
8 (worst quality)		Exact match & not 1st January	Exact match	Exact match

HES=Hospital Episode Statistics, NHS=National Health Service, ONS=Office for National Statistics

NHS Digital carries out the linkage and provides the linked ONS mortality data (with pseudonumised HESIDs for linkage with HES) to researchers.[143] ONS mortality data available for this PhD included all death records which have been matched to a HESID (death records which have not been linked to a HESID were not included).

NHS Digital also flags all hospital deaths recorded in HES (where the discharge method was recorded as 'died'). Deaths identified using HES only are also included in the provided ONS mortality data, even if no link to ONS mortality record was found. However, these deaths do not have any recorded causes of death.

3.4 Methods for developing an English birth cohort using HES

This section presents all steps taken to develop a representative birth cohort using linked HES and ONS mortality data (referred to as HES-ONS data). A de-identified extract (that is, not including the NHS Number, or exact postcode) was re-used with the permission of NHS Digital under a data sharing agreement for a programme of research on child mortality. Since the extract was de-identifiable, I did not require ethics approval to use the data.[150]

3.4.1 Identifying births in HES

3.4.1.1 Inclusion criteria

First, I extracted all HES episodes with an age at admission <7 days. I then applied broad selection criteria based on diagnostic and procedure codes, healthcare resource group codes and administrative variables recorded in HES (such as admission method or level of provided neonatal care) to identify birth episodes. Details of the selection criteria are presented in Appendix C, Table C.1.

3.4.1.2 Exclusion criteria

I excluded multiple births, terminations of pregnancy and stillbirths from birth episodes identified in Section 3.4.1.1. Exclusion criteria were based on diagnostic codes and admission fields recorded in HES (see Appendix C, Table C.2 for details).

I excluded multiple births since the strength of association between birth characteristics (such as preterm birth, low birth weight and chromosomal abnormalities) and the risk of death is different for singleton and multiple births.[54] Furthermore, same sex siblings in a multiple birth are more likely to be allocated the same HESID in the absence of NHS number.[148] As part of the HESID generating algorithm, records with the same sex, postcode and date of birth, where the NHS number is not conflicting (e.g., if it is missing in at least one record, see Box 3.1) are assumed to belong the same

individual. This could affect a high proportion of multiple births, particularly prior to 2002, before the implementation of the NN4B system.

A small number of births were flagged as terminations of pregnancy and were excluded. These were likely to be maternal records with miscoded age at admission. I excluded stillbirths to match the inclusion criteria in the extract of the Swedish Medical Birth Register (SMBR) which I used to define the birth cohort in Sweden, described in Chapter 4. The SMBR extract used for this PhD did not contain information on stillbirths.

In order to match inclusion criteria to the Swedish registers, I also needed to exclude non-English residents from the cohort. For this, accurate and complete information about postcode was required. Thus, this exclusion criterion is described in Section 3.4.3, where I describe cleaning and enhancing of variables through linkage of babies' and mothers' records in HES.

3.4.1.3 Data cleaning

I removed implausible recordings of risk factors at birth (birth weight, maternal age, gestational age) and dates (admission and discharge dates, episode start and end dates). All data cleaning rules are listed in Appendix C, Table C.3. I dropped unfinished episodes (that is, episodes of care which started in one financial year and finished in another), as they should not contain any clinical information.[136] I also removed false matches, which I defined as one HESID with two or more birth episodes with conflicting (non-missing) information on birth weight, gestational age, maternal age, or month and year of birth.

In remaining cases where one baby had several birth episodes, I assumed that these were consecutive episodes of care within a birth admission (e.g., if a baby was seen by more than one consultant around the time of birth). I kept the episode with the earliest admission and episode start dates as the birth episode. Finally, I excluded births outside the period from the 1st January 1998 to 31st December 2012, since information about deaths from ONS mortality data is only available from the 1st January 1998, and the extract of SMBR available for this study covered births until the 31st December 2012.

3.4.2 Longitudinal follow-up data until the fifth birthday

3.4.2.1 Hospital admission trajectories

I extracted all episodes of care where the age at admission was <5 years for HESIDs with birth episodes identified as described in Section 3.4.1. I then "cleaned the data". First, I removed episodes with no clinical information recorded (e.g., unfinished

episodes). I then validated date variables (such as admission and discharge dates) which can contain recording errors (e.g., if recorded admission date was after discharge). Where possible, I corrected the recording errors. Finally, I de-duplicated the episodes. Details of data cleaning rules are described in Table C.4 in Appendix C.

I then linked episodes into admissions to match information recorded in the Swedish Hospital Discharge Register. An admission was defined as the total time spent by a patient in one hospital, therefore hospital transfers were classified as separate inpatient admissions.

As mentioned in Section 3.3.3, the HESID generating algorithm relies heavily on the NHS number, which was likely to be missing in birth episodes of children born prior to 29th October 2002 when the NN4B system was introduced.[130] I examined trends over time in the proportion of children with at least one hospital admission after birth in first year of life in order to assess if the implementation of NN4B affected linkage of patient's longitudinal hospital admissions for births identified in HES.

3.4.2.2 Linkage to ONS mortality data

The ONS mortality data were linked to births identified in HES using the HESIDs provided by NHS Digital. I identified additional deaths not indicated by NHS Digital by flagging hospital admission records where the discharge method indicated death.

I excluded deaths occurring after a child's fifth birthday. To match the duration of follow-up available in the Swedish national registers, I also excluded deaths which occurred after the 31st December 2013. Further, I excluded false matches, which I defined as:

- records with a date of death before the birth date:
  where the difference was greater than one day, or if there was any difference between these dates for links with the poorest quality match rank (that is, where the NHS number was not required to match – match rank 8 in Box 3.2)
- records with subsequent hospital admissions after death:
  where there was >1 day's difference between the last admission date and death date (admissions one day after death could occur if for example, test results are released and recorded in the system after discharge and death)[151]
- records for in-hospital deaths where the difference in the date of discharge (and death) in admission record in HES and the date of death in ONS mortality records was >1 day:
  a difference of one day was deemed to be plausible if, for example, discharge

was not possible on the day of death or some test results emerged after the death of a patient [151]

To evaluate quality of linkage between the birth records identified in HES and ONS mortality data, I compared mortality rates in infancy (at 0-27 days and 28-364 days) based on the linked HES-ONS birth cohort with national child mortality statistics published by the ONS for England and Wales.[152,153]

As mentioned in Section 3.3.3.3, information about a death could be indicated in the HES inpatient admission record with no link to ONS mortality data. Such deaths do not have any recorded causes of death. I assessed completeness of recorded causes of death in the HES-ONS birth cohort to check the impact of these missed links for analyses of cause-specific mortality in Chapter 6.

3.4.3 Recording of key risk factors for child mortality

Risk factors of interest for this study recorded in HES included birth weight, gestational age and sex, maternal age, postcode and IMD score. I also developed an indicator of congenital anomalies using diagnoses recorded in longitudinal inpatient admission records in HES and causes of death recorded in ONS mortality data (described in Section 3.4.3.2).

3.4.3.1 Improving the completeness of risk factor variables using mother-baby linkage in HES

As outlined in Section 3.3.1.3 above, variables recorded in the "baby tail" are kept in two separate records in HES – a birth record for each baby and a delivery record for the mother.[132] Maternal delivery records are often more complete than birth records.[129] Therefore, completeness of recordings of birth weight, gestational age, maternal age, IMD scores and postcodes can be improved by replacing missing values in the baby record with complete recordings from the maternal delivery record. In this chapter, I refer to this process as "enhancing" the data, resulting in "enhanced" birth weight, gestational age, maternal age, IMD scores and postcodes.

Methods for linking mothers and babies were developed by Dr Katie Harron and are described in detail in Section C.3 of Appendix C. In brief, maternal delivery episodes and babies' birth episodes can be linked in the de-identified HES database as much of the information recorded in the two episodes overlap (such as variables describing maternal characteristics, pregnancy, delivery and birth outcomes which are recorded in the "baby tail" or residency details).[154] Harron *et al.*[154] developed methods for identifying births and deliveries in the HES dataset and linking mothers with their babies using deterministic and probabilistic methods. Deterministic linkage, such as the algorithm for linking HES inpatient admissions with ONS mortality data, requires an

exact or approximate agreement between a set of identifiers such as date of birth, postcode or sex to make a match (conflicting information is not permitted).[154] Probabilistic linkage methods allow calculating the likelihood of a match, given the agreement or disagreement in the set of observed identifiers amongst all possible pairs. The pair with the highest likelihood is identified as a match.

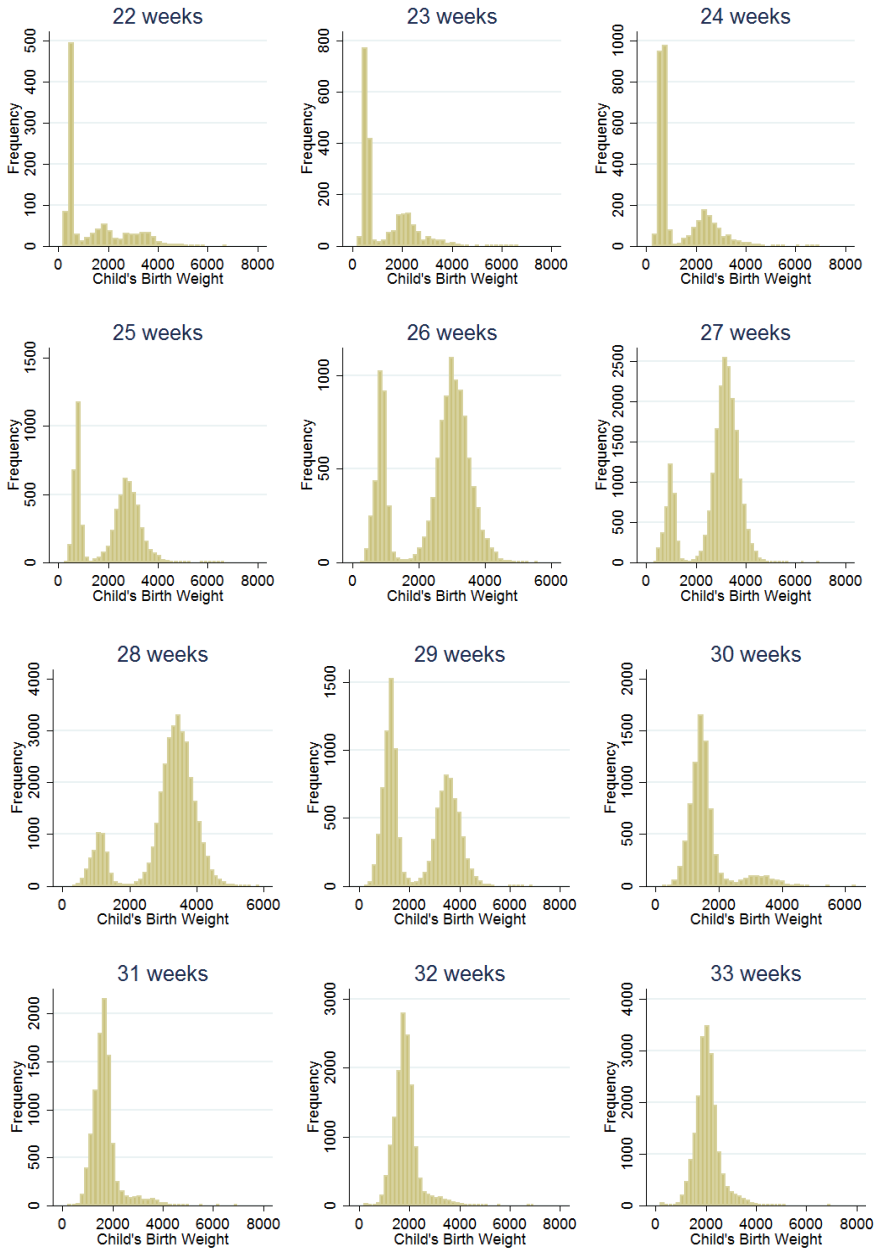
Harron et al.'s[154] linkage algorithm was replicated by my UCL colleague Dr Linda Wijlaars in the HES extract available for this study.

3.4.3.2 Cleaning birth characteristics variables

3.4.3.2.1 Birth weight and gestational age

Some hospitals in HES are known to report gestational age in days rather than in weeks in their maternity systems. The last digit consequently gets truncated by the HES cleaning algorithm; for example 280 days (40 weeks) would be recorded as 28 weeks.[155] Such errors lead to misclassification of term births as preterm, resulting in a bimodal distribution of birth weight at lower gestations (Figure 3.2), and biasing downward the estimates of child mortality in high risk babies born at early gestations as term babies have a much lower risk of death.

Figure 3.2 – Distribution of birth weight by week of gestation before removing implausible combinations of birth weight and gestational age in HES-ONS birth cohort



HES=Hospital Episode Statistics; ONS=Office for National Statistics.

To minimise the impact of these recording errors, I changed values of birth weight and gestational age to missing if the recorded birth weight fell outside +/-4 standard deviations (SD) of mean birth weight for each gestational age. To obtain birth weight centiles, I used LMSgrowth, a Microsoft Excel add-in with growth references for children in the UK, developed by Pan and Cole.[156] For preterm babies, I used birth weight centiles based on the *UK WHO preterm* reference, which was extrapolated to 22 weeks; for term babies born from 37 to 42 weeks I used *UK WHO term* reference. Data on 43-45 weeks was unavailable. Investigating birth weight curves from Australia and USA revealed that mean birth weight and centiles do not increase further after 42

weeks, so I used the values for 42 weeks as cut-offs for higher gestations.[157–159] The growth references are sex-specific. For a small number of records with missing recording for sex, I used values of birth weight centiles which were overlapping between boys and girls. That is, I used -4SD values for boys (as they were higher than for girls) and +4SD values for girls (as they were lower than for boys).

3.4.3.2.2 Sex

Where missing, information about sex of the baby was completed using longitudinal hospital admissions records and the ONS mortality records (where available) by taking the mode of recorded sexes across records.

3.4.3.2.3 Congenital anomalies

Presence of congenital anomalies may not be immediately obvious at birth, as it could take time for some of the anomalies to manifest and be diagnosed. Therefore, I indicated children as having a congenital anomaly if they had a relevant ICD-10 code recorded as any diagnosis within first two years of life, or as any cause of death recorded in the ONS mortality data. I used ICD-10 codes for congenital anomalies taken from a chronic condition code list developed by Hardelid et al. which identifies children that require medical follow-up for more than 12 months in 50% or more of cases.[58] I used only codes beginning with "Q", from Chapter 17 of ICD-10 "*Congenital malformations, deformations and chromosomal abnormalities*" included in the Hardelid et al.'s code list.[24]

3.4.3.3 Socio-economic factors

3.4.3.3.1 Maternal age

Maternal age was enhanced through mother-baby linkage using mother's age at admission for delivery.

3.4.3.3.2 IMD score and postcode

In the financial years 2007/8 - 2012/13, the patient postcode and all variables derived from the patient's postcode (including the IMD score) were missing from all birth episodes where the episode type was specified as birth.[160] In my cohort, this accounted for 85% of all singleton live births in 2007/8-2012/13. This was the result of an extraction error while processing HES extracts by NHS Digital.[160] It is possible that birth episodes before 2007/8 were also affected; however, issues with the quality of HES data identified by NHS Digital during data processing were not documented prior to 2007/8.

Maternal delivery records have near 100% completeness of postcode, thus enhancing the data through mother-baby linkage was crucial for obtaining information about the only measure of socio-economic status in HES – the IMD score. For babies' HESIDs which did not link to a maternity record, I copied the earliest recording of postcode and IMD scores from longitudinal hospital admissions in infancy to the birth record. This helped to maximise the completeness of recording the available information on postcode.

I then calculated quintiles of IMD scores amongst all pregnant women in a given calendar year, in order to derive a comparable indicator of SES to that available in the Swedish cohort (described in Chapter 4). To match inclusion criteria to the Swedish birth cohort (also described in Chapter 4) I used the enhanced information on postcode to exclude non-English residents from the birth cohort.

3.4.3.4 Cohort validation

I first evaluated rates of missing data for each risk factor of interest among live births and among deaths (by age at death). I compared mortality rates in infancy based on all births in the HES-ONS birth cohort ("whole cohort"), and in the "complete case" cohort, defined as cohort of births with complete information on all birth characteristics (birth weight, gestational age, sex) and socio-economic factors (maternal age and IMD score), with rates reported for England and Wales, published by the ONS (and freely available on the ONS website).[42,44,161,162]

I then validated the distribution of birth weight, gestational age and maternal age in live births in the whole and complete case cohorts against national statistics from ONS for singleton live births in England and Wales.[34,42,161] Finally, I compared mortality rates by age at death and categories of birth weight, gestational age and maternal age in the complete case HES-ONS birth cohort, with rates reported by ONS for England and Wales.[34,42,161]

3.4.4 Strategies for dealing with missing data

Mother-baby linkage substantially increased the completeness of risk factors at birth and socio-economic factors; however, 16.8% of records were still missing birth weight and 22.3% were missing gestational age. Thus, I explored two approaches to handling of missing data:

- Multiple imputation by chained equations
- Identifying a sub-cohort of hospitals which provide high quality of recorded data for complete case analyses

3.4.4.1 Multiple imputation (MI) using chained equations

3.4.4.1.1 Overview

Multiple imputation (MI) by chained equations is a simulation-based statistical method commonly used to analyse datasets with missing data for multiple variables of interest.[163,164] First, multiple copies of the original dataset are generated, with missing values being replaced with a set of plausible values based on the distribution of the observed data.[163,164] This is done in an iterative process as follows:

1. Initially, all missing values are filled at random (by sampling with replacement from the values observed in the dataset)
2. Next, the variable with the lowest proportion of missing values is regressed on all other variables in the imputation model. The imputation model should include:
    - all variables to be imputed
    - all variables to be used in the final analyses
    - any variables which could predict the patterns of missing data
3. Missing values for the given variable are then replaced by random draws from the posterior predictive distribution, based on estimates from the imputation model
4. Steps 2 and 3 are repeated for all variables with missing data in an iterative "cycle".
5. To generate one imputed dataset, the process is repeated for a number of cycles to stabilise the results.
6. Steps 1-5 are repeated *m* times to generate *m* imputed datasets

Next, each of the imputed datasets is analysed in an identical way and the results are pooled together into an overall estimate and variance-covariance matrix using Rubin's rules.[163,164]

3.4.4.1.2 MI in HES-ONS birth cohort

For this study, I aimed to impute birth weight, gestational age, maternal age, sex and IMD score. The imputation models additionally included:

- all variables for the analyses comparing child mortality between England and Sweden:
    - a binary indicator of congenital anomaly (yes/no)
    - a binary indicator of a death in the first five years of life (yes/no)
    - Nelson–Aalen estimator of cumulative hazard function recommended for multiple imputation of data used for Cox proportional hazards

regression).[165] It is a non-parametric estimator (i.e. no assumptions about the underlying distribution of cumulative hazard function are made) and it measures accumulated risk of death by the time t_i. Nelson–Aalen estimator is calculated as the sum of estimated probabilities of death in each time interval t_j:

$$\tilde{H}(t) = \sum_{j=1}^{k} \frac{d_j}{n_j},$$

where d_j denotes the number of deaths and n_j denotes the number individuals at risk of death (i.e. alive, not censored,) at each interval t_j.[166]

- possible predictors of incomplete data in the model:
    - a binary indicator of presence of a chronic condition diagnosed in first month of life (as it could be associated with low birth weight or preterm birth) generated by scanning admissions at age <28 days of life for any of the ICD-10 codes from the chronic condition code list developed by Hardelid et al.[58]
    - Categorised length of post-natal stay (0-6 days, 7-30 days, 31+ days). Longer stays are an indicator of child's poor health at birth
    - Financial year, as the rates of missing data varied by the financial year in which maternity data were sent to HES
    - A categorical variable indicating linkage outcome to a delivery record (whether the link was deterministic, probabilistic or missing as an overall indicator of quality of recorded data)

I generated five imputed datasets to begin with (since the process was computationally intensive on a big dataset like HES-ONS birth cohort (approximately 3GB in size). To test the imputation results, I calculated the proportions of births and infant mortality rates per 1000 births by categories of maternal age, birth weight and gestational age, based on each imputed dataset. I then pooled these together using Rubin's rules (that is, by taking an average).[163,164] I compared the pooled estimates with national statistics reported in official ONS publications for England and Wales.[34,42,161]

3.4.4.2 Restricting the cohort to hospitals providing high quality data on risk factors

The quality and completeness of recorded data on birth weight and gestational age in HES is known to vary between reporting hospitals. For example, some hospitals have standalone maternity systems and do not report any information to HES.[167,168] Some hospitals are known to record gestational age in days rather than weeks, leading to misclassification of term babies as preterm (see Section 3.4.3.2.1 above).[155] Thus, I

developed criteria for classifying hospitals based on the quality of recorded data on birth weight and gestational age in order to exclude hospitals reporting "poor" quality data.

The completeness and accuracy of information recorded in HES improved over time (as shown in the results Section 3.5.3), possibly due to changes in maternity systems used by hospitals to record variables in the "baby tail". Therefore, I assessed the quality of recorded data in each hospital separately for each financial year. As a result, a hospital could have been excluded in earlier financial year, but included for later years due to improvements in completeness of reported data.

A priori, I excluded hospitals which had:

- no births with recorded birth weight and gestational age in a given financial year, as missing data on both of these risk factors would indicate that they did not report any information from maternity systems to the "baby tail" in HES[167,168]
- <25% of records with complete birth weight and gestational age in a given financial year
- <500 births per financial year, as mortality rates in these hospitals were more prone to chance variation (I also excluded one particular hospital with 503 births in one financial year and <500 births in remaining years)

I explored a number of additional exclusion criteria by investigating correlations between hospital characteristics (by financial year) and hospital's rates of missing data in live births and in deaths. All investigated indicators are listed in Appendix C, Section C.4. For each indicator, I selected a cut-off for defining "high" and "poor" quality of recorded data by visual examination of histograms and scatter plots for the indicator against rates of missing data.

Selecting the final criteria was done in an iterative process, where I compared mortality rates in "whole" and "complete case" cohorts of births in selected hospitals, with official rates in England and Wales published by the ONS.[34,42,161] A sub-cohort of births with complete information on key risk factors at birth which matched most closely the distribution of mortality rates in England and Wales (as published by ONS) was chosen.

As a result, my final inclusion criteria were:

- More than 500 births per financial year
- More than 25% of records with complete birth weight and gestational age in a given financial year

- More than half of the deaths in the age range of 2 days-4 years recorded in a given hospital were linked to an ONS mortality record (as discussed in Section 3.3.3, information about a death could be only indicated in the HES admission record, with no link to ONS mortality data)
- If any of babies born in a given hospital (per financial year) died at the age of 28-364 days, at least one of these deaths had complete information about birth weight and gestational age
- If any of babies born in a given hospital (per financial year) died at the age of 7-27 days, at least one of these deaths had complete information about birth weight and gestational age
- If any of babies born in a given hospital (per financial year) died at the age of 2-6 days, at least one of these deaths had complete information about birth weight and gestational age

I validated the distribution of births and infant mortality rates (per 1000 births) by categories of maternal age, birth weight and gestational age from the "complete case" cohort of selected hospitals against published information for England and Wales from the ONS using the same steps as described in Section 3.4.3.4.[34,42,161]

3.5 Results

I had access to linked HES-ONS data for episodes of care from the 1st April 1997 to the 31st March 2014, with ONS mortality data available from the 1st January 1998 to the 31st March 2014.

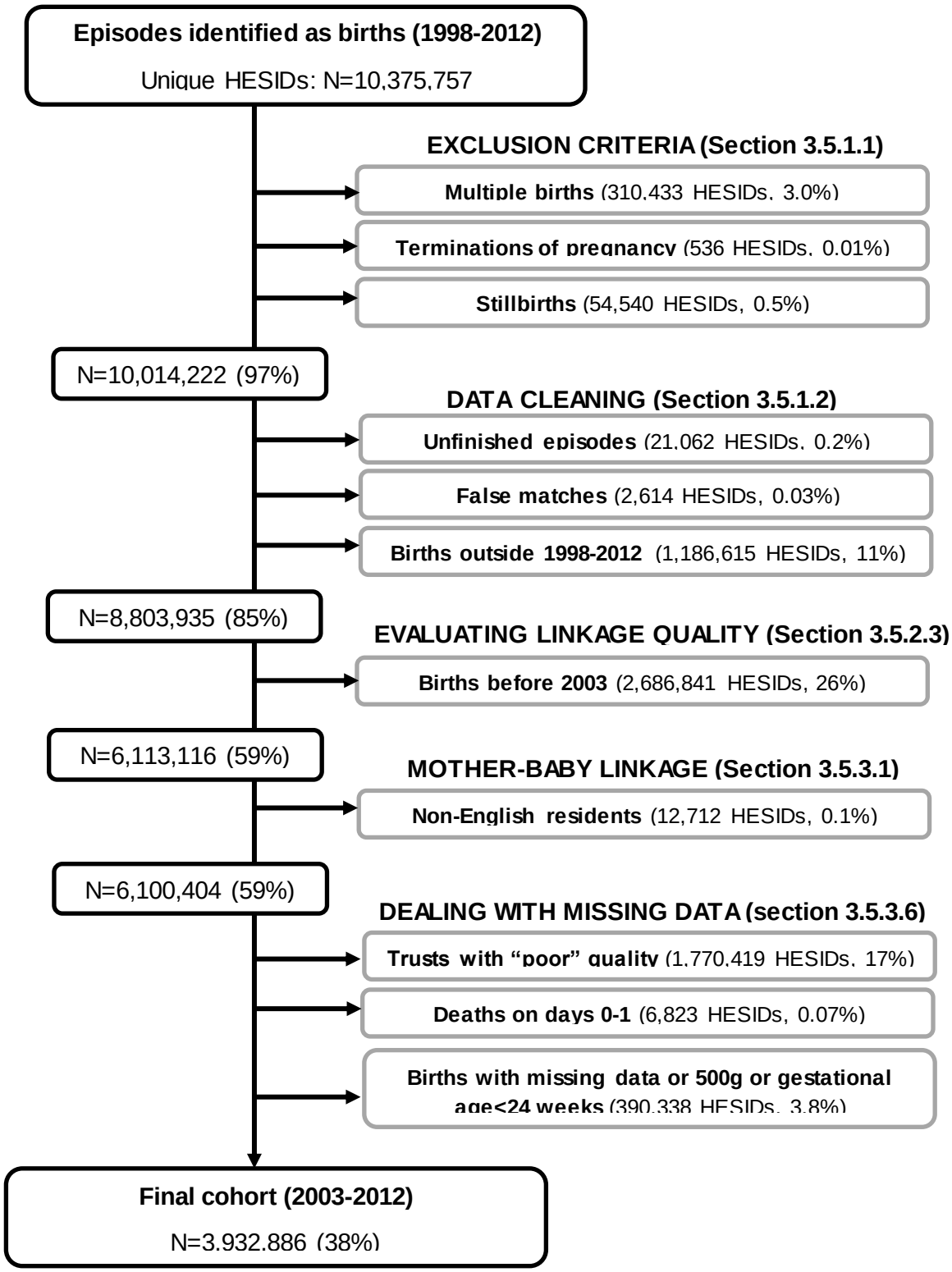
3.5.1 Identifying births

3.5.1.1 Inclusion and exclusion criteria

I identified 11,523,422 birth episodes for 10,375,757 unique HESIDs. I excluded 310,433 multiple births (3.0% of HESIDs) and 536 terminations of pregnancy (0.01% of HESIDs). A further 55,586 HESIDs were marked as stillbirths. However, 8.6% of these records were linked to a death record indicating either a false match between HES and ONS mortality data, or a miscoding of the birth episode as stillbirth in HES (e.g., an error in the discharge method field indicating a stillbirth). I assumed that high quality links between HES and ONS mortality data were live births miscoded as stillbirths, and I kept them in the cohort (1,046 HESIDs). High quality links had to have an exact agreement of NHS number, sex and date of birth (indicated by a match rank of 1 or 2, see Box 3.2). Thus, I excluded 54,540 stillbirths (0.5% of HESIDs). The numbers of excluded multiple births and stillbirths were consistent with numbers reported for England and Wales by ONS (3.0% and 0.5%).[44,162] All exclusions made are illustrated

in Figure 3.3. Overall, I identified 11,058,361 birth episodes for 10,014,226 singleton live births in the HES-ONS birth cohort between the 1st April 1997 and the 31st March 2014.

Figure 3.3 - Flow diagram showing exclusions made to develop a representative birth cohort using HES-ONS data



HES=Hospital Episode Statistics; ONS=Office for National Statistics. Data are number and % of births (unique HESIDs) removed at each stage. The final cohort covered 65% of all births in 2003-2012 and 38% of all births in 1998-2012 initially included in the cohort.

3.5.1.2 Data cleaning

I removed 21,062 HESIDs for which only unfinished HES episodes were available. These episodes were likely to be missed matches where a new HESID was generated for equivalent finished episodes recorded in the following financial year. I excluded 2,614 HESIDs identified as false matches. I then dropped 62,592 duplicated birth episodes and a further 952,714 consecutive episodes of care after birth (leaving one birth episode per baby). Finally, I excluded 1,186,615 HESIDs for babies born outside the period of available linked ONS mortality data and the Swedish birth cohort (1st January 1998 to 31st December 2012).

3.5.1.3 Cohort coverage

Following these exclusions, I identified 8,803,935 births in 1998-2012. Assuming that the ratio of singleton live births to all live births was the same in England as in England and Wales (97.0% in 1998-2012; data for England only was not available from ONS publications), the HES-ONS birth cohort covered 96.4% of all singleton live births in England in 1998-2012.[44,162]

Key finding from results Section 3.5.1:

- HES-ONS birth cohort covered 96.4% of singleton live births in England between 1998-2012

3.5.2 Longitudinal follow-up data until the fifth birthday

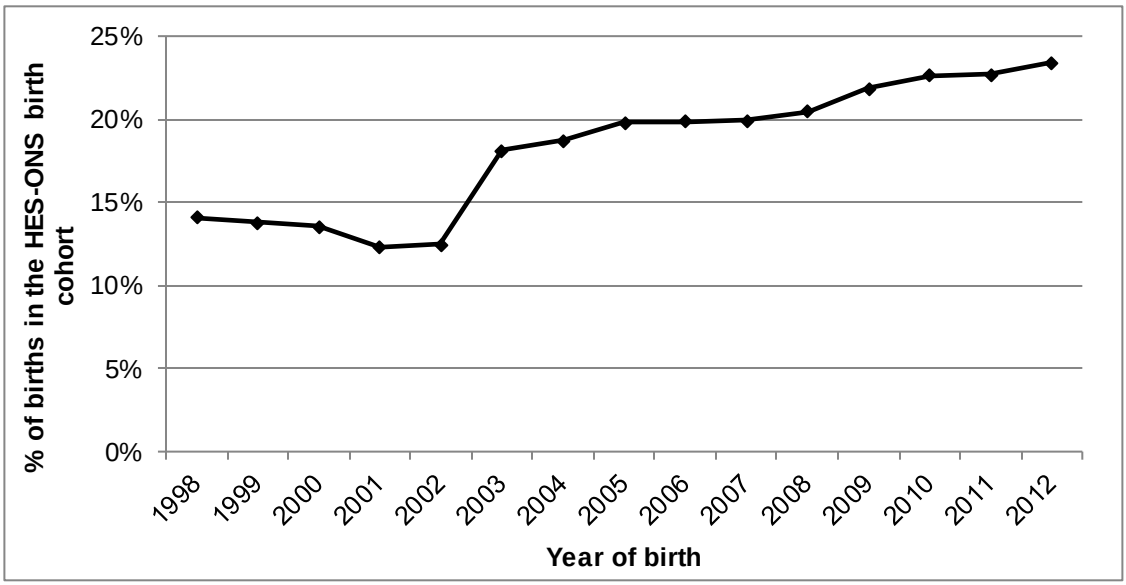
3.5.2.1 Hospital admission trajectories

Initially, there were 16,435,242 episodes of care (including birth episodes) with age at admission <5 years old identified for 8,803,935 singleton live births in the HES-ONS birth cohort. I removed 246,561 episodes during data cleaning as either duplicates or records with no recorded clinical information. The remaining 16,188,681 episodes were linked into 15,024,811 hospital admissions for 8,803,935 children born in 1998-2012, of which 6,220,870 were admissions for children after birth but before age of five years old.

In 1998-2002, 12.3-14.1% of babies had at least one hospital admission after birth in the first year of life (Figure 3.4). After 2002 when the NN4B service was implemented, the proportion rose to 18.1% in 2003, and increased annually, reaching 23.4% in 2012.

This shift suggests that there were more missed links between birth admissions and consecutive hospital admissions after birth due to increased chances of missing NHS number at birth prior to introduction of the NN4B system.

Figure 3.4 – Percentage of children in HES-ONS birth cohort with at least one hospital admission after birth in the first year of life by the year of birth



HES=Hospital Episode Statistics; ONS=Office for National Statistics.

3.5.2.2 Linkage to ONS mortality data

3.5.2.2.1 Checking the linkage between HES and ONS mortality data

Initially 43,491 deaths were linked to births in 1998-2012. I identified additional 371 in-hospital deaths recorded only in HES, where the discharge method in the hospital admission record indicated a death but there was no link to the ONS mortality data. All of these 'HES only' deaths were for births in 1998, when the linkage between HES and ONS mortality data was first introduced.

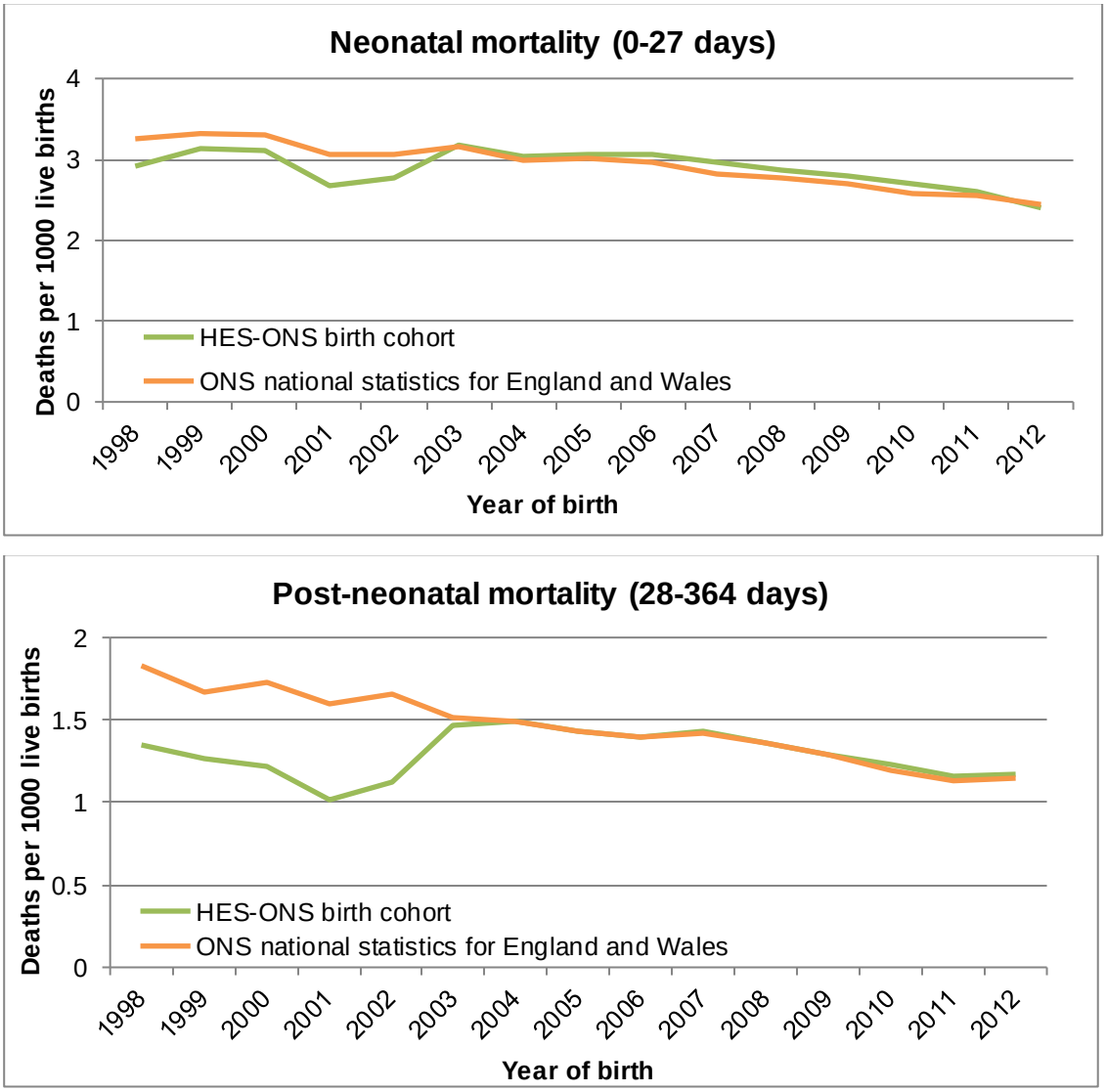
I excluded the following death records (i.e. I removed the link between death records and HES-ONS birth cohort and thus these birth records remained in the cohort):

- 1,997 deaths which occurred after child's fifth birthday (4.5% of all deaths),
- 80 deaths which occurred after the 31st December 2012 (0.2%)
- 4 deaths with date of death before birth date (0.01%)
- 149 deaths with subsequent hospital admissions after death (0.3%, see Section 3.4.2.2)
- 36 deaths for which the difference in the date of death according to HES and ONS mortality data was >1 day (0.1%)

After these exclusions, there were 41,616 child deaths in the HES-ONS birth cohort.

Similar to trends in the proportion of infants with at least one hospital admission after birth, infant mortality rates based on the HES-ONS birth cohort showed different patterns before and after implementation of NN4B system in October 2002. The rates were underestimated relative to rates reported for England and Wales for births before 2003, and closely matched rates reported by the ONS for births in 2003-2012 (Figure 3.5).

Figure 3.5 – Comparison of infant mortality rates per 1000 births based on HES-ONS birth cohort and singleton live births registered in England and Wales by age at death and year of birth



HES=Hospital Episode Statistics; ONS=Office for National Statistics. Data from ONS publications for England and Wales based on birth and death registration data.[34,42,161]

The differences in mortality rates calculated from the HES-ONS birth cohort and those reported by ONS for England and Wales before 2003 were larger for post-neonatal mortality compared to neonatal mortality (Figure 3.5). Children who die in the neonatal period are more likely to be cared for in a neonatal intensive care unit and therefore die in hospital; figure 3.5 suggests that these babies with longer post-natal hospital stays were more likely to have their NHS number added to their records during the birth admission, once the NHS number was allocated at birth registration, enabling linkage to an ONS mortality record. However, the vast majority of babies born before 2003 would be discharged shortly after birth and not get their NHS number updated in hospital records. If a birth episode did not contain an NHS number, it could only be linked to consecutive admissions in HES and to ONS mortality records using postcode, date of birth and sex (see Boxes 3.1 and 3.2) and no link would be established if, for example, the child changed their address. Thus, I considered linkage to ONS mortality data and to longitudinal hospital admission records in HES to be unreliable before 2003 due to a high risk of missed matches. Hence, I excluded 2,683,451 births prior to 2003 from the HES-ONS birth cohort from further analyses.

3.5.2.2.2 Recording of causes of death

In 2003-2012, 8.2% of all deaths were recorded in HES but did not link to an ONS mortality record (2,197 deaths). Missed links with ONS were the most common for early deaths – 66.5% of these "HES only" deaths were on days 0-1 of life (1,460 deaths), 26.9% at 2-27 days (591 deaths), 5.7% in the post-neonatal period (126 deaths), and only 0.01% of deaths were beyond infancy (25 deaths).

260 of the 39,419 deaths which did link to an ONS mortality record did not have any recorded causes of deaths, of which 98.1% were at age 28-30 days (255 deaths). They accounted for 78.5% of all deaths on days 28-30 days (325 deaths). It is likely that these deaths were certified using neonatal death certificates, which should be used for deaths in the neonatal period, and causes of deaths were removed by NHS Digital when processing the data. Data cleaning rules applied by NHS Digital to the ONS mortality data are not documented to confirm this. Thus, I restricted comparisons of cause-specific mortality between England and Sweden (presented in Chapter 6) to deaths beyond 30 days of life to ensure comparability.

Key findings for Section 3.5.2:

- Birth episodes recorded in HES prior to introduction of NN4B in 2002 were more likely to be missing the NHS number, leading to an increased risk of missed links to longitudinal hospital admission records and ONS mortality records. Thus, analyses have to be limited to 6,113,116 births in 2003-2012.
- Mortality rates in infancy based on the HES-ONS birth cohort were representative for England and Wales for births in 2003-2012.
- Comparisons of cause-specific mortality with Sweden should be restricted to deaths beyond 30 days of life due to the use of neonatal death certificate in England and high rates of missing data on causes of death at 28-30 days.

3.5.3 Recording of key risk factors for child mortality

3.5.3.1 Improving the completeness of risk factor variables using mother-baby linkage in HES

96% of births in the HES-ONS birth cohort linked to a delivery record (Table 3.1). Linkage results were comparable with those of Harron *et al.*[154]

Table 3.1 –Percentage of births in HES-ONS birth cohort which were linked to a maternal delivery record by year of birth and linkage method, compared to results reported by Harron *et al.*

	Deterministically linked		Probabilistically linked		Overall linkage rate	
Year	Harron *et al.*	My cohort	Harron *et al.*	My cohort	Harron *et al.*	My cohort
2003	24%	23%	73%	70%	97%	94%
2004	24%	23%	68%	72%	92%	95%
2005	25%	25%	70%	70%	95%	95%
2006	35%	35%	59%	61%	94%	96%
2007	24%	23%	71%	73%	95%	95%
2008	32%	31%	65%	65%	97%	96%
2009	36%	36%	61%	60%	98%	96%
2010	41%	41%	57%	55%	98%	96%
2011	43%	43%	55%	54%	99%	96%
2012	42%	41%	57%	55%	99%	96%
Total	**33.1%**	**32.5%**	**63.4%**	**63.0%**	**96.5%**	**95.5%**

HES=Hospital Episode Statistics; ONS=Office for National Statistics. Linkage rates for Harron et al.[154] were obtained via personal communication with Dr Harron.[169]

Mother-baby linkage led to substantial improvements in the completeness of risk factors at birth in the HES-ONS birth cohort. After linkage to maternal delivery records the proportion of births with recorded information increased from:

- 67% to 84% for birth weight
- 64% to 78% for gestational age
- 63% to 97% for maternal age
- 45% to 97% for IMD score

Importantly, the coverage of the complete case cohort has increased from only 18% (driven by high rates of missing IMD scores at birth) to 75% of all births in HES-ONS birth cohort. Using enhanced information on postcode, I excluded 12,712 births to non-English residents to match inclusion criteria to SMBR (Figure 3.3). As a result, 6,100,404 births remained in the HES-ONS birth cohort.

3.5.3.2 Cohort validation

3.5.3.2.1 Missing data

Rates of missing data were highest for gestational age and birth weight (missing for 22% and 17% of births respectively, Table 3.2). The rates of missing data were higher in children who died than for live births, and decreased with increasing age at death. Nearly half of deaths on days 0-1 of life did not have recorded gestational age, and a third did not have recorded birth weight. Due to an extraction error by NHS Digital, these early deaths were also more likely to have missing postcode and consequently, no IMD score (unless they had more than one hospital episode at birth). High rates of missing data in the "baby tail" (as indicated by higher rates of missing gestational age and birth weight) meant that these births were less likely to link to a delivery record, as these variables were part of linkage algorithm (see Appendix C.3).

Table 3.2 – Percentage of births and deaths by age at death recorded in the HES-ONS birth cohort in 2003-2012 with missing recording of risk factors of interest

		Deaths by age at death				
Risk factor	**Live births**	**0-1 days**	**2-6 days**	**7-27 days**	**28-364 days**	**1-4 years**
Number of births and deaths	6,100,404	9,679	3,626	4,089	8,161	3,632
Gestational age	22%	48%	37%	32%	29%	28%
Birth weight	17%	35%	32%	27%	23%	21%
Maternal age	3.2%	18%	14%	13%	8.2%	5.6%
Sex	0.10%	1.7%	0.43%	0.13%	0.011%	0%
IMD score	2.4%	27%	6.1%	6.0%	2.9%	1.3%
Any missing data	25%	62%	42%	37%	33%	31%

HES=Hospital Episode Statistics; IMD=Index of Multiple Deprivation; ONS=Office for National Statistics.

Complete case cohort covered 4,545,247 out of 6,100,404 births recorded in the HES-ONS birth cohort in 2003-2012 (75%). Mortality rates in infancy based on the whole HES-ONS birth cohort were comparable with rates reported for England and Wales (2.6 infant deaths/1000 live births vs 2.5/1000 live births, Table 3.3).[42,161] However, infant mortality rates in the complete case cohort were underestimated (2.3/1000 live births). This was primarily driven by highly underestimated mortality at 0-1 days (0.80 deaths/1000 live births vs 1.6/1000 births). Beyond the 1st day of life, the rates were underestimated by 0.07-0.10 deaths/1000 live births.

Table 3.3 - Comparison of crude mortality rates per 1000 births in whole and complete case HES-ONS birth cohorts, in England and Wales according to ONS national statistics in 2003-2012

Age at death	HES-ONS: whole birth cohort	HES-ONS: complete case birth cohort	England and Wales (ONS)
Number of births	6,100,404	4,545,247	6,604,156
0-1 days	1.59	0.80	1.61
2-6 days	0.59	0.46	0.54
7-27 days	0.67	0.57	0.64
28-364 days	1.34	1.22	1.33
0-364 days	2.60	2.25	2.51

HES=Hospital Episode Statistics; ONS=Office for National Statistics. Rates for England and Wales were obtained from ONS mortality publications for 2003-2012.[42,161] Mortality rate on days 0-1 in England and Wales was estimated by assuming that approximately ¾ of deaths on days 0-6 occur in the first two days.[170]

3.5.3.2.2 Distribution of risk factors in live births

The distributions of birth weight, gestational age, and maternal age for live births recorded in the HES-ONS birth cohort closely matched the distributions reported for England and Wales (both overall and for the complete case cohort, Table 3.4). Births at <24 weeks' gestation or weighing <1000g at birth contributed to a low proportion of births, but were highly underreported (by 20% in the whole cohort and by 35% in the complete case cohort for a birth weight <1000g, and by 63% in the whole cohort and 88% in the complete case cohort for births at <24 weeks' gestation).

Table 3.4 – Distribution of birth weight, maternal age, gestational age in whole and complete case HES-ONS birth cohorts, and in England and Wales according to ONS national statistics in 2003-2012

	HES-ONS: whole cohort	HES-ONS: complete case cohort	England and Wales (ONS)
Birth weight (g)			
<1000	0.34%	0.27%	0.42%
1000-1499	0.49%	0.46%	0.51%
1500-1999	1.0%	1.0%	1.0%
2000-2499	3.8%	3.8%	3.8%
2500-2999	16%	16%	16%
3000-3499	37%	37%	37%
3500-3999	30%	30%	30%
≥4000	11%	12%	12%
Maternal age (years)			
<20	6.3%	6.2%	6.3%
20-24	19%	19%	19%
25-29	27%	27%	27%
30-34	28%	28%	28%
35-39	16%	16%	16%
≥40	3.6%	3.6%	3.6%
Gestational age (weeks, 2010-2012)			
<24	0.033%	0.0087%	0.085%
24-27	0.24%	0.22%	0.27%
28-31	0.56%	0.53%	0.60%
32-36	4.7%	4.7%	4.7%
37-41	90%	90%	90%
≥42	4.5%	4.5%	4.2%

HES=Hospital Episode Statistics; ONS=Office for National Statistics. All data are % of all singleton live births per risk factor category. Information for England and Wales was obtained from ONS mortality publications for 2003-2012.[42,161] For gestational age tabulations, I used data from 2010-12, as the gestational age categories in ONS publications were sufficiently detailed only for these years.[34] Column totals may not add up to 100% due to rounding.

3.5.3.2.3 Distribution of risk factors by age at death

Mortality rates for a birth weight <1000g in the complete case cohort were severely underestimated at 0-6 days (150/1000 births compared to 250/1000 births in England and Wales, Table 3.5), and overestimated for deaths at 7-27 and 28-364 days (60/1000 births compared to 48/1000 births and 63/1000 births compared to 52/1000 births, respectively). For other birth weight categories, the rates were comparable with those reported for England and Wales. The differences in birth weight-specific mortality based on the whole HES-ONS birth cohort compared to rates in England and Wales were smaller than for complete case HES-ONS birth cohort (Table 3.5).

For maternal age, mortality rates in the complete case HES-ONS birth cohort were underestimated at 0-6 days for all maternal age categories, reflecting underrepresentation of deaths on days 0-1 of life in the complete case cohort compared to England and Wales (as shown in Table 3.3) due to under-recording of risk factors of interest for these early deaths (as shown in Table 3.2). Mortality rates by each maternal age category based on the whole HES-ONS birth cohort were underestimated relative to rates in England and Wales, but the differences were smaller than for the complete case cohort (Table 3.5).

For gestational age, mortality rates were underestimated at <24 and 24-27 weeks in the neonatal period, and overestimated in the post neonatal period relative to rates reported for England and Wales (for both whole and complete case HES-ONS birth cohorts). Mortality rates were representative for births at ≥28 weeks (Table 3.5).

Table 3.5 – Mortality rates per 1000 births by age at death and categories of birth weight, maternal age and gestational age in whole and complete case HES-ONS birth cohorts, and in England and Wales according to ONS national statistics in 2003-2012

Risk factor category	HES-ONS: whole birth cohort	HES-ONS: complete case birth cohort	England and Wales (ONS)	HES-ONS: whole birth cohort	HES-ONS: complete case birth cohort	England and Wales (ONS)	HES-ONS: whole birth cohort	HES-ONS: complete case birth cohort	England and Wales (ONS)
Birth weight (g)	**Early neonatal deaths (0-6 days)**			**Late neonatal deaths (7-27 days)**			**Post neonatal deaths (28-364 days)**		
<1000	230	150	250	53	60	48	56	63	52
1000-1499	30	30	32	11	11	12	14	14	16
1500-1999	12	11	12	3.8	3.9	4.2	8.2	8.2	9.1
2000-2499	3.7	3.4	3.7	1.6	1.7	1.6	3.9	3.9	4.1
2500-2999	1.1	1.0	1.2	0.58	0.57	0.57	1.6	1.6	1.6
3000-3499	0.51	0.48	0.52	0.26	0.25	0.26	0.8	0.79	0.81
3500-3999	0.36	0.34	0.37	0.16	0.16	0.16	0.5	0.5	0.51
≥4000	0.53	0.42	0.48	0.15	0.15	0.17	0.45	0.44	0.45
Maternal age (years)	**Early neonatal deaths (0-6 days)**			**Late neonatal deaths (7-27 days)**			**Post neonatal deaths (28-364 days)**		
<20	2.4	1.7	2.8	0.97	0.86	1.0	2.3	2.3	2.4
20-24	2.0	1.3	2.3	0.7	0.65	0.74	1.6	1.5	1.7
25-29	1.8	1.2	2.1	0.6	0.56	0.64	1.2	1.1	1.3
30-34	1.6	1.1	1.9	0.47	0.45	0.51	1.0	0.95	1.0
35-39	1.9	1.3	2.1	0.55	0.53	0.58	1.0	0.99	1.1
≥40	2.4	1.6	2.7	0.78	0.79	0.84	1.5	1.4	1.6
Gestational age (weeks, 2010-2012)				**Neonatal deaths (0-27 days)**			**Post neonatal deaths (28-364 days)**		
<24				800	510	860	50	93	27
24-27				190	190	180	56	56	57
28-31				40	39	38	12	12	12
32-36				6.4	6.0	6.4	3.8	3.8	3.7
37-41				0.77	0.73	0.74	0.76	0.77	0.74
≥42				0.78	0.71	0.85	0.58	0.57	0.53

HES=Hospital Episode Statistics; ONS=Office for National Statistics. Information for England and Wales was obtained from ONS mortality publications for 2003-2012.[42,161] For gestational age tabulations, I used data from 2010-12, as the gestational age categories in ONS publications were sufficiently detailed only for these years.[34]

Key findings from Section 3.5.3:

- The complete case HES-ONS birth cohort cannot be used for fair comparison of child mortality in England and in Sweden as mortality rates are underestimated compared to national figures reported for England and Wales by the ONS (especially on days 0-1 of life).
- Underestimated infant mortality rates in the complete case HES-ONS birth cohort were primarily driven by underreporting of gestational age and birth weight among the most vulnerable babies: born at <24 weeks, weighing <1000g at birth, or those who died shortly after birth.
- There was no "pattern in missingness" of maternal age; that is, infant mortality rates were underestimated in for all maternal age categories.

3.5.4 Strategies for dealing with missing data

3.5.4.1 Multiple imputation using chained equations

Following MI, the differences in proportions of live births by birth weight and gestational age categories relative to England and Wales remained, but were smaller than for the complete case cohort (Table 3.6). For example, a higher proportion of births in the imputed datasets had a birth weight <1000g (0.31% compared to 0.27% in the complete case cohort) and were born at <24 weeks' gestation (0.027% compared to 0.0087%). However, these proportions were still underreported by 26% and 63% relative to national statistics published by the ONS for England and Wales. The distribution of maternal age was representative for the population of children in England and Wales. This was expected since only 3.2% of records had missing maternal age (after linkage to the mothers' delivery records).

Table 3.6 – Distribution of birth weight, gestational age and maternal age among births in the complete case HES-ONS birth cohort, following MI, and in England and Wales according to ONS national statistics in 2003-2012

Risk factor category	HES-ONS: complete case birth cohort	HES-ONS: pooled results following MI	England and Wales (ONS)
Birth weight (g)			
<1000	0.27%	0.31%	0.42%
1000-1499	0.46%	0.51%	0.51%
1500-1999	1%	1.1%	1%
2000-2499	3.8%	3.8%	3.8%
2500-2999	16%	16%	16%
3000-3499	37%	37%	37%
3500-3999	30%	30%	30%
≥4000	12%	11%	12%
Maternal age (years)			
<20	6.2%	6.3%	6.3%
20-24	19%	19%	19%
25-29	27%	27%	27%
30-34	28%	28%	28%
35-39	16%	16%	16%
≥40	3.6%	3.6%	3.6%
Gestational age (weeks, 2010-2012)			
<24	0.0087%	0.027%	0.085%
24-27	0.22%	0.21%	0.27%
28-31	0.53%	0.58%	0.60%
32-36	4.7%	5.2%	4.7%
37-41	90%	90%	90%
≥42	4.5%	4.3%	4.2%

HES=Hospital Episode Statistics; MI=multiple imputation; ONS=Office for National Statistics. All data show % of all live births. The % were calculated separately for each imputed dataset and pooled together using Rubin's rules (that is, by taking an average).[163,164] Information for England and Wales was obtained from ONS mortality publications for 2003-2012.[42,161] For gestational age tabulations, I used data from 2010-12, as the gestational age categories in ONS publications were sufficiently detailed only for these years.[34] Column totals may not add up to 100% due to rounding.

Birth weight-specific mortality rates in the imputed datasets were not representative for England and Wales. The rates were overestimated for birth weight categories of: 1000-3499g for deaths at 0-6 days, <3000g at 7-27 days and <1499g at 28-364 days. Similarly, gestation-specific mortality rates did not match published rates for England and Wales (Table 3.7). Mortality rates by maternal age category in the imputed datasets, however, were representative of England and Wales compared to national statistics published by the ONS.

Table 3.7 – Mortality rates per 1000 births by age at death and birth weight, gestational age, and maternal age categories based on the complete case HES-ONS birth cohort, following MI, and in England and Wales according to ONS national statistics in 2003-2012

Risk factor category	HES-ONS: complete case birth cohort	HES-ONS: pooled results following MI	England and Wales (ONS)	HES-ONS: complete case birth cohort	HES-ONS: pooled results following MI	England and Wales (ONS)	HES-ONS: complete case birth cohort	HES-ONS: pooled results following MI	England and Wales (ONS)
Birth weight (g)	**Early neonatal deaths (0-6 days)**			**Late neonatal deaths (7-27 days)**			**Post neonatal deaths (28-364 days)**		
<1000	150	230	250	60	51	48	63	57	52
1000-1499	30	46	32	11	13	12	14	18	16
1500-1999	11	24	12	3.9	6.3	4.2	8.2	10	9.1
2000-2499	3.4	7.6	3.7	1.7	2.4	1.6	3.9	4.5	4.1
2500-2999	1.0	1.9	1.2	0.57	0.73	0.57	1.6	1.7	1.6
3000-3499	0.48	0.61	0.52	0.25	0.28	0.26	0.79	0.82	0.81
3500-3999	0.34	0.34	0.37	0.16	0.15	0.16	0.5	0.51	0.51
≥4000	0.42	0.46	0.48	0.15	0.13	0.17	0.44	0.42	0.45
Maternal age (years)	**Early neonatal deaths (0-6 days)**			**Late neonatal deaths (7-27 days)**			**Post neonatal deaths (28-364 days)**		
<20	1.7	2.8	2.8	0.86	1.0	1.0	2.3	2.3	2.4
20-24	1.3	2.3	2.3	0.65	0.75	0.74	1.5	1.7	1.7
25-29	1.2	2.2	2.1	0.56	0.68	0.64	1.1	1.3	1.3
30-34	1.1	1.9	1.9	0.45	0.54	0.51	0.95	1.1	1.0
35-39	1.3	2.1	2.1	0.53	0.6	0.58	0.99	1.1	1.1
≥40	1.6	2.7	2.7	0.79	0.84	0.84	1.4	1.5	1.6
Gestational age (weeks, 2010-2012)				**Neonatal deaths (0-27 days)**			**Post neonatal deaths (28-364 days)**		
<24				510	790	860	93	53	27
24-27				190	200	180	56	63	57
28-31				39	81	38	12	21	12
32-36				6.0	20	6.4	3.8	5.7	3.7
37-41				0.73	0.73	0.74	0.77	0.83	0.74
≥42				0.71	0.74	0.85	0.57	0.53	0.53

HES=Hospital Episode Statistics; MI=multiple imputation; ONS=Office for National Statistics. Mortality rates were calculated separately for each imputed dataset and pooled together using Rubin's rules[163,164] Information for England and Wales was obtained from ONS mortality publications for 2003-2012.[42,161] For gestational age tabulations, I used data from 2010-12, as the gestational age categories in ONS publications were sufficiently detailed only for these years.[34]

Since birth weight and gestational age were more likely to be missing for more vulnerable babies (with lower gestation and birth weight) and for babies who died shortly after birth, I concluded that the probability of missing data for birth weight and gestational age was associated with the health of a child at birth. I could not identify any additional variables needed to predict the missing values in birth weight and gestational age. Some of the variables that I would have liked to include in the imputation process, like ethnicity, level of provided neonatal intensive care, or resuscitation method were also missing via similar mechanisms (for 26.5%, 19.8% and 36.0% of births, respectively). Therefore, I concluded that birth weight and gestational age are likely to be Missing Not at Random (MNAR), which occurs when the missing values are related to the reason why the data are missing and this cannot be accurately captured by other variables in the data (e.g., birth weights and gestational ages missing in children with poor health at birth). Since multiple imputation on a big dataset like HES is computationally and time intensive (imputing five datasets took between 4 and 10 hours depending on the complexity of imputation models), I decided to try alternative methods for dealing with missing data.

3.5.4.2 Restricting the cohort to hospitals providing high quality data on risk factors

The derived cohort of selected hospitals reporting complete and high-quality data covered 4,329,985 of births, accounting for 71.0% of births in the whole HES-ONS birth cohort in 2003-2012. Complete case birth cohort based on these selected hospitals comprised 3,932,886 births, covering 64.5% of all births. The distributions of birth weight, gestational age and maternal age in the complete case birth cohort based on the selected hospitals matched the distributions reported by the ONS for England and Wales, but births at <24 weeks' gestation and weighing <1000g at birth remained underreported compared to population of England and Wales (Table 3.8).

Table 3.8 – Distribution of birth weight, maternal age and gestational age in the complete case birth cohort based on selected hospitals, and in England and Wales according to ONS national statistics in 2003-2012

	HES-ONS: complete case birth cohort based on selected hospitals	England and Wales (ONS)
Number of births	3,932,886	6,604,156
Birth weight (g)		
<1000	0.28%	0.42%
1000-1499	0.47%	0.51%
1500-1999	1.0%	1.0%
2000-2499	3.8%	3.8%
2500-2999	16%	16%
3000-3499	37%	37%
3500-3999	30%	30%
≥4000	11%	12%
Maternal age (years)		
<20	6.1%	6.3%
20-24	19%	19%
25-29	27%	27%
30-34	28%	29%
35-39	16%	16%
≥40	3.6%	3.6%
Gestational age (weeks, 2010-2012)		
<24	0.0088%	0.085%
24-27	0.23%	0.27%
28-31	0.54%	0.6%
32-36	4.7%	4.7%
37-41	90%	90%
≥42	4.5%	4.2%

HES=Hospital Episode Statistics; ONS=Office for National Statistics. All data are % of all singleton live births. Information for England and Wales was obtained from ONS mortality publications for 2003-2012.[42,161] For gestational age tabulations, I used data from 2010-12, as the gestational age categories in ONS publications were sufficiently detailed only for these years.[34] Column totals may not add up to 100% due to rounding.

Infant mortality rates in the complete case birth cohort based on the selected hospitals reporting complete and high-quality data matched the rates reported by the ONS for children in England and Wales, except for deaths on days 0-1 of life (Table 3.9). Therefore, I decided to exclude these early deaths from the comparison of child mortality presented in Chapter 5. Deaths on 2-6 days, 7-27 days, 28-364 days were underestimated by 0.3-0.5 deaths/1000 live births compared to rates for England and Wales. This was lower than for complete case HES-ONS birth cohort of all births, where the rates were underestimated by 0.7-0.11 deaths/1000 live births (Table 3.3). Similar underestimation between complete case cohort and the whole population was observed in the Swedish data (details are presented in the next chapter).

Table 3.9 –Comparison of crude mortality rates per 1000 births in whole and complete case birth cohorts in selected hospitals in HES, and in England and Wales according to ONS national statistics in 2003-2012

Age at death	HES-ONS: whole birth cohort based on selected hospitals	HES-ONS: complete case birth cohort based on selected hospitals	England and Wales (ONS)
Number of births	4,329,985	3,932,886	6,604,156
0-1 days	1.58	0.83	1.61
2-6 days	0.59	0.49	0.54
7-27 days	0.70	0.61	0.64
28-364 days	1.39	1.28	1.33
1-4 years	0.59	0.57	N/A
0-364 days	4.25	3.22	4.12

HES=Hospital Episode Statistics; N/A=not available; ONS=Office for National Statistics. All data are mortality rates per 1000 live births. Rates for England and Wales were obtained from ONS mortality publications for 2003-2012.[42,161] Mortality rate on days 0-1 in England and Wales was estimated by assuming that approximately ¾ of deaths on days 0-6 occur in the first two days.[170]

Birth weight-specific mortality rates were representative for England and Wales, apart from mortality rates for births weighing <1000g which were underestimated at 0-6 days and overestimated for deaths beyond the first week of life. Similarly, gestation-specific mortality rates were comparable with rates reported for England and Wales for all gestations apart from <24 weeks, which were underestimated at 0-27 days and overestimated at 28-364 days. The problem was not present for deaths in infants born at 24-27 weeks.

In 2005, only 64 out of 1,969 singleton live births at 24-27 weeks of gestation in England and Wales weighed <500g at birth (3.3% of births at 24-27 weeks) and 1,319 weighed 500-999g at birth (covering 70.0% of births); 5.2% of births had no information about birth weight.[171] Since the majority of births with gestational age of 24-27 weeks weighed 500-999g at birth, I assumed these two categories to be equivalent for singleton live births in England and Wales. Under this assumption, underestimated mortality rates for a birth weight <1000g were primarily driven by the underestimated mortality rates for a birth weight <500g (which is most common in births at <24 weeks of gestation), and mortality rates for a birth weight of 500-999g were representative for England and Wales (similarly to rates for a gestational age of 24-27 weeks).

Therefore, I limited the England-Sweden comparisons of child mortality in Chapters 5 and 6 to births with gestational age ≥24 weeks or birth weight ≥500g. The HES-ONS complete case selected hospital birth cohort excluding births at <24 weeks' gestation or with a birth weight <500g was representative for births in England in 2003-2012 in terms of the distribution of births and mortality rates by gestational age, maternal age, and birth weight.

Table 3.10 - Mortality rates per 1000 births by age at death and birth weight, gestational age, and maternal age categories based on the complete case HES-ONS birth cohort based on selected hospitals, and in England and Wales according to ONS national statistics in 2003-2012

Risk factor category	HES-ONS: complete case birth cohort based on selected hospitals	England and Wales (ONS)	HES-ONS: complete case birth cohort based on selected hospitals	England and Wales (ONS)	HES-ONS: complete case birth cohort based on selected hospitals	England and Wales (ONS)
Birth weight (g)	**Early neonatal deaths (0-6 days)**		**Late neonatal deaths (7-27 days)**		**Post neonatal deaths (28-364 days)**	
<1000	150	250	62	48	66	52
1000-1499	30	32	11	12	15	16
1500-1999	12	12	4.0	4.2	8.7	9.1
2000-2499	3.5	3.7	1.8	1.6	4.1	4.1
2500-2999	1.1	1.2	0.62	0.57	1.6	1.6
3000-3499	0.5	0.52	0.28	0.26	0.83	0.81
3500-3999	0.35	0.37	0.16	0.16	0.52	0.51
≥4000	0.44	0.48	0.16	0.17	0.46	0.45
Maternal age (years)	**Early neonatal deaths (0-6 days)**		**Late neonatal deaths (7-27 days)**		**Post neonatal deaths (28-364 days)**	
<20	1.7	2.8	0.91	1.0	2.4	2.4
20-24	1.4	2.3	0.7	0.74	1.6	1.7
25-29	1.3	2.1	0.61	0.64	1.2	1.3
30-34	1.2	1.9	0.48	0.51	1.0	1.0
35-39	1.4	2.1	0.57	0.58	1.0	1.1
≥40	1.7	2.7	0.87	0.84	1.5	1.6
Gestational age (weeks, 2010-2012)			**Neonatal deaths (0-27 days)**		**Post neonatal deaths (28-364 days)**	
<24			490	860	100	27
24-27			190	180	58	57
28-31			40	38	12	12
32-36			6.3	6.4	3.9	3.7
37-41			0.77	0.74	0.8	0.74
≥42			0.75	0.85	0.56	0.53

HES=Hospital Episode Statistics; ONS=Office for National Statistics. All data are rates per 1000 live births. Information for England and Wales was obtained from ONS mortality publications for 2003-2012.[42,161] *For gestational age tabulations, I used data from 2010-12, as the gestational age categories in ONS publications were sufficiently detailed only for these years.*[34]

Key findings from Section 3.5.4:

- The probability of birth weight and gestational age being missing is likely to be associated with the health of a child at birth. Babies with lower gestational ages and birth weights, and babies who died shortly after birth, were more likely to have missing data.
- HES does not contain sufficient additional variables to explain the missing data mechanisms and reliably impute the data.
- The sub-cohort of complete case births from hospitals with high quality and completeness of recorded data (covering 64.5% of all births) can be used to conduct a fair comparison of child mortality between countries if:
    - Deaths on days 0-1 of life are excluded
    - Births with a birth weight <500g or a gestational age <24 weeks are excluded

3.6 Discussion

3.6.1 Key findings

The HES-ONS data can be used to develop a nationally-representative birth cohort of singleton live births for births from 2003 onwards. The distribution of birth characteristics and socio-economic factors among live births is representative for the population of children in England and Wales. Thus, the HES birth cohort can be used as a denominator population for studies of child health outcomes. However, key risk factors at birth (birth weight, gestational age) are more likely to be missing in extremely low birth weight and extremely preterm babies, or infants who died shortly after birth, biasing any analyses of early life mortality. Linked HES-ONS data does not provide sufficient additional information about an infant's health at birth to reliably impute these variables using multiple imputation techniques. Instead, a cohort of selected hospitals with a high quality of recorded data can be used to study child deaths beyond days 0-1 of life, if births with a birth weight <500g or with a gestational age of <24 weeks are excluded.

3.6.2 Strengths

The complete case birth cohort based on selected hospitals with high quality of reported data in HES provides a rich and unique resource for studies of child mortality beyond days 0-1 of life in England. The long period of data collection and large sample size (3,932,886 births, covering 64.5% of all births recorded in HES) enable studying

child deaths according to risk factors with low prevalence in the population, such as extreme prematurity. Longitudinal follow-up via hospital admission and mortality records can be used to indicate congenital anomalies in all live born children (rather than only in children who died), which very few previous studies of child mortality in England accounted for. The cohort covered detailed information about birth characteristics (birth weight, gestational age, sex, and presence of congenital anomalies) and measures of socio-economic factors (maternal age and IMD score), and was representative for live births with birth weight ≥500g, gestational age ≥24 weeks, and for deaths at age >1 day in England. Therefore, the complete case birth cohort based on selected hospitals can be used to study child mortality adjusted for key risk factors at birth. Such adjusted comparisons can inform policy makers whether preventive strategies should focus on maternal health before and during pregnancy, or on improvements in the care received after birth.

The HES-ONS data can be used to develop a representative whole-country birth cohort for studies of child health outcomes which are not associated with mortality in the first week of life. The HES-ONS birth cohort covered 96.8% of all singleton live births in England in 2003-2012. Linkage between mothers' delivery and babies' birth records substantially increased the completeness of recording of birth weight, gestational age, maternal age, and IMD scores (from 18.1% of records with complete information to 75.4% among births between 2003 and 2012). The distribution of birth characteristics and socio-economic factors was representative for live births in England. Secondary use of this routinely collected dataset to examine health outcomes has the advantage of whole country coverage, minimising selection bias due to loss to follow-up.[121,122] Therefore, a whole-country birth cohort based on HES-ONS data provides an extremely cheap and time efficient alternative to birth cohorts involving de novo data collection (such as the Millennium Cohort Study). The HES-ONS birth cohort is currently being used within my research team to investigate socio-economic inequalities in waiting times for orchidopexy surgery,[172] to study risk factors for admissions for acute lower tract respiratory infections in infants[173] and for an international comparison of coding of congenital anomalies.[174]

Finally, linked HES-ONS data have the advantage of ongoing data collection, thus the HES-ONS birth cohort can be easily updated once more data become available (as of December 2017, I have updated the cohort to cover births until April 2017). I developed well annotated Stata do-files, which can be easily re-applied in de-identified HES-ONS data to generate cohorts for most recent years of data, or to replicate the cohort in other research centres. The Stata do-files will be made available in a public code repository such as GitHub (https://github.com/) for other researchers.

3.6.3 Limitations

While HES provides a unique resource for future studies, improvements in the quality of recorded data are needed. Firstly, I needed to exclude births prior to 2003, which were more likely to have incomplete follow-up since NHS numbers were not allocated at birth prior to the introduction of the NN4B programme in late 2002. This led to a higher rate of missed matches between birth admissions, and longitudinal hospital admissions and mortality records. Linking birth episodes to longitudinal hospital admissions prior to 2003 would provide a unique resource for birth cohort studies of health outcomes in adolescents with 15-20 years of follow-up after birth. For many of the HES birth records prior to 2003 both postcode and the NHS number were likely to be missing (exact proportion of births with missing NHS number cannot be derived from a de-identified HES extract). Improving the completeness of the postcode and NHS numbers on these records would require three steps. First, babies need to be linked to mothers to obtain information about postcode at birth, as postcode is 99% complete in mothers' delivery records in HES.[154] Second, using the date of birth, sex and complete postcode at delivery, birth episodes would need to be linked to the Personal Demographics Service (PDS), a national database of all patients who interact with the NHS (including all patients registered with a GP, babies who have received an NHS number at birth, as well as patients admitted to hospital via accident and emergency).[175] Thus, the PDS covers everyone who has an NHS number.[175] Finally, the NHS number obtained via linkage between HES birth records and PDS could be used to re-link HES birth episodes before 2003 to HES admissions after birth.

Missing data on birth characteristics (particularly birth weight and gestational age) in children who died is a further limitation of the HES-ONS birth cohort which I developed. The complete case cohort was not representative of all births in England and Wales, with higher rates of missing birth weight and gestational age in more vulnerable babies (with extremely low birth or gestational age, and those who died shortly after birth). It is possible that for children who are very unwell at birth there is less time for clinical staff to record birth weight and gestational age, whilst also working hard to prevent severe disability or death in the baby. Data for babies admitted to neonatal intensive care units are reported in a separate data collection stream (National Neonatal Research Database (NNRD) collected by The Neonatal Data Analysis Unit at Imperial College London),[176] and therefore might be less likely to be reported to HES. I concluded that birth weight and gestational age were likely to be missing not at random, violating the underlying assumption required for multiple imputation techniques. Therefore, I excluded 29% of births in hospitals with "poor" quality of recorded data on birth characteristics to carry out the analysis of child mortality presented in this thesis. I could also not investigate deaths on days 0-1 of life, which accounted for approximately

a quarter of child deaths in England between 2003 and 2012. A whole-country birth cohort with near 100% completeness of risk factors at birth and high quality of linkage to ONS mortality data could be developed by linking ONS birth registration, NHS birth notification data and HES records for mothers and babies.

Studies of cause-specific mortality based on the HES-ONS birth cohort need to be limited to deaths beyond 30 days of life, as 78.5% of deaths at 28-30 days which had an ONS mortality record did not have any recorded causes of death. These deaths were indicated in ONS mortality data, so it is likely that NHS Digital has removed recorded causes of death when processing the data (e.g., due to use of neonatal death certificate which should be used to certify deaths at <28 days of life). Further collaboration with NHS Digital is needed to determine why these data were missing.

Finally, in this thesis I focussed on singleton live births. This was due to differences in the strength of association between key risk factors (such as maternal age and gestational age) and the risk of child death between singleton and multiple births.[54] The number of multiple births identified in the cohort matched the numbers reported by ONS for England and Wales. However, episodes of care for multiple births have an increased missed match rate, especially for same sex siblings.[148] Further work is needed to evaluate the quality of recorded information on risk factors and linkage to longitudinal hospital admission and mortality records for multiple births in the HES-ONS cohort.

3.6.4 Implications for further research

Linked administrative datasets in England (introduced in Section 3.2.2) provide information on important risk factors at birth, socio-economic factors and longitudinal follow-up via hospital admission and mortality records. However, linkage between databases is needed to fully benefit from the collected data. ONS birth registration data provides information on accurate date of birth, high completeness of birth weight and maternal age.[128] Information about parental occupation is also available for 10% of births.[128] NHS birth notification data covers information on gestational age and ethnicity, complementing information collected at birth registration.[177] These two datasets are routinely linked to ONS mortality data with information about causes and timing of deaths to produce annual child mortality statistics.[34,76,177] Linkage to the baby's and mother's longitudinal hospital admission trajectories would enable development of additional risk factors such as presence of congenital anomalies in babies or chronic conditions during pregnancy in mothers or after birth in children. The feasibility of linking ONS birth registration, NHS birth notification data and HES records for mothers and babies has previously been demonstrated.[167,178] However, linkage for

births in 2005-2014 has only been achieved in 2016 (a year into my PhD), and is not updated on a regular basis.

3.6.5 Implications for this thesis

I have shown that a complete case birth cohort based on selected hospitals in HES with high quality of recorded data can be used to study child mortality beyond days 0-1 via linkage to longitudinal hospital admission and mortality records. This sub-cohort is representative of births in England and Wales and can be used for comparison of child mortality with Sweden after excluding births with a birth weight <500g and a gestational age <24 weeks. In the next chapter I describe Swedish datasets used for this investigation and evaluate the comparability of the English and Swedish birth cohorts.

Chapter 4. Comparability of national birth cohorts in England and in Sweden

What is already known:

- Individual-level data with information on multiple risk factors at birth are needed to identify the origins of inter-country differences in child mortality.
- A national birth cohort containing birth characteristics and socio-economic factors can be developed using an administrative hospital database in England.
- However, studies of child mortality in England need to be limited to a sub-cohort of hospitals with a high quality of recorded data, excluding deaths on days 0-1 of life and births at <24 weeks' gestation, with birth weights <500g, or missing information on any of the risk factors.

What this chapter adds:

- This chapter introduces administrative data sources available in Sweden for this study.
- I develop a birth cohort of Swedish children, comparable with the English birth cohort presented in Chapter 3.
- I evaluate potential sources of bias which could arise when comparing English and Swedish birth cohorts due to differences in coding and data artefacts.

4.1 Chapter overview

In the previous chapter, I presented methods for developing a birth cohort using an administrative hospital database in England. Due to the high rates of missing data in the English birth cohort, a comparison of child mortality in England and Sweden needs to be limited to deaths beyond days 0 and 1 of life. For England, I showed that a sub-cohort of births with complete information on all key risk factors, from hospitals with a high quality of recorded data, covering 64.5% of singleton live births recorded in Hospital Episode Statistics (HES) need to be used for analyses of child mortality.

This chapter presents further work towards objective 2: "*to develop comparable national birth cohorts using administrative linked datasets in England and in Sweden with information on birth characteristics and socio-economic factors*". I describe Swedish datasets which were available for this project and I present methods for deriving a Swedish birth cohort. I also discuss similarities and differences in the collection and recording of hospital and mortality data in England and in Sweden, which could bias inter-country comparisons of child mortality. The cohorts described in this and the previous chapter are used for analyses in Chapters 5 and 6.

The Swedish National Registers used in this chapter were accessed at Centre for Health Equity Studies (CHESS), at Stockholms Universitet/Karolinska Institutet. I received ethics approval to use the registers from the Regional Committee of Stockholm (no. 2016/1234-31/5, approved on 04/08/2016, a copy attached in the Appendix D).

4.2 Background

4.2.1 Swedish National Registers

Sweden has a long tradition of collecting administrative data in national registers for research purposes.[122] A birth cohort can be easily defined using the Swedish Medical Birth Register (SMBR), a medical birth register which covers information about maternal health during pregnancy, delivery details and birth characteristics for all births to mothers resident in Sweden.[112] Hospital admission records for mothers and babies are collected in the Swedish Hospital Discharge Register (SHDR); information about the causes and timing of deaths is recorded in the Swedish Cause of Death Register (SCDR). Accurate linkage between these databases, with low rate of linkage error, is done using a Personal Identity Number (PIN), a unique identifier allocated to every Swedish resident and used in many areas of Swedish society such as healthcare, taxation or education.[179] PubMed search for "Swedish Medical Birth Register" revealed that since 2000, almost 250 papers have been published using the SMBR (as of September 2017), covering a range of health outcomes in mothers, infants, children, and younger adults (via linkage to other registers). National registers in Sweden are vital resources for improving the health of Swedish population, and efforts are made by the data providers to maintain high coverage and completeness of the recorded data in the registers.[112,180,181]

4.2.2 International comparisons – ensuring comparability of the Englishz and Swedish cohorts

International comparisons of child health outcomes enable policy makers to identify the reductions in adverse health outcomes which should be achievable relative to a country

or countries with better health outcomes. However, for a fair comparison it is important to evaluate the comparability of the datasets used, to ensure that observed differences are due to tangible factors rather than data artefacts. For example, international comparisons of infant mortality are, in particular, prone to bias due to differences in registration practices for stillbirths and live births.[25,26,110] Cut-offs for inclusion to the birth cohort, based on gestational age or birth weight, should be used to ensure that data from compared countries capture infants with common definition of "viability".[25,26,110] Between-country differences in methods used for the calculation of gestational age could also bias a comparison – estimates of gestational age based on last menstrual period (LMP) lead to lower rates of preterm birth in the population than for ultrasound-based gestational age.[110] Finally, differences in national coding practices and in thresholds for hospital admissions could bias estimated prevalence of conditions identified using hospitalisation records, such as congenital anomalies. This chapter explores the comparability of birth cohorts in England and in Sweden.

4.2.3 Chapter aims

The aim of this chapter is to describe the development of a Swedish birth cohort using Swedish national registers, and to evaluate the comparability of information recorded in the English and Swedish birth cohorts. The results from this chapter provide information about potential biases arising in the analyses presented in Chapter 5 and Chapter 6.

4.3 Datasets used to develop the Swedish birth cohort

4.3.1 Swedish Medical Birth Register (SMBR)

4.3.1.1 Overview

The SMBR is a national register covering all hospital births (live or still) to mothers resident in Sweden.[112] It was established by the Swedish parliament in 1973 to enable research into the health and the quality of care of Swedish infants.[112] SMBR contains information from antenatal care clinics, delivery units and routine newborn examinations in hospital after birth.[182] It is collated and maintained by the National Board of Health and Welfare (NBHW).[182]

It is mandatory for all healthcare providers to submit information to SMBR.[112] Each year the number of births reported to SMBR is compared with the number of births registered in Sweden according to the Total Population Register (TPR).[112] TPR contains information about life events (births, deaths, marriages) and is collected by the government agency, Statistics Sweden, for publication of vital statistics (more details are described in Box 4.1).[112] If missing records are discovered in the SMBR, the NBHW

contacts the reporting hospitals to obtain the missing information.[112] Therefore, SMBR has a near whole population coverage – between 1973 and 1998, only 1.4% of births were missing from SMBR.[112]

Box 4.1 – Total Population Register and registration of vital events in Sweden

Total Population Register (TPR) is a dataset containing information on life events such as births, deaths, civil status, as well as family structure and migration status.[226] It is maintained by a government agency, Statistics Sweden, for publication of national statistics for Sweden.[226]

The data in TPR comes from the tax authorities, the Swedish Tax Agency, who are responsible for civil registration in Sweden.[226] Births, deaths, change of marital status, migration or change of address within Sweden have to be notified to the Swedish Tax Agency, who collates it in the Population Register (PR-tax) and also sends daily updates to Statistics Sweden for TPR.[226]

4.3.1.2 Information on risk factors of interest recorded in SMBR

The information on mothers and babies recorded in SMBR is more detailed than in HES. The recorded information includes:

- Mother's personal details (e.g., nationality, country of birth, year of immigration to Sweden, occupation, place of residence, marital status)
- Indicators of maternal health before and during pregnancy (e.g., weight before pregnancy and at delivery, the number of previous pregnancies and their outcomes, smoking/drug use before and during pregnancy, use of contraception)
- Details of the delivery (e.g., maternal diagnoses, operations at delivery, method of delivery, pregnancy duration, dates of admission and discharge for delivery)
- Birth details and indicators of the baby's health (e.g., birth date, live or stillbirth, sex, birth weight and length, head circumference, APGAR score, diagnoses and operations) [112].
- The PINs of the mother, the father and the child, enabling the linkage of parents and siblings into families in the SMBR.

Similar to birth episodes in HES, SMBR contains diagnoses recorded during the routine newborn examination. Since 1999, up to 12 diagnoses can be recorded, using a

Swedish adaptation of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) codes (in use since 1997).[112] The Swedish ICD-10 is based on the ICD-10 and includes several more in-depth codes (with an extra fifth letter).[183] Infants who are transferred to a neonatal ward are likely to be missing diagnoses at birth. However, additional information from the neonatal ward can be obtained through linkage of SMBR with SHDR, described in the next section.[112]

4.3.2 Swedish Hospital Discharge Register (SHDR)

4.3.2.1 Overview

The SHDR is the Swedish equivalent of HES data in England, covering more than 99% of all hospital discharges in Sweden from both privately and publicly funded physicians.[181] It was established in 1964, and has covered the whole country since 1987.[181] Information for SHDR is collected by the NBHW on a monthly basis.[184] Originally the register was used only to monitor the use of healthcare.[181] Since the beginning of the 1990s, it has also been used for the financing and management of hospitals.[185] This is based on Diagnosis-Related Group (DRG) codes which, like the Healthcare Resource Group (HRG) codes in HES, group together diagnoses and operations requiring comparable levels of hospital resources.[185]

4.3.2.2 Structure

Unlike in HES, a basic unit in SHDR is a hospital admission. Transfers between hospitals are kept as separate records.

4.3.2.3 Clinical information

The information collected in SHDR includes patient-related data (e.g., PIN, sex, age, county of residence), information about the caregiver (e.g., the type of hospital, hospital number), data about the admission (e.g., admission and discharge dates, mode of admission and discharge destination) and clinical information (e.g., diagnoses and procedures).[181,184]

Diagnoses have been coded using the Swedish ICD-10 since 1997, and each admission can have up to eight recorded diagnoses.[181] The coding is done in hospital by the physician responsible for discharging the patient.[181] Since 1997, procedures have been recorded using Swedish version of the Nordic Medico-Statistical Committee (NOMESCO) Classification of Surgical Procedures (NCSP), used in national registers in Nordic countries.[181] Up to 12 surgical procedures can be recorded per admission.[181]

Similar to HES, coding depth (in particular, the number of secondary diagnoses) has increased over time due to financial incentives with the introduction of DRGs in the

1990s.[181] For example, some hospitals have introduced compulsory coding of certain comorbidities known to generate additional funding (such as diabetes) as secondary diagnoses (if applicable).[181] Hospitals often have a designated physician who double-checks that no diagnosis was omitted before finalising and submitting the data to NBHW.[181] As in the case when using HES, trends in admission rates for particular conditions (especially those coded as secondary diagnoses) need to be interpreted with caution.

4.3.3 Swedish Cause of Death Register (SCDR)

4.3.3.1 Overview

The current electronic SCDR was established in 1961 and covers deaths going back to January 1952.[186] SCDR includes the deaths of all residents of Sweden, including deaths outside the country (in 2015, 816 deaths abroad were reported to SCDR).[186] Since 2012, non-Swedish residents who died in Sweden have also been included in SCDR (approximately 200-300 deaths per year).[186]

4.3.3.2 Data collection process

Unlike in England, there are no long delays in death registration. All deaths need to be notified to the tax authorities (Swedish Tax Agency, see Box 4.1) within one business day, and the death certificate, with causes of death, must be reported to the NBHW within 3 weeks.[180,186] Like with SMBR, the NBHW verifies the number of deaths recorded in SCDR with the number reported in the TPR.[180] If discrepancies or missing data are detected, the NBHW contacts an appropriate medical institution for details.[180]

4.3.3.3 Death certification in Sweden

Unlike in England, all deaths (including neonatal deaths) are registered using a death certificate compatible with the international template recommended by the World Health Organization (WHO). It consists of two parts, with part I covering the underlying cause of death and a sequence of conditions which led directly to death; and part II detailing a set of contributing conditions which were present at death, but not part of the terminal sequence of conditions.[186]

4.3.3.4 Recorded information

SCDR contains information on all causes of death, the underlying cause of death, the date of death, and whether an autopsy was conducted.[186,187] Additional demographic data (such as age, sex, civil status, place of residence, nationality and country of birth) is fed to the SCDR from the TPR by the NBHW.[187] Causes of death are coded using ICD-10 (since 1997), and up to 48 contributing causes can be reported additionally to

the underlying cause.[186] Like in England, the underlying cause of death is identified using automated software compatible with ICD-10 guidelines.[186]

4.3.4 Longitudinal Integration Database for Health Insurance and Labour Market Studies (LISA)

The Longitudinal Integration Database for Health Insurance and Labour Market Studies (LISA) database combines existing information from national registers kept by Statistics Sweden, the Social Insurance Agency and the Swedish Agency for Innovative Systems.[188] Collected information includes an individual's income, details of employment history and education for all individuals aged 16 years and above who are registered as Swedish residents by 31st December every year since 1990.[188] In this PhD, linkage to LISA was used to obtain disposable income per family member to calculate quintile of socio-economic status (SES).

4.3.5 Following a patient in the Swedish registers – PIN

High-quality, deterministic linkage between the national registers in Sweden is possible using PIN, a unique identifier for all Swedish residents (i.e. nationals who get a PIN at birth, and immigrants who intend to stay in Sweden for longer than one year).[179] Since 1947, it has been used in all areas of Swedish society including healthcare, migration, taxation, education, civil registration, income and social security.[179] All PIN numbers are replaced by unique serial numbers before data are delivered to researchers.[179]

False matches or missed matches in PIN allocations are rare. Individuals with an incorrect PIN (e.g., where the day of birth or the month of birth are outside of the plausible range of values) are allocated a new corrected PIN upon discovery.[179] Approximately 1000 incorrect PINs were identified in 2004-2009.[179] As of January 2008, 15,887 people with a re-used PIN have been identified (approximately 0.16% of the population), primarily among residents born abroad in the 1950s and 60s.[179] These errors did not affect the results presented in this thesis because I investigated births after 2003. A small proportion of records in the SMBR (0.2% of singleton live births in 1998-2010[112]) and admissions in the SHDR (2.9% between 1964 and 2008[181]) do not have PIN, and therefore cannot be linked to other databases. These individuals are not included in the data extracts available for this study.

4.4 Methods for developing a Swedish birth cohort

I had access to the four Swedish national registers described in Sections 4.3.1-4.3.4 for births between 2003 and 2012 in SMBR, and follow-up in SHDR and SCDR until 2013. Developing a Swedish birth cohort took a month in total. I first visited CHESS (for a week) in 2015 to learn about the datasets, translate the variables to English and

develop Stata do-files for data cleaning. In 2016, I spent a further two weeks generating a birth cohort for analyses. In 2017, I returned to CHESS to finalise the cohort and run additional sensitivity analyses. The completeness and validity of recorded risk factors was high, although some underreporting of risk factors in deaths remained, especially for deaths in the first week of life.

4.4.1 Developing a Swedish birth cohort

4.4.1.1 Identifying births to match inclusion criteria to the English cohort

Inclusion to the Swedish birth cohort was based on birth records in SMBR. As in England, I excluded multiple births from the cohort. SMBR does not include terminations of pregnancy or births to non-Swedish residents and the extract available for this study did not include stillbirths. SMBR covers one record per individual, therefore de-duplication was not necessary.

4.4.1.2 Longitudinal follow-up

Longitudinal hospital admission data were extracted from SHDR. I removed admissions where the recorded admission date was before the date of birth (as recorded in SMBR). Mortality data were extracted from SCDR. I removed deaths for children aged over five years old. To match the inclusion criteria in the HES-ONS cohort, I also removed deaths which occurred abroad – follow-up for these children was censored on their date of death.

4.4.1.3 Deriving risk factors of interest

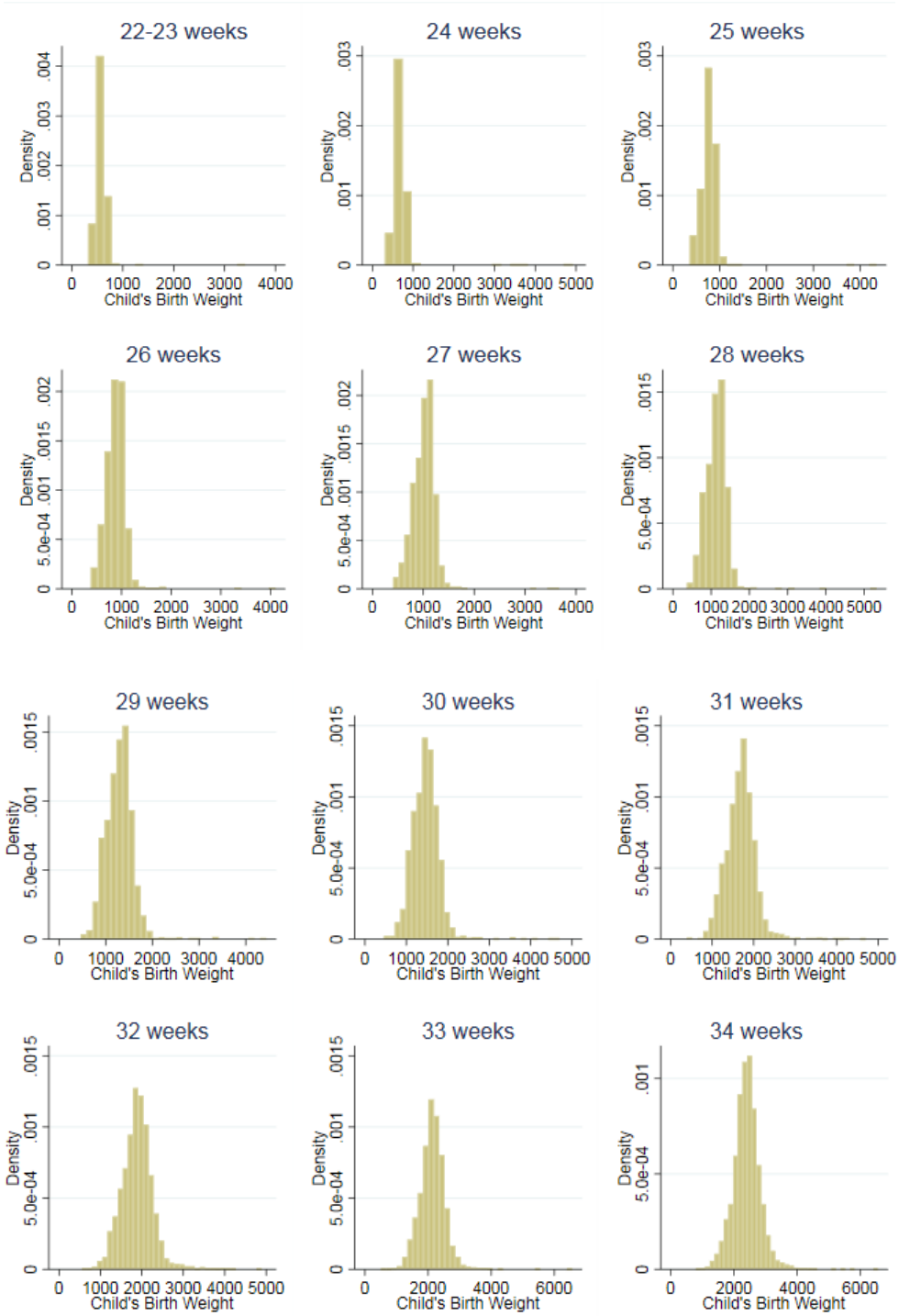
4.4.1.3.1 Birth characteristics

Sex, birth weight and gestational age were obtained from SMBR. Gestational age was calculated in days, therefore, I converted it into completed weeks (e.g., 36 weeks would cover 36 weeks +0 days to 36 weeks + 6 days) to match how the gestational age data were recorded in the English birth cohort.

Previous studies based on SMBR have reported a bimodal distribution of birth weight for gestational ages <30 weeks due to recording errors, for example, by mistyping 29-31 weeks instead of 39-41 weeks.[189] These errors results in a misclassification of term births as preterm (as illustrated by long tails of distribution of birth weight per week of gestation for gestational age less than 33 in Figure 4.1). Therefore, as in the English birth cohort, I removed implausible combinations of birth weights for each week of gestation. I used a set of criteria developed at CHESS based on birth weight, gestational age and birth length (presented in Appendix D). Figure 4.1 illustrates that implausible birth weight for gestational age is a much smaller problem than in the

English birth cohort based on HES-ONS (as shown in Figure 3.2 in Section 3.4.3.2) – I identified approximately 30 implausible combinations in the Swedish cohort.

Figure 4.1 – Distribution of birth weight (in grams) by week of gestation before removing implausible combinations of birth weight per week of gestation in the cohort of singleton live births from Swedish Medical Birth Register in 2003-2012.



Note that there are approximately 30 implausible combinations of birth weight for gestational age in the tails of birth weight distributions (on the right-hand side, as indicated by the scale of the x-axis), which cannot be seen due to a very small numbers of observations.

An indicator of the presence of one or more congenital anomalies was developed using diagnoses recorded at birth in SMBR, in hospital admissions up to the age of two recorded in SHDR, and using causes of death until the age of five recorded in SCDR. To identify a congenital anomaly, I used the same code lists as for the English data, described in Chapter 3 (Section 3.4.3.2).

4.4.1.3.2 Socio-economic indictors

Maternal age was obtained from the SMBR. A second measure of SES was obtained from LISA. I used disposable income per family member a year before birth to calculate quintiles of SES amongst all pregnant women in a given calendar year. Income per family member was defined as household income (after tax) divided by the number of family members, where adults were given a weight of 1 and children were given a weight of 0.7.

4.4.1.3.3 Missing data

I tabulated the rates of missing data by risk factor recorded at birth among births and deaths. I also compared mortality rates in the "complete case" cohort, defined as the cohort of births with a complete recording of all risk factors of interest in this study, with rates in the whole cohort (used as a gold standard since SMBR is thoroughly validated by NBHW against all births registered in TPR during data collection process).

4.4.2 Collating the Swedish and English birth cohorts

To enable analyses of combined birth cohorts from England and Sweden, I derived tables of births and deaths from the two birth cohorts by categories of birth weight, gestational age, sex, presence of congenital anomalies, maternal age, quintile of socio-economic status, and length of available follow-up.

4.4.3 Comparability of Swedish and English birth cohorts

4.4.3.1 Definitions of still- and live births

National rates of early life mortality are strongly influenced by registration practices for stillbirths, live births and early neonatal deaths.[25,26] Registration criteria for live births are the same in England and Sweden. In both countries all births showing signs of life, irrespective of gestational age or birth weight, need to be registered, in line with the WHO definition.[190] The registration criteria for stillbirths, however, differ: in England, all stillbirths at ≥24 weeks of gestation need to be registered;[23] in Sweden, the registration threshold was set at ≥28 completed weeks until June 2008, when it was changed to ≥22 weeks.[191] These differences were unlikely to affect the comparisons of child mortality in England and Sweden presented in Chapters 5 and 6, since deaths on days

0-1 and births with birth weight <500g or gestation <24 weeks were excluded from the analyses. Therefore, I did not examine the effect of change in the definition of stillbirths on registration practices further.

4.4.3.2 Diagnostic coding depth in the Swedish and English birth cohorts

Diagnoses and causes of death recorded in the HES-ONS (Office for National Statistics) birth cohort and in the Swedish national registers were used to develop an indicator of congenital anomalies for analyses in Chapters 5 and 6. These diagnostic codes are also used to define respiratory tract-infection (RTI)-related deaths and deaths from sudden unexpected deaths in infancy (SUDI) in Chapter 6. Therefore, it is important to compare coding depth in the two countries.

4.4.3.2.1 Recorded diagnoses at birth and during hospital admissions

The recording of diagnostic information differs in administrative hospital databases in England and Sweden. HES contain a higher number of diagnostic fields available per episode than SHDR (Table 4.1). Furthermore, one admission can consist of more than one episode in HES. Coding practices are likely to be more standardised between hospitals in England due to national accreditation training for the coders. However, coding in Sweden is done by the physician responsible for discharge so it could be more accurate, as they can interpret the discharge notes. In England, the coders who translate medical documentation into ICD-10 codes can only record information explicitly stated in the notes. However, there has been no formal comparison between the countries in terms of coding quality.

Table 4.1 – Summary of differences in recording of diagnostic information in Hospital Episode Statistics in England and in Medical Birth Register and Hospital Discharge Register in Sweden

	England	Sweden
Diagnostic information recorded at birth		
Source of information	**Hospital Episode Statistics**	**Medical Birth Register**
Number of diagnostic fields available	April 2002- March 2007: 14 fields Since April 2007: 20 fields	Since January 1999 – 12 fields
Coding	ICD-10 since March 1995	Swedish ICD-10: since January 1997
Longitudinal hospital admission data for children		
Source of information	**Hospital Episode Statistics**	**Hospital Discharge Register**
Number of diagnostic fields available	2002-2007: 14 fields Since April 2007: 20 fields	1997-2009 – 8 fields Since 2010 – unlimited (however, the NBHW will generally only provide the primary diagnosis plus first 7 additional diagnoses to researchers)[181]
Coding	ICD-10: since March 1995	Swedish ICD-10: since January 1997

ICD-10=International Statistical Classification of Diseases and Related Health Problems 10th Revision; NBHW=National Board of Health and Welfare

I compared the number of unique diagnoses (based on the first three letters of ICD-10 codes) recorded per birth in SMBR and per birth admission in HES (that is, during any episode of care contributing to the birth admission). I then compared the coding depth per hospital admission in SHDR and per hospital admission after birth in HES. To ensure comparability, an admission in HES was defined as all episodes of care with the same admission date and hospital code (i.e. hospital transfers were treated as separate admissions, like in SHDR). I counted each unique diagnosis only once per admission, even if the ICD-10 code appeared in multiple episodes of the same admission. Finally, I compared the proportion of babies with at least one hospital admission in infancy (with admission starting at age 0-364 days) and the mean number of hospital admissions per baby in first year of life as proxies for admission thresholds in the two countries.

4.4.3.2.2 Coding of causes of death

There are differences in registration practices for causes of death in England and in Sweden (Table 4.2). Unlike Sweden, England is one of the few countries in the world using the neonatal death certificate recommended by the World Health Organization (WHO) for deaths occurring before 28 days of life (details are described in Section 3.4.2.). For deaths after 28 days of life, both countries use a death certificate compatible with the international template recommended by the WHO, and the underlying cause of death is selected and coded using automated software based on ICD-10 rules (Table 4.2).[23,180]

Table 4.2 – Summary of differences in recording of causes of death in ONS Mortality Data in England and in Cause of Death Register in Sweden

	England ONS Mortality Data	Sweden The Cause of Death Register
Death certificate follows WHO recommended format with two parts	Yes	Yes
Separate perinatal death certificate	Yes	No
Automated underlying cause of death selection since	1993	1987
Coding systems used	ICD-10 since January 2001	ICD-10 since January 1997
Number of recorded causes of death	Underlying cause plus up to 15 additional causes	Underlying cause plus up to 48 additional causes

ICD-10=International Statistical Classification of Diseases and Related Health Problems 10th Revision; ONS=Office for National Statistics.

To compare the depth of coding at death certification in the two countries, I calculated the mean number of ICD-10 codes recorded per death in SCDR and in ONS mortality data. As shown in Section 3.4.2 of Chapter 3, 8.2% of all deaths in 2003-2012 in the English cohort were recorded only in HES (that is, there was no link to an ONS mortality record thus no recorded causes of death were available); 78.5% of deaths on days 28-30 had no recorded causes of death despite an existing link to ONS mortality record. Therefore, in the English birth cohort, I based the mean number of causes of death only on deaths which had at least one recorded cause of death.

4.4.4 Recording of risk factors of interest

4.4.4.1 Gestational age measurement

In both countries, gestational age was derived using one of three methods: an estimated date of delivery from ultrasound measurement, based on date of LMP or a

clinical assessment (in the absence of the other two measures).[112,136] In 1998 in Sweden, 81.8% of gestational age measurements were based on an ultrasound.[112] Equivalent statistics were not available for HES. However, in both countries >90% of women attend antenatal care before 20 weeks of gestation, and a high proportion of women receive early ultrasound to measure gestational age (estimated >95% in Sweden and 75-95% in England in 2016).[51] Therefore, I assumed that these ultrasound-based estimates are captured in SMBR and HES, and I did not conduct additional analyses to compare the methods used to record gestational age in England and Sweden.

4.4.4.2 Congenital anomaly indicator

Congenital anomalies are an important risk for child mortality. However, few previous international comparisons of child mortality have been able to allow for differences in the prevalence of congenital anomalies. I developed an indicator of the presence of one or more congenital anomalies using longitudinal follow-up in administrative linked datasets used for this study. A congenital anomaly was indicated if an appropriate ICD-10 code was recorded at birth, in hospital admission data up to the age of two, or as any cause of death up to the age of five (details are described in Chapter 3, Section 3.4.3.2 and in Section 4.4.3.1).

To assess the validity of the indicator, I compared the prevalence of congenital anomalies estimated from the Swedish and English birth cohorts, with the prevalence reported by the EUROCAT network. Then, to learn more about possible differences in diagnostic process and types of the anomalies captured in each country, I compared age at first diagnosis and the most commonly diagnosed anomalies in each country. Age at diagnosis was based on the difference in birth date and date of hospital admission when the anomaly was recorded for the first time, or age at death for anomalies only certified at death.

4.4.4.3 Quintile of socio-economic status in England and Sweden

Recorded information about socio-economic status was not directly comparable in England and Sweden. For England, I used Index of Multiple Deprivation score, an area level indicator (described in Chapter 3, Section 3.4.3.3) and for Sweden I used income per household member a year before birth (described in Section 4.4.1.3). To ensure maximum comparability, I calculated quintiles of each measure relative to the population of pregnant women in a given year. I compared the two measures by plotting the distribution of SES quintiles within each of maternal age categories in the two countries (<20, 20-24, 25-29, 30-34, 35-39, ≥40 years). I used maternal age as it is the only comparable SES indicator recorded in both countries.

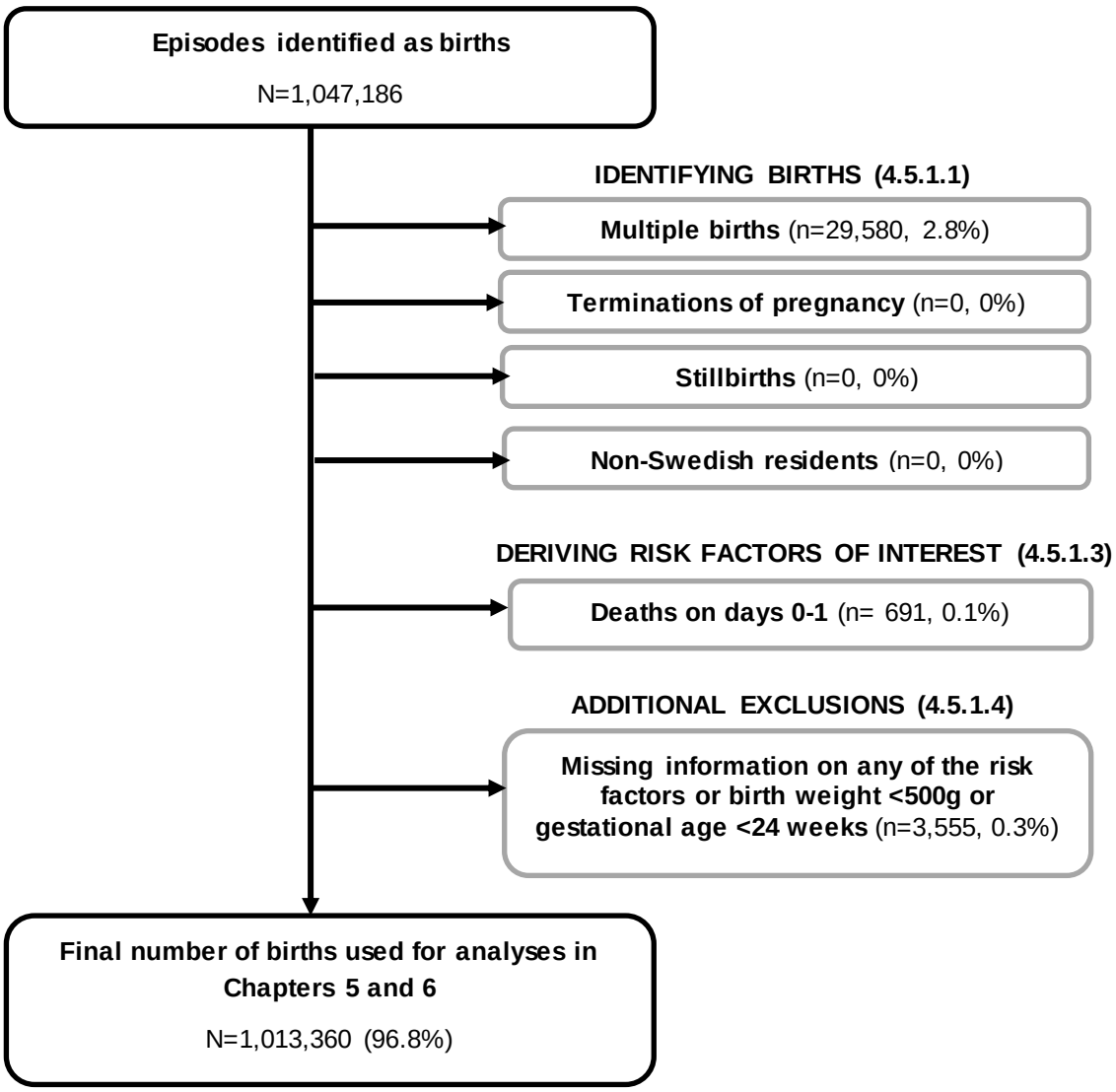
4.5 Results

4.5.1 Developing a Swedish birth cohort

4.5.1.1 Identifying births to match inclusion criteria to the English cohort

There were 1,047,186 births in the SMBR between 2003 and 2012. To match exclusions from the English cohort I only needed to exclude multiple births from SMBR. Multiple births accounted for 3% of all births, a similar proportion as in the English birth cohort (Figure 4.2).

Figure 4.2 - Flow diagram showing exclusions made to develop a representative birth cohort in Sweden using the Swedish Medical Birth Register, covering births in 2003-2012



Data are number of births removed at each stage and % of all births identified.

4.5.1.2 Longitudinal follow-up

4.5.1.2.1 Hospital admissions

There were 490,927 hospital admissions with an age at admission of less than five years old recorded in SHDR for babies identified in the Swedish birth cohort. I removed 30 admissions where the recorded admission date was before the date of birth.

4.5.1.2.2 Mortality registration data

Initially, there were 2,849 deaths which linked to the birth cohort. I removed 29 deaths which had occurred abroad (1.0%) and 27 deaths in children aged over five years old among children in the birth cohort (1.0%). This left 2,793 deaths included in the Swedish birth cohort.

4.5.1.3 Deriving risk factors of interest

The proportion of missing data for the key risk factors was much lower in the Swedish birth cohort than in the English birth cohort. Only 0.4% of records had missing information on any of birth weight, gestational age, sex, maternal age, or quintile of income (Table 4.3). Rates of missing data were higher among deaths than among births, especially for recordings of birth weight for children who died in the first week of life.

Table 4.3 - Percentage of births and deaths by age at death with missing data on risk factors of interest in the Swedish birth cohort in 2003-2012

Risk factor	Live births	Deaths by age				
		0-1 days	2-6 days	7-27 days	28-364 days	1-4 years
All births	1,444,103	1,081	587	634	1,266	753
Birth weight	0.23%	9.4%	11%	6.3%	2.5%	1.1%
Gestational age	0.06%	1.2%	0.68%	1.1%	0.55%	0.13%
Sex	0.0004%	0	0	0	0	0
Maternal age	0.0005%	0	0	0	0	0
Quintile of SES	0.15%	0.83%	0.17%	0.63%	0.16%	0.13%
At least one risk factor	0.45%	11%	12%	7.4%	3.0%	1.2%

SES=socio-economic status. All data are % of all births/deaths in a given age-at-death category.

Child mortality in in the complete case cohort was lower than in the whole cohort (2.3/1000 births compared to 2.5/1000 births, Table 4.4). This reflected under-reported mortality on days 0-1 of life due to higher rates of missing risk factors for these early deaths (0.7/1000 births compared to 0.8/1000 births). Deaths on days 0-1 of life

needed to be excluded from the analyses due to data quality issues in the English cohort (as discussed in Chapter 3). I excluded 691 deaths on days 0-1 (Figure 4.2)

In both England and Sweden, mortality rates beyond days 0-1 of life based on the complete case cohort were under-reported by 0.3-0.5 deaths/1000 births compared to the gold standard (that is, the rates in the whole cohort, Table 4.4). Because the degree of underreporting in the complete case cohorts was comparable in England and in Sweden, I chose not to take any further steps to minimise the effect of missing data in the Swedish cohort.

Table 4.4 – Mortality rates (per 1000 births) in the HES-ONS complete case cohort from selected hospitals and from national publications for England and Wales (ONS), and in the whole Swedish birth cohort and in the Swedish "complete case cohort" by age category

Age at death	HES ONS complete case cohort from selected hospitals	Gold Standard: England and Wales (ONS)	Complete case Swedish cohort	Gold Standard: Whole Swedish cohort
0-1 days	0.83	1.61	0.67	0.75
2-6 days	0.49	0.54	0.36	0.41
7-27 days	0.61	0.64	0.41	0.44
28-364 days	1.28	1.33	0.85	0.88
1-4 years	0.57	N/A	0.52	0.52
0-364 days	3.22	4.12	2.29	2.47

HES=Hospital Episode Statistics; N/A=not available; ONS=Office for National Statistics. Data from ONS publications for England and Wales based on birth and death registration data[34,42,161]

4.5.1.4 Additional exclusions

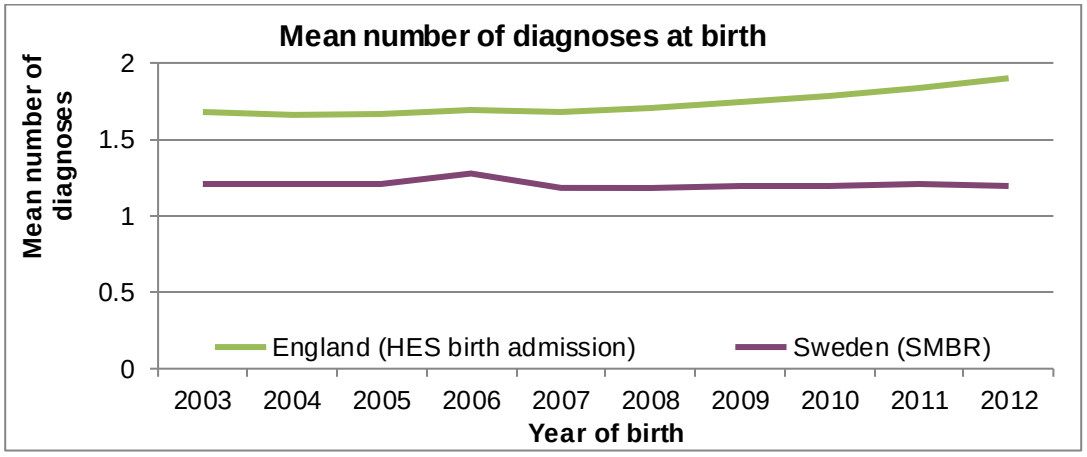
To match exclusion criteria imposed by the missing data in the English birth cohort, I further excluded births with any missing information on risk factors of interest, and with birth weight <500g or gestational age <24 weeks. This led to 3,555 births being excluded (Figure 4.2). Thus, the final cohort covered 1,013,360 births in Sweden in 2003-2012.

4.5.2 Comparability of Swedish and English birth cohorts

4.5.2.1.1 Recorded diagnoses at birth and during hospital admissions

In 2003-2012, babies born in England had, on average, 0.4-0.7 more diagnoses recorded at birth than babies born in Sweden (Figure 4.3). This could reflect the fact that babies admitted to neonatal intensive care units after birth were more likely to have missing diagnoses in SMBR (as their data were recorded in SHDR instead of SMBR).[112]

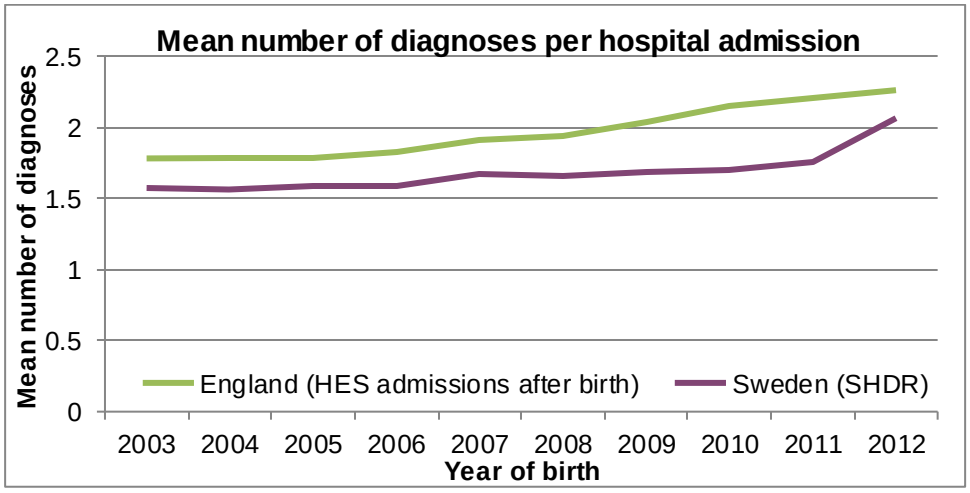
Figure 4.3 – Mean number of diagnoses recorded in SMBR per birth* and in HES per birth admission in 2003-2012**



*HES=Hospital Episode Statistics; SMBR=Swedish Medical Birth Register; SHDR=Swedish Hospital Discharge Register. * Based on births with non-missing diagnoses. Infants transferred to neonatal wards are more likely to have missing diagnoses, which are recorded in SHDR rather than SMBR [112] ** Based on unique diagnoses in any episode of a birth admission in HES.*

The recording of diagnoses in administrative hospital databases was comparable between the two countries, with only 0.2-0.4 more diagnoses on average recorded per hospital admission in HES compared to SHDR (Figure 4.4). In both countries, depth of coding in administrative hospital databases (HES and SHDR) increased over time, most likely as a result of the introduction of financial incentives.

Figure 4.4 – Mean number of diagnoses in SHDR and HES per hospital admission* after birth in 2003-2012



*HES=Hospital Episode Statistics; SHDR=Swedish Hospital Discharge Register. *Based on unique diagnoses in any episode of an admission in HES. An admission was defined as all episodes of care which shared hospital code and admission date.*

A higher proportion of infants in England had at least one hospital admission in the first year of life (25.5% of infants in England vs 17.6% in Sweden), and the mean number of admissions per child in infancy was also higher in 2003-2012 (on average, 2.9 admissions in England compared to 1.5 in Sweden, Table 4.5). This means that there were more opportunities to record comorbidities in HES than in SHDR. This is likely to reflect lower thresholds for hospital admissions in England. However, the differences may also, in part, be explained by to the higher prevalence of adverse birth characteristics, leading to higher hospitalisation rates due to poor health in England.

Table 4.5 – Comparison of trends in hospital admissions in infancy (0-364 days) in England and in Sweden in 2003-2012

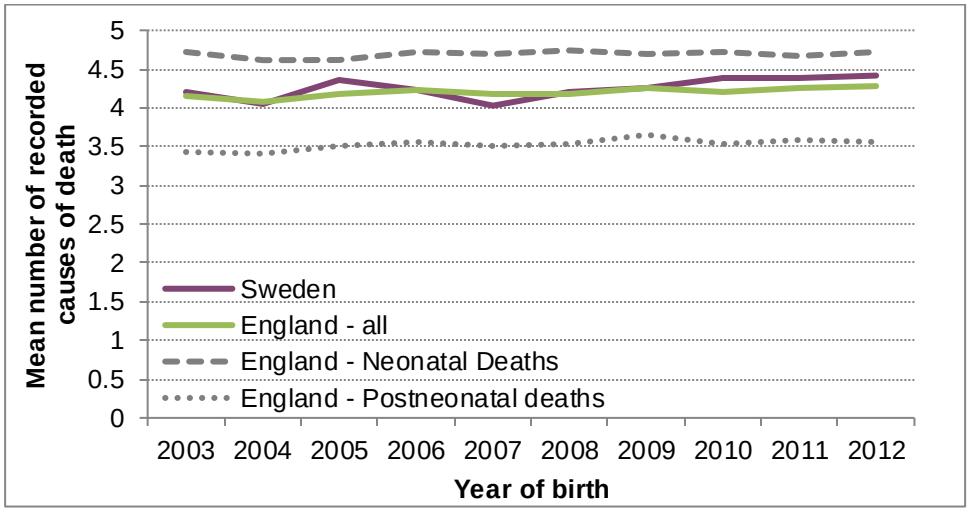
	% of infants with a hospital admission in infancy		Mean number of hospital admissions per child in infancy	
Year of birth	**Sweden**	**England**	**Sweden**	**England**
2003	16.6%	23.5%	1.44	2.86
2004	16.0%	24.2%	1.42	2.87
2005	17.5%	25.2%	1.41	2.89
2006	16.9%	25.3%	1.43	2.87
2007	17.8%	25.2%	1.45	2.89
2008	16.4%	25.5%	1.47	2.89
2009	16.8%	26.9%	1.47	2.90
2010	18.2%	26.9%	1.44	2.85
2011	19.5%	26.0%	1.47	2.81
2012	20.2%	25.5%	1.44	2.73
Total	17.6%	25.5%	1.45	2.85

Data in second and third column are % of all infant deaths. Data in fourth and fifth column are mean numbers of hospital admission per child (based on births from the birth cohorts)

4.5.2.1.2 Coding of causes of death

The coding depth for causes of death was comparable between England and Sweden (Figure 4.5). On average, four causes of death were recorded both in England and in Sweden. In England, the coding depth was higher in the neonatal period, possibly due to the inclusion of additional maternal conditions which contributed to the neonatal death on the Neonatal Death Certificate.[22] In the post-neonatal period, the mean number of causes of death recorded in England was lower by 0.5 recorded causes than in Sweden.

Figure 4.5 – Mean number of causes of deaths recorded on deaths certificate in England and in Sweden for babies aged 0-364 days in 2003-2012

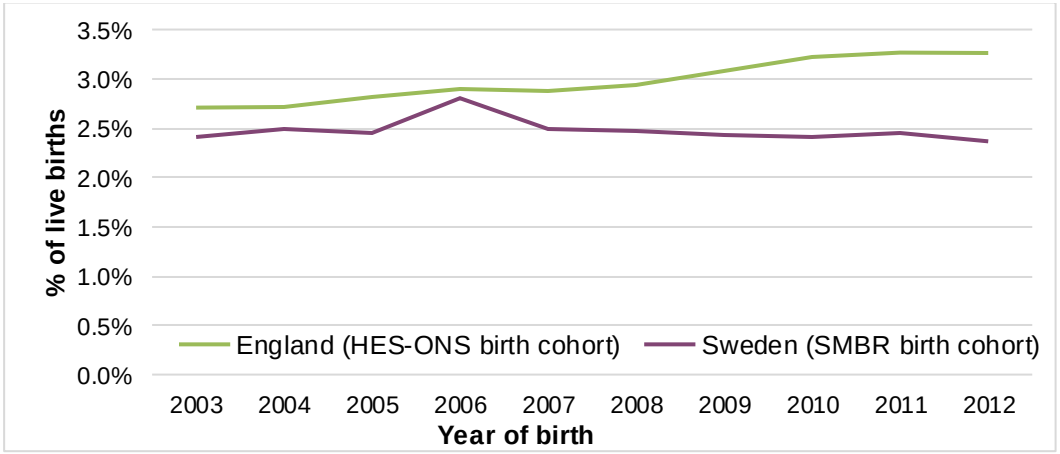


4.5.3 Recording of risk factors of interest

4.5.3.1 Congenital anomaly indicator

Overall, the prevalence of congenital anomalies was higher in England than in Sweden (2.9% compared to 2.4%, Figure 4.6). For both countries, the prevalence was higher than that reported by the EUROCAT network: 1.8% for Sweden in 2007-2012 and 2.0% in England in 2003-2012.[46] These differences could reflect differences in coverage (whole country coverage in HES vs regional registration data in EUROCAT for England, not including London), differences in ICD-10 codes used to classify an anomaly, and differences in data collection process.

Figure 4.6 – Prevalence of congenital anomalies in England and in Sweden in 2003-2012 based on HES-ONS birth cohort (described in Section 3.5) and Swedish birth cohort (described in Section 3.5.1)

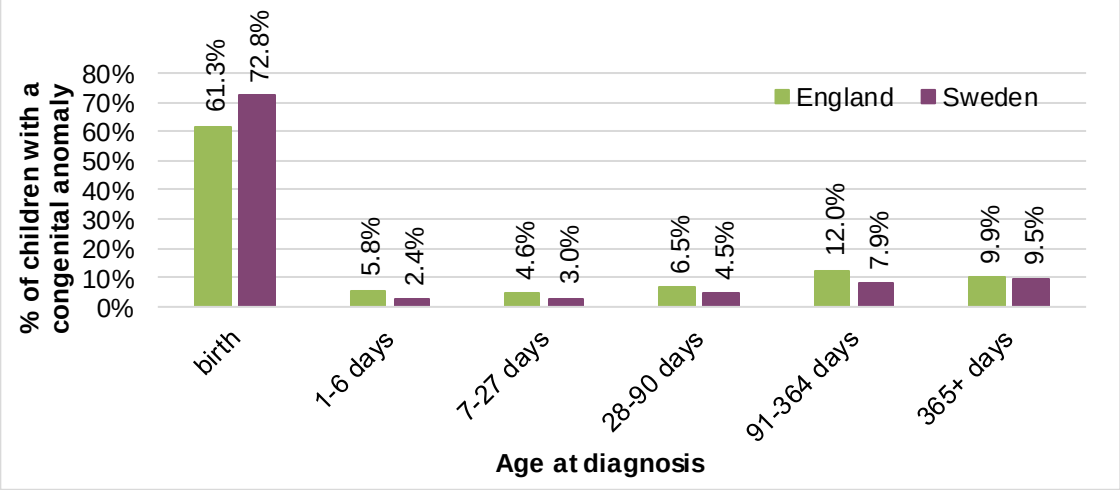


HES=Hospital Episode Statistics; ONS=Office for National Statistics; SMBR=Swedish Medical Birth Register.

In England, the prevalence of congenital anomalies recorded in HES-ONS birth cohort increased over time (from 2.7% in 2003, to 3.3% in 2012). The increase in the prevalence of congenital anomalies observed in the HES data could partly reflect improved coding depth due to financial incentives (as indicated by an increase in the mean number of recorded diagnoses illustrated in Figure 4.4), or improvements in diagnosis of congenital anomalies. Some of the differences could also reflect changes in the death certification practices, leading to a higher number of congenital anomalies being identified as a cause of death – the proportion of deaths beyond 30 days of life with a congenital anomaly recorded as any cause of death increased from 23.7% in 2003-5 to 27.4% in 2010-12. Finally, it could be a true increase in the prevalence, as the proportion of mothers aged over 40 years old (who have an increased risk of pregnancy with chromosomal abnormalities) also increased over the study period (from 3.0% of mothers of singleton live births in 2003 to 4.0% in 2012 in England and Wales).[42,161]

Congenital anomalies were diagnosed at an earlier age in Sweden than in England. In Sweden, 75% of children with a congenital anomaly were diagnosed in the first week of life (based on admission date or date of death), compared to 67% in England, (Figure 4.7). The proportion of children who had a congenital anomaly diagnosis recorded in the first week of life remained constant over the study period in both countries. Overall, the most commonly recorded anomalies were similar in England and Sweden (Table 4.6). Therefore, I concluded that despite differences in the coding of diagnoses in administrative hospital databases, the indicator of congenital anomalies was comparable between the two countries.

Figure 4.7 – Age at first diagnosis of a congenital anomaly in 2003-2012 by country



Diagnosis at birth was identified based on diagnoses in Swedish Medical Birth Register or at birth admission in Hospital Episode Statistics. Age at diagnosis is based on the difference in birth date and date of admission when a congenital anomaly was recorded for the first time, or age at death for anomalies only certified at death.

Table 4.6 – Most commonly recorded congenital anomalies in England and in Sweden and % of children with a given ICD-10 code out of all children diagnosed with at least one congenital anomaly

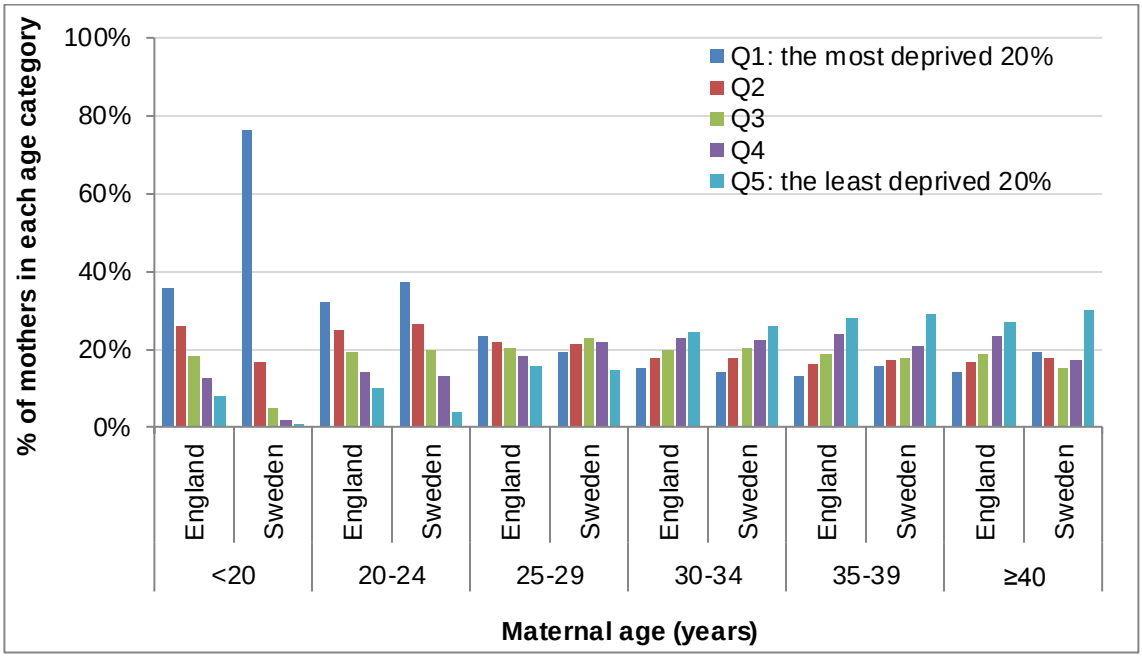
England			Sweden		
ICD-10 code	**Description**	**% of children with a CA**	**ICD-10 code**	**Description**	**% of children with a CA**
Q21	Congenital malformations of cardiac septa	24.3%	Q21	Congenital malformations of cardiac septa	32.4%
Q25	Congenital malformations of great arteries	22.6%	Q25	Congenital malformations of great arteries	12.9%
Q54	Hypospadias	14.9%	Q54	Hypospadias	11.9%
Q62	Congenital obstructive defects of renal pelvis and congenital malformations of ureter	10.5%	Q62	Congenital obstructive defects of renal pelvis and congenital malformations of ureter	7.0%
Q65	Congenital deformities of hip	7.2%	Q24	Other congenital malformations of heart	6.3%
Q24	Other congenital malformations of heart	5.9%	Q65	Congenital deformities of hip	5.3%
Q31	Congenital malformations of larynx	5.6%	Q90	Down syndrome	5.0%
Q75	Other congenital malformations of skull and face bones	4.7%	Q22	Congenital malformations of pulmonary and tricuspid valves	3.4%
Q04	Congenital hydrocephalus	4.4%	Q35	Cleft palate	3.3%
Q63	Other congenital malformations of kidney	4.1%	Q75	Other congenital malformations of skull and face bones	3.2%

CA=congenital anomaly; ICD-10=the International Statistical Classification of Diseases and Related Health Problems version 10. Data are % of children with any identified congenital anomaly.

4.5.3.2 Quintile of socio-economic status in England and Sweden

The distribution of maternal age by quintile of SES differed between England and Sweden (Figure 4.8). In Sweden, 76.5% of teenage mothers (<20 years old) were in the most deprived 20% of mothers, compared to only 35.8% in England. Overall, the distribution of SES quintiles for mothers aged ≥20 years old was comparable between the two countries; however, the differences were narrower in England, reflecting the use of an area-level indicator in England. Therefore, the observed differences in child mortality by SES in England are likely to be attenuated compared to using individual level measure of socio-economic status in Sweden.

Figure 4.8 – Percentage of mothers in each quintile of socio-economic status by maternal age category* in England and in Sweden in 2003-2012



4.6 Discussion

4.6.1 Key findings

Administrative datasets in England and national registers in Sweden can be used to develop comparable birth cohorts of singleton live births for the investigation of child mortality after day one of life. The datasets in both countries contain comparable information on birth weight, gestational age, sex and maternal age. Despite differences in the coding of diagnostic information and rates of hospital admissions, a comparable indicator of congenital anomalies can be generated. However, the only available measure of SES is not directly comparable between the two countries.

4.6.2 Strengths

Swedish national registers provide an extremely rich resource for epidemiological research. They cover all residents in the country and can be linked with low error rates using the PIN. The data collected in these registers is intended for research and undergo quality assurance checks in the process of data collection. Developing a birth cohort with longitudinal follow-up can be therefore done relatively quickly, in contrast to the English birth cohort. Developing the English birth cohort was a drawn out and complex process which took a year and a half in total. By contrast, my work to develop a comparable Swedish birth cohort took a month in total.

4.6.3 Limitations

A higher proportion of children in England had at least one hospital re-admission in infancy, and the mean number of admissions was also higher in England than in Sweden. Children in England therefore had more opportunities for the recording of congenital anomalies. It remains unclear whether the apparent increase in the prevalence of congenital anomalies in England reflected improved diagnosing and recording of congenital anomalies in England, changes to death registration practices or a true increase in the prevalence. Both in England and in Sweden, most anomalies were diagnosed at birth or within the first week of life and the most commonly recorded anomalies were similar. Therefore, I concluded that the indicators of congenital anomaly were comparable between England and Sweden.

The measures of SES available in England and Sweden were not directly comparable – I used an area-level measure in England and individual-level measure in Sweden. Maternal education level is considered the most comparable SES indicator for inter-country comparisons of health outcomes,[192] but such a variable was not available in England. No area-level measures of SES were available in LISA or the other Swedish registers. Calculating quintiles of SES amongst all pregnant women helped to standardise the indicator of SES. However, the effect of SES on child mortality in England is likely to be underestimated in the analyses presented in Chapters 5 and 6.

4.6.4 Implications for this thesis

I developed birth cohorts for comparing child mortality in England and in Sweden, including hospital singleton live births to resident mothers between the 1st January 2003 and 31st December 2012, with follow-up through hospital admissions and mortality databases until 31st December 2013. Both cohorts covered information on birth characteristics (birth weight, gestational age, sex and presence of congenital anomalies) and socio-economic factors (maternal age and quintile of socio-economic status). The cohorts included deaths beyond days 0-1 of life occurring within each country. Births with missing data, birth weight <500 or gestational age <24 weeks were excluded. Although I used the same code list and definition to derive an indicator of presence of congenital anomalies for both cohorts, differences in coding depth and hospital admission rates in infancy and early childhood between countries may introduce bias. Therefore, an indicator of more severe congenital anomalies is used for sensitivity analyses in Chapter 5. Differences in child mortality by SES are likely to be attenuated in England relative to Sweden, due to the use of an area-level indicator. However, this was the only indicator of SES available in the English cohort.

Chapter 5. Comparison of child mortality in England and Sweden

What is already known:

- Individual-level data containing information on multiple risk factors at birth are needed to identify the origins of child mortality differences between countries.
- Comparable, nationally-representative birth cohorts created using linked administrative databases are available for England and Sweden, excluding deaths beyond the first day of life, births at <24 weeks, with birth weight <500g or with missing information on any of the key risk factors.

What this chapter adds:

- I compare child mortality in England and Sweden using individual-level data.
- I quantify the contribution of birth characteristics and socio-economic factors to the excess risk of death in England relative to Sweden.
- My results can be used to inform policies to reduce child mortality in England relative to Sweden.

5.1 Chapter overview

As discussed in Chapter 1, the risk of child death is associated with a child's health at birth, which in turn is determined by maternal health and socio-economic circumstances before and during pregnancy. To better understand why inter-country differences in child mortality arise, we need to disentangle the contribution of risk factors operating before and during pregnancy (as manifested by adverse birth characteristics), and risk factors operating after birth.

This chapter presents work towards objective 3: "*to compare the risk of child mortality in England and Sweden using individual-level data and to determine to what extent the differences can be explained by birth characteristics and socio-economic factors.*" To overcome the limitations of using aggregate data, I analysed data from national birth cohorts from England and Sweden (described in Chapters 3 and 4 of this thesis). Such an approach enabled me to quantify the contribution of birth characteristics and socio-economic factors to the excess child mortality in England relative to Sweden.

I have presented some of the work described in this chapter at the 2016 International Population Data Linkage Conference (Swansea, United Kingdom (UK)), the 2017 Administrative Data Research Network Conference (Edinburgh, UK) and at the 2017 IEA World Congress of Epidemiology (Saitama, Japan). A manuscript based on the main analyses from this chapter has been accepted for publication at the Lancet.

5.2 Background

5.2.1 Overview of child mortality in the UK and Sweden

As discussed in Chapter 1, the UK has some of the highest child mortality rates in Western Europe, while Sweden has some of the lowest (Figure 1.1 in Chapter 1). In 2013, child mortality in the UK was almost twice as high as that in Sweden (4.9 deaths/1000 births compared to 2.7/1000 births, respectively).[3]

Previous comparisons attributed these differences to wider socio-economic inequalities in the UK relative to Sweden and differences in the provision of healthcare.[6] Wider socio-economic inequalities lead to an increased prevalence of preterm birth, and thus higher rates of prematurity-related deaths (138.5 deaths/100,000 births compared to 10.1/100,000 births in 2006-8).[6] High rates of mortality due to infections in the UK relative to Sweden (63.9 deaths/100,000 births compared to 34.8/100,000 births in 2006-8) were attributed to delays in the diagnosis of acute life-threatening infections.[6] In particular, it was argued that the introduction of mandatory paediatric training for GPs and better integration of primary care and paediatric services could ensure more timely diagnosis and treatment of infections in children, possibly preventing some of the infection-related mortality.[8,10,193]

5.2.2 Limitations of previous comparisons

Previous comparisons of child mortality in the UK relative to Sweden did not account for differences in the prevalence of adverse birth characteristics associated with increased risk of child death, such as congenital anomalies, preterm birth and low birth weight. As detailed in Chapter 1, the comparisons were based on unadjusted mortality rates or data on the underlying cause of death. It was not possible to determine, whether increased child mortality reflected the UK's high prevalence of adverse birth characteristics, or differences in care received after birth, given a child's characteristics at birth. Such distinction is crucial to inform policy makers when and how to target interventions to prevent the largest number of child deaths. Should they focus on addressing maternal health and socio-economic factors before and during pregnancy, or on improving the care of babies and children after birth? As shown in Chapter 2, comparisons of child mortality based on aggregate data tabulated by a key risk factor (such as gestational age) provide limited insights into the origins of inter-country

differences in child mortality. For a fair comparison, analyses adjusted for multiple risk factors at birth are needed.

5.2.3 Chapter aims

This chapter compared child mortality between England (where 85% of all births in the UK occur)[1] and Sweden using birth cohorts described in Chapters 3 and 4. I determined how much of the excess child mortality in England relative to Sweden was explained by differences in birth characteristics and socio-economic factors. Presented results can inform policy makers in England as to which preventive strategies would most effectively reduce child mortality rates relative to Sweden.

5.3 Methods

5.3.1 Datasets

I used nationally-representative comparable birth cohorts derived from administrative linked datasets in England and Sweden, described in Chapters 3 and 4. The cohorts included singleton live births born in hospital to resident mothers between 1st January 2003 and 31st December 2012. Due to problems with the completeness of risk factors recorded in the English birth cohort (outlined in Chapter 3), 29% of births in hospitals with "poor" quality of recorded data on birth characteristics and socio-economic factors in England were excluded from the analyses. I also excluded births with birth weight <500g or gestational age <24 weeks in both countries.

5.3.2 Outcomes

The outcomes of interest were all-cause mortality rates at age 2-27 days, 28-364 days and 1-4 years. I looked at these age-at-death categories separately since the effect of birth characteristics (e.g., birth weight) and socio-economic factors (e.g., maternal age) on the risk of death is different in the first 27 days of life, at 28-364 days and at 1-4 years.[194] I excluded deaths on days 0-1 of life from the analyses due to a high proportion of missing data on risk factors of interest for these deaths in the English cohort (described in detail in Section 3.5.4.2 in Chapter 3), and to reduce bias from inter-country differences in registration practices for stillbirths, live births and early neonatal deaths. Children were followed up until their fifth birthday, death, or the 31st December 2013, whichever occurred first.

5.3.3 Risk factors

Birth characteristics of interest included birth weight (categorised as 500-999, 1000-1499, 1500-2499, 2500-3499, ≥3500g), gestational age (grouped as 24-27, 28-31, 32-34, 35-36, 37-38, ≥39 weeks), sex and presence of congenital anomalies recorded during hospital admissions in the first two years of life or as any cause of death (coded

as a binary variable: absent vs. present, details of the method are presented in Section 3.4.3.2.3 in Chapter 3, and a comparison of congenital anomalies coding in England and Sweden is presented in in Section 4.5.3.1 in Chapter 4). Socio-economic factors included maternal age (categorised as <20, 20-24, 25-29, 30-34, 35-39, ≥40 years old) and quintile of socio-economic status (SES), measured using Index of Multiple Deprivation scores in England (described in detail in Section 3.3.1.3.4 in Chapter 3) and family's disposable income a year before pregnancy in Sweden (described in Section 4.3.4. in Chapter 4). All analyses were based on a cohort of births with complete information on all risk factors. The number of births and deaths excluded at each stage of cohort specification are presented in Figure 5.1 in Section 5.3 of this chapter.

5.3.4 Statistical analyses

5.3.4.1 Exploratory analyses

I derived the numbers and proportions of live births and deaths, tabulated by each risk factor of interest, to compare the characteristics of children who were born and who died in England and Sweden. I also calculated unadjusted child mortality rates per 100,000 child years, overall and by risk factor category for each country.

For each of the birth characteristics, I examined the inter-country differences in timing of deaths by plotting Kaplan-Meier failure curves. I then calculated the average number of "excess" deaths per year, attributable to the increased mortality in England relative to Sweden at age 1 and 5 years old. The excess deaths were calculated by multiplying the number of singleton live births in England in 2003-2012 by the difference in the proportion of children who died in England and Sweden (by their first and fifth birthday), and dividing it by ten to get the average number of excess deaths per year. For the calculation, I used the number of all singleton live births identified in the English birth cohort in 2003-2012, before excluding 29% of births from hospitals reporting data of poor quality (*n*=6,100,404). This whole-country birth cohort covered 96.0% of singleton live births overall in England in 2003-2012, and 99.0% of singleton live births in hospitals.[44,162]

To assess whether socio-economic inequalities in child mortality showed similar patterns in England and Sweden, I also plotted unadjusted mortality rate ratios for each quintile of SES relative to the least deprived 20% of the population and for maternal age categories relative to mothers aged 30-34.

5.3.4.2 Cox proportional hazards models

Mortality is an example of a time-to-event outcome, where we are interested in both the outcome (death) and the time when it was observed (age at death). Therefore, for the analyses in this chapter I used Cox proportional hazards (PH) regression models, the most common statistical method for analysing time-to-event data.[166] I fitted Cox PH regression models to estimate hazard ratios (HRs) for mortality in England relative to Sweden (baseline) at 2-27 days, 28-364 days and 1-4 years. I first fitted unadjusted models, including only country of birth as a covariate. Next, I added birth weight, gestational age and sex. Then, I adjusted the models for all birth characteristics, including presence of congenital anomalies. This enabled me to observe the contribution of congenital anomalies to the HR for low birth weight, illustrating the importance of including an indicator of congenital anomalies in international comparisons of child mortality. Finally, I added socio-economic factors (maternal age and SES quintile) to the models.

5.3.4.3 Percentage of excess risk mediated

To quantify the contribution of inter-country differences in birth characteristics and socio-economic factors to the increased risk of death in England relative to Sweden, I calculated percentage excess risk mediated (PERM) – the proportional reduction in the HR for England relative to Sweden after adjusting for all birth characteristics and socio-economic risk factors. PERM was calculated as per Equation 5.1.[195,196]

Equation 5.1 – Percentage excess risk of death in England relative to Sweden mediated by birth characteristics and socio-economic factors

$$\text{PERM}=\frac{\text{Hazard Ratio}\left(\begin{matrix}\text{adjusted for}\\ \text{country only}\end{matrix}\right)-\text{Hazard Ratio}\left(\begin{matrix}\text{adjusted for country,}\\ \text{birth characteristics,}\\ \text{and socio-economic factors}\end{matrix}\right)}{\text{Hazard Ratio}\left(\begin{matrix}\text{adjusted for}\\ \text{country only}\end{matrix}\right)-1}\times 100$$

PERM=percentage excess risk mediated.

I then used PERM as an approximate mean to partition the contribution of birth characteristics and socio-economic factors (independently of their effect on birth characteristics) to the increased risk of child death in England relative to Sweden by splitting it into three components (Equation 5.2). Component (1) represented the excess risk of death mediated by birth weight, gestational age, and sex; component (2) reflected the independent contribution of congenital anomalies, beyond what was already accounted for by other birth characteristics; component (3) reflected a further independent contribution of socio-economic factors, given birth characteristics. PERM was calculated only for age group models where the HR for England in the fully

adjusted model remained statistically significant, i.e. where the Wald test for the country parameter was $p<0.05$.[166] For the calculation of PERM I assumed that there are no unmeasured confounders.

Equation 5.2 – Breakdown of PERM into three components describing independent contribution of (1) birth weight, gestational age and sex, (2) presence of congenital anomalies given other birth characteristics and (3) socio-economic factors, given all birth characteristics to the percentage excess risk of death in England relative to Sweden

$$PERM = PERM_{\text{Birth characteristics}} + PERM_{\text{Congenital anomalies}} + PERM_{\text{Socio-economic factors}} =$$

$$= \frac{\text{Hazard Ratio}\left(\begin{matrix}\text{adjusted for}\\ \text{country only}\end{matrix}\right) - \text{Hazard Ratio}\left(\begin{matrix}\text{adjusted for country, birth weight,}\\ \text{gestational age and sex}\end{matrix}\right)}{\text{Hazard Ratio}\left(\begin{matrix}\text{adjusted for}\\ \text{country only}\end{matrix}\right) - 1} \times 100 + \quad (1)$$

$$+ \frac{\text{Hazard Ratio}\left(\begin{matrix}\text{adjusted for country, birth weight,}\\ \text{gestational age and sex}\end{matrix}\right) - \text{Hazard Ratio}\left(\begin{matrix}\text{adjusted for country, birth weight,}\\ \text{gestational age, sex}\\ \text{and congenital anomalies}\end{matrix}\right)}{\text{Hazard Ratio}\left(\begin{matrix}\text{adjusted for}\\ \text{country only}\end{matrix}\right) - 1} \times 100 + \quad (2)$$

$$+ \frac{\text{Hazard Ratio}\left(\begin{matrix}\text{adjusted for country, birth weight,}\\ \text{gestational age, sex}\\ \text{and congenital anomalies}\end{matrix}\right) - \text{Hazard Ratio}\left(\begin{matrix}\text{adjusted for country, birth weight,}\\ \text{gestational age, sex,}\\ \text{congenital anomalies}\\ \text{and socio-economic factors}\end{matrix}\right)}{\text{Hazard Ratio}\left(\begin{matrix}\text{adjusted for}\\ \text{country only}\end{matrix}\right) - 1} \times 100 \quad (3)$$

PERM=percentage excess risk mediated.

5.3.4.4 Subgroup analyses

Mortality in children with no adverse birth characteristics can be seen as an indicator of the care received after birth, whether in the healthcare setting or at home. These low-risk babies are born with no underlying risk factors which could increase their susceptibility to poor health (such as preterm birth, low birth weight or presence of congenital anomalies). Therefore, their risk of death is associated with risk factors operating throughout their life. For example, in both England and Sweden, sudden infant death syndrome (SIDS) is one of the most common causes of death in non-malformed infants born at term.[32,34] As described in Chapter 1, the risk of SIDS is associated with key aspects of care at home, including safe sleeping practices, or parental smoking.[89,197]

To determine the contribution of risk factors operating after birth to the differences in child mortality in England and Sweden, I compared mortality in low-risk children in the two countries. I defined low-risk children as born with birth weight ≥2500g, at full term (39-41 weeks), with no congenital anomalies. I compared mortality at 2-27 days, 28-364 days and 1-4 years using Cox PH model. For each age at death category, I first fitted an unadjusted model, including country as the only covariate. I then added birth characteristics (gestational age by week, birth weight by 500g categories and sex), and

socio-economic factors (SES quintile and maternal age). I calculated PERM statistics for all models where the Wald test for the country parameter was $p<0.05$.[166]

5.3.4.5 Sensitivity analyses

I repeated all analyses with a stricter definition for severe congenital anomalies based on a code list of paediatric complex chronic conditions developed by Feudtner *et al.*[198] The list was restricted to the 10th Revision International Statistical Classification of Diseases and Related Health Problems (ICD-10) codes from Chapter 17 ("Congenital malformations, deformations and chromosomal abnormalities"[24], beginning with Q).[24]

The underlying assumption of the Cox PH model is that the HRs for all covariates remain constant over time.[166] This assumption was tested using the Grambsch and Therneau test of PH, based on scaled Schoenfeld residuals. In brief, Schoenfeld residuals are the difference in the covariate value and the expected value of the covariate for each observed failure (i.e. death).[166] Residuals are calculated for all deaths in the sample and for all covariates in Cox PH model, and weighted using the inverse of the covariance matrix.[166] The Grambsch and Therneau test checks whether the slope in a generalized linear regression of the scaled Schoenfeld residuals over time is non-zero (globally and for each individual risk factor). This is equivalent to testing that the logarithm of the HR function is constant over time.[166]

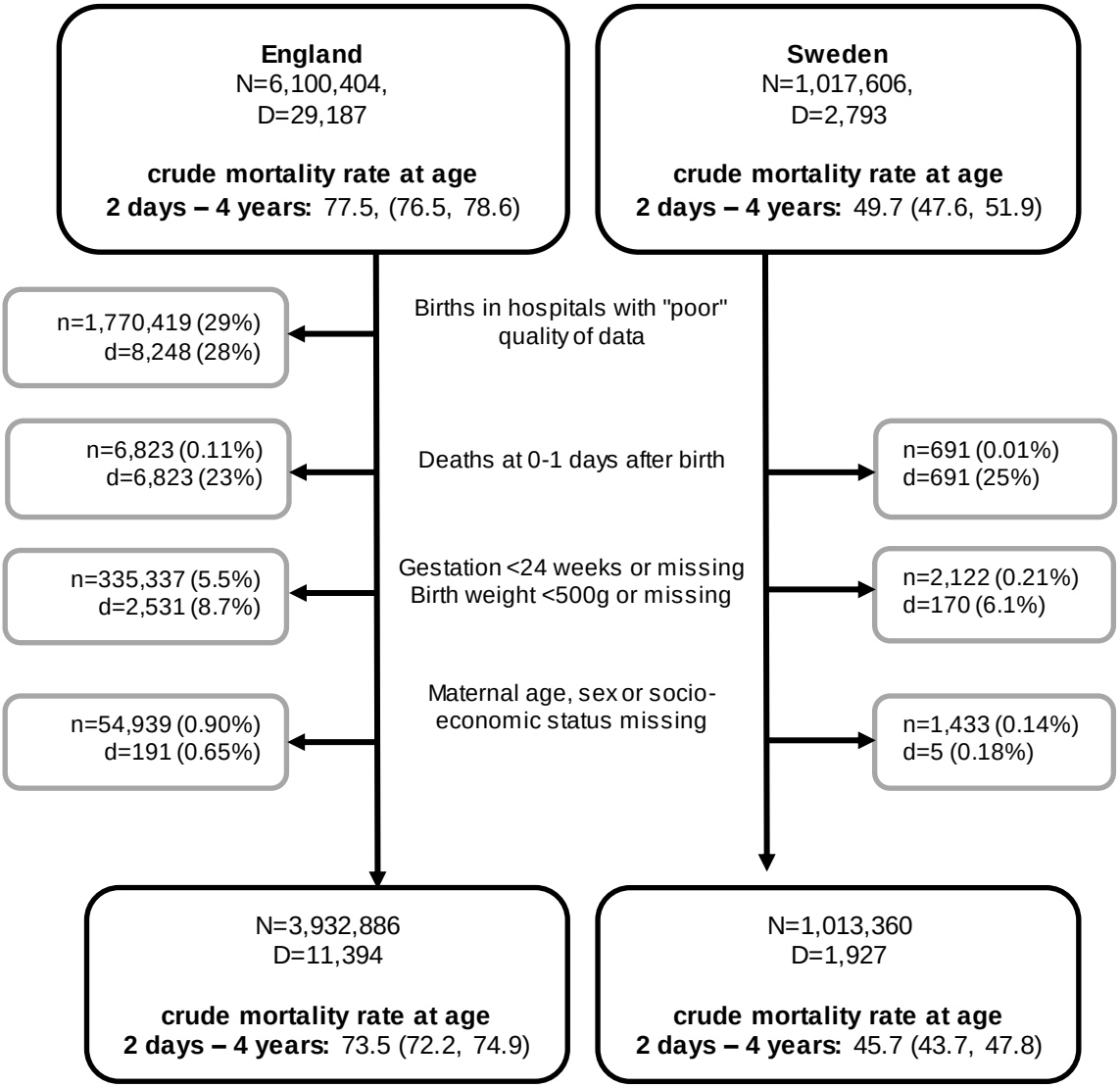
Given the large sample size in this study, even small changes in the slope could result in statistically significant *p*-values. Thus, if the Grambsch and Therneau test indicated that PH assumption was violated, I plotted scaled Schoenfeld residuals (on the y-axis) versus follow-up time (on the x-axis) for visual examination.[199] If the pattern of residuals over time (assessed using a smoothed fitted line for ease of interpretation) did not show a straight line with a slope of 0, I fitted additional Cox PH regression models with an interaction between the given covariate and survival time.

5.4 Results

5.4.1 Comparable birth cohorts in England and Sweden

The cohorts used for the analyses comprised 3,932,886 births and 11,392 deaths in the English cohort and 1,013,360 births and 1,927 deaths in the Swedish cohort. Numbers of births and deaths excluded at each stage are presented in Figure 5.1. The cohorts represented 64.5% of all singleton live births and 58.4% of all deaths at age 2 days – 4 years in England; and 99.8% of births and 91.7% of deaths in Sweden. The derivation of these cohorts was explained in detail in Chapters 3 and 4.

Figure 5.1 – Flow diagram showing steps taken to develop comparable and representative birth cohorts in England and Sweden in 2003-2012



The numbers of live births (n) and deaths (d) are presented. For each exclusion criterion, the percentage of all live births and all deaths is shown in brackets. Crude mortality rates at age 2 days – 4 years per 100,000 child years are presented for each country before and after applying all exclusion criteria.

5.4.2 Characteristics of births and children who died in England and Sweden

England had a less favourable distribution of birth weight than Sweden – the prevalence of low birth weight was higher (5.5% vs 3.0%) and a lower proportion of births weighed ≥3500g (41% compared to 55% in Sweden, Table 5.1). Rates of preterm birth and congenital anomalies were 5.7% and 2.9%, respectively in England, compared to 4.8% and 2.4% in Sweden. Mothers in England were four times more likely to give birth aged <20 years old (6.1% vs 1.6% in Sweden). In both countries a birth weight of <2500g and gestational age of <32 weeks were more common in the most deprived 20% of mothers; in England the most deprived 20% of mothers also

experienced higher rates of birth at 32-38 weeks or with a congenital anomaly (see Appendix E, Table E.1).

Table 5.1 – Socio-demographic characteristics of singleton live births in England and Sweden in 2003-2012

Risk factor	England n (%)	Sweden n (%)
Birth weight (g)		
500-999	9,458 (0.24%)	1,742 (0.17%)
1000-1499	18,288 (0.47%)	3,102 (0.31%)
1500-2499	190,299 (4.8%)	25,817 (2.5%)
2500-3499	2,090,583 (53%)	429,107 (42%)
≥3500	1,624,258 (41%)	553,592 (55%)
Gestational age (weeks)		
24-27	8,806 (0.22%)	1,769 (0.17%)
28-31	22,327 (0.57%)	4,354 (0.43%)
32-34	56,093 (1.4%)	11,764 (1.2%)
35-36	137,046 (3.5%)	30,295 (3.0%)
37-38	726,907 (18%)	191,130 (19%)
≥39	2,981,707 (76%)	774,048 (76%)
Sex		
Boy	2,016,683 (51%)	520,985 (51%)
Girl	1,916,203 (49%)	492,375 (49%)
Congenital anomaly		
No	3,817,789 (97%)	988,681 (98%)
Yes	115,097 (2.9%)	24,679 (2.4%)
Maternal age (years)		
<20	241,503 (6.1%)	16,160 (1.6%)
20-24	758,596 (19%)	129,240 (13%)
25-29	1,064,469 (27%)	295,905 (29%)
30-34	1,110,202 (28%)	356,356 (35%)
34-39	617,394 (16%)	178,992 (18%)
≥40	140,722 (3.6%)	36,707 (3.6%)
Quintile of socio-economic status		
Q1: most deprived	852,422 (22%)	201,613 (20%)
Q2	804,432 (20%)	200,440 (20%)
Q3	768,484 (20%)	202,670 (20%)
Q4	763,076 (19%)	204,215 (20%)
Q5: least deprived	744,472 (19%)	204,422 (20%)

All data are numbers of all singleton live births in each country (percentage). Column totals may not add up to 100% due to rounding.

Characteristics of children who died were largely similar between the two countries (Table 5.2). Preterm birth and low birth weight accounted for approximately half of all deaths at 2-27 days, a third of deaths at 28-364 days and a seventh of deaths at 1-4 years in England and Sweden. Beyond the first month of life, the highest proportion of deaths occurred among babies with normal birth weight (≥2500g) or born at ≥39 weeks, which account for the largest number of births overall and are not typically

considered “high-risk”. In both countries, over 40% of deaths at 2-27 days and 28-364 days, and a lower proportion at 1-4 years (38% in England and 25% in Sweden) occurred to children with a congenital anomaly. Deaths at all ages were more common in the most deprived 20% of births, than for other quintiles.

Table 5.2 – Socio-demographic characteristics of children who died in England and Sweden by age at death in 2003-2012

	Deaths at 2-27 days		Deaths at 28-264 days		Deaths at 1-4 years	
	England	**Sweden**	**England**	**Sweden**	**England**	**Sweden**
Number of deaths	4207	648	4964	803	2223	476
Birth weight (g)						
500-999	25%	20%	13%	9.5%	2.5%	0.84%
1000-1499	9.3%	9.7%	5.6%	5.2%	2.3%	2.1%
1500-2499	18%	21%	19%	17%	14%	9.9%
2500-3499	34%	31%	46%	43%	55%	45%
≥3500	13%	19%	16%	26%	27%	42%
Gestational age (weeks)						
24-27	25%	21%	12%	9.7%	2.3%	0.63%
28-31	11%	9.9%	6.3%	4.9%	2.1%	1.5%
32-34	7.6%	9.7%	6.4%	6.2%	3.2%	5.0%
35-36	8.3%	11%	9.0%	10%	6.5%	5.9%
37-38	17%	18%	23%	24%	23%	21%
≥39	32%	31%	43%	45%	63%	66%
Sex						
Boy	57%	57%	57%	57%	54%	55%
Girl	43%	43%	43%	43%	46%	45%
Congenital anomalies						
No	56%	56%	55%	59%	62%	75%
Yes	44%	44%	45%	41%	38%	25%
Maternal age (years)						
<20	8.9%	1.7%	12%	4.2%	8.5%	3.4%
20-25	21%	16%	24%	19%	25%	15%
25-30	27%	25%	25%	29%	26%	28%
30-35	23%	34%	22%	27%	25%	33%
35-40	15%	17%	13%	16%	13%	17%
≥40	4.8%	6.5%	4.1%	4.9%	3.1%	4.2%
Quintile of socio-economic status						
Q1: most deprived	30%	26%	32%	31%	28%	24%
Q2	23%	20%	25%	21%	22%	24%
Q3	18%	14%	17%	13%	19%	20%
Q4	16%	14%	15%	16%	16%	18%
Q5: least deprived	13%	27%	11%	18%	14%	14%

All data are % of all singleton live births or deaths in each country (n). Column totals may not add up to 100% due to rounding.

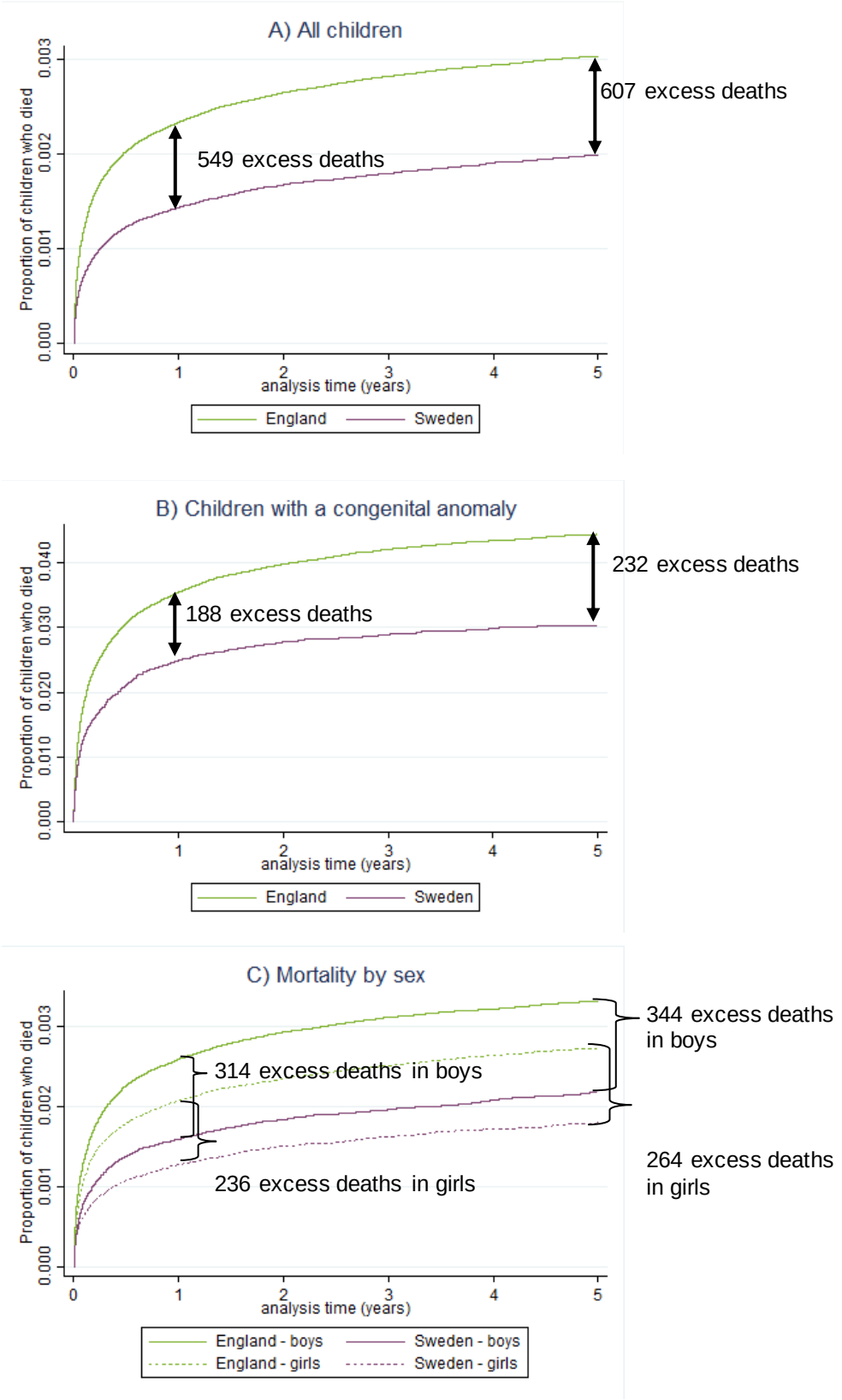
5.4.3 Comparison of unadjusted mortality rates in England and Sweden

Unadjusted child mortality rate was almost twice as high in England than in Sweden (74 vs 46/100,000 child-years). This difference was equivalent to approximately 607 excess deaths per year in the whole of England relative to Sweden (out of the total of 1,767 deaths at 2 days-4 years per year, based on all birth in the English cohort before exclusions, Figure 5.1); 549 of these deaths were attributable to differences in mortality at 2-364 days (Figure 5.2, Plot A).

In both countries, mortality was highest for birth characteristics typically considered to be “high risk”: babies weighing 500-1499g at birth, born at 24-32 weeks’ gestation, or with congenital anomalies (as indicated by higher values on the y-axes of Kaplan-Meier plots for these characteristics). Inter-country differences in mortality for these high-risk birth characteristics accounted for an excess of 106 child deaths per year for birth weight of 500-1499g, 134 excess deaths at gestational age of 24-31 weeks and 232 excess deaths in children with congenital anomalies.

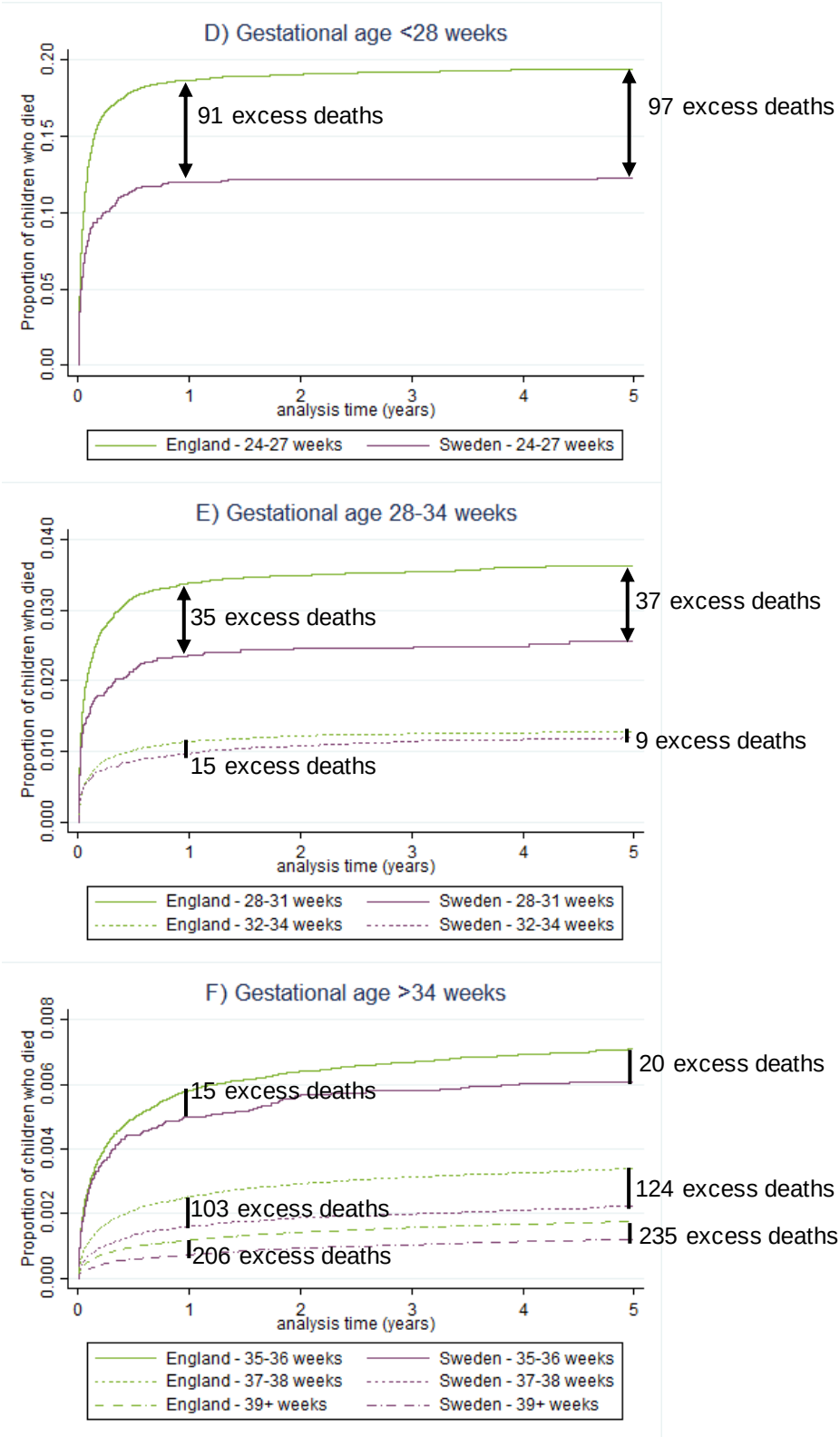
The highest numbers of excess deaths per birth characteristic were attributable to categories covering the largest numbers of births, typically associated with low risk of child mortality (as detailed in Chapter 1): normal birth weight (≥2500g, 249 excess deaths per year) and gestational age of ≥39 weeks (235 excess deaths per year). Boys accounted for 30% more excess deaths than girls.

Figure 5.2 – Kaplan-Meier failure curves comparing mortality at 2 days-4 years in England and Sweden overall and by birth characteristics



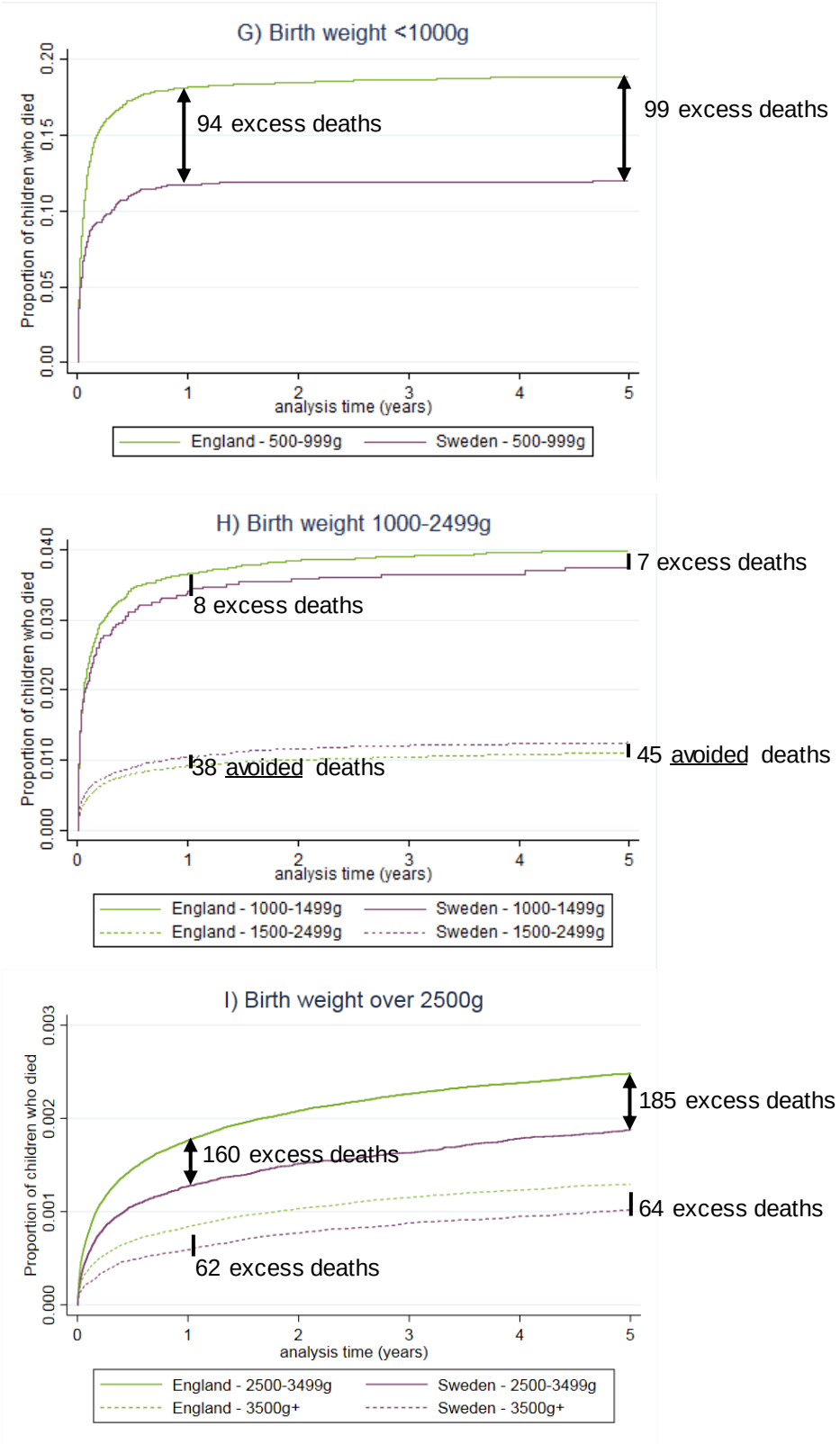
Each plot presents Kaplan-Meier failure curves for England (green) and Sweden (purple). Note that the y-axis differs between the plots. Excess deaths at age 1 and age 5 were calculated by multiplying the difference in proportion of children who died by the first and fifth birthday in England and Sweden by the number of births in the English cohort based on all births (before exclusions, n=6,100,404), and dividing by ten to get average values per year. Figure continues overleaf.

Figure 5.2 (continued) – Kaplan-Meier failure curves comparing mortality at 2 days-4 years in England and Sweden overall and by selected risk factors at birth.



Each plot presents Kaplan-Meier failure curves for England (green) and Sweden (purple). Note that the y-axis differs between the plots. Excess deaths at age 1 and age 5 were calculated by multiplying the difference in proportion of children who died by the first and fifth birthday in England and Sweden by the number of births in the English cohort based on all births (before exclusions, n=6,100,404), and dividing by ten to get average values per year. Figure continues overleaf.

Figure 5.2 (continued) – Kaplan-Meier failure curves comparing mortality at 2 days-4 years in England and Sweden overall and by selected risk factors at birth.

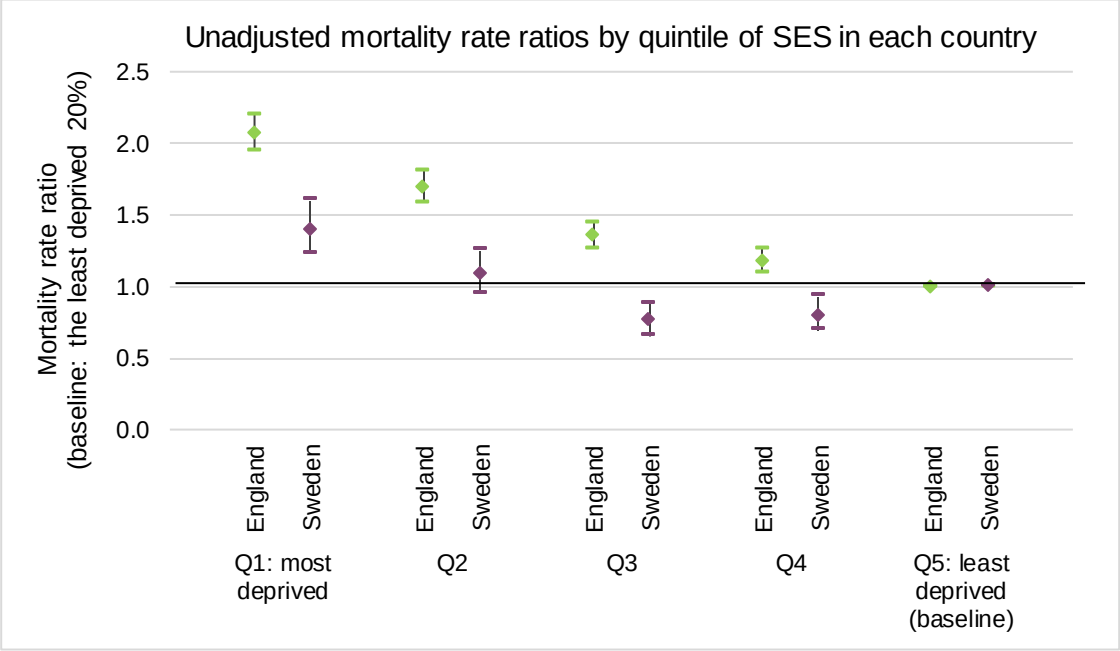


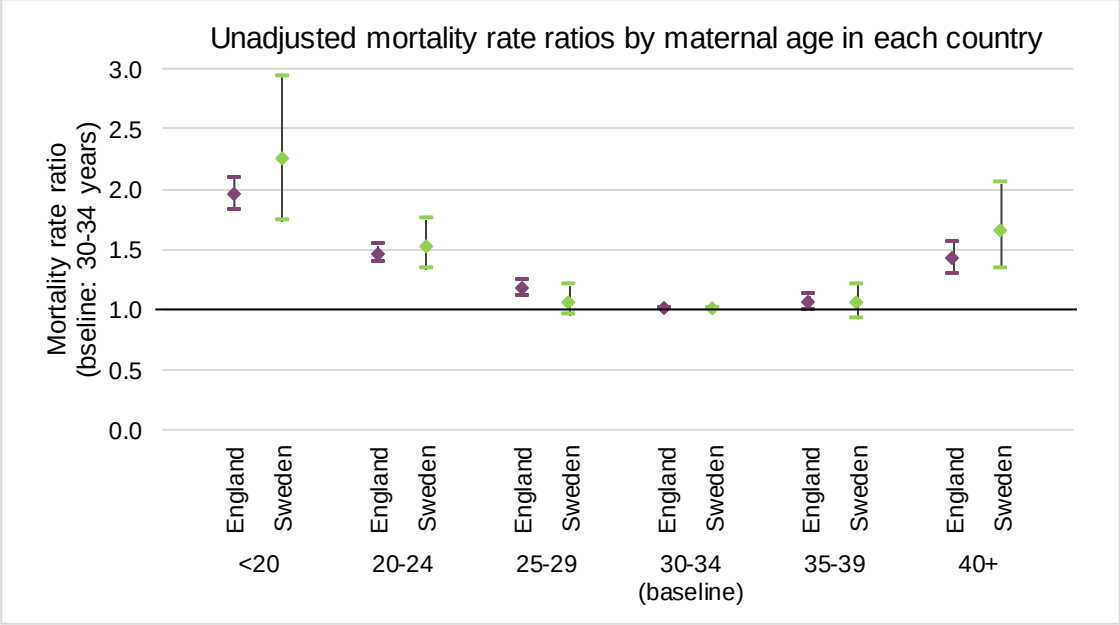
Each plot presents Kaplan-Meier failure curves for England (green) and Sweden (purple). Note that the y-axis differs between the plots. Excess deaths at age 1 and age 5 were calculated by multiplying the difference in proportion of children who died by the first and fifth birthday in England and Sweden by the number of births in the English cohort based on all births (before exclusions, n=6,100,404), and dividing by ten to get average values per year.

The differences in mortality rate ratios according to quintile of SES relative to the least deprived 20% of the population were greater in England than in Sweden (Figure 5.3 A). In England, the most deprived 20% of children were twice as likely to die as the least deprived 20%, and there was a clear gradient by quintile of SES. In Sweden, the most deprived 20% of the population were 50% more likely to die than the least deprived 20%. Children in the third and fourth quintiles had lower mortality rates than the least deprived 20%. The measures of SES used in England (an area-level indicator) and in Sweden (an individual-level indicator) were not directly comparable (as discussed in Chapter 4) and the area-level indicator of SES in England was likely to underestimate the true differences between individuals in the quintiles.

Mortality rate ratios for children grouped by maternal age showed similar patterns in the two countries (Figure 5.3 B). However, unadjusted child mortality rates per category of maternal age were higher in England than in Sweden (see Appendix E, Table E.2).

Figure 5.3 – Comparison of social inequalities in mortality at 2 days – 4 years in England and Sweden in 2003-2012





SES=socio-economic status. The plots present unadjusted mortality rate ratios for England (green) and Sweden (purple) by quintile of SES (baseline: Q5, the least deprived 20%) and by maternal age (baseline: 30-34 years).

5.4.4 Comparison of mortality adjusted for birth characteristics and socio-economic factors

5.4.4.1 Neonatal mortality (2-27 days)

The unadjusted HR for England relative to Sweden at 2-27 days was 1.66 (95% confidence interval (CI): 1.53, 1.81; Table 5.3). After adjusting for birth weight, gestational age and sex, the HR for England decreased to 1.37 (1.26, 1.48). Further adjustment for congenital anomalies and socio-economic factors reduced the HR to 1.15 (1.06, 1.25) and to 1.13 (1.04, 1.23), respectively. Between-country differences in the distribution of birth characteristics explained 77% of the excess risk of death in England relative to Sweden (with congenital anomalies independently accounting for 33%). A further 3% was explained by socio-economic factors, over and above their effect on birth characteristics. The risk of death was highest for babies with a birth weight of <1500g, gestation <28 weeks and one or more congenital anomaly.

Table 5.3 – Unadjusted and adjusted Cox PH models for all-cause mortality at 2-27 days in England relative to Sweden in 2003-2012

	Model 1	Model 2	Model 3	Model 4
Country				
England	1.66 (1.53, 1.81)	1.37 (1.26, 1.48)	1.15 (1.06, 1.25)	1.13 (1.04, 1.23)
Sweden (baseline)	1	1	1	1
Birth weight (g)				
500-999		31.3 (24.8, 39.5)	16.4 (13.0, 20.6)	15.5 (12.3, 19.6)
1000-1499		11.9 (9.6, 14.7)	7.7 (6.2, 9.5)	7.3 (5.9, 9.1)
1500-2499		6.0 (5.3, 6.9)	5.2 (4.6, 6.0)	5.0 (4.4, 5.7)
2500-3499		1.82 (1.66, 2.00)	1.81 (1.65, 1.98)	1.77 (1.61, 1.94)
≥3500 (baseline)		1	1	1
Gestational age (weeks)				
24-27		15.4 (12.3, 19.2)	7.6 (6.1, 9.5)	7.8 (6.2, 9.7)
28-31		5.5 (4.5, 6.7)	3.89 (3.21, 4.72)	3.95 (3.26, 4.80)
32-34		3.39 (2.91, 3.94)	2.91 (2.50, 3.39)	2.95 (2.54, 3.43)
35-36		2.70 (2.38, 3.07)	2.45 (2.16, 2.78)	2.46 (2.17, 2.80)
37-38		1.62 (1.48, 1.76)	1.53 (1.40, 1.67)	1.53 (1.40, 1.68)
≥39 (baseline)		1	1	1
Sex				
Boy		1.27 (1.20, 1.35)	1.19 (1.13, 1.26)	1.19 (1.13, 1.26)
Girl (baseline)		1	1	1
Congenital anomalies				
Yes			7.2 (6.7, 7.7)	7.1 (6.6, 7.7)
No			1	1
Maternal age (years)				
<20				1.23 (1.09, 1.38)
20-24				1.14 (1.04, 1.24)
25-29				1.10 (1.02, 1.19)
30-34 (baseline)				1
35-39				1.06 (0.97, 1.16)
≥40				1.32 (1.15, 1.52)
Quintile of socio-economic status				
Q1: most deprived				1.24 (1.13, 1.36)
Q2				1.11 (1.01, 1.22)
Q3				1.00 (0.91, 1.11)
Q4				0.94 (0.85, 1.04)
Q5: least deprived (baseline)				1

PH=proportional hazards. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model.

5.4.4.2 Post-neonatal mortality (28-364 days)

At 28-364 days, the unadjusted HR for England relative to Sweden was 1.59 (1.47, 1.71, Table 5.4). The HR declined to 1.32 (1.22, 1.42) after adjusting for birth weight, gestational age and sex, to 1.19 (1.10, 1.28) after further adjustment for congenital anomalies, and finally to 1.12 (1.04, 1.21) after adjusting for socio-economic factors. Between-country differences in the distribution of birth characteristics accounted for 68% of the excess risk of death in England relative to Sweden; socio-economic factors independently explained a further 11%. Children with a congenital anomaly and birth weight <1500g had the highest risk of death at 28-364 days.

Table 5.4 – Unadjusted and adjusted Cox PH models for all-cause mortality at 28-364 days in England relative to Sweden in 2003-2012

	Model 1	Model 2	Model 3	Model 4
Country				
England	1.59 (1.47, 1.71)	1.32 (1.22, 1.42)	1.19 (1.10, 1.28)	1.12 (1.04, 1.21)
Sweden (baseline)	1	1	1	1
Birth weight (g)				
500-1499		27.3 (22.8, 32.8)	11.7 (9.7, 14.0)	10.4 (8.7, 12.6)
1500-2499		7.0 (6.2, 7.8)	5.3 (4.7, 5.9)	4.7 (4.3, 5.3)
2500-3499		2.00 (1.85, 2.15)	1.94 (1.80, 2.09)	1.84 (1.71, 1.98)
≥3500 (baseline)		1	1	1
Gestational age (weeks)				
24-31		3.09 (2.60, 3.67)	1.58 (1.33, 1.89)	1.65 (1.38, 1.97)
32-34		1.63 (1.42, 1.88)	1.33 (1.15, 1.52)	1.38 (1.20, 1.59)
35-36		1.92 (1.72, 2.14)	1.63 (1.46, 1.82)	1.66 (1.49, 1.85)
37-38		1.53 (1.43, 1.64)	1.39 (1.30, 1.50)	1.41 (1.32, 1.52)
≥39 (baseline)		1	1	1
Sex				
Boy		1.33 (1.26, 1.40)	1.18 (1.12, 1.24)	1.17 (1.11, 1.24)
Girl (baseline)		1	1	1
Congenital anomalies				
Yes			15.4 (14.5, 16.3)	15.2 (14.4, 16.2)
No			1	1
Maternal age (years)				
<20				1.72 (1.56, 1.90)
20-24				1.32 (1.22, 1.42)
25-29				1.10 (1.02, 1.19)
30-34 (baseline)				1
35-39				0.99 (0.90, 1.08)
≥40				1.20 (1.05, 1.38)
Quintile of socio-economic status				
Q1: most deprived				1.66 (1.52, 1.81)
Q2				1.49 (1.36, 1.64)
Q3				1.14 (1.04, 1.26)
Q4				1.12 (1.01, 1.24)
Q5: least deprived (baseline)				1

PH=proportional hazards. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Some birth weight and gestational age categories were merged due to small numbers.

5.4.4.3 Early childhood (1-4 years)

At 1-4 years, the unadjusted HR for England relative to Sweden was 1.27 (1.15, 1.40, Table 5.5). The HR declined to 1.10 (1.00, 1.22) after adjusting for all birth characteristics. After a further adjustment for socio-economic factors, there were no statistically significant differences in child mortality between England and Sweden (1.06, 95% CI: 0.96, 1.18). Congenital anomalies were the single most important risk factor for deaths at 1-4 years, increasing the risk of death 17 times, followed by low-birth weight (<2500g) with a four-fold increase.

Table 5.5 – Unadjusted and adjusted Cox PH models for all-cause mortality at 1-4 years in England relative to Sweden in 2003-2012

	Model 1	Model 2	Model 3	Model 4
Country				
England	1.27 (1.15, 1.40)	1.14 (1.03, 1.26)	1.10 (1.00, 1.22)	1.06 (0.96, 1.18)
Sweden (baseline)	1	1	1	1
Birth weight (g)				
500-1499		10.7 (8.3, 13.7)	3.04 (2.36, 3.93)	2.86 (2.22, 3.69)
under 1500		4.3 (3.63, 5.0)	3.12 (2.67, 3.66)	2.92 (2.49, 3.43)
1500-2499		1.53 (1.40, 1.67)	1.48 (1.36, 1.62)	1.43 (1.31, 1.57)
≥3500 (baseline)		1	1	1
Gestational age (weeks)				
<37		1.05 (0.89, 1.23)	0.84 (0.71, 0.99)	0.86 (0.73, 1.02)
37-38		1.18 (1.07, 1.29)	1.07 (0.97, 1.18)	1.08 (0.98, 1.20)
≥39 (baseline)		1	1	1
Sex				
gender		1.18 (1.10, 1.28)	1.05 (0.97, 1.13)	1.04 (0.97, 1.13)
Girl (baseline)		1	1	1
Congenital anomalies				
Yes			17.1 (15.8, 18.6)	17.1 (15.7, 18.6)
No			1	1
Maternal age (years)				
<20				1.26 (1.07, 1.47)
20-24				1.25 (1.12, 1.40)
25-29				1.01 (0.91, 1.12)
30-34 (baseline)				1
35-39				0.92 (0.81, 1.05)
≥40				0.92 (0.74, 1.15)
Quintile of socio-economic status				
Q1: most deprived				1.39 (1.23, 1.58)
Q2				1.29 (1.13, 1.47)
Q3				1.20 (1.05, 1.37)
Q4				1.10 (0.96, 1.26)
Q5: least deprived (baseline)				1

PH=Proportional hazards. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Some birth weight and gestational age categories were merged due to small numbers.

Table 5.6 – Summary of differences in child mortality between England and Sweden in 2003-2012 attributable to birth characteristics and socio-economic factors at 2-27 days, 28-364 days and 1-4 years.

Variables included in the model:	2-27 days		28-364 days		1-4 years	
	HR	PERM	HR	PERM	HR	PERM
Country	1.66 (1.53, 1.81)	-	1.59 (1.47, 1.71)	-	1.27 (1.15, 1.40)	-
Country + birth weight, gestational age and sex	1.37 (1.26, 1.48)	44%	1.32 (1.22, 1.42)	46%	1.14 (1.03, 1.26)	N/A
Country + birth weight, gestational age and sex +congenital anomalies	1.15 (1.06, 1.25)	33%	1.19 (1.10, 1.28)	22%	1.10 (1.00, 1.22)	N/A
Country + birth weight, gestational age and sex +congenital anomalies +socio-economic factors (maternal age and SES)	1.13 (1.04, 1.23)	3%	1.12 (1.04, 1.21)	11%	1.06 (0.96, 1.18)	N/A

HR=hazard ratio for England relative to Sweden (the baseline); PERM=percentage excess risk mediated; N/A=not applicable (PERM was calculated only if the HR for England versus Sweden in the fully adjusted model remained statistically significant). Data are adjusted HRs (95% confidence interval) and PERM, based on models presented in tables 5.3-5.5.

5.4.5 Subgroup analyses

In the English cohort, low-risk babies accounted for 69% of births (2,710,895 births), 27% of deaths at 2-27 days (1,119 deaths), 26% of deaths at 28-364 days (1,298 deaths) and 39% of deaths at 1-4 years (867 deaths). In the Swedish cohort, the corresponding numbers were 6% (687,946), 27% (173), 27% (220) and 49% (233), respectively.

Mortality in low-risk babies in the first year of life was substantially higher in England relative to Sweden (unadjusted HRs were 1.87 (1.51, 2.32) at 2-27 days, 1.48 (1.28, 1.70) at 28-364 days, Table 5.7). These differences only partly reduced after adjusting for birth characteristics and socio-economic factors. The fully adjusted HRs were 1.64 (1.32, 2.03) at 2-27 days and 1.19 (1.03, 1.38) at 28-364 days. Birth characteristics accounted for only 19% and 29% of the excess risk of death in England relative to Sweden at 2-27 days and 28-364 days, respectively. Socio-economic factors independently explained a further 8% and 31%, respectively.

The differences in mortality beyond the first year of life were negligible in this low-risk group; unadjusted HR was 1.00 (0.86, 1.15). Full results for all models for low-risk children are shown in Appendix E, Tables E.3-E.5.

Table 5.7 – Differences in child mortality between England and Sweden attributable to birth characteristics and socio-economic factors at 2-27 days, 28-364 days and 1-4 years for low-risk babies

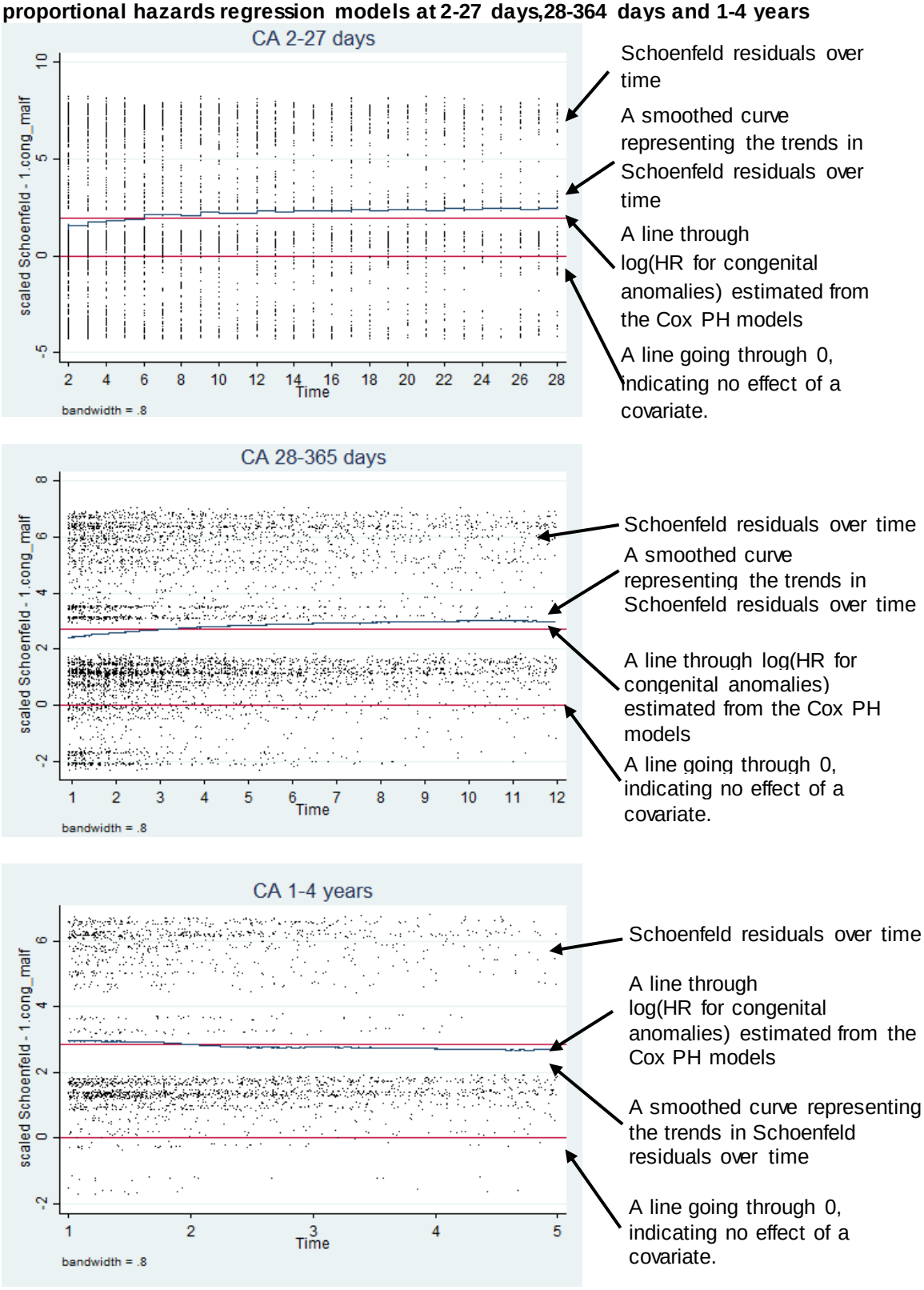
	2-27 days		28-364 days		1-4 years	
	England	**Sweden**	**England**	**Sweden**	**England**	**Sweden**
Number of deaths (% of all deaths)	1,119 (26.6%)	173 (26.7%)	1,298 (26.2%)	220 (27.4%)	867 (39.0%)	233 (49.0%)
Variables included in the model:	**HR**	**PERM**	**HR**	**PERM**	**HR**	**PERM**
Country	1.87 (1.51, 2.32)	-	1.48 (1.28, 1.70)	-	1.00 (0.86, 1.15)	-
Country + gestational week, birth weight category (by 500g), and sex	1.71 (1.38, 2.12)	19%	1.34 (1.16, 1.55)	29%	0.94 (0.81, 1.09)	N/A
Country + gestational week, birth weight category (by 500g), and sex + maternal age and SES	1.64 (1.32, 2.03)	8%	1.19 (1.03, 1.38)	31%	0.91 (0.78, 1.05)	N/A

HR=hazard ratio for England relative to Sweden (baseline); PERM=percentage excess risk mediated; N/A=not applicable (PERM was calculated only if the HR for England versus Sweden in the fully adjusted model remained statistically significant). Low-risk babies were defined as those born at full term (39-41 weeks), with normal birth weight (>2500g), and no congenital anomaly. Data are adjusted HRs (95% confidence interval) and PERM, based on models presented in Appendix E, Tables E.3-E.5.

5.4.6 Sensitivity analyses

The results did not change appreciably in sensitivity analyses using an indicator for severe congenital anomaly (see Appendix E, Tables E.6-E.8). Grambsch and Therneau test of PH assumption indicated that all models violated the PH assumption, and this effect was driven by the indicator of congenital anomaly. Smoothed lines representing trends in Schoenfeld residuals for the indicator of congenital anomaly over time showed that the HRs increased over time at 2-27 days and 28-364 days and decreased at 1-4 years (Figure 5.4). However, including an effect modification term with age for congenital anomaly gave near identical results (see Tables E.9-E.11 in Appendix E).

Figure 5.4 – Schoenfeld residual plots for an indicator of congenital anomaly from Cox proportional hazards regression models at 2-27 days,28-364 days and 1-4 years



CA=congenital anomalies; HR=hazard ratio; PH=proportional hazards. Each plot shows Schoenfeld residual plots for an indicator of CA from the fully adjusted Cox PH model at 2-27 days, 28-364 days and 1-4 years. A smoothed line representing the trends in residuals over time has been superimposed to aid interpretation (blue). If the PH assumption holds, the smoothed line should be horizontal and around the coefficient for the indicator of congenital anomalies estimated from the Cox PH models (log(HR), red line). I also superimposed a horizontal line going through 0, indicating no effect of a covariate.

5.5 Discussion

5.5.1 Key findings

Child mortality was substantially higher at all ages in England relative to Sweden. This difference was largely explained by inter-country differences in distribution of birth characteristics and socio-economic factors.

The less favourable distribution of birth characteristics in England (i.e. a higher prevalence of preterm birth, low birth weight and congenital anomalies) accounted for 77% and 68% of the excess risk of death at 2-27 days and 28-364 days, respectively. Socio-economic factors accounted for a further 3% and 11% of the excess risk of death in these two age groups, respectively. After adjusting for birth characteristics and socio-economic factors, the risk of death in England relative to Sweden remained 13% higher at 2-27 days and 12% higher at 28-364 days.

Mortality for low-risk babies in infancy was also substantially higher in England than in Sweden. Birth characteristics and socio-economic factors explained only 29% of the excess risk of death at 2-27, and the risk of death remained 67% higher in England relative to Sweden after adjustment for all risk factors. At 28-364 days, birth characteristics and socio-economic factors each accounted for approximately 30% of the observed excess risk of death in England.

The differences between Sweden and England in child mortality beyond the first year of life were negligible in the fully adjusted model. For low-risk babies there were also no differences in mortality in early childhood.

5.5.2 Strengths

The datasets available for this comparison were the biggest strength of this study. I used individual-level data from nationally-representative birth cohorts, with detailed information about characteristics of children and mothers at birth. Information from hospital admissions records coded using internationally-standardised coding systems enabled me to develop a congenital anomaly indicator, which no previous international comparison of child mortality has used. The large sample sizes and long follow-up periods allowed me to investigate the effect of rare risk factors, such as congenital anomalies or extreme prematurity. Furthermore, the results were robust to all sensitivity analyses. Thus, this study can serve as an example for future comparisons of child health outcomes in countries with administrative linked datasets.

5.5.3 Limitations

5.5.3.1 Missing data in the English cohort

Due to high rates of missing data on key birth characteristics in the English birth cohort, I had to exclude one third of births in England from the analyses. As mentioned in Chapter 3, a national birth cohort with near 100% completeness of risk factors at birth, high quality of linkage to mortality data and to hospital admission trajectories for mothers and babies could be developed by linking ONS birth registration, National Health Service (NHS) birth notification data and HES records for mothers and babies.[167,178] Researchers from the City University of London showed that such linkage is possible; however, accessing the data requires seeking further permissions (such as application to the Confidentiality Advisory Group),[133] and the linkage is not routinely updated. Nonetheless, the sub-cohort of hospitals with good quality of recorded data used in this chapter for England was thoroughly validated and representative for the population of children in England and Wales.

Due to incomplete recording of birth characteristics I also excluded deaths on days 0-1 of life, which accounted for one quarter of child deaths in England and Sweden. Further research using a more complete birth dataset in England is required to examine inter-country differences in these early deaths. Such a comparison would require data on both stillbirths and live births to allow for between-country differences in definitions and mortality registration practices. Such a comparison, based on total births, would also minimise the 'live birth' bias, which arises when the same prenatal exposures are associated with the outcome of interest and the risk of foetal death.[200]

5.5.3.2 Lack of information about additional maternal risk factors

Measures of SES were not directly comparable and showed a different association with the risk of child death in the two countries. I used an area-level measure in England and individual-level measure in Sweden. Maternal education level would be the most comparable SES indicator for inter-country comparisons,[192] but such a variable is not available in any dataset in England. A comparison based on parental occupation could be conducted by linking HES to ONS birth registration data, as comparable information is available in the Swedish National Registers. This variable is collected for only 10% of the population,[128] but combining births in 2003-2012 would provide a large sample size of approximately 600,000 births.

This study would have benefitted from including additional information on maternal risk factors during pregnancy such as, smoking or body mass index (BMI), which were available in the Swedish Medical Birth Register (SMBR) but not in HES. Both maternal smoking and obesity are associated with an increased risk of low birth weight,[74]

preterm birth,[63,74] and some major congenital anomalies.[62,74] The prevalence of these factors was much higher in England than in Sweden: in 2010, 12.0% of mothers in England smoked during pregnancy,[201] compared to only 6.5% of mothers who smoked in the first trimester in Sweden.[12] One in eight Swedish mothers were obese (BMI ≥30), compared to one in five in England (based on all females aged 16-44 years).[66] Information about these risk factors would help to determine which preventive strategies to address adverse exposures during pregnancy might be most effective for reducing child mortality in England relative to Sweden.

5.5.4 Interpretation and further work

The differences in child mortality rates in England relative to Sweden were largely driven by the differences in distribution of birth characteristics in the two countries – after adjustment for birth characteristics the HR for England relative to Sweden reduced from 1.66 to 1.15 at 2-27 days, from 1.59 to 1.19 at 28-364 days and from 1.27 to 1.10 at 1-4 years. The prevalence of adverse birth characteristics such as low birth weight, preterm birth and congenital anomalies was higher in England than in Sweden, and a higher proportion of children born in Sweden weighed ≥3500g at birth, a range associated with the lowest risk of death in infancy.[41] As discussed in Chapter 1, healthy development *in utero* is strongly associated with the circumstances of the mother during pregnancy. For example, maternal smoking, being obese or underweight, and young or old age are associated with an increased risk of low birth weight,[41,202] preterm birth,[28,202] and some major congenital anomalies.[60,62,74] Therefore, policies to reduce child mortality in England need to focus on improving maternal health and well-being before and during pregnancy.

Further research is needed to assess which interventions would be most effective in reducing adverse birth characteristics. Most previously tested interventions for prematurity and intrauterine growth restriction, such as nutritional supplementation or treatment of infections during pregnancy, did not show consistent benefits.[41,203] Smoking cessation programs for pregnant women have been effective at increasing mean birth weight and reducing the rate of low birth weight.[204] Reducing other maternal risk factors, like teenage pregnancy or obesity, require multi-agency responses including sectors outside healthcare and welfare policy. Not all risk factors can be modified. For example, some of the increased mortality due to congenital anomalies in England might reflect lower rates of terminations of pregnancy (TOP) for chromosomal anomalies than in Sweden,[48] possibly due to differences in cultural attitudes to TOP or in timing of detection of the anomaly. High rates of consanguineous marriages amongst couples of Pakistani origin in England could also contribute to differences in the prevalence of congenital anomalies in the two countries.[60]

Reductions in child mortality rates in England will require addressing socio-economic factors. Social disadvantage is associated with an increased prevalence of preterm birth,[14,93] low birth weight[14,93] and congenital anomalies,[13] and these risk factors were more common among the most deprived 20% of mothers in the English birth cohort. Social gradients are also observed in maternal risk factors associated with higher rates of adverse birth characteristics such as maternal obesity,[66] young maternal age,[56] bacterial vaginosis, alcohol and drug use, and smoking.[14] Further comparisons using causal mediation methods could determine the total effect of socio-economic disadvantage on the risk of death in England, including both the effect mediated by low birth weight, preterm birth and congenital anomalies, and the direct effect.

Socio-economic disadvantage also determines the types of risks the child is exposed to after birth. In this study, socio-economic factors explained a further 11% of the excess risk of death overall and 31% of excess risk in low-risk babies at 28-364 days. This could reflect inter-country differences in mortality from socially-patterned causes, such as infections or SIDS.[95] These causes are investigated further in the next chapter. At 2-27 days, when mortality is strongly influenced by the quality of healthcare, socio-economic factors accounted for only 3% of excess risk of death in all children, and 8% in low-risk babies (independently of the effect of SES on birth characteristics). These results indicate that children from all socio-economic backgrounds receive the same level of care in the NHS.

Low-risk babies accounted for over a quarter of infant deaths in England and Sweden. The risk of death at 2-27 days remained over 60% higher in England than in Sweden after full adjustment for available risk factors, with birth characteristics explaining only 19% of excess risk of death. Some of the unexplained differences could reflect variation in the provision of obstetric and neonatal care. However, it is more likely that some of included infants suffered from other neonatal morbidity such as birth asphyxia, which were not indicated by birth weight, gestational age or presence of congenital anomalies. Information about APGAR score could help to validate this hypothesis; however, such information was only available in SMBR and not in HES. Maternal hospital admissions trajectory could be used to identify low-risk babies from uncomplicated pregnancies, for a more fair comparison of mortality in children with no underlying health conditions at birth.

In the post-neonatal period, the risk of death in low-risk babies remained 19% higher in England than in Sweden after full adjustment for all risk factors. Some of these differences could be due to differences in care in the home setting, such as exposure to tobacco smoke, breastfeeding rates or safe sleeping practices, which were not accounted for (and were unlikely to be confounded by the area-level SES indicator

available in the English data). Low risk babies comprised almost 40% and 50% of deaths at 1-4 years in England and Sweden, and there were no differences in mortality for these children.

5.5.5 Policy implications

The biggest reductions in child mortality in England relative to Sweden could be achieved by reducing the prevalence of adverse birth characteristics. Policies to reduce child mortality in England should focus on universal strategies to improve the health of women before and during pregnancy, and on reducing socio-economic disadvantage.

5.5.6 Implications for this thesis

Some of the excess child mortality in England relative to Sweden remained unexplained in the fully adjusted models and could reflect differences in the care received after birth, given birth characteristics. Thus, in the next chapter I compared mortality from causes which could be amenable to risk factors operating after birth, namely respiratory tract infections (RTI), which are amenable to healthcare through vaccination and antibiotics treatment,[18] and sudden unexpected deaths in infancy (SUDI), which are amenable to public health interventions, such as advice on safe sleeping practices or smoking cessation programs.[89]

Chapter 6. Comparison of mortality from preventable causes in England and Sweden

What is already known:

- Differences in the distribution of birth characteristics accounted for 77% and 68% of excess risk of death in England relative to Sweden at 2-27 days and 28-364 days, respectively.
- Socio-economic factors contributed a further 11% (over and above effect on birth characteristics) to the gap in mortality at 28-364 days.
- The risk of death in first year of life remained 12-13% higher in England relative to Sweden after adjustment for birth characteristics and socio-economic factors.
- The differences in mortality beyond infancy, however, were not statistically significant after adjusting for all risk factors.

What this chapter adds:

- In this chapter I look at mortality from two causes potentially preventable through public health or healthcare interventions to determine the contribution of modifiable factors after birth:
    - Respiratory tract infection (RTI)-related deaths, which are amenable to healthcare,
    - Sudden unexpected deaths in infancy (SUDI) deaths, which are amenable to public health interventions.

6.1 Chapter overview

In the previous chapter, I compared all-cause mortality in England and Sweden to determine the contribution of birth characteristics and socio-economic factors to the excess mortality in England. Small but statistically significant differences in child mortality remained after accounting for all risk factors, and could reflect differences in the quality of care received after birth. A comparison of cause-specific mortality adjusted for birth characteristics can help to determine the contribution of specific risk factors in the quality of care after birth to the excess mortality in England relative to Sweden, given a child's characteristics at birth.

This chapter presents work towards objective 4: "*to compare the risk of child mortality from causes which could be potentially preventable by improving the quality of care received after birth"*. I compare mortality from two causes. First, I look at respiratory tract infection (RTI)-related mortality, since RTIs are amenable to healthcare through vaccination (particularly of high-risk groups) and prompt antibiotic treatment.[18] Second, I compare mortality due to sudden unexpected deaths in infancy (SUDI), amenable to public health interventions such as advice on safe sleeping practices or smoking cessation programs.[89] This comparison uses birth cohorts from England and Sweden presented in Chapters 3 and 4 of this thesis.

6.2 Background

Increased child mortality in England relative to Sweden in 2003-2012 can be largely explained by England's high prevalence of adverse birth characteristics, and to a lesser extent, by socio-economic factors (independent of their effect on birth characteristics). Thus, the largest reductions in child mortality can be achieved by improving maternal health before and during pregnancy. Small, but statistically significant differences in mortality in the first year of life remained unexplained.

Unexplained excess mortality in England relative to Sweden after adjustment for all risk factors could reflect differences in the care that children receive after birth, given their birth characteristics. This care could be either by healthcare providers, or by parents. A comparison of child mortality from potentially preventable causes could, therefore, indicate the contribution of modifiable risk factors operating after birth to the excess mortality in England relative to Sweden.

6.2.1 Healthcare amenable mortality

Infections are the third most common cause of child deaths both in the United Kingdom (UK) and in Sweden (as detailed in Chapter 1).[6] In 2006-08, infection-related mortality rate was 50% higher at 28 days-4 years in the UK compared to Sweden (36/100,000 births compared to 23/100,000 births).[6]

Previous comparisons attributed these differences to delays in the diagnosis of acute life-threatening infections.[6] Child health professionals have called for changes in the provision of healthcare to reduce infection-related mortality in the UK. For example, better integration of primary care and paediatric services and mandatory paediatrics training for general practitioners (GPs) would enable early response to serious childhood infections.[8–10]

As outlined in Chapter 1, previous comparisons of child mortality between the UK and Sweden were limited by the use of data tabulated by the underlying cause of death. They have not accounted for differences in the distribution of birth characteristics

between the two countries. Birth characteristics such as low birth weight or prematurity, however, can increase an infant's susceptibility to infection. For example, preterm babies have increased risk of infection-related deaths[21] and hospital admissions due to respiratory syncytial virus (RSV),[20] while babies with extremely low birth weight (<1000g) are more likely to be readmitted to hospital after birth due to respiratory illnesses (including lower-RTIs).[20] A comparison between England and Sweden adjusted for birth characteristics is needed to determine whether differences in infection-related mortality reflect higher rates of adverse health outcomes or failure of services to respond adequately to prevent or treat infections.

In this chapter, I focus on mortality from RTIs, which are one of the most common causes of emergency hospital admissions in children in England, thus they can be reliably captured in hospital admissions datasets.[205,206] RTIs account for almost a quarter of non-injury deaths at 28 days-4 years in England and Wales (22% in 2001-2010)[18], despite being amenable to healthcare through vaccination (e.g., for pneumococcal infection, Haemophilus influenza type b (Hib), pertussis and influenza) and prompt antibiotic treatment in primary and secondary care.

6.2.2 Mortality amenable to public health interventions

Sudden Infant Death Syndrome (SIDS) is an example of a cause of death amenable to preventive public health measures. SIDS is defined as deaths under the age of one, which occur suddenly and unexpectedly, for which no cause of death can be identified following an autopsy, a death scene investigation, and a review of clinical history.[89,90] Rates of SIDS have declined in many countries since the introduction of public health campaigns in the 1990s, which recommended that infants be put to sleep on their backs ('Back to sleep' campaigns).[95]

Mortality from SIDS has reduced in England, but remained higher in the most deprived families, indicating that there is scope for further reductions.[91] A cohort study from England (covering Bristol, Bath and surrounding areas, with approximately 10,000 live births per year) showed that in 1984-1988 approximately half of SIDS cases (72 out of 153) occurred in the most disadvantaged families, compared to 75% of SIDS deaths in 1999-2003 (25 out of 34 cases overall).[95] Risk factors for SIDS such as unsafe sleeping practices and parental smoking[89,91] could be contributing to these remaining differences and could be addressed by public health interventions.[107]

Previous comparisons of SIDS rates in the UK and Sweden found no statistically significant differences between the two countries (27/100,000 live births in the UK compared to 26/100,000 live births in Sweden).[6] However, SIDS is certified as the cause of death through an exclusion of other causes following a death scene investigation, an autopsy and a review of clinical history.[89,90] Therefore, inter-country

differences in investigative practices and autopsy protocols could lead to a variation in International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) codes used to denote SIDS.[90]

To ensure comparability, in this chapter I compared rates of sudden unexpected infant deaths (SUDI) rather than SIDS. SUDI covers all unexpected deaths in infancy, including deaths from unexplained causes (such as SIDS, or "other ill-defined and unspecified causes of mortality"),[24] and from explained causes (such as accidental suffocation and strangulation in bed).[90] I used a definition of SUDI recommended for international comparisons which covers seven ICD-10 codes which are likely to account for inter-country differences in investigative and diagnostic practices.[90]

6.2.3 Chapter aims

The aim of this chapter is to compare child mortality in England and Sweden due to causes of death which are amenable to healthcare intervention (RTI-related deaths) or preventable through public health interventions (SUDI). I used the birth cohorts from England and Sweden described in Chapters 3 and 4. I determined to what extent the differences in mortality from these causes were explained by birth characteristics and socio-economic factors.

6.3 Methods

6.3.1 Study participants

Analyses in this chapter were based on nationally-representative, comparable birth cohorts developed in Chapters 3 and 4 using administrative linked datasets. The cohorts included 65% of all singleton live births in hospital in England, and 99.6% of all singleton live births in hospital in Sweden in 2003-2012. As explained in Chapter 3, 29% of records in the English cohort were removed due to high rates of missing data. I also excluded deaths before 31 days of age due to inter-country differences in recoding of causes of death in that period. Deaths at ≤27 days of life were certified using a standard death certificate in Sweden,[180] and using a neonatal death certificate in England.[22] Different coding rules apply to recording of causes of death for the two types of death certificates; in particular, an underlying cause of death cannot be identified using the neonatal death certificate.[24] Death registration practices were comparable after 27 days of age in the two countries. However, 75% of deaths at 28-30 days did not have any recorded causes of death in the English birth cohort, and were therefore excluded from the analyses. Finally, I excluded births with birth weight <500g, gestational age <24 weeks or missing data on any of the risk factors of interest.

6.3.2 Outcomes

The outcomes of interest in this study were RTI-related mortality at 31-364 days and 1-4 years, and SUDI mortality at 31-364 days. I did not investigate SUDI beyond first year of life since these deaths are the most common at 2-4 months of life, and become rare beyond infancy.[89] Children were followed up until their fifth birthday, death, or 31st December 2013, whichever occurred first.

6.3.2.1 RTI-related deaths

RTI-related deaths were defined using ICD-10 code list developed by Hardelid *et al.*, listed in Table 6.1.[18] RTI-related deaths were non-injury deaths, with a relevant ICD-10 code recoded as any cause of death or diagnosis from a hospital admission within 30 days before death (based on the hospital admission date). Excluded deaths from injury were defined by any code for the underlying cause of death taken from Chapter 20 of ICD-10, "External causes of morbidity and mortality".[24]

Table 6.1 – ICD-10 codes proposed by Hardelid *et al.*[18] to capture RTI-related deaths in hospital admission records

ICD-10 code	Definition
A15	Respiratory tuberculosis, bacteriologically and histologically confirmed
A16	Respiratory tuberculosis, not confirmed bacteriologically or histologically
A19	Miliary tuberculosis
B97.4	Respiratory syncytial virus as the cause of diseases classified to other chapters
A37	Whooping cough
J00-J06	Acute upper respiratory infections
J09-J18	Influenza and pneumonia
J20-22	Other acute lower respiratory infections

ICD-10=International Statistical Classification of Diseases and Related Health Problems 10th Revision; RTI=respiratory tract infection.

6.3.2.2 SUDI

I used a broad definition of SUDI, recommended for international comparisons to minimise bias from differences in coding and investigative practices.[90] A death was indicated as SUDI if any of the ICD-10 codes listed in Table 6.2 were used as the underlying cause of death.

Table 6.2 – ICD-10 codes proposed by Taylor *et al.*[90] to capture SUDI deaths.

ICD-10 code	Definition
R95	Sudden infant death syndrome
R96	Other sudden death, cause unknown
R98	Unattended death
R99	Other ill-defined and unspecified causes of mortality
W75	Accidental suffocation and strangulation in bed
W78	Inhalation of gastric contents
W79	Inhalation and ingestion of food causing obstruction of respiratory tract

ICD-10=International Statistical Classification of Diseases and Related Health Problems 10th Revision; SUDI=sudden unexpected deaths in infancy.

6.3.3 Risk factors

Birth characteristics included birth weight (categorised as 500-1499g, 1500-2499g, 2500-3499g and ≥3500g), gestational age (grouped as 24-34, 35-36, 37-38, ≥39 weeks), sex and presence of congenital anomalies. Socio-economic factors included maternal age (categorised as <20, 20-24, 25-29, 30-34, ≥35 years) and quintile of socio-economic status, measured using Index of Multiple Deprivation (IMD) scores in England and household income per family member a year before pregnancy in Sweden. Compared to analyses presented in Chapter 5, categories of birth weight, gestational age and maternal age had to be merged due to the small number of events.

6.3.4 Statistical analyses

6.3.4.1 Exploratory analyses

I derived the numbers and percentages of babies who survived to age 31 days of life and beyond in England and Sweden by each risk factor category. I compared the characteristics of RTI-related deaths at age 31-364 days and 1-4 years, and SUDI at 31-364 days in the two countries by deriving the numbers and percentages of deaths and the unadjusted mortality rates per 100,000 child years by each risk factor category.

6.3.4.2 Cox proportional hazards models

As discussed in Chapter 5, mortality is an example of a time-to-event outcome. Therefore, I fitted Cox proportional hazards (PH) regression models to estimate hazard ratios (HR) for cause-specific mortality in England relative to Sweden (the baseline country in the statistical models). For RTI-related deaths, I fitted separate models for mortality at 31-364 days and 1-4 years. As in Chapter 5, I first fitted unadjusted models including only a covariate for a country of birth. I then added birth characteristics and

socio-economic factors. For some of the models, risk factor categories had to be further merged due to the small number of events per category. For example, all RTI-related deaths in children of mothers aged <25 and RTI-related deaths at 1-4 years in preterm babies had to be grouped together.

6.3.4.3 Percentage of excess risk mediated (PERM)

For each cause of death, I quantified the contribution of inter-country differences in birth characteristics and socio-economic factors to the increased risk of death in England relative to Sweden using percentage excess risk mediated (PERM, described in Section 5.3.4 in Chapter 5). PERM was calculated only for models where the HR for England in the fully adjusted model remained statistically significant, i.e. where the *p*-value for the Wald test for the country parameter was <0.05.

6.3.4.4 Subgroup analyses

Adverse birth characteristics such as preterm birth or low birth weight can increase child's susceptibility to infection. For such children, it is difficult to determine whether RTI was the underlying cause of death, or a final complication associated with a poor health at birth. For a fairer comparison of mortality which could be attributed to the quality of care received after birth, I repeated all analyses on a subgroup of low-risk babies with no underlying perinatal risk factors which could increase their susceptibility to infection. I defined these babies as: born with birth weight ≥2500g, at term (37-41 weeks), with no congenital anomalies.

6.3.4.5 Sensitivity analyses

6.3.4.5.1 PH assumption

The underlying assumption of the Cox PH model is that the HRs for each covariate remain constant over time. The PH assumption was assessed using Schoenfeld residual plots for each covariate (described in more detail in Section 5.3.4.5 in Chapter 5).[199] If the PH assumption is met, these plots should show a straight line with a slope of 0. Where the assumption was violated, I fitted additional Cox PH regression models with an interaction between the given covariate and survival time.

6.3.4.5.2 Inter-country differences in coding of SUDI

I compared coding practices in England and Sweden by deriving the proportion of SUDI deaths identified using each ICD-10 code from Table 6.2. I repeated all analyses using standard ICD-10 code for SIDS ("R95"). I also repeated all analyses in low risk babies (defined in Section 6.3.4.4) to assess whether differences in coding of deaths due to e.g., congenital anomalies or consequences of prematurity, affect the results.

6.3.4.5.3 Sensitivity to the choice of congenital anomaly indicator

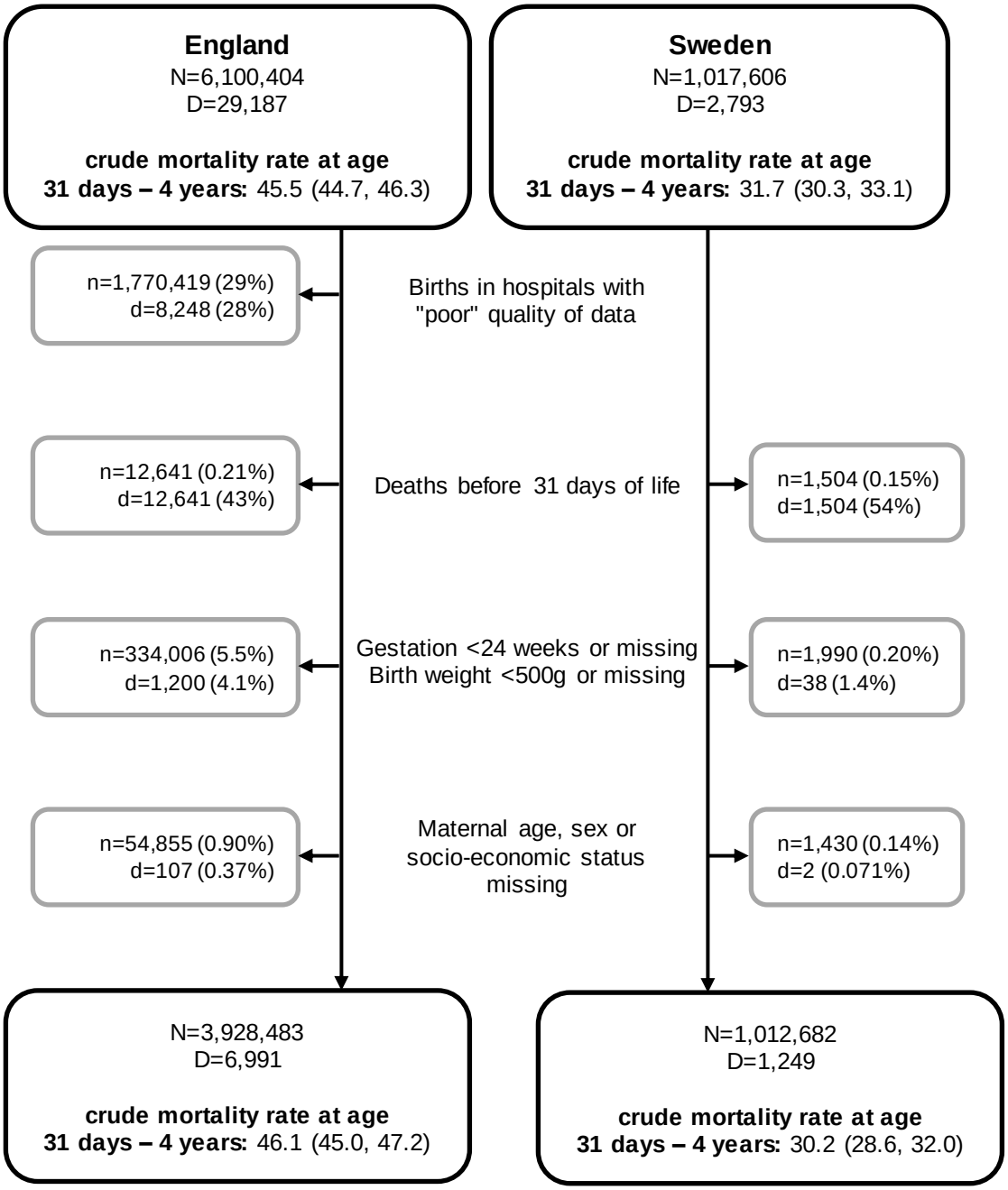
In Chapter 5, I conducted additional sensitivity analyses using a stricter definition for severe congenital anomalies. Since the choice of congenital anomaly indicator had only a marginal effect on the results in Chapter 5, I did not conduct additional sensitivity analyses using this indicator in this chapter.

6.4 Results

6.4.1 Comparable birth cohorts in England and Sweden

The study population comprised 3,928,483 births in England and 1,012,682 in Sweden. There were 4,768 deaths at 31-264 days and 2,223 deaths at 1-4 years in the English cohort and 774 and 476 deaths in the Swedish cohort, respectively. RTI-related deaths contributed 807 of deaths at 31-364 days (17%) and 691 deaths at 1-4 years (31%) in England. In Sweden, the corresponding figures were 139 RTI-related deaths at 31-364 days (18%) and 118 at 1-4 years (25%). SUDI accounted for 24% of all deaths at 31-364 days in both countries (1,166 in England and 189 in Sweden).

Figure 6.1 - Flow diagram showing steps taken to develop comparable and representative birth cohorts in England and Sweden for births in 2003-2012 who survived beyond 30 days of life.



The numbers of live births (n) and deaths (d) are presented. For each exclusion criterion, the percentage of all live births and all deaths is shown in brackets. Crude mortality rates at age 31 days – 4 years per 100,000 person years are presented for each country before and after applying all exclusion criteria.

6.4.2 Characteristics of children alive at 31 days in England and Sweden

The characteristics of children who were alive at 31 days of life were largely similar to characteristics of live born babies who survived beyond days 0-1 of life, described in Chapter 5. Infants in England weighed less than infants in Sweden – a higher proportion had low birth weight (5.5% vs 3.0% in Sweden), a lower proportion weighed ≥3500g at birth (41% vs 55% in Sweden, see Table 6.3). Babies born in England were also more likely to be born prematurely (5.7% vs 4.7% in Sweden), with a congenital anomaly (2.9% vs 2.4%) or to a mother aged less than 25 years old (25% vs 14%).

Table 6.3 – Socio-demographic characteristics of children who survived beyond 30 days of life in England and Sweden

	England n (%)	Sweden n (%)
Birth weight (g)		
500-1499	26,221 (0.67%)	4,646 (0.46%)
1500-2499	189,481 (4.8%)	25,674 (2.5%)
2500-3499	2,089,095 (53%)	428,894 (42%)
≥3500	1,623,686 (41%)	553,468 (55%)
Gestational age (weeks)		
24-34	85,340 (2.2%)	17,616 (1.7%)
35-36	136,681 (3.5%)	30,220 (3%)
37-38	726,161 (18%)	191,005 (19%)
39+	2,980,301 (76%)	773,841 (76%)
Sex		
Boy	2,014,198 (51%)	520,597 (51%)
Girl	1,914,285 (49%)	492,085 (49%)
Congenital anomalies		
No	3,815,315 (97%)	988,298 (98%)
Yes	113,168 (2.9%)	24,384 (2.4%)
Maternal age (years)		
<20	241,111 (6.1%)	16,149 (1.6%)
20-25	757,667 (19%)	129,130 (13%)
25-30	1,063,293 (27%)	295,730 (29%)
30-35	1,109,174 (28%)	356,128 (35%)
35+	757,238 (19%)	215,545 (21%)
Quintile of socio-economic status		
Q1: most deprived	851,119 (22%)	201,434 (20%)
Q2	803,435 (20%)	200,305 (20%)
Q3	767,678 (20%)	202,580 (20%)
Q4	762,378 (19%)	204,117 (20%)
Q5: least deprived	743,873 (19%)	204,246 (20%)

All data are number of all singleton live births who survived beyond 30 days of life (%). Column totals may not add up to 100% due to rounding.

6.4.3 Comparison of RTI-related mortality

6.4.3.1 Characteristics of RTI-related deaths

Children who died from an RTI-related death had similar characteristics in England and Sweden (Table 6.4). In both countries, RTI-related deaths were more common in children with a normal birth weight (≥2500g), which accounted for almost 70% of RTI-related deaths at 31-364 days and 80% at 1-4 years. In Sweden, however, a higher proportion of children who died weighed ≥3500g at birth, reflecting differences in the distribution of birth weight in live births between the two countries (as shown in Table 6.3). Births at ≥39 weeks' gestation contributed approximately 50% of RTI-related deaths at 31-364 days and 60% at 1-4 years. Boys accounted for a higher number of deaths than girls. Congenital anomalies contributed a higher proportion of RTI-related deaths in England than in Sweden (50% in England compared to 43% and 36% in Sweden at 31-364 days and 1-4 years, respectively). In both countries, RTI-related deaths were more common in the most deprived 20% of children.

Table 6.4 –Socio-demographic characteristics of children born in 2003-2012 who died at 31 days – 4 years, from any cause, and from an RTI-related death in England and Sweden

	All deaths at 31-364 days		All deaths at 1-4 years		RTI-related deaths at 31-364 days		RTI-related deaths at 1-4 years	
	England	Sweden	England	Sweden	England	Sweden	England	Sweden
Number of deaths	4,768	774	2,223	476	807	139	691	118
Birth weight (g)								
500-1499	18%	14%	4.8%	2.9%	9.8%	10%	5.4%	5.1%
1500-2499	19%	16%	14%	9.9%	22%	22%	16%	15%
2500-3499	46%	43%	55%	45%	52%	37%	56%	44%
≥3500	17%	26%	27%	42%	16%	31%	23%	36%
Gestational age (weeks)								
24-34	24%	20%	7.6%	7.1%	17%	19%	8%	11%
35-36	9.1%	9.9%	6.5%	5.9%	9.8%	7.2%	7.4%	9.3%
37-38	23%	24%	23%	21%	24%	25%	27%	19%
≥39	44%	46%	63%	66%	49%	48%	58%	60%
Sex								
Boy	57%	56%	54%	55%	56%	58%	53%	56%
Girl	43%	44%	46%	45%	44%	42%	47%	44%
Congenital anomalies								
Yes	55%	59%	62%	75%	50%	57%	50%	64%
No	45%	41%	38%	25%	50%	43%	50%	36%
Maternal age (years)								
<25	36%	23%	33%	19%	33%	28%	31%	16%
25-30	25%	29%	26%	28%	27%	32%	26%	27%
30-35	22%	27%	25%	33%	22%	23%	27%	38%
≥35	17%	21%	16%	21%	18%	17%	16%	19%
Quintile of socio-economic status								
Q1: most deprived	32%	31%	28%	24%	32%	37%	29%	25%
Q2	25%	21%	22%	24%	26%	20%	21%	19%
Q3	17%	14%	19%	20%	17%	11%	18%	23%
Q4	15%	16%	16%	18%	14%	15%	16%	17%
Q5: least deprived	11%	18%	14%	14%	11%	17%	16%	16%

RTI=respiratory tract infections. Data are % of all deaths and RTI-related deaths at 31 days – 4 years. Column totals may not add up to 100% due to rounding.

Unadjusted RTI-related mortality in England was 50% higher at 31-364 days than in Sweden (22.5 RTI-related deaths/100,000 births vs 15.0/100,000 births) and 60% higher at 1-4 years (6.0/100,000 births vs 3.7/100,000 births, Table 6.5). Mortality rates for the two countries were comparable for births with high-risk birth characteristics such as birth weight of 500-1499g, or gestational age of 24-34 weeks. Children with a congenital anomaly had higher mortality in England than in Sweden (however, confidence intervals were overlapping). RTI-related mortality rates were also higher in England for more prevalent, low-risk characteristics such as birth weight of 2500-3499g and gestational age ≥39 weeks, both at age 31-364 days and 1-4 years.

Table 6.5 – Unadjusted RTI-related mortality at 31-364 days and 1-4 years by birth characteristics and socio-economic factors in England and Sweden in 2003-2012

	RTI-related mortality at 31- 364 days		RTI-related mortality at 1-4 years	
	England	**Sweden**	**England**	**Sweden**
Overall	23 (21, 24)	15 (13, 18)	6 (5.5, 6.4)	3.7 (3.1, 4.4)
Birth weight (g)				
500-1499	340 (270, 420)	340 (200, 570)	49 (35, 67)	41 (18, 92)
1500-2499	100 (88, 120)	130 (93, 190)	19 (16, 23)	22 (14, 35)
2500-3499	22 (20, 24)	13 (10, 17)	6.3 (5.7, 6.9)	3.8 (2.9, 5.0)
≥3500	8.9 (7.5, 11)	8.5 (6.3, 11)	3.3 (2.8, 3.9)	2.4 (1.8, 3.2)
Gestational age (weeks)				
24-34	180 (150, 210)	170 (120, 250)	22 (17, 28)	23 (14, 40)
35-36	63 (51, 79)	36 (19, 67)	13 (9.5, 16)	11 (6.4, 21)
37-38	29 (25, 33)	20 (14, 28)	8.5 (7.4, 9.9)	3.8 (2.5, 5.7)
≥39	15 (13, 16)	9.5 (7.4, 12.0)	4.6 (4.2, 5.0)	2.9 (2.3, 3.7)
Sex				
Boy	24 (22, 27)	17 (14, 21)	6.2 (5.6, 6.9)	4.0 (3.1, 5.1)
Girl	20 (18, 23)	13 (10, 17)	5.7 (5.1, 6.4)	3.3 (2.5, 4.4)
Congenital anomalies				
Yes	400 (360, 440)	270 (210, 350)	110 (98, 120)	56 (42, 76)
No	12 (11, 13)	8.7 (7.0, 11)	3.1 (2.8, 3.4)	2.4 (1.9, 3.0)
Maternal age (years) (years)				
<25	29 (26, 33)	29 (21, 40)	7.1 (6.2, 8.2)	4.2 (2.7, 6.5)
25-30	23 (20, 26)	16 (12, 22)	5.7 (5.0, 6.7)	3.4 (2.4, 4.8)
30-35	17 (15, 20)	9.8 (6.9, 14)	5.8 (5.0, 6.6)	3.9 (2.9, 5.3)
≥35	21 (18, 24)	12 (8.2, 18)	5 (4.2, 6.0)	3.3 (2.1, 5.0)
Quintile of socio-economic status				
Q1: most deprived	33 (29, 37)	28 (21, 37)	7.9 (6.9, 9.1)	4.6 (3.2, 6.6)
Q2	28 (24, 32)	15 (11, 22)	6.1 (5.2, 7.2)	3.6 (2.4, 5.4)
Q3	20 (17, 24)	8.1 (4.9, 13)	5.6 (4.7, 6.6)	4.2 (2.9, 6.1)
Q4	17 (14, 20)	11 (7.3, 17)	4.8 (4, 5.8)	3.1 (2, 4.8)
Q5: least deprived	13 (10, 16)	12 (8.2, 19)	5.1 (4.3, 6.2)	2.9 (1.9, 4.6)

RTI=respiratory tract infections. Data are unadjusted mortality rates per 100,000 child-years (95% confidence intervals).

6.4.3.2 Adjusted comparison of RTI-related mortality

At 31-364 days, the unadjusted HR for RTI-related mortality in England relative to Sweden was 1.50 (95% confidence interval (CI): 1.25, 1.80, Table 6.6). After adjusting for birth characteristics, the HR for England relative to Sweden reduced to 1.16 (0.97, 1.39); further adjustment for socio-economic factors reduced the HR to 1.11 (0.92, 1.33). In the fully adjusted model, congenital anomalies were associated with the highest risk of RTI-related death, followed by low birth weight.

Table 6.6 - Unadjusted and adjusted Cox PH models for RTI-related mortality at 31-364 days in England relative to Sweden in 2003-2012

Risk factor	Model 1	Model 2	Model 3
Country			
England	1.50 (1.25, 1.80)	1.16 (0.97, 1.39)	1.11 (0.92, 1.33)
Sweden (baseline)	1	1	1
Birth weight (g)			
<1500		5.7 (3.8, 8.4)	5.1 (3.4, 7.5)
1500-2499		5.6 (4.3, 7.3)	5.1 (3.9, 6.6)
2500-3499		2.06 (1.73, 2.47)	1.96 (1.64, 2.34)
≥3500 (baseline)		1	1
Gestational age (weeks)			
24-34		1.27 (0.93, 1.73)	1.32 (0.97, 1.80)
35-36		1.35 (1.04, 1.75)	1.38 (1.06, 1.79)
37-38		1.23 (1.04, 1.46)	1.24 (1.05, 1.47)
≥39 (baseline)		1	1
Sex			
Boy		1.11 (0.98, 1.27)	1.11 (0.97, 1.26)
Girl (baseline)		1	1
Congenital anomaly			
Yes		23.5 (20.4, 26.9)	23.3 (20.3, 26.7)
No (baseline)		1	1
Maternal age (years)			
<25			1.40 (1.17, 1.68)
25-29			1.25 (1.04, 1.50)
30-34 (baseline)			1
≥35			1.13 (0.92, 1.38)
Quintile of socio-economic status			
Q1: most deprived			1.85 (1.48, 2.32)
Q2			1.65 (1.31, 2.07)
Q3			1.22 (0.95, 1.56)
Q4			1.15 (0.90, 1.48)
Q5: least deprived (baseline)			1

PH=Proportional hazards, RTI=respiratory tract infections. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Some birth weight and gestational age categories were merged due to small numbers.

At 1-4 years, the hazard of an RTI-related death in England was 1.58 times higher relative to Sweden (Table 6.7). After adjusting for birth characteristics, the HR for England vs Sweden reduced to 1.32 (1.09, 1.61). The HR did not change substantially after adjusting for socio-economic factors (1.30, 95% CI: 1.07, 1.59). Birth characteristics explained 44.4% of the excess risk of death in England relative to Sweden, socio-economic factors explained a further 3.8%, independent of the effect on birth characteristics.

Table 6.7 – Unadjusted and adjusted Cox PH models for RTI-related mortality at 1-4 years in England relative to Sweden in 2003-2012

Risk factor	Model 1	Model 2	Model 3
Country			
England	1.58 (1.30, 1.92)	1.32 (1.09, 1.61)	1.30 (1.07, 1.59)
Sweden (baseline)	1	1	1
Birth weight (g)			
<1500		3.12 (2.02, 4.83)	2.99 (1.93, 4.63)
1500-2499		3.67 (2.78, 4.84)	3.50 (2.65, 4.62)
2500-3499		1.70 (1.43, 2.02)	1.66 (1.39, 1.97)
≥3500 (baseline)		1	1
Gestational age (weeks)			
<37		0.85 (0.64, 1.13)	0.86 (0.65, 1.15)
37-38		1.18 (1.00, 1.41)	1.20 (1.01, 1.42)
≥39 (baseline)		1	1
Sex			
Boy		0.99 (0.86, 1.13)	0.98 (0.85, 1.13)
Girl (baseline)		1	1
Congenital anomaly			
Yes		28.4 (24.6, 32.9)	28.4 (24.6, 32.8)
No (baseline)		1	1
Maternal age (years)			
<25			1.04 (0.86, 1.25)
25-29			0.92 (0.76, 1.11)
30-34 (baseline)			1
≥35			0.81 (0.66, 1.00)
Quintile of socio-economic status			
Q1: most deprived			1.28 (1.02, 1.60)
Q2			1.08 (0.85, 1.36)
Q3			1.07 (0.84, 1.35)
Q4			0.93 (0.73, 1.19)
Q5: least deprived (baseline)			1

PH=Proportional hazards, RTI=respiratory tract infections. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Some birth weight and gestational age categories were merged due to small numbers.

6.4.3.3 Subgroup analyses

Low-risk babies accounted for 38% (564) of all RTI-related deaths in England and 44% (113) in Sweden. Unadjusted differences in RTI-related mortality in low-risk babies between England and Sweden were narrower than when comparing RTI-related mortality in all children (HR of 1.28 at 31-364 days compared to 1.50 in all children; HR of 1.37 at 1-4 years compared to 1.58 in all children).

After adjusting the models for birth characteristics and socio-economic factors, the HR for an RTI-related death at 31-364 days in England relative to Sweden reduced to 1.11. The risk of an RTI-related death was highest for babies with a lower birth weight (2500-3499g), born to mothers aged <25 and to the most deprived 20% of families (Table 6.8).

At 1-4 years, the fully adjusted HR for England relative to Sweden remained high (1.30, 95% CI: 0.97-1.74, Table 6.9), but not statistically significant (likely due to reduced sample size for low-risk babies). This result was comparable with the adjusted HR for the whole cohort (see Table 6.7)

Table 6.8 – Unadjusted and adjusted Cox PH models for RTI-related mortality at 31-364 days in low-risk children in England relative to Sweden in 2003-2012

Risk factor	Model 1	Model 2	Model 3
Country			
England	1.28 (0.96-1.70)	1.20 (0.90-1.59)	1.11 (0.84-1.49)
Sweden (baseline)	1	1	1
Birth weight (g)			
2500-3499		1.80 (1.42-2.28)	1.69 (1.34-2.14)
≥3500g (baseline)		1	1
Gestational age (week)			
37-38		1.36 (1.06-1.73)	1.39 (1.09-1.77)
39-41 (baseline)		1	1
Sex			
Boy		1.49 (1.20-1.84)	1.47 (1.18-1.82)
Girl (baseline)		1	1
Maternal age (years)			
<25			1.99 (1.47-2.70)
25-29			1.35 (0.98-1.85)
30-34 (baseline)			1
≥35			1.35 (0.95-1.90)
Quintile of socio-economic status			
Q1: most deprived			1.65 (1.15-2.36)
Q2			1.38 (0.95-2.00)
Q3			1.27 (0.87-1.86)
Q4			1.11 (0.75-1.64)
Q5: least deprived (baseline)			1

PH=Proportional hazards, RTI=respiratory tract infections. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Some birth weight and gestational age categories were merged due to small numbers. Low-risk children were defined as born with birth weight ≥2500g, at term (37-41 weeks), with no congenital anomalies.

Table 6.9 – Unadjusted and adjusted Cox PH models for RTI-related mortality at 1-4 years in low-risk children in England relative to Sweden in 2003-2012

Risk factor	Model 1	Model 2	Model 3
Country			
England	1.37 (1.02-1.82)	1.31 (0.98-1.76)	1.30 (0.97-1.74)
Sweden (baseline)	1	1	1
Birth weight (g)			
2500-3499		1.46 (1.16-1.83)	1.42 (1.13-1.80)
≥3500g (baseline)		1	1
Gestational age (week)			
37-38		1.36 (1.06-1.75)	1.38 (1.07-1.77)
39-41 (baseline)		1	1
Sex			
Boy		1.01 (0.81-1.25)	1.00 (0.80-1.25)
Girl (baseline)		1	1
Maternal age (years)			
<25			0.96 (0.71-1.28)
25-29			0.82 (0.62-1.10)
30-34 (baseline)			1
≥35			0.70 (0.50-0.97)
Quintile of socio-economic status			
Q1: most deprived			1.32 (0.94-1.86)
Q2			1.03 (0.72-1.48)
Q3			0.96 (0.66-1.38)
Q4			1.01 (0.70-1.44)
Q5: least deprived (baseline)			1

PH=Proportional hazards, RTI=respiratory tract infections. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Some birth weight and gestational age categories were merged due to small numbers. Low-risk children were defined as born with birth weight ≥2500g, at term (37-41 weeks), with no congenital anomalies.

6.4.3.4 Sensitivity analyses: PH assumption

The PH assumption was not met for an indicator of congenital anomaly for RTI-related deaths at 31-364 days. Thus, I allowed the HR for congenital anomaly to have different value at 1-2 months, 2-3 months and 3-12 months to meet the PH assumption. The results were robust to sensitivity analyses (Appendix F, Table F.1). The HR for congenital anomalies increased over time (11.7 at 1-2 months, 15.2 at 2-3 months and 29.7 at 3-12 months).

6.4.4 Comparison of SUDI mortality

6.4.4.1 Characteristics of SUDI deaths

SUDI deaths were more common in babies with the most prevalent birth characteristics (when looking at each risk factor independently of others). For example, deaths in infants with normal birth weight accounted for 79% of SUDI deaths in England and 83% in Sweden; deaths in children born at ≥39 weeks accounted for 55% and 56% of SUDI in England and Sweden, respectively (see Table 6.10). Children who died and had any congenital anomalies accounted for over 40% of all deaths at 31-364 days, but for only 4% of SUDI deaths in England and 13% in Sweden. In both countries, boys were more likely to die from SUDI than girls, and SUDI deaths disproportionately occurred in the most deprived 20% of the population.

Table 6.10 - Socio-demographic characteristics of children born in 2003-2012 who died from SUDI at 31-364 days in England and Sweden

	England	Sweden
Number of deaths	1,166	189
Birth weight (g)		
500-1499	4.5%	9.0%
1500-2499	17%	7.9%
2500-3499	56%	49%
≥3500	23%	34%
Gestational age (weeks)		
24-34	10%	12%
35-36	9.9%	8.5%
37-38	25%	24%
≥39	55%	56%
Sex		
Boy	63%	62%
Girl	37%	38%
Congenital anomalies		
Yes	4.2%	13%
No	96%	87%
Maternal Age (years)		
<20	19%	6.3%
20-24	31%	25%
25-29	24%	29%
30-34	15%	23%
≥35	11%	16%
Quintile of socio-economic status		
Q1: most deprived	34%	37%
Q2	27%	22%
Q3	18%	12%
Q4	14%	14%
Q5: least deprived	7.5%	16%

SUDI=sudden unexpected death in infancy. All data are % of all SUDI deaths at 31-364 days. Column totals may not add up to 100% due to rounding.

Unadjusted SUDI rates were 60% higher in England than in Sweden (32.4/100,000 child years vs 20.3/100,000, see Table 6.11). SUDI mortality was higher in England for most of the categories of risk factors, apart from for children with congenital anomalies and birth weight <1500g, for who SUDI mortality was two times higher in Sweden.

Table 6.11 – Unadjusted rates of SUDI per 100,000 child years by birth characteristics and socio-economic factors in England and Sweden in 2003-2012

	England	Sweden
Overall	32 (31, 34)	20 (18, 24)
Birth weight (g)		
500-1499	230 (170, 300)	410 (250, 650)
1500-2499	110 (99, 130)	64 (39, 110)
2500-3499	34 (32, 37)	23 (19, 29)
≥3500	18 (16, 20)	13 (10, 16)
Gestational age (weeks)		
24-34	150 (130, 180)	140 (90, 210)
35-36	93 (77, 110)	58 (35, 95)
37-38	44 (39, 49)	26 (19, 34)
≥39	23 (22, 25)	15 (12, 18)
Sex		
Boy	40 (37, 43)	25 (20, 29)
Girl	25 (22, 27)	16 (13, 20)
Congenital anomalies		
Yes	48 (36, 63)	110 (73, 160)
No	32 (30, 34)	18 (16, 21)
Maternal age (years)		
<20	100 (88, 110)	81 (46, 140)
20-24	52 (47, 57)	41 (31, 54)
25-30	29 (26, 33)	20 (16, 26)
30-35	17 (15, 20)	13 (10, 18)
≥35	18 (16, 22)	15 (11, 22)
Quintile of socio-economic status		
Q1: most deprived	50 (46, 56)	37 (30, 47)
Q2	43 (38, 48)	23 (17, 31)
Q3	30 (26, 34)	12 (7.8, 18)
Q4	23 (20, 27)	14 (9.5, 20)
Q5: least deprived	13 (10, 16)	16 (11, 23)

SUDI=sudden unexpected death in infancy. All data SUDI rates per 100,000 child years at 31-364 days (with 95% confidence intervals).

6.4.4.2 Adjusted comparison of SUDI rates

Overall, children born in England had 59% higher risk of SUDI death than children born in Sweden (1.59, see Table 6.12). After adjusting for birth characteristics, the HR for England vs. Sweden reduced to 1.40, and to 1.19 after further adjustment for socio-economic factors. The differences in birth characteristics explained 32% of the excess risk of SUDI in England relative to Sweden. Socio-economic factors explained a further 35% of increased risk of SUDI, independent of birth characteristics. The most important risk factors associated with the risk of SUDI were a low birth weight and young maternal age.

Table 6.12 – Unadjusted and adjusted Cox PH models for SUDI mortality at 31-364 days in England relative to Sweden in 2003-2012

Risk factor	Model 1	Model 2	Model 3
Country			
England	1.59 (1.36, 1.85)	1.40 (1.20, 1.63)	1.19 (1.02, 1.39)
Sweden (baseline)	1	1	1
Birth weight (g)			
500-1499		8.6 (5.8, 12.7)	6.8 (4.6, 10.1)
1500-2499		4.2 (3.33, 5.2)	3.33 (2.66, 4.17)
2500-3499		1.80 (1.57, 2.06)	1.59 (1.39, 1.82)
≥3500 (baseline)		1	1
Gestational age (weeks)			
24-34		1.81 (1.35, 2.43)	1.98 (1.48, 2.66)
35-36		2.12 (1.71, 2.63)	2.25 (1.82, 2.78)
37-38		1.44 (1.26, 1.65)	1.52 (1.33, 1.74)
≥39 (baseline)		1	1
Sex			
Boy		1.68 (1.51, 1.88)	1.66 (1.49, 1.86)
Girl (baseline)		1	1
Congenital anomaly			
Yes		1.09 (0.85, 1.40)	1.07 (0.83, 1.37)
No		1	1
Maternal age (years)			
<20			4.4 (3.7, 5.4)
20-24			2.45 (2.08, 2.90)
25-29			1.49 (1.26, 1.77)
30-34 (baseline)			1
≥35			1.06 (0.86, 1.30)
Quintile of socio-economic status			
Q1: most deprived			2.19 (1.78, 2.70)
Q2			2.05 (1.66, 2.53)
Q3			1.54 (1.23, 1.93)
Q4			1.41 (1.12, 1.78)
Q5: least deprived (baseline)			1

PH=proportional hazards, SUDI=sudden unexpected death in infancy. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Some birth weight and gestational age categories were merged due to small numbers.

6.4.4.3 Sensitivity analyses

6.4.4.3.1 PH Assumption

Cox PH models for SUDI met the PH assumption and no interactions with time were therefore fitted.

6.4.4.3.2 Inter-country differences in coding practices

R95 code denoting SIDS was the most commonly used ICD-10 code to denote a SUDI death (Table 6.13). It accounted for a higher proportion of SUDI deaths in Sweden (74.6%) than in England (62.7%). The second most frequently used code was R99 ("*Other ill-defined and unspecified causes of mortality*"), covering 32.6% of SUDI deaths in England and 21.7% in Sweden. W75 ("*Accidental suffocation and strangulation in bed*") accounted for a further 3.5% and 2.1% of SUDIs in England and Sweden, respectively.

Table 6.13 – The most commonly used ICD-10 code to denote SUDI death in England and in Sweden

ICD-10 code	England	Sweden
R95	63%	75%
R96	0.1%	0
R98	0	0.5%
R99	33%	22%
W75	3.5%	2.1%
W78	0.8%	0
W79	0.3%	1.1%

ICD-10=International Statistical Classification of Diseases and Related Health Problems, SUDI=sudden unexpected death in infancy

Sensitivity analyses based only on the ICD-10 code for SIDS (R95) showed narrower differences in mortality between England and Sweden (HR of 1.33 for England relative to Sweden, Appendix F, Table F.2). This difference became negligible after adjustment for birth characteristics and socio-economic factors (HR of 0.98). As in the models for SUDI, the risk of death was highest for children weighing <2500g at birth and born to teenage mothers (<20).

Finally, I compared SUDI rates in low-risk children, who accounted for 69% of all SUDIs in England and 65% in Sweden. Between-country differences in SUDI mortality in low-risk children were larger than when using the whole cohort (unadjusted HR for England vs Sweden was 1.70 95% CI: 1.41, 2.06, see Appendix F, Table F.3). The HR reduced to 1.61 (1.33, 1.95) after adjusting for birth characteristics and to 1.35 (1.12, 1.65) after further adjustment for socio-economic factors. Birth characteristics accounted for 13% of excess risk of SUDI in "low-risk" babies in England relative to Sweden; socio-economic factors accounted for a further 36%, independent of the effect

on birth characteristics. Children born to mothers aged <20 years old were five times more likely to die from SUDI than those of mothers aged 30-34, and children of mothers aged 20-24 years old were three times more likely to die.

6.5 Discussion

6.5.1 Key findings

6.5.1.1 RTI-related mortality

The excess RTI-related mortality in England relative to Sweden at 31-364 days was largely explained by the differences in birth characteristics. RTI-related deaths contributed 17% of all deaths at 31-364 days in England and 18% in Sweden. The risk of an RTI-related death was 50% higher in England relative to Sweden, and decreased to 16% after adjusting for birth characteristics, and to 11% after further adjustment for socio-economic factors.

At 1-4 years, over half of the excess RTI-related mortality remained unexplained in England relative to Sweden. RTI-related deaths accounted for 31% and 25% of all deaths at 1-4 years in England and Sweden, respectively. The risk of an RTI-related death was 58% higher in England relative to Sweden. Birth characteristics contributed 45% of the excess risk of death in England, and socio-economic factors contributed a further 3%. After full adjustment, the risk of an RTI-related death remained 30% higher in England relative to Sweden.

6.5.1.2 SUDI

The differences in the distribution of birth characteristics and socio-economic factors contributed equally to the increased risk of SUDI death in England relative to Sweden. SUDI accounted for 24% of all deaths at 31-364 days in both countries. Children born in England had 59% higher risk of SUDI death than children born in Sweden. Differences in the distribution of birth characteristics contributed to 32% of the excess risk of death in England; socio-economic factors contributed a further 35%.

6.5.2 Strengths

The use of individual-level data enabled me to overcome the limitations of previous comparisons of cause-specific mortality, which were based on aggregate data tabulated by the underlying cause of death. First, the comparisons presented in this chapter accounted for differences in the distribution of birth characteristics and socio-economic factors in England and Sweden. Such an approach enabled me to quantify the contribution of risk factors operating before and during pregnancy to the increased risk of an RTI-related death and a SUDI death in England relative to Sweden.

Second, I used a broad definition of RTI-related mortality to reduce possible bias from inter-country differences in selection of the underlying cause of death. Unlike in previous comparisons, the definition of RTI-related death used in this chapter included a mention of RTIs as any cause of death on a death certificate, rather than only as the underlying cause.[6–8] I also included cases where RTI was severe enough to require hospitalisation a month prior to death, even if RTI was not recorded on a death certificate, to account for reported discrepancies between hospital admission records and death records.[207,208] In England this method was shown to increase the number of identified RTI deaths in children by 34% compared to using causes of death on the death certificate alone.[18]

Third, I used a definition of SUDI recommended for international comparisons to account for inter-country differences in investigative practices and autopsy protocols,[90] which were illustrated by the sensitivity analyses: ICD-10 code for SIDS ("R95") was more commonly used in Sweden than in England, leading to smaller inter-country differences in SIDS rates than for SUDI rates. The use of the broader definition also helped to account for a reported variation in certification of SIDS diagnosis (ICD-10 code "R95") versus "unascertained death" diagnosis (ICD-10 code "R99") between pathologists in England.[209]

6.5.3 Limitations

6.5.3.1 Unrecorded risk factors of interest

Almost 90% of RTI-related deaths occur to children with at least one chronic condition in England,[18] which I did not adjust for. The risk of an RTI-related death at 1-4 years remained 30% higher in England than in Sweden after adjustment for birth characteristics and socio-economic factors. These differences could be explained by differences in the prevalence of chronic conditions, which may originate *in utero*, during the neonatal period or later in early childhood, such as cerebral palsy, bronchopulmonary dysplasia, or cancer. An indicator of the presence of chronic conditions could be derived using children's hospital admission records and recorded causes of death from death certificates. Classifications of chronic conditions in children based on ICD-10 codes exist.[58,198] However, some conditions might be managed only in primary care (and not be coded in secondary care records) in one country, but not the other. Further work is needed to identify conditions which are likely to be treated and recorded in a similar way between the two countries before including such an indicator in the analyses.

Information about known risk factors for SUDI, such as smoking during pregnancy, parental smoking and alcohol use, breastfeeding, sleeping position or co-sleeping was

not available for this comparison [89]. Variation in the prevalence of these risk factors could account for some of the unexplained differences in SUDI rates between England and Sweden which remained after adjusting for birth characteristics and socio-economic factors. Further comparative analyses are possible using more detailed data from the Nordic Epidemiological SIDS study in 1992-1995 [197] and data from Bristol and surrounding areas collected for the CESDI SUDI Studies 1993-1996 [91]. However, these datasets are over 20 years old and might not be representative for SUDI cases now.

6.5.3.2 Exclusion of deaths on days 28-30

I could not include deaths on days 28-30, because a high proportion of these deaths in England did not have any recorded causes of death. It is likely, that these fields were removed by NHS Digital during processing of the hospital admission data. It is possible that if a neonatal death certificate was used for infants who died aged >27days, it was assumed to be an error. Deaths on 28-30 days of life account for only 4% of deaths at 28-364 days in England and 3.6% of deaths in Sweden. However, this data processing error limits future studies of cause-specific infant mortality (including deaths in first month of life).

6.5.4 Interpretation of the results

6.5.4.1 RTI-related mortality

The increased RTI-related mortality rates in England relative to Sweden at 31-364 days were largely explained by inter-country differences in the distribution of birth characteristics. The HR for RTI-related death in England relative to Sweden decreased from 1.50 to 1.16 for deaths at 31-364 days after adjustment for birth characteristics, and from 1.58 to 1.32 for deaths at 1-4 years. Therefore, preventive strategies aimed at reducing the prevalence of adverse birth characteristics by addressing maternal health before and during pregnancy could reduce RTI-related mortality in England relative to Sweden, especially in the first year of life.

At 1-4 years, children born in England had a 30% higher risk of an RTI-related death than children born in Sweden, even in the model adjusted for birth characteristics and socio-economic factors. Further analyses including a comparable indicator of chronic conditions are needed to determine whether this excess risk of death reflects differences in provision of healthcare (e.g., differences in the timing of diagnosis of serious infections in the primary care setting, as has been suggested previously),[6] or an increased prevalence of chronic conditions in England relative to Sweden. Such an indicator should include conditions which are treated and recorded in a similar way in hospital admissions datasets in the two countries. Additional information from primary care records could be included to account for inter-country differences in thresholds for

hospital admissions and increased rates of hospital admissions of children in England. However, primary care data are not collected in Sweden.

6.5.4.2 SUDI

Differences in distribution of birth characteristics accounted for a third of increased risk of SUDI death in England relative to Sweden. Therefore, some reductions in SUDI deaths in England could be achieved by reducing the prevalence of adverse birth characteristics.

A further third of the excess risk of SUDI in England relative to Sweden was explained by socio-economic factors, independent of birth characteristics, both overall and in the subset of low-risk babies. Young maternal age and low birth weight were the most important risk factors contributing to the risk of SUDI. Both of these factors are associated with smoking during pregnancy. Information about smoking during pregnancy was not available for this project, however, smoking is more prevalent among women in England than in Sweden (in 2010, 12% of women in England smoked during pregnancy, compared to 6.5% in Sweden).[12] Parental smoking was identified as a modifiable factor for child deaths in England in a review of 71% of all child deaths (under the age of 18) which occurred between April 2008 and March 2011.[107] Thus, smoking cessation programs could be effective at reducing excess SUDI deaths in England.

The risk of SUDI in England relative to Sweden remained approximately 20% higher after adjusting for all risk factors. Other reasons for the differences in the observed SUDI rate could reflect inter-country differences in safe sleeping practices, which are unlikely to be accounted for by family characteristics included in the models. However, information about differences in prevalence of unsafe sleeping practices in England and Sweden is needed to confirm this hypothesis.

6.5.5 Policy recommendations

Rates of RTI-related deaths and SUDI in England could be reduced relative to Sweden by improving maternal health before and during pregnancy to reduce prevalence of adverse birth characteristics. Further work is needed to determine whether the unexplained excess in RTI-related mortality at 1-4 years in England reflected a higher prevalence of underlying chronic conditions in England, or failure of services to diagnose and treat RTIs in a timely manner.

Socio-economic factors contributed to a third of the observed differences in SUDI rates between England and Sweden. Thus, further reductions in SUDI could be achieved by reducing some of the modifiable factors known to be socially patterned such as parental smoking or teenage pregnancy. Differences in safe sleeping practices could

further contribute to the differences; however, information about the prevalence of unsafe sleeping practices in England and in Sweden is needed to confirm this. Children from the most socio-economically deprived families were at the highest risk of SUDI death, thus this high-risk group stand to benefit most from campaigns addressing high SUDI rates in England.

Chapter 7. Summary of findings, implications and conclusions

7.1 Summary of research

7.1.1 Rationale and thesis aims

The United Kingdom (UK) has the second highest child mortality rate in Western Europe, while Sweden has the third lowest (see Figure 1.1, Chapter 1). In 2013, child mortality in the UK was almost twice as high as in Sweden.[3] These differences have previously been attributed to wider socio-economic inequalities in the UK, leading to higher rates of preterm birth, and to differences in the provision of healthcare in the two countries.[6–10]

Previous comparisons of child mortality in the UK relative to Sweden did not account for differences in the distribution of birth characteristics and socio-economic factors in the two countries. Without adjustment for birth characteristics it was not possible to determine whether the increased child mortality in the UK reflected differences in exposures during pregnancy (leading to a higher prevalence of adverse birth characteristics) or in the care received after birth, given a child's characteristics at birth. Therefore, any suggested explanations for increased child mortality rates in the UK relative to Sweden remained speculative.

This thesis aimed to quantify the contribution of birth characteristics and socio-economic factors to the increased child mortality rates in England relative to Sweden. The analyses focused on England, as England is the biggest and the most diverse of the four UK countries. Furthermore, data for England was available from the start of my PhD.

7.1.2 Key findings

Aggregate data tabulated by a key risk factor at birth (e.g., gestational age) can provide insights into the origins of inter-country differences in infant mortality compared to data presented by age-at-death only (objective 1).

Crude and gestation-standardised infant mortality can be used to develop two metrics for making international comparisons of infant mortality more relevant for policy makers than relying on crude infant mortality rates alone:

- Metric 1 is the within-country difference in crude and standardised mortality, and it reflects the influence of distribution of gestational age on inter-country differences in infant mortality.

- Metric 2 is the between-country difference in gestation-standardised mortality. It reflects excess mortality due to differences in the quality of infant care received after birth.

In England and Wales, the two metrics contributed almost equally to the difference in crude infant mortality rates relative to Sweden. This indicated that preventive strategies need to address both maternal health before and during pregnancy, as well as the care after birth. However, differences in the prevalence of congenital anomalies or low birth weight could have contributed to metric 2.

Administrative linked datasets in England and Sweden can be used to develop nationally-representative, comparable birth cohorts (objective 2).

Hospital Episode Statistics (HES) linked to Office for National Statistics (ONS) mortality data can be used to develop a national birth cohort of singleton live births. Longitudinal follow-up data were available via linkage to hospital admission trajectories and mortality records for births after 2003. High completeness of recording of birth characteristics and socio-economic factors was achieved by linking maternal delivery episodes and birth episodes. The distribution of birth characteristics and socio-economic factors among live births was representative for the population of children in England and Wales. However, the cohort could not be used to investigate early life mortality, as key risk factors at birth (birth weight, gestational age) were more likely to be missing in extremely low birth weight and extremely preterm babies, or infants who died shortly after birth. Linked HES-ONS data did not provide sufficient additional information about an infant's health at birth to reliably impute these variables using multiple imputation techniques.

For comparisons of child mortality in England and Sweden, I developed a sub-cohort of selected hospitals with high quality of recorded data. The cohort excluded deaths on days 0-1 of life, and individuals with birth weight <500g, gestational age <24 weeks, or missing information on any risk factor of interest. Infant mortality and the distribution of birth characteristics in this sub-cohort were representative for the population of children in England and Wales. However, due to missing data on causes of death at 28-30 days, international comparisons of cause-specific mortality could be carried out for deaths after 30 days of life only.

A comparable Swedish birth cohort was developed using linked Swedish national registers. Birth cohorts in both countries covered information on birth weight, gestational age, sex and maternal age. An indicator of the presence of congenital anomalies was developed using mortality records and diagnostic information from

hospital admission trajectories. However, the only available measure of socio-economic status (SES) was not directly comparable between the two countries.

Higher child mortality rates in England relative to Sweden in 2003-2012 were primarily driven by the differences in distribution of birth characteristics, and to a lesser extent by an independent effect of socio-economic factors (objective 3).

The risk of death was 66% higher at 2-27 days and 59% higher at 28-364 days in England relative to Sweden. Unfavourable distribution of birth characteristics in England (i.e. a higher prevalence of preterm birth, congenital anomalies and low birth weight, as well as a lower mean birth weight) accounted for 77% and 68% of the excess risk of death at 2-27 days and 28-364 days, respectively. Socio-economic factors independently contributed to a further 3% and 11% of excess risk of death at 2-27 days and 28-364 days, respectively (over and above the effect of socio-economic factors on birth characteristics). Small, but statistically significant differences in infant mortality remained after adjustment for these factors.

The risk of death in low-risk babies was 87% higher at 2-27 days and 47% higher at 28-364 days in England relative to Sweden. Birth characteristics and socio-economic factors explained only 29% of excess risk of death at 2-27 days, and the risk of death remained 67% higher in England after adjustment for all risk factors. Some of this difference could potentially be explained by a higher prevalence of neonatal morbidity (such as birth asphyxia and other birth trauma) in England, which was not accounted for by the included birth characteristics and socio-economic factors. At 28-364 days, birth characteristics and socio-economic factors each accounted for approximately 30% of the observed excess risk of death in England. Small, but statistically significant differences in infant mortality remained after adjustment for birth characteristics and socio-economic factors.

The risk of child death at 1-4 years was 27% higher in England relative to Sweden and the difference in mortality became negligible in the fully adjusted model. For low-risk babies there were no differences in mortality beyond infancy.

Birth characteristics contributed to the increased risk of respiratory tract infection (RTI)-related death and of death from sudden unexpected death in infancy (SUDI) in England relative to Sweden. Socio-economic factors further contributed to the excess mortality from SUDI in England (objective 4).

RTI-related deaths accounted for approximately 20% of all deaths at 31 days-4 years in both countries. At 31-364 days, the risk of an RTI-related death was 50% higher in

England relative to Sweden, and decreased to 16% after adjusting for birth characteristics, and to 11% after further adjustment for socio-economic factors. At 1-4 years, children born in England had 58% higher risk of an RTI-related death; birth characteristics explained 44% of the excess risk of death in England relative to Sweden, socio-economic factors independently contributed a further 4%.

SUDI accounted for a quarter of all deaths at 31-364 days in England and in Sweden. Children born in England had 59% higher risk of SUDI death than children born in Sweden. Birth characteristics and socio-economic factors each contributed to a third of excess risk of SUDI overall in England relative to Sweden.

7.2 Strengths

The datasets available for this comparison were a major strength in this study. I developed and validated comparable, nationally-representative birth cohorts for both England and Sweden using individual-level data, with information about birth characteristics and socio-economic factors at birth. The internationally standardised coding systems used in both countries enabled me to develop a comparable congenital anomaly indicator, which no previous international comparison of child mortality has used. The English birth cohort was based on a subsample of 64.5% of hospital births, which was thoroughly validated against national statistics published for England and Wales by the ONS. The Swedish birth cohort had whole-country coverage, with high completeness of recorded variables; the complete case cohort covered 99.6% of all singleton live births in Sweden. The large sample sizes and long follow-up periods allowed me to investigate the effect of rare risk factors, such as congenital anomalies or extreme prematurity on mortality, which in itself is a rare outcome among children in Western Europe. Analysing combined tables of data derived from the birth cohorts in England and Sweden enabled me to quantify the contribution of risk factors at birth to the overall differences in child mortality, overcoming the limitations of previous comparisons of child mortality in the UK and in Sweden. Furthermore, the results presented in Chapters 5 and 6 were robust to all sensitivity analyses.

7.3 Limitations and future directions

7.3.1 Comparison of mortality around the time of birth and stillbirths

I did not compare mortality on days 0-1 of life in England and Sweden. Children who died around the time of birth were more likely to have missing recording of birth weight or gestational age in the English birth cohort; mortality rates based on the complete case cohort remained underestimated even after enhancing the completeness of

recorded risk factors through linkage to maternal records. There was not sufficient additional information recorded in HES to reliably impute birth weight and gestational age using multiple imputation techniques. Therefore, I could not compare mortality around the time of birth in England and Sweden. These deaths accounted for one quarter of child deaths in England and in Sweden.

Further comparison using a more complete birth dataset in England is required to examine inter-country differences in deaths around the time of birth. Such a comparison would also need to include stillbirths to account for possible bias due to inter-county differences in registration practices for still- and live births. Furthermore, many risk factors for stillbirth and neonatal deaths are similar (such as, maternal obesity, smoking or socio-economic deprivation),[12,109] so analyses including both stillbirths and live births would better present the full potential benefits from reducing the prevalence of such risk factors on child mortality and early life survival. Finally, a comparison based on total births would minimise the 'live birth' bias, which arises when the same prenatal exposures are associated with the outcome of interest and the risk of foetal death.[200]

A comparison of early life mortality (including stillbirths) could be based on a whole-country birth cohort from linked ONS birth registration, National Health Service (NHS) birth notification data and longitudinal hospital admission records for mothers and babies from HES. ONS birth registration data have a high completeness of birth weight and maternal age, and additional individual-level socio-economic indicators such as parental country of birth and occupation; however, these are only coded for 10% of births.[128] NHS birth notification complements the ONS birth registration data with gestational age and ethnicity.[167,178] Longitudinal hospital admission data (HES) can be used to derive comorbidities in mothers and babies. The feasibility of such linkage has previously been demonstrated.[167,178] However, it was only achieved in 2016 and is not routinely updated. Further funding is needed to update this valuable resource and make it available to other researchers.

7.3.2 Estimating the total effect of socio-economic factors on increased child mortality in England relative to Sweden

In this PhD, I only accounted for the contribution of socio-economic factors to the increased risk of child death in England relative to Sweden over and above the effect of social deprivation on the increased risk of adverse birth characteristics. However, a family's SES is associated with an increased risk of preterm birth,[14,93] low birth weight,[14,93] or presence of a congenital anomaly,[13] which I did not account for. Therefore, the contribution of socio-economic factors on the differences in child mortality between England and Sweden presented in this PhD is underestimated.

Further work is required to determine the total contribution of socio-economic factors to the differences in mortality between England and Sweden. Causal mediation methods could be used to determine the total effect of socio-economic factors (operating before and after birth) to the excess mortality in England relative to Sweden, including the effect of SES mediated by adverse birth characteristics.

Ideally, such a comparison would be based on comparable, individual-level measures of SES. In this study, I used an area-level measure in England and an individual-level measure in Sweden. Calculating quintiles of SES among all pregnant women helped to standardise the indicator of SES. However, the differences between SES quintiles were likely to be underestimated for England in the comparisons in Chapters 5 and 6. Maternal education level is recognised as the most internationally comparable indicator of SES,[192] but information on maternal education was not available in HES in England. A comparison based on parental occupation could be conducted by linking HES to ONS birth registration data, as comparable information is available for Sweden; however, this variable is collected by ONS for only 10% of the population,[128] so the study sample size would be reduced to approximately 600,000 births in 2003-2012.

7.3.3 Adjusted comparison of child mortality by ethnic groups in England

As outlined in Chapter 1, mortality rates and the prevalence of adverse birth characteristics vary between ethnic groups in England. These differences reflect a complex interplay between socio-economic disadvantage[210] and cultural factors (e.g., differences in attitudes to termination of pregnancy (TOP) for foetal anomalies)[13,77] over and above some biological factors (e.g., the effect of maternal stature on birth weight).[78,79] I could not determine how much of the difference in child mortality between England and Sweden could be explained by the differences in the ethnic make-up of the two populations, since ethnicity is not recorded in any of the registers in Sweden (as detailed in Chapter 1). However, ethnicity is not a modifiable risk factor, and determining the origins of inequalities in child mortality between ethnic groups within England would be more relevant for policy than comparing the contribution of differences in ethnic make-up of the populations in England and in Sweden.

Further work requires a comparison of child mortality rates between ethnic groups in England, adjusted for birth characteristics and socio-economic factors. Such a comparison would inform policy as to how to best address the observed inequalities in child health outcomes between ethnic groups: by addressing inequalities in maternal health and socio-economic circumstances before and during pregnancy, or by improving care received after birth.

HES birth cohort derived in Chapter 3 could be used for such a comparison because ethnic category has been recorded in HES since 1995, and it has been a mandatory return to HES for all episodes of care (including births) since 2009.[136] According to Harron *et al.,*[154] approximately 60% of babies and 75% of mothers had ethnicity recorded in 2003/4, and the completeness increased over time to more than 90% in 2012/13. Ethnic group classification used in HES is based on ethnic group classification from the 1991 Census and includes 9 groups (expanded to 16 groups in the 2001 Census),[136] reflecting the colonial history and migration patterns to the UK.

7.3.4 Additional maternal risk factors

This study would have benefitted from allowing for additional maternal risk factors during pregnancy in the analyses, as outlined in Chapter 1. These variables include smoking during pregnancy and obesity, which have a higher prevalence in England than in Sweden. In 2010, 12.6% of mothers in Sweden were obese (defined as BMI≥30),[12] compared to approximately 20% in England (based on population of all women in all females aged 16-44 years old);[66] 6.5% of mothers smoked in the 1st trimester in Sweden, compared to 12% who smoked at any point during pregnancy in England.[12] These variables were available in Medical Birth Register in Sweden (SMBR), but not in HES.

In the future, a comparison accounting for smoking during pregnancy and maternal BMI could be based on the new Maternity and Children's Data Set (MCDS) in England, collected since 2015.[126] However, time is needed to achieve complete coverage of all births (as of June 2017, only 88% of hospitals contribute data on births to MCDS),[127] and improve completeness of recorded variables (in June 2017, BMI was missing for 14-24% of records and smoking status was missing for 8-17% of mothers, depending on the reporting region).[127]

7.3.5 Comparison of mortality in uncomplicated, low-risk pregnancies

Further research is needed to determine the origins of unexplained excess neonatal mortality in low-risk babies in England relative to Sweden. It is likely that some of the included infants had other neonatal morbidity which was not indicated by birth weight, gestational age or the presence of congenital anomalies. This could include birth asphyxia, which is one of the most common causes of neonatal mortality in term, non-malformed babies,[32] and complications related to severe asphyxia (such as meconium aspiration or neonatal seizures).[64,65] Hospitalisation records of mothers and babies could be used to identify uncomplicated, low-risk pregnancies for a more fair comparison of mortality in low-risk babies.

7.3.6 Indicator of chronic conditions in children

In England, almost 90% of RTI-related deaths occur in children with at least one chronic condition.[18] Therefore, some of the unexplained excess mortality at 1-4 years in England relative to Sweden could reflect differences in the prevalence of chronic conditions in England. A comparison adjusted for an indicator of presence of chronic conditions could determine whether remaining differences in RTI-related mortality reflected an increased prevalence of underlying adverse health outcomes in England, or a failure of services to adequately diagnose and treat RTIs.

An indicator of the presence of chronic conditions for future studies could be derived using diagnoses recorded in children's longitudinal hospital admissions and causes of death recorded on death certificates. Classifications of chronic conditions in children using the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) exist.[58,198] However, further work is needed to identify conditions which are likely to be treated in a similar way and recorded in hospital admissions databases in the two countries.

7.4 Implications for child mortality data collection collation, and linkage

7.4.1 Routine collection of aggregated data by risk factor at birth in high-income countries

Aggregate data tabulated by one key risk factor at birth (such as birth weight or gestational age) can provide some insights into the origins of inter-country differences in infant mortality. The EURO-PERISTAT project has shown that many European countries (18 out of 31) record relevant information on births and deaths tabulated by birth weight and gestational age.[12,211] Such data are also collected in perinatal registers in regions of Australia,[118] Canada,[119] and the United States of America (USA).[120] Thus, in order to allow more policy-relevant international comparisons of early life mortality, counts of live births, stillbirths, neonatal and infant deaths tabulated by birth weight and/or gestational age categories should be routinely collated and published by international agencies such as the World Health Organization (WHO) or the Organisation for Economic Co-operation and Development (OECD).

More funding is needed, both in-country and for international collaborations such as the EURO-PERISTAT, to ensure that such data are available in all countries, and collected regularly using similar definitions of stillbirths and live births.[26,109] Improvements in the completeness of recorded data are also required, as the EURO-PERISTAT project showed that children who died *in utero* or during infancy were less likely to have complete information on birth weight and gestational age at birth.

7.4.2 Need for a national birth register in England and improvements in the HES-ONS data quality

This PhD illustrated the importance of developing a national register of births in England, with details of antenatal, obstetric, and neonatal care and key characteristics of mothers and babies, equivalent to the Medical Birth Registers in the Nordic countries.[123] Since 2015, maternity and child health services in England are required to contribute data on medical and clinical details about the birth, mother and baby to the MCDS.[126] However, this dataset will take time to achieve completeness of coverage and recording of risk factors at birth (as mentioned in Section 7.3.4), and build up follow-up data. Alternatively, high quality information on risk factors at birth, socio-economic factors and longitudinal follow-up for all births in England could be obtained through the linkage of ONS birth registration, NHS birth notification, HES records for mothers and babies and ONS mortality dataset (described in section 7.3.1), however such linkage is not yet routinely provided.

In the meantime, a birth cohort based on linked HES-ONS mortality data provides a unique resource for future studies of child health in England. In this PhD, I showed that the HES-ONS dataset can be used to develop a nationally-representative birth cohort of singleton live births for births in or after 2003. The cohort could not be used to investigate early life mortality, as birth weight and gestational age were more likely to be missing in extremely low birth weight and extremely preterm babies, or infants who died shortly after birth. However, the distribution of birth characteristics and socio-economic factors among live births was representative for the population of children in England and Wales. The HES birth cohort can, therefore, be used for studies of child health outcomes, which are not associated with mortality in first two days of life. As of December 2017, the cohort has already been used for other studies in my research team, including an international comparison of the coding of congenital anomalies,[174] a study of socio-economic inequalities in waiting times for orchidopexy surgery,[172] and a PhD investigating risk factors for admissions for acute lower RTIs in infants.[173] Linked HES-ONS data have the advantage of an ongoing data collection; therefore, the HES-ONS birth cohort can be easily updated once more data become available (as of December 2017, I have updated the cohort to cover births until April 2017). Therefore, investments in this dataset would be worthwhile.

Re-linking birth episodes to the hospital admission records prior to 2003 in HES is crucial for creating accurate healthcare use trajectories. For many of these pre-2003 birth records, both postcode and the NHS number are missing. This would, therefore, require first linking birth and delivery episodes to obtain information about the baby's postcode at birth.[154] Using the date of birth, sex and postcode, historical birth episodes

in HES could be linked to the Personal Demographics Service (PDS) to obtain the baby's NHS number, which could then be used to re-link birth episodes to consecutive admissions after birth.

Improvements in the completeness of risk factors at birth are also needed. Mother-baby linkage substantially increased the completeness of birth weight, gestational age, maternal age, postcode and Index of Multiple Deprivation (IMD) scores (from 18.1% of records with complete information to 75.4%) and should be conducted by NHS Digital on a routine basis. However, babies who died shortly after birth were less likely to link to their mother. Detailed information on pregnancy and birth recorded in the HES "baby tail" should, therefore, be mandated returns in the new MCDS dataset, and until this dataset is set up, in the HES "baby tail" to ensure completeness of recorded risk factors.

7.5 Implications for policy and practice in England

7.5.1 Reducing the prevalence of adverse birth characteristics

The differences in mortality at age of 2 days-4 years in England relative to Sweden were primarily driven by a higher prevalence of adverse birth characteristics in England. The healthy development of a foetus in the womb is strongly associated with the health of the mother during pregnancy. For example, maternal smoking, obesity, underweight, and young and old age are associated with an increased risk of low birth weight,[41,202] congenital anomalies,[60,64,74,77] or preterm birth.[28,63,74,202] Therefore, policies to reduce child mortality in England should focus on improving the health of women before and during pregnancy to reduce the prevalence of these adverse birth characteristics.

In Chapter 1, I presented a list of maternal characteristics which could contribute to an increased risk of adverse birth characteristics. These included maternal age, BMI, parity, maternal health status, risky behaviours during pregnancy (such as smoking, or alcohol consumption), ethnicity, and socio-economic disadvantage. Further inter-country comparisons of the prevalence of adverse birth characteristics adjusted for these maternal characteristics and underlying health conditions are needed to determine which of the maternal risk factors contribute most to the increased prevalence of adverse birth characteristics in England relative to Sweden. Large differences in the prevalence of maternal smoking, maternal obesity and teenage pregnancy were observed in England and in Sweden. Therefore, in this section I review the interventions which could address these maternal risk factors.

7.5.2 Reducing rates of maternal smoking during pregnancy

Maternal smoking is associated with an increased risk of preterm birth,[12,74] low birth weight[12,74] and some congenital anomalies in England.[12,74] Furthermore, smoking during pregnancy is an independent risk factor for sudden infant death syndrome (SIDS).[89] The prevalence of smoking is higher in England than in Sweden: in 2010, 12% of women smoked during pregnancy in England, compared to 6.5% of mothers who smoked in the 1st trimester in Sweden.[12] Therefore, reducing rates of maternal smoking (and exposure to second hand smoke) could lead to reductions in England's child mortality relative to Sweden. During the study period, smoke-free legislation banning smoking in public places was introduced in England in 2007.[212] Since then, the prevalence of maternal smoking at the time of delivery in England has reduced from 15% in 2006/7, to approximately 11% in 2016/17.[213] This is still higher than in Sweden, and further reductions could be achieved.

To ensure the best health outcomes for children, future mothers should be encouraged to quit smoking before pregnancy and universal smoking cessation interventions could be applicable to them. The Cochrane Tobacco Addiction Group reviewed overall 61 different smoking cessation interventions; however, evidence for many of these interventions remains inconclusive. Some that have shown potential benefits include mobile phone apps supporting smoking cessation,[214] and advice or counselling given by nurses.[215] In 2016, the UK introduced standardised tobacco packaging.[216] The public health effects of this policy change have not yet been evaluated in the UK, but standardised packaging led to decrease in tobacco use in Australia.[217]

Preventing tobacco use in young people and the future generation of mothers is of equal importance. Cochrane Tobacco Addiction Group reviewed eight interventions for preventing young people from starting to smoke, however there was not sufficient evidence to support any of them.

Smoking cessation during pregnancy can reduce some but not all of the risks of adverse birth characteristics. Smoking cessation in the 1st trimester reduces the risk of prematurity, stillbirth and low birth weight close to the risks observed for non-smoking mothers (odds ratios adjusted for maternal age, parity, sex and SES were 1.07, 1.01, 1.09, respectively); however, an increased risk remains for the presence of congenital anomalies.[74] Smoking cessation late in pregnancy can lead to increased birth weight and reductions in the rate of low birth weight, and psychosocial interventions are effective at increasing the proportion of women who stop smoking late in pregnancy.[204] Therefore, while it would be most beneficial for the foetus if a mother quit smoking before pregnancy, women at any stage of pregnancy should be encouraged to stop smoking.

7.5.3 Reducing maternal obesity

Maternal obesity is more common in England than in Sweden and its increased prevalence could be another factor contributing to the increased prevalence of adverse birth characteristics in England. In 2010, 12.6% of mothers in Sweden were obese,[12] compared to approximately 20% in England.[66] Maternal obesity is associated with an increased risk of congenital anomalies,[61,62] or preterm birth.[63] Obese mothers also have higher rates of pregnancy complications such as pre-eclampsia or gestational diabetes,[12] which are associated with an increased risk of neonatal mortality in their children.[64] The increased prevalence of maternal obesity in England could also contribute to the unexplained excess neonatal mortality in low-risk babies in England relative to Sweden. According to a study of mothers in Sweden, term infants of obese mothers have an increased risk of birth asphyxia and severe asphyxia related complications (such as meconium aspiration or neonatal seizures).[64,65]

The evidence is inconclusive about the effectiveness of weight loss interventions targeting specifically obese women before or during pregnancy on the prevalence of adverse birth characteristics. According to a Cochrane review, there have been no trials for preconception health programs and interventions directed specifically at overweight women to improve their pregnancy outcomes (as of 2015).[218] There is not sufficient evidence to show the benefits of aerobic exercise during pregnancy on the health of mothers or the baby.[219] Dietary interventions encouraging a balanced diet (as dieting for weight loss could harm development of the foetus *in utero*)[66] could be beneficial for reducing maternal comorbidities (such as pre-eclampsia), but there is no evidence of a positive effect on birth characteristics or neonatal mortality.[220] Policies targeting obesity in the whole population are likely to be the most effective at reducing obesity among mothers.

The current government in the UK has introduced some legislation to tackle the problem of obesity, especially in children. From 2018, sugar-sweetened drinks will be additionally taxed. The government has also set a target for reducing children's sugar intake by 20% by 2020.[221] However, these interventions will take time to show potential benefits.

7.5.4 Reducing teenage pregnancy and social determinants of health

Children of teenage mothers (<20 years old) have an increased risk of adverse birth characteristics such as preterm birth, or low birth weight.[56,57] Teenage conception rates have declined in England since the implementation of a 10-year Teenage Pregnancy Strategy in 1999.[222] The strategy focused on improved sex and relationships education, increasing access to effective contraception, better support for young parents, and

campaigns at a local and a national level.[222] As a result, conception rates in women aged <18 years old have halved by 2014 compared to 1999.[222] However, the proportion of teenage mothers was still much higher in England than in Sweden in the study period (6.1% vs 1.6%), indicating that further reductions could be achieved. Reducing teenage pregnancy is likely to require a multiagency approach across government and the society.

Young maternal age is strongly associated with social disadvantage, and early childbearing could limit a mother's educational and employment opportunities. Therefore, further reductions in teenage pregnancy rates will require addressing social determinants of health.[223] As outlined in Chapter 1, socio-economic disadvantage is associated with increased risks of adverse birth characteristics (such as preterm birth,[14,93] with low birth weight,[14,93] or a congenital anomaly)[13] and associated maternal characteristics (maternal obesity,[66] short maternal stature and low pregnancy weight gain,[14] smoking[224]), as well as the risk of child death after birth. The relative difference in incomes of the most deprived and the least deprived 20% of the population in the UK is almost twice that of Sweden.[97] Policies focussed on reducing child poverty and increasing welfare/social support for the most deprived mothers could, therefore, lead to reductions in adverse birth characteristics in England relative to Sweden. Further comparisons using causal mediation methods could determine the total effect of socio-economic factors on the risk of child death in England relative to Sweden (i.e. both including the effect mediated by low birth weight, preterm birth and congenital anomalies, and the direct effect).

7.6 Concluding remarks

The biggest reductions in child mortality in England relative to Sweden could be achieved by reducing the prevalence of adverse birth characteristics. Policies to reduce child mortality in England should focus on universal strategies to improve the health of women and on reducing socio-economic disadvantage before and after birth. This thesis has emphasised the importance of international comparisons using nationally-representative birth cohorts developed using administrative databases to determine how child mortality can most effectively be reduced.

Appendix A. Literature search terms for comparisons of child mortality in the UK and in Sweden

I searched PubMed for international comparisons of child mortality published from 2000, which included England, Great Britain or United Kingdom (UK) and Sweden in the analyses (and these countries were mentioned in study titles or abstracts). I identified additional studies by reviewing references of identified papers. Finally, I reviewed key papers on child mortality in the UK for comparisons with Sweden, and publications based on the data from the EURO-PERISTAT project, a collaboration between countries in the European Union (EU), which aimed to design and collect internationally comparable indicators of maternal and perinatal health.[12] I excluded studies which only used data from before 2000. I identified 14 studies. I then excluded 9 studies which did not attempt to identify the origins of differences in child mortality between the UK and Sweden.

Search terms:

(England[tiab] OR English[tiab] OR United Kingdom[tiab] OR UK[tiab]) AND (Sweden[tiab] OR Swedish[tiab]) AND (infant[tiab] OR neonatal[tiab] OR x`post-neonatal[tiab] OR child[tiab] OR childhood[tiab] OR under-5[tiab]) AND (death[tiab] OR mortality[tiab] OR dying[tiab] OR survival[tiab])

Table A.1 – Summary of the prevalence of key birth characteristics in live births in England and Wales by ethnic group category

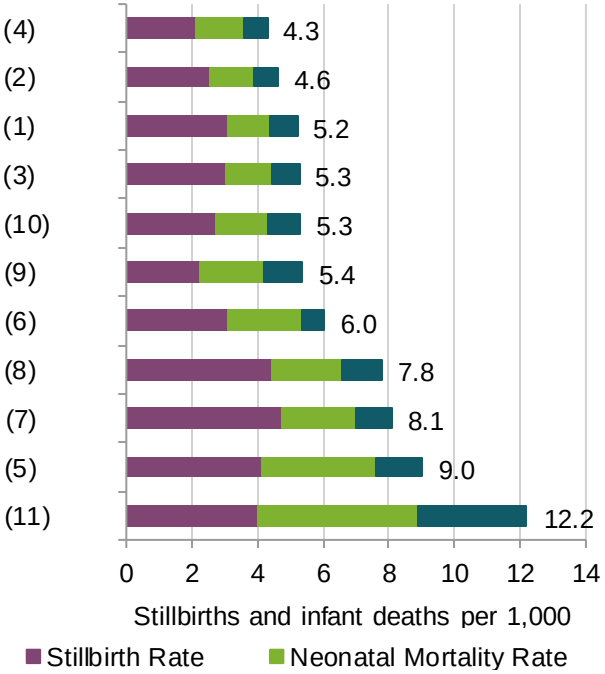
	All	White	Bangladeshi	Indian	Pakistani	African	Caribbean	All others	Not stated
% of all live births in 2010*		74.2%	1.3%	3.0%	3.7%	3.5%	1.0%	9.3%	4.0%
% of all infant deaths in 2010*		65.1%	1.6%	4.0%	8.0%	5.3%	1.9%	10.0%	4.2%
infant mortality (per 1000 live births) in 2010*	4.1	3.6	5.0	5.5	8. 8	6.2	7.8	4.4	4.3
% of preterm births*	7.0%	6.9%	7.1%	7.5%	6.9%	7.6%	9.5%	7.1%	7.2%
% of births with low birth weight (data from 2005)**	6.1%	6.0%	10.0%	10.5%	9.8%	7.4%	1.1%	7.0%	5.9%
Infant deaths from congenital anomalies per 1000 births in 2005**	1.3	1.0	1.6	1.8	4.8	1.7	0.9	1.6	1.4

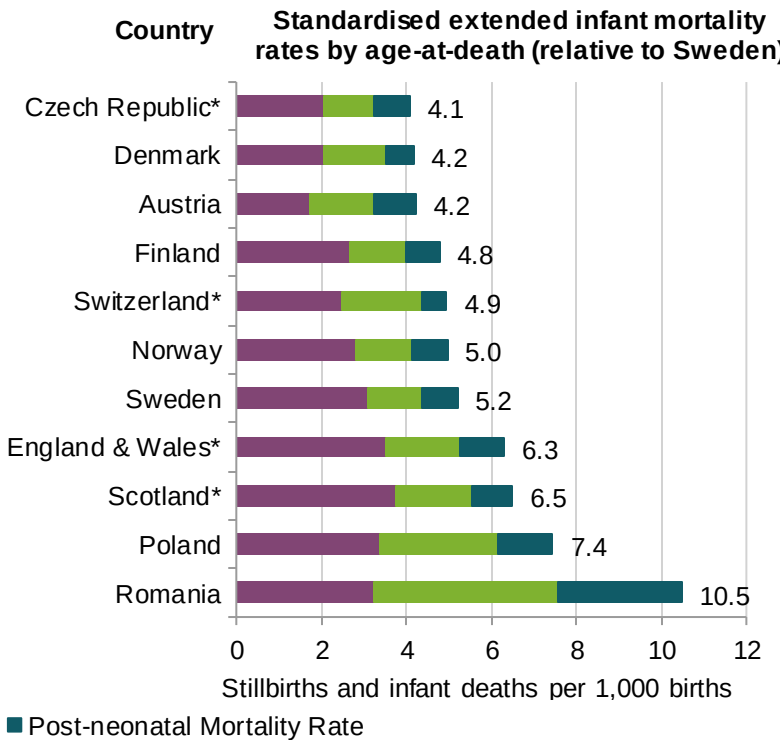
**Information from Office for National Statistics publication "Gestation-specific mortality[34] **Data from Moser et al.[76]*

Appendix B. Comparison of extended infant mortality in Europe using aggregate data tabulated by birth weight

Figure B.1 – Rankings of countries based on crude and birth weight-standardised extended infant mortality rates by age at death (low to high mortality rates)

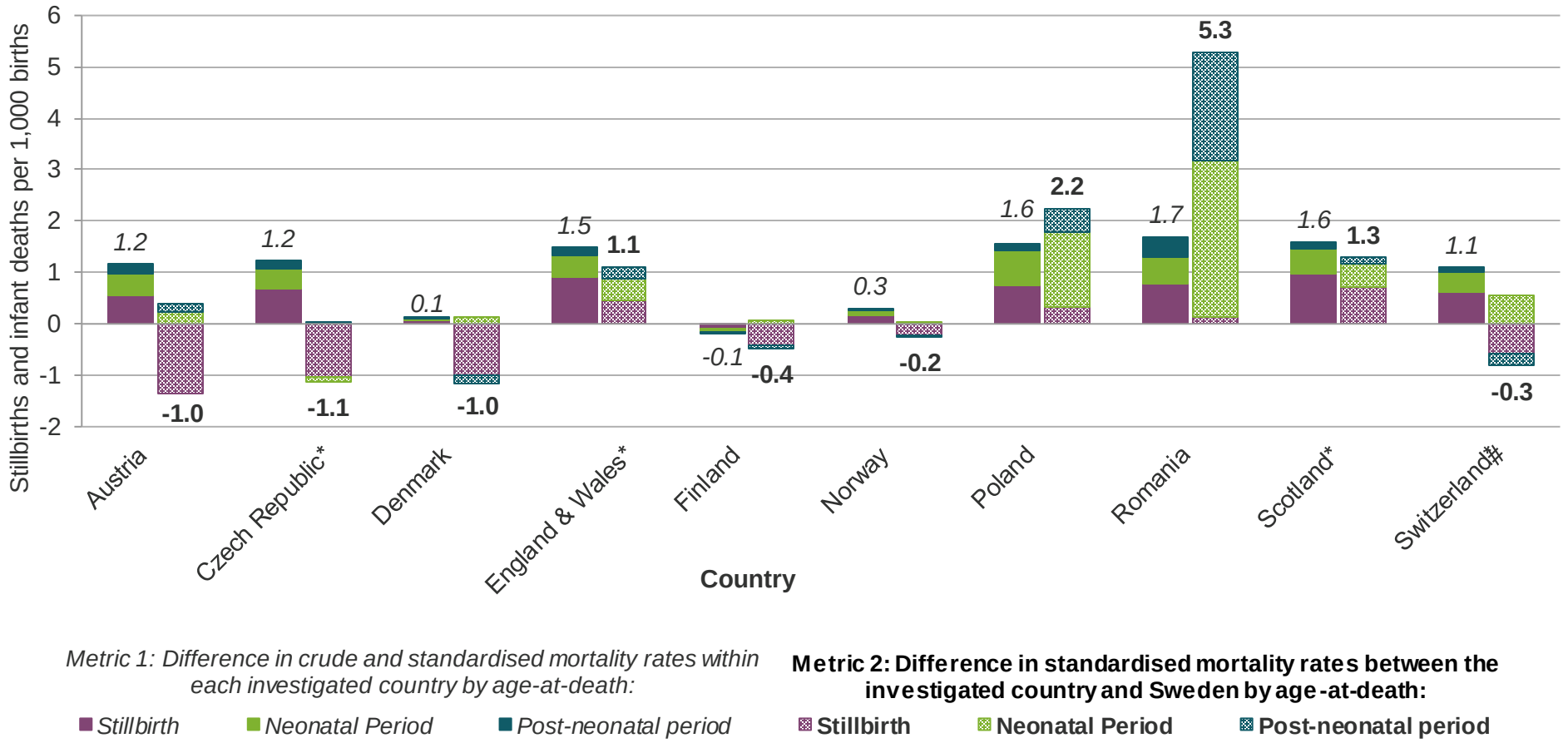
Country	Low birth weight rate (ranking)
Denmark	5.2% (4)
Finland	4.4% (2)
Sweden	4.3% (1)
Norway	5.1% (3)
Czech Republic*	7.7% (10)
Austria	7.1% (9)
Switzerland*	6.7% (6)
England & Wales*	7.1% (8)
Scotland*	6.8% (7)
Poland	5.9% (5)
Romania	8.2% (11)





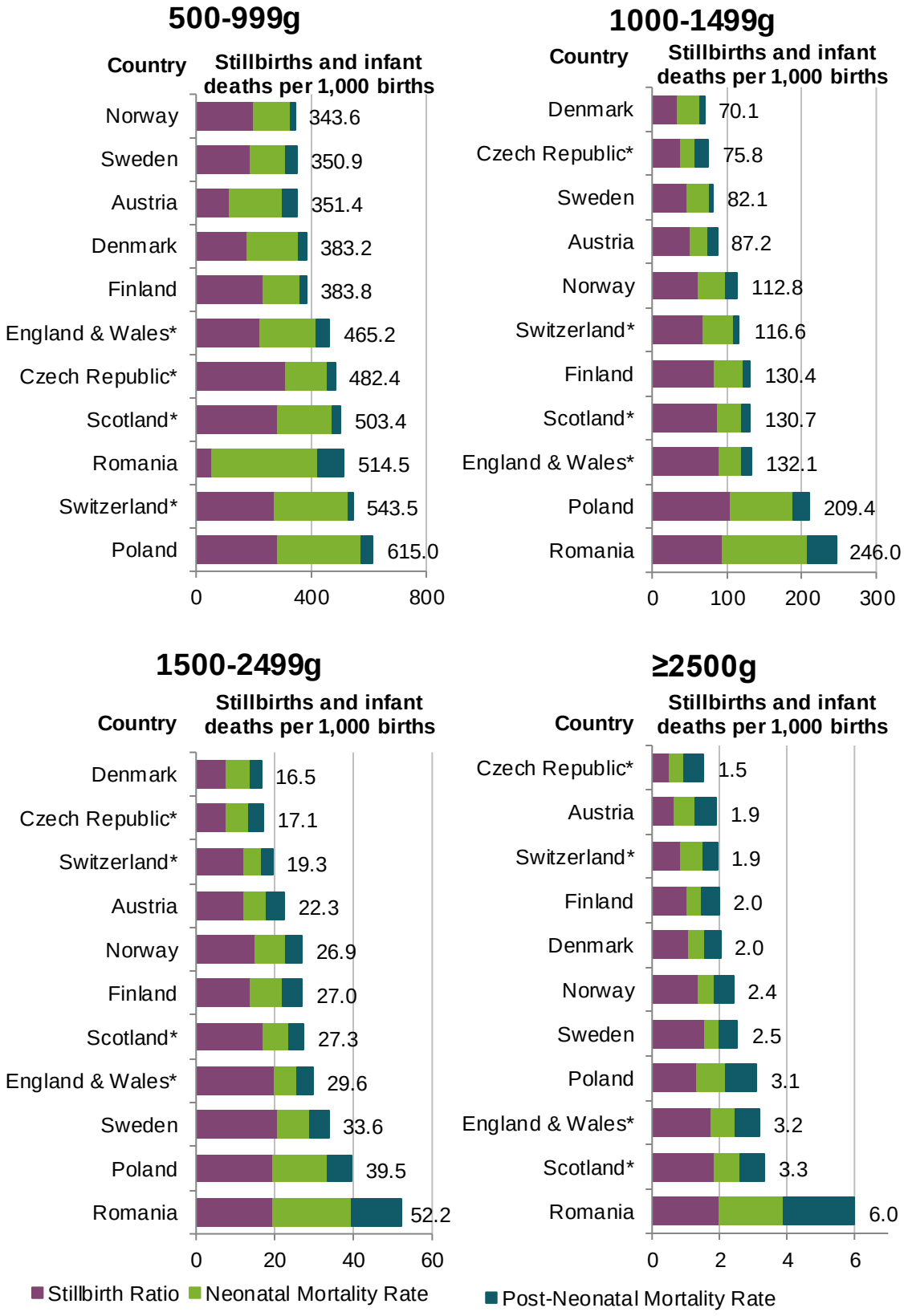
*Extended infant mortality was defined as the number of stillbirths and infant deaths per 1000 total births (live or still). The second column shows the proportion of total births with low birth weight (<2500g). In Poland, access to terminations of pregnancy (TOP) was restricted. Countries with * included TOP in their counts of stillbirths. England & Wales and Scotland included terminations of pregnancy and stillbirths only after 24 weeks. All calculations were done given birth weight was non-missing and ≥500g.*

Figure B.2 – Decomposition of the difference in crude extended infant mortality rates between each country and Sweden



*Extended infant mortality was defined as the number of stillbirths and infant deaths per 1000 total births (live or still). Bars on the left-hand side represent metric 1; bars on the right-hand side represent metric 2. In Poland, access to terminations of pregnancy (TOP) was restricted. Countries with * included TOP in their counts of stillbirths. England & Wales and Scotland included terminations of pregnancy and stillbirths only after 24 weeks. All calculations were done given birth weight was non-missing and ≥500g.*

Figure B.3 – Birth weight-specific extended infant mortality rates in each country by age at death



*Extended infant mortality was defined as the number of stillbirths and infant deaths per 1000 total births (live or still). In Poland, access to terminations of pregnancy (TOP) was restricted. Countries with * included TOP in their counts of stillbirths. England & Wales and Scotland included terminations of pregnancy and stillbirths only after 24 weeks. All calculations were done given birth weight was non-missing and ≥500g.*

Table B.1– Distribution of birth weight in total births in 11 compared countries in 2010

Country	500-1499g	1500-2499g	2500-4499g	≥4500g
Denmark	0.9%	4.3%	91.8%	3.0%
Finland	0.8%	3.7%	93.0%	2.5%
Norway	0.9%	4.2%	91.8%	3.2%
Sweden	0.8%	3.5%	92.1%	3.6%
Austria	1.2%	5.9%	92.0%	0.9%
Czech Republic	1.2%	6.5%	91.4%	0.9%
England & Wales	1.2%	5.9%	91.2%	1.7%
Poland	1.0%	4.9%	92.5%	1.5%
Romania	0.9%	7.3%	91.1%	0.7%
Scotland	1.2%	5.6%	91.1%	2.1%
Switzerland	1.0%	5.7%	92.5%	0.8%

Information comes from the EURO-PERISTAT project[12]

Appendix C. Supporting information for developing a birth cohort using HES-ONS dataset

C.1. Identifying births in HES

Table C.1 – Criteria for identifying all birth episodes in HES

Variable used		Inclusion Criteria (value recorded in HES and explanation)
Diagnostic codes (ICD-10)		Z38: Liveborn infants according to place of birth and type of delivery Z37: Outcome of delivery
Healthcare Resource Group Codes	**version 3.5**	N01: Neonates - Died <2 days old N02: Neonates with Multiple Minor Diagnoses N03: Neonates with one Minor Diagnosis N04: Neonates with Multiple Major Diagnoses N05: Neonates with one Major Diagnosis
	version 4.0 (in use since financial year 2011/12)	PB01Z: Major Neonatal Diagnoses PB02Z: Minor Neonatal Diagnoses PB03Z: Healthy Baby
HES Specific Fields	**Episode type**	3: Birth episode 6: Other birth event
	Patient classification	5: Mothers and babies using only delivery facilities
	Admission method	82: Other: babies born in health care provider 83: Other: babies born outside the health care provider, except when born at home as intended 2C: Baby born at home as intended (available from 2013/14)
	Neonatal Care	0: Normal care 1: Special care 2: Level 2 intensive care (high dependency intensive care) 3: Level 1 intensive care (maximal intensive care)

HES=Hospital Episode Statistics; ICD-10=International Statistical Classification of Diseases and Related Health Problems. Financial years in England run from 1st April to 31st March the following year.

Table C.2 – Exclusion criteria for multiple births, stillbirths and terminations of pregnancy

	Variable used		Exclusion criteria (value recorded in HES and explanation)
Multiple Births		Diagnostic codes (ICD-10)	Z372: Twins, both liveborn Z373: Twins, one liveborn and one stillborn Z374: Twins, both stillborn Z375: Other multiple births, all liveborn Z376: Other multiple births, some liveborn Z377: Other multiple births, all stillborn Z383: Twin, born in hospital Z384: Twin, born outside hospital Z385: Twin, unspecified as to place of birth Z386: Other multiple, born in hospital Z387: Other multiple, born outside hospital Z388: Other multiple, unspecified as to place of birth
	HES Specific Fields	Birth order	greater than 1 (birordr>1)
		Number of babies	more than 1 (numbaby>1)
Termination of pregnancy		Diagnostic codes (ICD-10)	P964: Termination of pregnancy, affecting foetus and newborn
Stillbirth		Diagnostic codes (ICD-10)	P95: Foetal death of unspecified cause Z371: Single stillbirth Z373: Twins, one liveborn and one stillborn Z374: Twins, both stillborn Z376: Other multiple births, some liveborn Z377: Other multiple births, all stillborn
	HES Specific Fields	Discharge method	5: Baby was stillborn
		Birth status	2: Stillbirth: ante-partum 3: Stillbirth: intra-partum 4: Stillbirth: indeterminate

HES=Hospital Episode Statistics; ICD-10=International Statistical Classification of Diseases and Related Health Problems. Financial years in England run from 1st April to 31st March the following year.

Table C.3 – Cleaning rules for episodes of care identified as births

Criterion	Action
Gestation<22 or >45 weeks	Change to missing
Birth weight <200g or >7000g	Change to missing
Maternal Age <10 or >60 years	Change to missing
Exact duplicates	Drop duplicates in terms of: HESID, age at start and end of admission, month and year of birth, sex, county of residence, IMD rank, all recorded diagnoses, all recorded operations, cause, birth weight, gestation, maternal age, provider code, episode start and end dates, episode order, admission and discharge dates
Admission date missing	Replace to episode start date if epiorder=1 Replace to admission date from episode with closest episode start date if epiorder!=1 Else, replace with episode start date
Episode start date missing	No such cases
Episode start> episode end	Replace episode start with admission date if the issue is with the recording of episode start (episode start > episode end ≥ admission date) Replace episode end with episode start date if the issue is with the recording of episode end (episode start = admission date > episode end) Switch episode start with episode end, and admission date with discharge date if they were incorrectly recorded (episode start > episode end & admission date > discharge date where discharge date is not missing)
Admission date > episode start	Replace episode start date with admission date
Admission date > episode end	Replace episode end date with episode start date
Age at start of episode > age at end of episode	Switch age at start with age at end of episode

HES=Hospital Episode Statistics; IMD=Index of Multiple Deprivation.

C.2 Longitudinal follow-up data until fifth birthday

Table C.4 – Cleaning rules for HES longitudinal records

	Criterion	Action
Drop episodes with no clinical information recorded	Unfinished episodes	Drop (as usually more complete record was available)
	Missing episode end date	Drop (as usually more complete record was available)
	Only clinical information recorded was diagnosis "R69" – "*Illness, unspecified*"	Drop (as usually more complete record was available)
	No recorded diagnoses	Drop (as usually more complete record was available)
Validate and correct date variables (admission and discharge dates, episode start and end dates)	Admission date missing	Replace to episode start date for the first episode of the admission (epiorder=1) Else, replace to admission date from episode with episode order smaller by 1 and closest episode start date Else, replace with episode start date
	Episode start date missing	Replace to admission date for the first episode of the admission (epiorder=1) Else, replace to episode end date from another episode with the same admission date and lower episode order
	Episode end date missing	Removed as part of exclusion criteria
	Episode start> episode end	Replace episode start with admission date if the issue is with the recording of episode start (episode start > episode end ≥ admission date) Replace episode end with episode start date if the issue is with the recording of episode end (episode start = admission date > episode end) Switch episode start with episode end, and admission date with discharge date if they were incorrectly recorded (episode start > episode end & admission date > discharge date where discharge date is not missing)
	Admission date > episode start	Replace episode start date with admission date
	Admission date > episode end	Replace episode end date with episode start date

HES=Hospital Episode Statistics; IMD=Index of Multiple Deprivation. The table continues overleaf.

Table C.4 (continued) – Cleaning rules for HES longitudinal records

	Criteria	Action
Validate and correct date variables (admission and discharge dates, episode start and end dates)	Discharge date missing	Discharge date is recorded only on the last episode of care. Therefore, I generated a maximum discharge date by HESID and admission date as the "complete" discharge date. If "complete" discharge date was missing (when discharge date was not recorded for an admission), I replaced it with maximum episode end date by HESID and admission date. If "complete" discharge date was smaller than the maximum episode end date, I replaced it with the maximum episode end date.
	Episode ends in a different year than it starts	Drop if the difference is greater than or equal to two. It seems impossible to be seen by only one consultant while staying in the hospital for 2 years so it must be a data error.
	Missing episode start age	No such observations
	Missing episode end date	Generate an age using episode start and end dates for episodes with startage=7001 ("less than 1 day")
	Age at start of episode > age at end of episode	Switch age at start with age at end of episode
	Epistart – Epiend > 365	Episodes that lasted more than 1 year were assumed to be recording errors and dropped as it is unlikely that a patient would be seen by just one consultant for that long.
Drop duplicates	Exact duplicates	Drop duplicates in terms of: HESID, age at start and end of admission, month and year of birth, gender, post code, start and end date of the episode, episode order, admission and discharge dates, provider code,all diagnoses and operations and cause of injury

HES=Hospital Episode Statistics; IMD=Index of Multiple Deprivation.

C.3. Improving the completeness of risk factor variables using mother-baby linkage in HES

C.3.1. Identifying mothers and babies in HES

Harron *et al.*[154] identified maternal delivery admissions by searching for records for women aged 12-50 with a diagnosis indicating birth, OPCS codes indicating delivery procedures, or two or more complete and valid fields recorded in the baby tails. Selection criteria for identifying birth episodes largely overlapped with my selection criteria. However, there were small differences in how we specified the birth cohorts (listed in table C.5), due to different aims of our cohorts. For example, I was interested in an accurate date of birth and in ensuring that all likely links to ONS mortality data were included in the HES-ONS birth cohort. Therefore, I excluded episodes of care with age at admission >6 days a priori, which were included by Harron *et al.*[154] if they were indicated as births using any of the criteria. I also included misclassified stillbirths (if they were linked to an ONS mortality record with a high-quality match rank), which were not included in the cohort of Harron *et al.*[154]

Table C.5 – Comparison of inclusion criteria in my cohort and in Harron *et al.*[154]

Codeset	Value	Inclusion criteria*	Harron *et al.*	My cohort
Diagnoses	Z37	Outcome of delivery	YES	YES
	Z38	Live born infant	YES	YES
HES specific fields	epitype	3: Birth event	YES	YES
		6: Other birth event	YES	YES
	admimeth	82: Other: babies born in health care provider	YES	YES
		83: Other: babies born outside the health care provider, except when born at home as intended	YES	YES
		2C: Baby born at home as intended (available from 2013/14)	NO	YES
	startage	7001: <1 day	YES	Condition of inclusion in the cohort
		7002: 1-6 days	YES	
	neocare	0: Normal care	YES	YES
		1: Special care	YES	YES
		2: Level 2 intensive care	YES	YES
		3: Level 1 intensive care	YES	YES
HRG version 3.5	N01	Neonates – died <2 days old	YES	YES
	N02	Neonates with multiple minor diagnoses	YES	YES
	N03	Neonates with one minor diagnosis	YES	YES
	N04	Neonates with multiple major diagnoses	YES	YES
	N05	Neonates with one major diagnosis	YES	YES
HRG version 4.0	PB01Z	Major Neonatal Diagnoses	NO	YES
	PB02Z	Minor Neonatal Diagnoses	NO	YES
	PB03Z	Healthy Baby	NO	YES

HES=Hospital Episode Statistics; HRG=Healthcare Resource Group.

To ensure that the linkage was done correctly, Dr Wijlaars replicated the birth cohort developed by Harron *et al.*[154] and used that cohort for linkage. I then validated the number of births and linkage rate against results of Harron *et al.*[154] in the replicated cohort, and merged the linked maternal records with my HES-ONS birth cohort using baby's HESID. This meant that linkage was not attempted for 1,040 records from my HES-ONS birth cohort. These were primarily records for misclassified stillbirths included in my cohort, and not included in birth cohort of Harron *et al.*[154] These

misclassified stillbirths were unlikely to link with a maternal delivery record due to missing postcode (and no longitudinal hospital admissions to enhance completeness of the postcode), and poor recording of variables in the baby tail.

C.3.2. Deterministic linkage

Theory

Deterministic linkage requires an exact or approximate agreement between a set of identifiers such as date of birth, postcode or sex. Records are matched if the identifiers agree and not matched if they disagree. It is usually unlikely that two individuals will have the same set of identifiers, therefore deterministic linkage produces a low rate of false matches (which occurs if two individuals are identified as one based on their identifiers).[154]

Deterministic linkage of mothers and babies in HES

Mothers and babies in HES were deterministically linked if both records had identical information on GP practice, maternal age, birth weight, gestation, birth order or sex. The records would be linked if there was missing data, given that at least 3 of the variables agreed and there were no disagreements.[154]

C.3.3. Probabilistic linkage

Theory

Probabilistic linkage looks at the likelihood of a given pair of records belonging to the same individual. Probabilistic linkage often uses a larger number of potential identifiers than deterministic linkage and accounts for differences in their discriminative value (e.g., the NHS number can better distinguish between individuals than their sex).

The likelihood is estimated by match weights, calculated for all combinations of pairs of records. To calculate match weights, each identifier is first assigned two probabilities:

- a probability of agreement between records, given that they belong to the same individual (M-probability), which is estimated during linkage process and updated as more links are made
- a probability of agreement between records, given that they belong to different individuals (U-probability), which can be approximated as a probability of chance agreement. For example, there is 50% chance that two records have the same sex.

The two probabilities are then combined as $log_2(m/u)$ for each identifier and summed over all identifiers to produce an overall match weight for a given pair of records. The

higher the match weights, the larger the likelihood that two records belong to the same individual. Usually the record with the highest match weight is kept. Records with lower match weights are classified as links or non-links given a pre-specified cut-off. [154]

Probabilistic linkage of mothers and babies in HES

Probabilistic linkage of mothers and babies in HES was based on 23 variables including maternal and child details (such as ethnicity, postcode, maternal age, birth weight, gestational age, sex of the baby) and details of delivery and birth (e.g., intended delivery place, hospital, delivery date). For categorised gestational age, intended delivery place, status of person conducting delivery, first letter of postcode district and ethnic category, Harron *et al.* used frequency-based match weights, which give higher weights to agreement of rare values (e.g., low gestational age vs term birth). For linkage, we used M- and U-probabilities and cut-off for identifying links estimated by Harron *et al.*[154]

C.3.4. Using blocks for making linkage more computationally efficient

Theory

Record linkage compares all possible pairs of records in two datasets. Even in two small datasets, the number of all possible combinations get very large and make linkage computationally intensive. For example, two small datasets with 1,000 individuals each, would contribute to 1,000x1,000=1,000,000 pairs. Running linkage algorithm within mutually exclusive blocks of data can be used to narrow down the number of pairs (known as candidate matches) and speed up the linkage process.

Blocks used for mother-baby linkage in HES

Mother-baby linkage was done within each HES year. Within each year, Harron *et al.*[154]proposed blocking the records by hospital and including only mother-baby pairs where the estimated dates of delivery and dates of birth were plausible (e.g., where the date of birth was not earlier than date of delivery). This strategy was then relaxed for any remaining pairs of unlinked mothers and babies.

C.4. Restricting the cohort to hospitals providing high quality data on risk factors

I explored a number of additional exclusion criteria listed below. For each, I examined histograms to learn about the distribution of the indicator amongst all hospitals (in each financial year). I looked at scatter plots of proportions of recorded births and deaths with complete information on birth weight and gestational age against each of the indicators to see if they explained the missing data patterns. Based on the visual examination, I selected cut-off points for each indicator, to distinguish between hospitals with "good" and "bad" quality of data.

First, I investigated indicators for the quality of recorded identifiers and linkage with ONS mortality data in each hospital:

- The proportion of infant deaths only recorded in HES, not linked ONS death record (beyond days 0-1 of death)
- The rate of missing sex in birth records per hospital
  Sex of a patient is one of variables recorded in all episodes of care in HES. A high rate of missing data on sex could indicate systematic data quality issues.
- Proportion of records with present and validated NHS number
  I used *nhsnoind* variable in HES, however this measure is not perfect, as this field was missing for some of the hospitals.

Next, I investigated indicators for the quality of recorded birth weight and gestational age:

- The proportion of records with implausible birth weights for given gestational age
- The proportion of births with low or extremely low birth weight (<2500g and <1000g, respectively), born preterm (<37 weeks) or extremely preterm (<28 weeks) to identify hospitals where there could be bias in recording variables for more vulnerable children. A similar strategy was adopted in a study of parity, where hospitals with good quality of recorded data were selected based on the within-hospital ratio of primiparous to multiparous women – hospitals with the ratio outside the expected range of values were excluded.[225]
- The proportion of babies admitted to neonatal intensive care units
  I hypothesised that these infants might be less likely to have a recording of birth weight or gestational age, as these data are likely to be reported to the Neonatal Research Database (NNRD) kept by the Neonatal Data Analysis Unit instead of HES.[176] I indicated these children using *neocare* variable in HES, however, not all hospitals report this variable.

- The proportion of births that did not link to delivery records (indicating poor recording of variables in the baby tail)
- The proportion of "vulnerable" children defined as children with congenital anomalies (identified using methods explained in chapters 3 and 4) or with a chronic condition diagnosed in the neonatal period (defined using codelist developed by Hardelid et al[58]).
- Proportion of babies with complete information that were transferred to another hospital following birth admission – I hypothesised that they might be less likely to have a recording of birth weight or gestational age.
- Completeness of birth weight and gestational age in deaths on days 2-6, 7-27 and 28-364. I looked separately at these age at death categories as they showed different missing data patterns, thus the mechanisms behind missing data might also be different.

Appendix D. Developing a birth cohort using Swedish National Registers

D.1. Ethics approval to use the registers from the Regional Committee of Stockholm

Regionala etikprövningsnämnden i Stockholm | PROTOKOLL Avdelning 5 | 2016/5:7 2016-08-04

Ordförande
Agneta Eberhardt

Ledamöter med vetenskaplig kompetens
Claes-Robert Julander *(företagsekonomi)*, vetenskaplig sekreterare
Stephan Hau *(psykologi)*, deltar inte i ärende 2016/1208, 2016/1431
Katrin Goldstein Kyaga *(etnicitet)*
Sven Ove Hansson *(filosofi)*, deltar inte i ärende 2016/1295
Gert Helgesson *(medicinsk etik)*
Staffan Marklund *(arbetsliv)*, deltar inte i ärende 2016/1353
Jerzy Sarnecki *(kriminologi)*
Marianne Sundström *(nationalekonomi)*

Ledamöter som företräder allmänna intressen
Bo Bängtsson
Gilbert de Wendel
Margaretha Hertelius
Maria Modig
Marianne Upmark

Administrativ sekreterare
Helena Hallgren Lönn

§ 1 Ordföranden förklarar sammanträdet öppnat.

§ 2 Ansökningar om etisk granskning av forskningsprojekt, se **Bilaga**.

§ 3 Med stöd av 27 § tredje stycket etikprövningslagen beslutas att lämna över till ordförande och den vetenskaplige sekreteraren att ompröva beslut enligt 27 § förvaltningslagen (1986:223)

§ 4 Ordföranden förklarar mötet avslutat och meddelar att nästa sammanträde i avdelning 5 äger rum torsdagen den 8 september 2016.

Agneta Eberhardt
Ordförande

Claes-Robert Julander
Protokollförare, vetenskaplig sekreterare

Postadress FE 289, 171 77 STOCKHOLM
Org. nummer 202200 1578
Paket-/Besöksadress Tomtebodavägen 18A, plan 3, 171 65 SOLNA
Telefon 08-524 870 00
Fax 08-524 866 99
E-Post kansli@stockholm.epn.se
Webb www.epn.se

Diarienummer:
2016/1234-31/5
Föredragande:
Sven Ove Hansson

Sökande: Stockholms universitet
Behörig företrädare: Jenny Eklund
Projekt: Varför är risken att dö för barn under fem år högre i Storbritannien än i Sverige?
Forskare som genomför projektet: Anders Hjern

BESLUT
Nämnden godkänner forskningen.

Postadress
FE 289
171 77 STOCKHOLM

Paket-/Besöksadress
Tomtebodavägen 18A plan 3
171 65 SOLNA

Telefon
08-524 800 00

E-Post
kansli@stockholm.epn.se

Org. nummer
202200 1578

Fax
08-524 866 99

Webb
www.epn.se

D.2. Cleaning rules for identifying implausible combinations of birth weight for gestational age

Implausible combinations of birth weight for gestational age were defined as:

- Birth weight <500g (and non-missing) and birth length ≥42 cm (and non-missing) and gestational age ≥34 weeks (and non-missing)
- Birth weight <1000g (and non-missing) and birth length ≥45 cm (and non-missing) and gestational age ≥37 weeks (and non-missing)
- Birth weight <1500g (and non-missing) and birth length ≥48 cm (and non-missing) and gestational age ≥38 weeks (and non-missing)
- Birth weight ≥1500g (and non-missing) and birth length ≥35 cm (and non-missing) and gestational age <24 weeks (and non-missing)
- Birth weight ≥2000g (and non-missing) and birth length ≥40 cm (and non-missing) and gestational age <26 weeks (and non-missing)
- Birth weight ≥2500g (and non-missing) and birth length ≥43 cm (and non-missing) and gestational age <28 weeks (and non-missing)
- Birth weight ≥3000g (and non-missing) and birth length ≥45 cm (and non-missing) and gestational age <29 weeks (and non-missing)
- Birth weight ≥3500g (and non-missing) and birth length ≥46 cm (and non-missing) and gestational age <30 weeks (and non-missing)
- Birth weight >2500g (and non-missing) and birth length ≥48 cm and gestational age of 36 weeks' gestation

Appendix E. Supporting information for the comparison of child mortality in England and Sweden

Table E.1 – Distribution of live births within each risk factor category by quintiles of socio-economic status in England and in Sweden in 2003-2012

	England						Sweden					
Risk factor	**Number of births**	**Q1: most deprived**	**Q2**	**Q3**	**Q4**	**Q5: least deprived**	**Number of births**	**Q1: most deprived**	**Q2**	**Q3**	**Q4**	**Q5: least deprived**
Birth weight (g)												
500-999	9,458	29%	24%	19%	16%	13%	1,742	27%	20%	16%	18%	18%
1000-1499	18,288	28%	23%	19%	16%	14%	3,102	25%	19%	18%	19%	19%
1500-2499	190,299	29%	23%	19%	16%	13%	25,817	25%	19%	17%	19%	20%
2500-3499	2,090,583	24%	21%	19%	18%	17%	429,107	23%	19%	19%	19%	20%
≥3500	1,624,258	18%	19%	20%	21%	22%	553,592	17%	20%	21%	21%	20%
Gestational age (weeks)												
24-27	8,806	29%	24%	19%	16%	13%	1,769	27%	19%	16%	19%	19%
28-31	22,327	27%	23%	19%	17%	15%	4,354	24%	19%	18%	19%	19%
32-34	56,093	26%	22%	19%	17%	16%	11,764	21%	18%	18%	21%	21%
35-36	137,046	25%	22%	19%	18%	16%	30,295	21%	19%	19%	21%	20%
37-38	726,907	23%	21%	19%	19%	18%	191,130	21%	20%	20%	20%	20%
≥39	2,981,707	21%	20%	20%	20%	19%	774,048	20%	20%	20%	20%	20%

All data are % of live births in a given risk factor category. Column totals may not add up to 100% due to rounding. Table continues overleaf.

Table E.1 (continued) – Distribution of live births in each risk factor category by quintiles of socio-economic status in England and in Sweden in 2003-2012

	England						Sweden					
Risk factor	**Number of births**	**Q1: most deprived**	**Q2**	**Q3**	**Q4**	**Q5: least deprived**	**Number of births**	**Q1: most deprived**	**Q2**	**Q3**	**Q4**	**Q5: least deprived**
Sex												
Boy	2,016,683	22%	20%	20%	19%	19%	520,985	20%	20%	20%	20%	20%
Girl	1,916,203	22%	20%	20%	19%	19%	492,375	20%	20%	20%	20%	20%
Congenital anomaly												
No	3,817,789	22%	20%	20%	19%	19%	988,681	20%	20%	20%	20%	20%
Yes	115,097	24%	21%	19%	18%	18%	24,679	21%	20%	19%	20%	20%
Maternal age (years)												
<20	241,503	36%	26%	18%	12%	7.8%	16,160	77%	17%	4.7%	1.5%	0.74%
20-25	758,596	32%	25%	19%	14%	9.8%	129,240	37%	26%	20%	13%	3.8%
25-30	1,064,469	23%	22%	20%	18%	16%	295,905	19%	21%	23%	22%	15%
30-35	1,110,202	15%	18%	20%	23%	25%	356,356	14%	18%	20%	22%	26%
35-40	617,394	13%	16%	19%	24%	28%	178,992	16%	17%	18%	21%	29%
≥40	140,722	14%	17%	19%	23%	27%	36,707	19%	18%	15%	17%	30%

All data are % of live births in a given risk factor category. Column totals may not add up to 100% due to rounding.

Table E.2 – Unadjusted mortality rates per 100,000 child-years (95% confidence intervals) overall and by risk factors at birth in England and Sweden in 2003-2012

	2-27 days		28-364 days		1-4 years	
	England	**Sweden**	**England**	**Sweden**	**England**	**Sweden**
Overall	1500 (1500, 1600)	920 (850, 990)	140 (130, 140)	86 (80, 92)	19 (18, 20)	15 (13, 16)
Birth weight (g)						
500-999	180000 (170000, 190000)	110000 (94000, 130000)	9000 (8300, 9700)	5300 (4200, 6600)	230 (180, 310)	81 (30, 220)
1000-1499	31000 (28000, 35000)	30000 (23000, 38000)	1700 (1500, 1900)	1500 (1100, 2100)	98 (75, 130)	93 (49, 180)
1500-2499	5800 (5400, 6300)	7400 (6200, 8800)	550 (520, 590)	570 (490, 680)	54 (48, 61)	58 (44, 77)
2500-3499	980 (930, 1000)	680 (590, 780)	120 (110, 120)	87 (78, 96)	20 (19, 21)	16 (14, 18)
≥3500	490 (450, 530)	310 (260, 370)	54 (50, 58)	41 (35, 47)	12 (12, 14)	11 (9.8, 13)
Gestational age (weeks)						
24-27	190000 (170000, 200000)	120000 (98000, 140000)	9000 (8300, 9800)	5400 (4300, 6700)	240 (180, 320)	60 (19, 190)
28-31	29000 (26000, 32000)	21000 (17000, 27000)	1600 (1400, 1700)	990 (720, 1400)	71 (53, 95)	52 (25, 110)
32-34	8200 (7300, 9100)	7700 (6000, 9900)	620 (560, 690)	470 (360, 620)	43 (34, 54)	62 (41, 93)
35-36	3600 (3300, 4000)	3400 (2700, 4200)	350 (320, 380)	290 (230, 360)	35 (30, 41)	29 (20, 42)
37-38	1400 (1300, 1500)	880 (730, 1100)	170 (160, 180)	110 (93, 120)	24 (22, 26)	17 (14, 20)
≥39	650 (610, 680)	370 (320, 420)	78 (75, 82)	51 (46, 57)	16 (15, 17)	13 (11, 14)
Sex						
Boy	1700 (1600, 1800)	1000 (910, 1100)	150 (150, 160)	96 (87, 100)	20 (19, 21)	16 (14, 18)
Girl	1400 (1300, 1400)	810 (720, 920)	120 (120, 130)	76 (69, 85)	18 (17, 19)	14 (12, 16)

Table continues overleaf.

Table E.2 continued – Unadjusted mortality rates per 100,000 child-years (95% confidence intervals) overall and by risk factors at birth in England and Sweden in 2003-2012

	2-27 days		28-364 days		1-4 years	
	England	**Sweden**	**England**	**Sweden**	**England**	**Sweden**
Overall	1500 (1500, 1600)	920 (850, 990)	140 (130, 140)	86 (80, 92)	19 (18, 20)	15 (13, 16)
Congenital anomalies						
No	890 (860, 930)	530 (480, 590)	77 (74, 80)	52 (48, 57)	12 (12, 13)	11 (10, 13)
Yes	23000 (22000, 24000)	17000 (15000, 19000)	2200 (2100, 2300)	1500 (1300, 1600)	260 (250, 280)	150 (130, 190)
Maternal age (years)						
<20	2200 (2000, 2400)	980 (540, 1800)	260 (240, 280)	230 (160, 320)	25 (22, 29)	31 (19, 50)
20-25	1700 (1600, 1800)	1100 (930, 1400)	170 (160, 180)	130 (110, 150)	25 (23, 27)	18 (14, 23)
25-30	1500 (1400, 1600)	790 (680, 920)	130 (120, 130)	86 (75, 97)	18 (17, 20)	14 (12, 16)
30-35	1300 (1200, 1300)	890 (780, 1000)	110 (100, 120)	67 (59, 77)	17 (16, 19)	13 (11, 16)
35-40	1500 (1400, 1600)	860 (720, 1000)	110 (100, 120)	76 (64, 91)	16 (14, 18)	14 (11, 18)
≥40	2100 (1800, 2400)	1600 (1200, 2200)	160 (140, 180)	120 (84, 160)	17 (13, 22)	18 (11, 27)
Quintile of socio-economic status						
Q1: most deprived	210 (200, 220)	120 (100, 140)	20 (19, 21)	14 (12, 15)	2.5 (2.3, 2.7)	1.8 (1.5, 2.1)
Q2	170 (160, 180)	92 (78, 110)	17 (16, 17)	9.1 (7.8, 11)	2.1 (1.9, 2.2)	1.8 (1.5, 2.1)
Q3	140 (130, 150)	62 (50, 77)	12 (11, 13)	5.7 (4.7, 6.9)	1.8 (1.7, 2.0)	1.5 (1.2, 1.8)
Q4	120 (120, 130)	65 (53, 80)	10 (9.7, 11)	7 (5.9, 8.3)	1.7 (1.5, 1.9)	1.3 (1.0, 1.6)
Q5: least deprived	110 (100, 120)	120 (100, 140)	8.3 (7.6, 9.0)	7.7 (6.5, 9.0)	1.4 (1.3, 1.6)	1.1 (0.83, 1.3)

Table E.3 – Unadjusted and adjusted Cox PH models for all-cause mortality at 2-27 days in low-risk babies in England relative to Sweden in 2003-2012

Risk factor	Model 1	Model 2	Model 3
Country			
England	1.87 (1.51, 2.32)	1.71 (1.38, 2.12)	1.64 (1.32, 2.03)
Sweden (baseline)	1	1	1
Birth weight (g)			
2500-2999		2.60 (1.97, 3.43)	2.40 (1.82, 3.17)
3000-3499		1.44 (1.12, 1.85)	1.37 (1.07, 1.77)
3500-3999		1.06 (0.82, 1.36)	1.03 (0.80, 1.34)
4000-4499 (baseline)		1	1
≥4500		1.29 (0.78, 2.13)	1.30 (0.79, 2.15)
Gestational age (weeks)			
39 (baseline)		1	1
40		1.20 (1.01, 1.42)	1.18 (1.00, 1.40)
41		1.40 (1.16, 1.68)	1.38 (1.14, 1.66)
Sex			
Boy		1.37 (1.19, 1.58)	1.36 (1.19, 1.57)
Girl (baseline)		1	1
Maternal age (years)			
<20			1.57 (1.18, 2.09)
20-24			1.45 (1.18, 1.78)
25-29			1.12 (0.92, 1.37)
30-34 (baseline)			1
35-39			1.11 (0.88, 1.41)
≥40			1.42 (0.97, 2.09)
Quintile of socio-economic status			
Q1: most deprived			1.32 (1.06 1.65)
Q2			1.09 (0.87, 1.37)
Q3			0.93 (0.73, 1.19)
Q4			1.00 (0.79, 1.27)
Q5: least deprived (baseline)			1

PH=proportional hazards. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Low-risk babies were defined as born at full term (39-41 weeks), with normal birth weight (>2500g), with no congenital anomaly.

Table E.4 – Unadjusted and adjusted Cox PH models for all-cause mortality at 28-364 days in low-risk babies in England relative to Sweden in 2003-2012

Risk factor	Model 1	Model 2	Model 3
Country			
England	1.48 (1.28, 1.70)	1.34 (1.16, 1.55)	1.19 (1.03, 1.38)
Sweden (baseline)	1	1	1
Birth weight (g)			
2500-2999		2.66 (2.17, 3.28)	2.19 (1.78, 2.69)
3000-3499		1.60 (1.33, 1.93)	1.42 (1.17, 1.71)
3500-3999		1.09 (0.89, 1.32)	1.03 (0.85, 1.25)
4000-4499 (baseline)		1	1
≥4500		0.94 (0.61, 1.45)	0.97 (0.63, 1.49)
Gestational age (weeks)			
39 (baseline)		1	1
40		0.99 (0.88, 1.12)	0.96 (0.85, 1.08)
41		0.95 (0.83, 1.09)	0.91 (0.79, 1.04)
Sex			
Boy		1.51 (1.36, 1.67)	1.48 (1.34, 1.64)
Girl (baseline)		1	1
Maternal age (years)			
<20			3.45 (2.88, 4.13)
20-24			2.05 (1.76, 2.39)
25-29			1.33 (1.14, 1.55)
30-34 (baseline)			1
35-39			1.00 (0.82, 1.22)
≥40			1.39 (1.01, 1.90)
Quintile of socio-economic status			
Q1: most deprived			1.79 (1.49, 2.14)
Q2			1.63 (1.36, 1.96)
Q3			1.30 (1.07, 1.57)
Q4			1.21 (1.00, 1.48)
Q5: least deprived (baseline)			1

PH=proportional hazards. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Low-risk babies were defined as born at full term (39-41 weeks), with normal birth weight (>2500g), with no congenital anomaly.

.

Table E.5 – Unadjusted and adjusted Cox PH models for all-cause mortality at 1-4 years in low-risk babies in England relative to Sweden in 2003-2012

Risk factor	Model 1	Model 2	Model 3
Country			
England	1.00 (0.86, 1.15)	0.94 (0.81, 1.09)	0.91 (0.78, 1.05)
Sweden (baseline)	1	1	1
Birth weight (g)			
2500-2999		2.00 (1.55, 2.57)	1.84 (1.43, 2.38)
3000-3499		1.54 (1.24, 1.92)	1.47 (1.18, 1.83)
3500-3999		1.31 (1.05, 1.63)	1.29 (1.03, 1.61)
4000-4499 (baseline)		1	1
≥4500		0.95 (0.57, 1.56)	0.95 (0.58, 1.57)
Gestational age (weeks)			
39 (baseline)		1	1
40		0.93 (0.81, 1.07)	0.92 (0.80, 1.06)
41		0.99 (0.84, 1.15)	0.97 (0.83, 1.14)
Sex			
Boy		1.14 (1.01, 1.29)	1.13 (1.01, 1.28)
Girl (baseline)		1	1
Maternal age (years)			
<20			1.46 (1.15, 1.86)
20-24			1.13 (0.95, 1.35)
25-29			0.98 (0.83, 1.15)
30-34 (baseline)			1
35-39			0.95 (0.78, 1.15)
≥40			1.05 (0.73, 1.49)
Quintile of socio-economic status			
Q1: most deprived			1.58 (1.29, 1.93)
Q2			1.43 (1.17, 1.75)
Q3			1.25 (1.01, 1.54)
Q4			1.21 (0.98, 1.50)
Q5: least deprived (baseline)			1

PH=proportional hazards. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Low-risk babies were defined as born at full term (39-41 weeks), with normal birth weight (>2500g), with no congenital anomaly.

Table E.6 – Unadjusted and adjusted Cox PH models for all-cause mortality at 2-27 days in England relative to Sweden in 2003-2012 using an indicator of severe congenital anomalies

	Model 1	Model 2	Model 3	Model 4
Country				
England	1.66 (1.53, 1.81)	1.37 (1.26, 1.48)	1.17 (1.07, 1.27)	1.14 (1.05, 1.24)
Sweden (baseline)	1	1	1	1
Birth weight (g)				
500-999		31.3 (24.8, 39.5)	18.4 (14.6, 23.2)	17.5 (13.9, 22.1)
1000-1499		11.9 (9.6, 14.7)	8.5 (6.9, 10.4)	8.1 (6.6, 10.0)
1500-2499		6.0 (5.3, 6.9)	5.2 (4.5, 5.9)	5.0 (4.4, 5.7)
2500-3499		1.82 (1.66, 2.00)	1.80 (1.64, 1.97)	1.76 (1.61, 1.93)
≥3500 (baseline)		1	1	1
Gestational age (weeks)				
24-27		15.4 (12.3, 19.2)	10.8 (8.7, 13.4)	11.0 (8.8, 13.6)
28-31		5.5 (4.5, 6.7)	4.6 (3.80, 5.6)	4.7 (3.86, 5.7)
32-34		3.39 (2.91, 3.94)	2.99 (2.57, 3.48)	3.04 (2.61, 3.53)
35-36		2.70 (2.38, 3.07)	2.45 (2.15, 2.78)	2.46 (2.17, 2.79)
37-38		1.62 (1.48, 1.76)	1.53 (1.40, 1.67)	1.53 (1.40, 1.67)
≥39 (baseline)		1	1	1
Sex				
Boy		1.27 (1.20, 1.35)	1.19 (1.12, 1.26)	1.18 (1.12, 1.25)
Girl (baseline)		1	1	1
Congenital anomalies				
Yes			9.4 (8.8, 10.1)	9.4 (8.7, 10.0)
No			1	1
Maternal age (years)				
<20				1.22 (1.08, 1.37)
20-24				1.15 (1.05, 1.25)
25-29				1.11 (1.02, 1.20)
30-34 (baseline)				1
35-39				1.04 (0.95, 1.14)
≥40				1.25 (1.09, 1.43)
Quintile of socio-economic status				
Q1: most deprived				1.20 (1.09, 1.31)
Q2				1.09 (0.99, 1.19)
Q3				0.98 (0.89, 1.09)
Q4				0.95 (0.86, 1.05)
Q5: least deprived (baseline)				1

PH=proportional hazards. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model.

Table E.7 – Unadjusted and adjusted Cox PH models for all-cause mortality at 28-364 days in England relative to Sweden in 2003-2012 using an indicator of severe congenital anomalies

	Model 1	Model 2	Model 3	Model 4
Country				
England	1.59 (1.47, 1.71)	1.32 (1.22, 1.42)	1.17 (1.09, 1.26)	1.10 (1.02, 1.19)
Sweden (baseline)	1	1	1	1
Birth weight (g)				
500-1499		27.3 (22.8, 32.8)	13.8 (11.5, 16.6)	12.4 (10.3, 14.9)
1500-2499		7.0 (6.2, 7.8)	5.3 (4.7, 5.9)	4.8 (4.3, 5.3)
2500-3499		2.00 (1.85, 2.15)	1.94 (1.80, 2.09)	1.84 (1.71, 1.98)
≥3500 (baseline)		1	1	1
Gestational age (weeks)				
24-31		3.09 (2.60, 3.67)	2.10 (1.77, 2.50)	2.19 (1.84, 2.60)
32-34		1.63 (1.42, 1.88)	1.37 (1.19, 1.58)	1.43 (1.24, 1.65)
35-36		1.92 (1.72, 2.14)	1.61 (1.45, 1.80)	1.65 (1.48, 1.84)
37-38		1.53 (1.43, 1.64)	1.38 (1.29, 1.48)	1.40 (1.31, 1.50)
≥39 (baseline)		1	1	1
Sex				
Boy		1.33 (1.26, 1.40)	1.20 (1.14, 1.27)	1.20 (1.14, 1.26)
Girl (baseline)		1	1	1
Congenital anomalies				
Yes			19.6 (18.4, 20.8)	19.3 (18.2, 20.5)
No			1	1
Maternal age (years)				
<20				1.70 (1.54, 1.87)
20-24				1.33 (1.23, 1.43)
25-29				1.11 (1.03, 1.20)
30-34 (baseline)				1
35-39				0.97 (0.89, 1.07)
≥40				1.15 (1.00, 1.32)
Quintile of socio-economic status				
Q1: most deprived				1.62 (1.48, 1.77)
Q2				1.47 (1.34, 1.61)
Q3				1.13 (1.02, 1.24)
Q4				1.12 (1.02, 1.24)
Q5: least deprived (baseline)				1

PH=proportional hazards. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Some birth weight and gestational age categories were merged due to small numbers.

Table E.8 – Unadjusted and adjusted Cox PH models for all-cause mortality at 1-4 years in England relative to Sweden in 2003-2012 using an indicator of severe congenital anomalies

	Model 1	Model 2	Model 3	Model 4
Country				
England	1.27 (1.15, 1.40)	1.14 (1.03, 1.26)	1.06 (0.96, 1.17)	1.03 (0.93, 1.13)
Sweden (baseline)	1	1	1	1
Birth weight (g)				
500-1499		10.7 (8.3, 13.7)	4.2 (3.27, 5.4)	3.96 (3.08, 5.1)
under 1500		4.3 (3.63, 5.0)	3.05 (2.61, 3.57)	2.85 (2.43, 3.34)
1500-2499		1.53 (1.40, 1.67)	1.47 (1.35, 1.61)	1.42 (1.30, 1.56)
≥3500 (baseline)		1	1	1
Gestational age (weeks)				
<37		1.05 (0.89, 1.23)	0.83 (0.71, 0.98)	0.85 (0.72, 1.01)
37-38		1.18 (1.07, 1.29)	1.04 (0.94, 1.15)	1.05 (0.96, 1.16)
≥39 (baseline)		1	1	1
Sex				
gender		1.18 (1.10, 1.28)	1.06 (0.98, 1.15)	1.06 (0.98, 1.14)
Girl (baseline)		1	1	1
Congenital anomalies				
Yes			27.0 (24.8, 29.4)	26.9 (24.7, 29.3)
No			1	1
Maternal age (years)				
<20				1.24 (1.06, 1.45)
20-24				1.25 (1.12, 1.40)
25-29				1.01 (0.91, 1.13)
30-34 (baseline)				1
35-39				0.92 (0.81, 1.04)
≥40				0.88 (0.70, 1.10)
Quintile of socio-economic status				
Q1: most deprived				1.37 (1.21, 1.56)
Q2				1.28 (1.12, 1.45)
Q3				1.19 (1.05, 1.36)
Q4				1.10 (0.96, 1.26)
Q5: least deprived (baseline)				1

PH=proportional hazards. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Some birth weight and gestational age categories were merged due to small numbers.

Table E.9 – Unadjusted and adjusted PH models for all-cause mortality at 2-27 days in England relative to Sweden in 2003-2012 including an effect modification term with time for congenital anomaly indictor

Risk factor	Model 1	Model 2	Model 3	Model 4
Country				
England	1.66 (1.53, 1.81)	1.37 (1.26, 1.48)	1.15 (1.06, 1.25)	1.13 (1.04, 1.23)
Sweden (baseline)	1	1	1	1
Birth weight (g)				
500-999		31.3 (24.8, 39.5)	16.4 (13.0, 20.7)	15.6 (12.3, 19.6)
1000-1499		11.9 (9.6, 14.7)	7.7 (6.2, 9.5)	7.3 (5.9, 9.1)
1500-2499		6.0 (5.3, 6.9)	5.2 (4.6, 6.0)	5.0 (4.4, 5.7)
2500-3499		1.82 (1.66, 2.00)	1.81 (1.65, 1.98)	1.77 (1.61, 1.94)
3500 (baseline)		1	1	1
Gestational age (weeks)				
24-27		15.4 (12.3, 19.2)	7.6 (6.1, 9.5)	7.8 (6.2, 9.7)
28-31		5.5 (4.5, 6.7)	3.89 (3.20, 4.7)	3.95 (3.25, 4.8)
32-34		3.39 (2.91, 3.94)	2.91 (2.50, 3.39)	2.95 (2.54, 3.43)
35-36		2.70 (2.38, 3.07)	2.45 (2.16, 2.78)	2.46 (2.17, 2.80)
37-38		1.62 (1.48, 1.76)	1.53 (1.40, 1.67)	1.53 (1.40, 1.68)
≥39 (baseline)		1	1	1
Sex				
Boy		1.27 (1.20, 1.35)	1.19 (1.13, 1.26)	1.19 (1.13, 1.26)
Girl (baseline)		1	1	1
Congenital anomaly				
Yes			5.2 (4.7, 5.8)	5.2 (4.7, 5.8)
effect with time (days)			1.03 (1.02, 1.04)	1.03 (1.02, 1.04)
No			1	1
Maternal age (years)				
<20				1.23 (1.09, 1.38)
20-24				1.14 (1.04, 1.24)
25-29				1.10 (1.02, 1.19)
30-34 (baseline)				1
35-39				1.06 (0.97, 1.16)
≥40				1.32 (1.15, 1.52)
Quintile of socio-economic status				
Q1: most deprived				1.24 (1.13, 1.36)
Q2				1.11 (1.01, 1.22)
Q3				1.00 (0.91, 1.11)
Q4				0.94 (0.85, 1.04)
Q5: least deprived (baseline)				1

PH=proportional hazards. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Some birth weight and gestational age categories were merged due to small numbers.

Table E.10– Unadjusted and adjusted Cox PH models for all-cause mortality at 28-364 days in England relative to Sweden in 2003-2012 including an effect modification term with time for congenital anomaly indictor

Risk factor	Model 1	Model 2	Model 3	Model 4
Country				
England	1.59 (1.47, 1.71)	1.32 (1.22, 1.42)	1.19 (1.10, 1.28)	1.12 (1.04, 1.21)
Sweden (baseline)	1	1	1	1
Birth weight (g)				
500-1499		27.3 (22.8, 32.8)	11.7 (9.7, 14.1)	10.4 (8.7, 12.6)
1500-2499		7.0 (6.2, 7.8)	5.3 (4.7, 5.9)	4.7 (4.3, 5.3)
2500-3499		2.00 (1.85, 2.15)	1.94 (1.80, 2.09)	1.84 (1.71, 1.98)
≥3500 (baseline)		1	1	1
Gestational age (weeks)				
<32		3.09 (2.60, 3.67)	1.58 (1.33, 1.89)	1.65 (1.38, 1.97)
32-34		1.63 (1.42, 1.88)	1.33 (1.15, 1.52)	1.38 (1.20, 1.59)
35-36		1.92 (1.72, 2.14)	1.63 (1.46, 1.82)	1.66 (1.49, 1.85)
37-38		1.53 (1.43, 1.64)	1.39 (1.30, 1.50)	1.41 (1.32, 1.52)
≥39 (baseline)		1	1	1
Sex				
Boy		1.33 (1.26, 1.40)	1.18 (1.12, 1.24)	1.17 (1.11, 1.24)
Girl (baseline)		1	1	1
Congenital anomaly				
Yes			14.6 (13.3, 16.1)	14.4 (13.0, 15.9)
effect with time (months)			1.01 (0.99, 1.03)	1.05 (0.98, 1.13)
No			1	1
Maternal age (years)				
<20				1.72 (1.56, 1.90)
20-24				1.32 (1.22, 1.42)
25-29				1.10 (1.02, 1.19)
30-34 (baseline)				1
35-39				0.99 (0.90, 1.08)
≥40				1.20 (1.05, 1.38)
Quintile of socio-economic status				
Q1: most deprived				1.66 (1.52, 1.81)
Q2				1.49 (1.36, 1.64)
Q3				1.14 (1.04, 1.26)
Q4				1.12 (1.01, 1.24)
Q5: least deprived (baseline)				1

PH=proportional hazards. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Some birth weight and gestational age categories were merged due to small numbers.

Table E.11– Unadjusted and adjusted Cox PH models for all-cause mortality at 1-4 years in England relative to Sweden in 2003-2012 including an effect modification term with time for congenital anomaly indictor

	Model 1	Model 2	Model 3	Model 4
Country				
England	1.27 (1.15, 1.40)	1.14 (1.03, 1.26)	1.10 (1.00, 1.22)	1.06 (0.96, 1.18)
Sweden (baseline)	1	1	1	1
Birth weight (g)				
500-1499		10.7 (8.3, 13.7)	3.04 (2.36, 3.92)	2.86 (2.22, 3.68)
1500-2499		4.3 (3.6, 5.0)	3.12 (2.67, 3.66)	2.92 (2.49, 3.42)
2500-3499		1.53 (1.40, 1.67)	1.48 (1.36, 1.62)	1.43 (1.31, 1.57)
≥3500 (baseline)		1	1	1
Gestational age (weeks)				
<37		1.05 (0.89, 1.23)	0.84 (0.71, 0.99)	0.86 (0.73, 1.02)
37-38		1.18 (1.07, 1.29)	1.07 (0.97, 1.18)	1.08 (0.98, 1.20)
≥39 (baseline)		1	1	1
Sex				
Boy		1.18 (1.10, 1.28)	1.05 (0.97, 1.13)	1.04 (0.97, 1.13)
Girl (baseline)		1	1	1
Congenital anomalies				
Yes			26.1 (21.6, 31.5)	26.0 (21.5, 31.4)
effect with time (years)			0.83 (0.77, 0.90)	0.83 (0.77, 0.90)
No			1	1
Maternal age (years)				
<20				1.26 (1.07, 1.47)
20-24				1.25 (1.12, 1.40)
25-29				1.01 (0.91, 1.12)
30-34 (baseline)				1
35-39				0.92 (0.81, 1.05)
≥40				0.92 (0.74, 1.15)
Quintiles of socio-economic status				
Q1: most deprived				1.39 (1.22, 1.58)
Q2				1.29 (1.13, 1.47)
Q3				1.20 (1.05, 1.37)
Q4				1.10 (0.96, 1.26)
Q5: last deprived (baseline)				1

PH=proportional hazards. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Some birth weight and gestational age categories were merged due to small numbers.

Appendix F. Supporting information for the comparison of cause-specific mortality in England and Sweden

Table F.1 – Unadjusted and adjusted Cox PH models for RTI-related mortality at 31-364 days in England relative to Sweden in 2003-2012 including an effect modification term with time for congenital anomaly indictor

Risk factor	Model 1	Model 3	Model 4
Country			
England	1.50 (1.25, 1.80)	1.16 (0.97, 1.39)	1.11 (0.92, 1.33)
Sweden (baseline)	1	1	1
Birth weight (g)			
500-1499		5.7 (3.8, 8.4)	5.1 (3.4, 7.5)
1500-2499		5.6 (4.3, 7.3)	5.1 (3.9, 6.6)
2500-3499		2.06 (1.73, 2.47)	1.96 (1.64, 2.34)
≥3500 (baseline)		1	1
Gestational age (weeks)			
24-34		1.27 (0.93, 1.72)	1.32 (0.97, 1.80)
35-36		1.35 (1.04, 1.75)	1.38 (1.06, 1.79)
37-38		1.23 (1.04, 1.46)	1.24 (1.05, 1.47)
≥39 (baseline)		1	1
Sex			
Boy		1.11 (0.98, 1.27)	1.11 (0.97, 1.26)
Girl (baseline)		1	1
Congenital anomaly			
Yes: 1-2 months		11.8 (8.4, 16.7)	11.7 (8.4, 16.6)
Yes: 2-3 months		15.3 (10.8, 21.5)	15.2 (10.7, 21.4)
Yes: 3-12 months		29.9 (25.4, 35.1)	29.7 (25.2, 34.9)
No		1	1
Maternal age (years)			
<25			1.40 (1.17, 1.68)
25-29			1.25 (1.04, 1.50)
30-34 (baseline)			1
≥35			1.13 (0.92, 1.38)
Quintile of socio-economic status			
Q1: Most deprived			1.85 (1.48, 2.32)
Q2			1.65 (1.31, 2.07)
Q3			1.22 (0.95, 1.56)
Q4			1.15 (0.90, 1.48)
Q5: Least deprived (baseline)			1

PH=proportional hazards, RTI=respiratory tract infections. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Some birth weight and gestational age categories were merged due to small numbers.

Table F.2 – Unadjusted and adjusted Cox PH models for SIDS at 31-364 days in England relative to Sweden in 2003-2012

Risk factor	Model 1	Model 3	Model 4
Country			
England	1.33 (1.11, 1.60)	1.17 (0.98, 1.41)	0.98 (0.81, 1.18)
Sweden (baseline)	1	1	1
Birth weight (g)			
500-1499		11.0 (6.8, 17.8)	9.0 (5.6, 14.5)
1500-2499		4.2 (3.2, 5.6)	3.43 (2.59, 4.56)
2500-3499		1.81 (1.53, 2.15)	1.62 (1.37, 1.92)
≥3500 (baseline)		1	1
Gestational age (weeks)			
24-34		1.76 (1.21, 2.55)	1.93 (1.33, 2.80)
35-36		2.12 (1.62, 2.78)	2.26 (1.73, 2.95)
37-38		1.53 (1.29, 1.80)	1.62 (1.37, 1.92)
≥39 (baseline)		1	1
Sex			
Boy		1.92 (1.67, 2.21)	1.90 (1.65, 2.19)
Girl (baseline)		1	1
Congenital anomaly			
Yes		0.86 (0.62, 1.21)	0.85 (0.61, 1.18)
No		1	1
Maternal age (years)			
<20			5.0 (3.9, 6.3)
20-24			2.75 (2.23, 3.38)
25-29			1.47 (1.19, 1.83)
30-34 (baseline)			1
≥35			1.01 (0.78, 1.31)
Quintile of socio-economic status			
Q1: Most deprived			1.67 (1.29, 2.15)
Q2			1.92 (1.49, 2.48)
Q3			1.49 (1.14, 1.95)
Q4			1.46 (1.10, 1.92)
Q5: Least deprived (baseline)			1

PH=proportional hazards, SIDS=Sudden Infant Death Syndrome. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Some birth weight and gestational age categories were merged due to small numbers.

Table F.3 – Unadjusted and adjusted Cox PH models for SUDI deaths at 31-364 days in low-risk babies in England relative to Sweden in 2003-2012

Risk factor	Model 1	Model 2	Model 3
Country			
England	1.70 (1.41, 2.06)	1.61 (1.33, 1.95)	1.36 (1.12, 1.65)
Sweden (baseline)	1	1	1
Birth weight (g)			
2500-3499		1.70 (1.47, 1.97)	1.47 (1.27, 1.70)
≥3500 (baseline)		1	1
Gestational age (weeks)			
37-38		1.43 (1.20, 1.71)	1.57 (1.32, 1.87)
39		1.10 (0.93, 1.31)	1.17 (0.98, 1.39)
40 (baseline)		1	1
41		0.89 (0.73, 1.09)	0.89 (0.73, 1.08)
Sex			
Boy		1.91 (1.66, 2.19)	1.87 (1.63, 2.15)
Girl (baseline)		1	1
Maternal age (years)			
<20			5.5 (4.4, 7.0)
20-24			3.02 (2.46, 3.72)
25-29			1.74 (1.41, 2.15)
30-34 (baseline)			1
≥35			1.01 (0.77, 1.31)
Quintile of socio-economic status			
Q1: Most deprived			2.13 (1.66, 2.73)
Q2			1.92 (1.49, 2.48)
Q3			1.49 (1.14, 1.95)
Q4			1.46 (1.11, 1.92)
Q5: Least deprived (baseline)			1

PH=proportional hazards, SUDI=Sudden Unexpected Death in Infancy. Data are adjusted hazard ratios (95% confidence intervals). Each column represents a separate Cox PH model. Low-risk babies were defined as born with birth weight ≥2500g, at term (37-41 weeks), with no congenital anomalies.

References

1. Office for National Statistics. Births by mothers' usual area of residence in the UK, 2012. 2013. Available at https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/datasets/birthsbyareaofusualresidenceofmotheruk. Accessed on 10/8/2017.

2. Rajaratnam JK, Marcus JR, Flaxman AD, et al. Neonatal, postneonatal, childhood, and under-5 mortality for 187 countries, 1970-2010: a systematic analysis of progress towards Millennium Development Goal 4. *Lancet* 2010; 375(9730): 1988-2008.

3. Wang H, Liddell CA, Coates MM, et al. Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2014; 384(9947): 957-79.

4. Viner RM, Hargreaves DS, Coffey C, Patton GC, Wolfe I. Deaths in young people aged 0-24 years in the UK compared with the EU15+ countries, 1970-2008: analysis of the WHO Mortality Database. *Lancet* 2014; 384(9946): 880-92.

5. OECD. National Accounts at a Glance 2015. Organisation for Economic Co-operation and Development; 2015.

6. Tambe P, Sammons HM, Choonara I. Why do young children die in the UK? A comparison with Sweden. *Arch Dis Child* 2015; 100(10): 928-31.

7. Wolfe I, Thompson M, Gill P, et al. Health services for children in western Europe. *Lancet* 2013; 381(9873): 1224-34.

8. Wolfe I, Cass H, Thompson MJ, et al. Improving child health services in the UK: insights from Europe and their implications for the NHS reforms. *BMJ* 2011; 342: d1277.

9. Wolfe I, Donkin A, Marmot M, Macfarlane A, Cass H, Viner R. UK child survival in a European context: recommendations for a national Countdown Collaboration. *Arch Dis Child* 2015; 100(10): 907-14.

10. Wolfe I, Macfarlane A, Donkin A, Marmot M, Viner R. Why children die : death in infants , children and young people in the UK Part A. UK: Royal College of Paediatrics and Child Health, National Children's Bureau, British Association for Child and Adolescent Public Health; May 2014.

11. Office for National Statistics. Births in England and Wales: 2013. 2013; Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthsummarytablesenglandandwales/2014-07-16. Accessed on 10/8/2017.

12. Euro-Peristat project with SCPE and Eurocat. *European Perinatal Health Report. The health and care of pregnant women and babies in Europe in 2010*. May 2013.

13. Smith LK, Budd JL, Field DJ, Draper ES. Socioeconomic inequalities in outcome of pregnancy and neonatal mortality associated with congenital anomalies: population based study. *BMJ* 2011; 343: d4306.

14. Kramer MS, Seguin L, Lydon J, Goulet L. Socio-economic disparities in pregnancy outcome: why do the poor fare so poorly? *Paediatr Perinat Epidemiol* 2000; 14(3): 194-210.

15. Pritchard C, Wallace MS. Comparing UK and Other Western Countries' Health Expenditure, Relative Poverty and Child Mortality: Are British Children Doubly Disadvantaged? *Children & Society* 2015; 29(5): 462-72.

16. Olatunde O, Windsor-Shellard B, Campbell A. Revised Definition of Avoidable Mortality. May ed. Office for National Statistics; 2016.

17. Nolte E, McKee M. Variations in amenable mortality--trends in 16 high-income nations. *Health Policy* 2011; 103(1): 47-52.

18. Hardelid P, Dattani N, Cortina-Borja M, Gilbert R. Contribution of respiratory tract infections to child deaths: a data linkage study. *BMC Public Health* 2014; 14: 1191.

19. World Health Organization. WHO Mortality Database. http://apps.who.int/healthinfo/statistics/mortality/whodpms/. Accessed on 19/11/2015.

20. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *Lancet* 2008; 371(9608): 261-9.

21. Kramer MS, Demissie K, Yang H, Platt RW, Sauve R, Liston R. The contribution of mild and moderate preterm birth to infant mortality. Fetal and Infant Health Study Group of the Canadian Perinatal Surveillance System. *JAMA* 2000; 284(7): 843-9.

22. Office for National Statistics. Child Mortality Statistics : Metadata. March 2015.

23. Office for National Statistics. Mortality Statistics: Metadata. July 2015.

24. World Health Organization (WHO). International statistical classification of diseases and related health problems, 10th revision. Volume 2: instruction manual. 2010 Edition. Geneva: WHO Library Cataloguing-in-Publication Data International; 2011.

25. Kramer MS, Platt RW, Yang H, Haglund B, Cnattingius S, Bergsjo P. Registration artifacts in international comparisons of infant mortality. *Paediatr Perinat Epidemiol* 2002; 16(1): 16-22.

26. Joseph KS, Liu S, Rouleau J, et al. Influence of definition based versus pragmatic birth registration on international comparisons of perinatal and infant mortality: population based retrospective study. *BMJ* 2012; 344: e746.

27. Sidebotham P, Fraser J, Covington T, et al. Understanding why children die in high-income countries. *Lancet* 2014; 384(9946): 915-27.

28. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. *Lancet* 2008; 371(9606): 75-84.

29. Srinivasjois R, Nembhard W, Wong K, Bourke J, Pereira G, Leonard H. Risk of Mortality into Adulthood According to Gestational Age at Birth. *The Journal of Pediatrics* 2017; 190: 185,191.e1.

30. Field D, Boyle E, Draper E, et al. Towards reducing variations in infant mortality and morbidity: a population-based approach. (Programme Grants for Applied Research, No. 4.1.) ed. Southampton (UK): NIHR Journals Library; 2016.

31. Stephens AS, Lain SJ, Roberts CL, Bowen JR, Nassar N. Association of Gestational Age and Severe Neonatal Morbidity with Mortality in Early Childhood. *Paediatr Perinat Epidemiol* 2016; 30(6): 583-93.

32. Altman M, Edstedt Bonamy AK, Wikstrom AK, Cnattingius S. Cause-specific infant mortality in a population-based Swedish study of term and post-term births: the contribution of gestational age and birth weight. *BMJ Open* 2012; 2(4): 10.1136/bmjopen,2012-001152. Print 2012.

33. Rosenstein MG, Cheng YW, Snowden JM, Nicholson JM, Caughey AB. Risk of stillbirth and infant death stratified by gestational age. *Obstet Gynecol* 2012; 120(1): 76-82.

34. Office for National Statistics. Gestation-specific Infant Mortality in England and Wales. 2007-2012.

35. Roberts CL, Algert CS, Morris JM, Ford JB. Increased planned delivery contributes to declining rates of pregnancy hypertension in Australia: a population-based record linkage study. *BMJ Open* 2015; 5(10): e009313,2015-009313.

36. Olesen AW, Westergaard JG, Olsen J. Perinatal and maternal complications related to postterm delivery: A national register-based study, 1978-1993. *American Journal of Obstetrics and Gynecology* 2003; 189: 222-7.

37. National Institute for Health and Care Excellence. Inducing labour. Clinical guideline (available at nice.org.uk/guidance/cg70). 23 July 2008; .

38. Grunewald C, Hakansson S, Saltvedt S, Kallen K. Significant effects on neonatal morbidity and mortality after regional change in management of post-term pregnancy. *Acta Obstet Gynecol Scand* 2011; 90(1): 26-32.

39. Hjern A. Children's health: Health in Sweden: The National Public Health Report 2012. Chapter 2. *Scand J Public Health* 2012; 40(9 Suppl): 23-41.

40. Watkins WJ, Kotecha SJ, Kotecha S. All-Cause Mortality of Low Birthweight Infants in Infancy, Childhood, and Adolescence: Population Study of England and Wales. *PLOS Medicine* 2016; 13(5): e1002018.

41. Spencer N. Weighing the Evidence: How is Birthweight Determined? Oxford: Radcliffe Medical Press Ltd; 2003.

42. Office for National Statistics. Birth cohort tables for infant deaths, 2007-2012. 2007-2012; .

43. Socialstyrelsen (The National Board of Health and Welfare). Statistikdatabas för graviditeter, förlossningar och nyfödda. 2017; ([accessed on 11/11/2017]).

44. Office for National Statistics. Characteristics of Birth 1, England and Wales. 2009-2012. 2014; Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/datasets/characteristicsofbirth1englandandwales, 2016.

45. Spellacy WN, Miller S, Winegar A, Peterson PQ. Macrosomia--maternal characteristics and infant complications. *Obstet Gynecol* 1985; 66(2): 158-61.

46. EUROCAT. EUROCAT Prevalence Tables (2003-2012). ; Available at: http://www.eurocat-network.eu/accessprevalencedata/prevalencetables. Accessed 03/09, 2016.

47. Tennant PW, Pearce MS, Bythell M, Rankin J. 20-Year Survival of Children Born with Congenital Anomalies: a Population-Based Study. *Lancet* 2010; 375(9715): 649-56.

48. Gissler M, Fronteira I, Jahn A, et al. Terminations of pregnancy in the European Union. *BJOG* 2012; 119(3): 324-32.

49. EUROCAT. Special report: Prenatal Screening Policies in Europe. 2010; 1(available at http://www.eurocat-network.eu/content/Special-Report-Prenatal-Screening-Policies.pdf): accessed on 11/11/2017.

50. Department of Health. Abortion Statistics, England and Wales: 2013. Summary information from the abortion notification forms returned to the Chief Medical Officers of England and Wales. . June 2014; Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/319460/Abortion_Statistics__England_and_Wales_2013.pdf.

51. Kelly LE, Shah PS, Hakansson S, et al. Perinatal health services organization for preterm births: a multinational comparison. *J Perinatol* 2017; 37(7): 762-8.

52. Sawyer CC. Child mortality estimation: estimating sex differences in childhood mortality since the 1970s. *PLoS Med* 2012; 9(8): e1001287.

53. Mage DT, Donner EM. Is excess male infant mortality from sudden infant death syndrome and other respiratory diseases X-linked? *Acta Paediatr* 2014; 103(2): 188-93.

54. Anthony S, Jacobusse GW, van der Pal-de Bruin KM, Buitendijk S, Zeitlin J, EURO-PERISTAT Working Group on Risk Factors. Do differences in maternal age, parity and multiple births explain variations in fetal and neonatal mortality rates in Europe?--Results from the EURO-PERISTAT project. *Paediatr Perinat Epidemiol* 2009; 23(4): 292-300.

55. Tucker J, McGuire W. Epidemiology of preterm birth. *BMJ* 2004; 329(7467): 675-8.

56. Chen XK, Wen SW, Fleming N, Demissie K, Rhoads GG, Walker M. Teenage pregnancy and adverse birth outcomes: a large population based retrospective cohort study. *Int J Epidemiol* 2007; 36(2): 368-73.

57. Haldre K, Rahu K, Karro H, Rahu M. Is a poor pregnancy outcome related to young maternal age? A study of teenagers in Estonia during the period of major socio-economic changes (from 1992 to 2002). *Eur J Obstet Gynecol Reprod Biol* 2007; 131(1): 45-51.

58. Hardelid P, Dattani N, Davey J, Pribramska I, Gilbert R. Overview of child deaths in the four UK countries. . 2013; .

59. Chevalier A, Viitanen TK, Viitanen TK. The long-run labour market consequences of teenage motherhood in Britain. *Journal of Population Economics* 2003; 16(2): 323-43.

60. Kurinczuk JJ, Hollowell J, Boyd PA, Oakley L, Brocklehurst P, Gray R. Inequalities in infant mortality project briefing paper 4. The contribution of congenital anomalies to infant mortality. 2010; .

61. Stothard KJ, Tennant PW, Bell R, Rankin J. Maternal overweight and obesity and the risk of congenital anomalies: a systematic review and meta-analysis. *JAMA* 2009; 301(6): 636-50.

62. Persson M, Cnattingius S, Villamor E, Soderling J, Pasternak B, Stephansson O, Neovius M. Risk of major congenital malformations in relation to maternal overweight and obesity severity: cohort study of 1.2 million singletons. *BMJ* 2017; 357: j2563.

63. Cnattingius S, Villamor E, Johansson S, Edstedt Bonamy AK, Persson M, Wikstrom AK, Granath F. Maternal obesity and risk of preterm delivery. *JAMA* 2013; 309(22): 2362-70.

64. Johansson S, Villamor E, Altman M, Bonamy AK, Granath F, Cnattingius S. Maternal overweight and obesity in early pregnancy and risk of infant mortality: a population based cohort study in Sweden. *BMJ* 2014; 349: g6572.

65. Persson M, Johansson S, Villamor E, Cnattingius S. Maternal overweight and obesity and risks of severe birth-asphyxia-related complications in term infants: a population-based cohort study in Sweden. *PLoS Med* 2014; 11(5): e1001648.

66. Public Health England. Maternal obesity. December 2015; Available at: https://khub.net/c/document_library/get_file?uuid=a5768682-fb3d-4fda-ab4a-937a8d80f855&groupId=31798783, 2017.

67. Bai J, Wong FWS, Bauman A, Mohsin M. Parity and pregnancy outcomes. *American Journal of Obstetrics and Gynecology* 2002; 186: 274-8.

68. EUROCAT. A Review of Environmental Risk Factors for Congenital Anomalies . 29 April 2004; 1(available at http://www.eurocat-network.eu/content/Special-Report-Env-Risk-I-and-II.pdf): accessed on 11/11/2018.

69. Yogev Y, Visser GHA. Obesity, gestational diabetes and pregnancy outcome. *Seminars in Fetal and Neonatal Medicine* 2009; 14: 77-84.

70. Roberts CL, Ford JB, Algert CS, et al. Population-based trends in pregnancy hypertension and pre-eclampsia: an international comparative study. *BMJ Open* 2011; 1(1): e000101,2011-000101.

71. Farrar D, Simmonds M, Griffin S, et al. The identification and treatment of women with hyperglycaemia in pregnancy: an analysis of individual participant data, systematic reviews, meta-analyses and an economic evaluation. Chapter 4, Prevalence of gestational diabetes in the UK and Republic of Ireland: a systematic review. . Health Technology Assessment, No. 20.86. ed. Southampton (UK): NIHR Journals Library; 2016 Nov.

72. Bramham K, Parnell B, Nelson-Piercy C, Seed PT, Poston L, Chappell LC. Chronic hypertension and pregnancy outcomes: systematic review and meta-analysis. *BMJ* 2014; 348: g2301.

73. Hutcheon JA, Lisonkova S, Joseph KS. Epidemiology of pre-eclampsia and the other hypertensive disorders of pregnancy. *Best Practice & Research Clinical Obstetrics & Gynaecology* 2011; 25: 391-403.

74. Raisanen S, Sankilampi U, Gissler M, Kramer MR, Hakulinen-Viitanen T, Saari J, Heinonen S. Smoking cessation in the first trimester reduces most obstetric risks, but not the risks of major congenital anomalies and admission to neonatal care: a population-based cohort study of 1,164,953 singleton pregnancies in Finland. *J Epidemiol Community Health* 2014; 68(2): 159-64.

75. Davies H, Gilbert R, Johnson K, et al. Neonatal drug withdrawal syndrome: cross-country comparison using hospital administrative data in England, the USA, Western Australia and Ontario, Canada. *Arch Dis Child Fetal Neonatal Ed* 2016; 101(1): F26-30.

76. Moser K, Stanfield KM, Leon DA. Birthweight and gestational age by ethnic group, England and Wales 2005: introducing new data on births. *Health Stat Q* 2008; (39)(39): 22,31, 34-55.

77. Sheridan E, Wright J, Small N, et al. Risk factors for congenital anomaly in a multiethnic birth cohort: an analysis of the Born in Bradford study. *Lancet* 2013; 382(9901): 1350-9.

78. Griffiths LJ, Dezateux C, Cole TJ. Differential parental weight and height contributions to offspring birthweight and weight gain in infancy. *Int J Epidemiol* 2007; 36(1): 104-7.

79. Kelly Y, Panico L, Bartley M, Marmot M, Nazroo J, Sacker A. Why does birthweight vary among ethnic groups in the UK? Findings from the Millennium Cohort Study. *J Public Health (Oxf)* 2009; 31(1): 131-7.

80. Hawkins SS, Lamb K, Cole TJ, Law C, Millennium Cohort Study Child Health Group. Influence of moving to the UK on maternal health behaviours: prospective cohort study. *BMJ* 2008; 336(7652): 1052-5.

81. Ekeus C, Cnattingius S, Essen B, Hjern A. Stillbirth among foreign-born women in Sweden. *Eur J Public Health* 2011; 21(6): 788-92.

82. Urquia ML, Qiao Y, Ray JG, Liu C, Hjern A. Birth outcomes of foreign-born, native-born, and mixed couples in Sweden. *Paediatr Perinat Epidemiol* 2015; 29(2): 123-30.

83. Marmot Review Team. The Health Impacts of Cold Homes and Fuel Poverty. London: Friends of the Earth & the Marmot Review Team; 2011.

84. DiFranza JR, Aligne CA, Weitzman M. Prenatal and Postnatal Environmental Tobacco Smoke Exposure and Children's Health. *Pediatrics* 2004; 113(Supplement 3): 1007-15.

85. Wettergren B, Blennow M, Hjern A, Soder O, Ludvigsson JF. Child Health Systems in Sweden. *J Pediatr* 2016; 177S: S187-202.

86. NHS Digital. Statistics on Smoking, England - 2016. June 2017; Available at: https://digital.nhs.uk/catalogue/PUB20781, 2017.

87. Victora CG, Bahl R, Barros AJ, et al. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. *Lancet* 2016; 387(10017): 475-90.

88. Ip S, Chung M, Raman G, et al. Breastfeeding and maternal and infant health outcomes in developed countries. *Evid Rep Technol Assess (Full Rep)* 2007; (153)(153): 1-186.

89. Fleming PJ, Blair PS, Pease A. Sudden unexpected death in infancy: aetiology, pathophysiology, epidemiology and prevention in 2015. *Arch Dis Child* 2015; 100(10): 984-8.

90. Taylor BJ, Garstang J, Engelberts A, et al. International comparison of sudden unexpected death in infancy rates using a newly proposed set of cause-of-death codes. *Arch Dis Child* 2015; 100(11): 1018-23.

91. Blair PS, Sidebotham P, Berry PJ, Evans M, Fleming PJ. Major epidemiological changes in sudden infant death syndrome: a 20-year population-based study in the UK. *Lancet* 2006; 367(9507): 314-9.

92. Blair PS, Sidebotham P, Evason-Coombe C, Edmonds M, Heckstall-Smith EM, Fleming P. Hazardous cosleeping environments and risk factors amenable to change: case-control study of SIDS in south west England. *BMJ* 2009; 339: b3666.

93. Kramer MS, Goulet L, Lydon J, et al. Socio-economic disparities in preterm birth: causal pathways and mechanisms. *Paediatr Perinat Epidemiol* 2001; 15 Suppl 2: 104-23.

94. Smith LK, Manktelow BN, Draper ES, Springett A, Field DJ. Nature of socioeconomic inequalities in neonatal mortality: population based study. *BMJ* 2010; 341: c6654.

95. Wood AM, Pasupathy D, Pell JP, Fleming M, Smith GC. Trends in socioeconomic inequalities in risk of sudden infant death syndrome, other causes of infant mortality, and stillbirth in Scotland: population based study. *BMJ* 2012; 344: e1552.

96. Kurinczuk J, Hollowell J, Brocklehurst P, Gray R. Inequalities in infant mortality project briefing paper 1. Infant mortality: overview and context. 2009; .

97. Pritchard C, Williams R. Poverty and child (0-14 years) mortality in the USA and other Western countries as an indicator of "how well a country meets the needs of its children" (UNICEF). *Int J Adolesc Med Health* 2011; 23(3): 251-5.

98. UNICEF Office of Research. Child Well-being in Rich Countries: A comparative overview. Innocenti Report Card 11 ed. Florence: UNICEF Office of Research; 2013.

99. Organisation for Economic Co-operation and Development. OECD Health Statistics 2015. http://stats.oecd.org/index.aspx?DataSetCode=HEALTH_STAT. [Accessed on 19/11/2015].

100. Lundberg O, Yngwe MA, Stjarne MK, et al. The role of welfare state principles and generosity in social policy programmes for public health: an international comparative study. *Lancet* 2008; 372(9650): 1633-40.

101. OECD. Family Database (2012). ; Available at: https://stats.oecd.org/Index.aspx?DataSetCode=FAMILY, 2017.

102. Wolfe I, Sigfrid L, Chanchlani N, Lenton S. Child Health Systems in the United Kingdom (England). *J Pediatr* 2016; 177S: S217-42.

103. Hallsworth M, Farrands A, Oortwijn WJ, Hatziandreu E. The provision of neonatal services. Data for international comparisons. Technical Report . *National Audit Office* 2007; .

104. Chen A, Oster E, Williams H. Why Is Infant Mortality Higher in the United States Than in Europe? *Am Econ J Econ Policy* 2016; 8(2): 89-124.

105. MacDorman MF, Mathews TJ. Behind international rankings of infant mortality: how the United States compares with Europe. *NCHS Data Brief* 2009; (23)(23): 1-8.

106. MacDorman MF, Matthews TJ, Mohangoo AD, Zeitlin J. International comparisons of infant mortality and related factors: United States and Europe, 2010. *Natl Vital Stat Rep* 2014; 63(5): 1-6.

107. Department for Education. CHILD DEATH REVIEWS: YEAR ENDING 31 MARCH 2011. 2011; July.

108. Blondel B, Cuttini M, Hindori-Mohangoo AD, et al. How do late terminations of pregnancy affect comparisons of stillbirth rates in Europe? Analyses of aggregated routine data from the Euro-Peristat Project. *BJOG* 2017; .

109. Flenady V, Middleton P, Smith GC, et al. Stillbirths: the way forward in high-income countries. *Lancet* 2011; 377(9778): 1703-17.

110. Lack N, Zeitlin J, Krebs L, Kunzel W, Alexander S. Methodological difficulties in the comparison of indicators of perinatal health across Europe. *Eur J Obstet Gynecol Reprod Biol* 2003; 111 Suppl 1: S33-44.

111. Mohangoo AD, Blondel B, Gissler M, Velebil P, Macfarlane A, Zeitlin J, Euro-Peristat Scientific Committee. International comparisons of fetal and neonatal mortality rates in high-income countries: should exclusion thresholds be based on birth weight or gestational age? *PLoS One* 2013; 8(5): e64869.

112. The National Board of Health and Welfare (NBHW). The Swedish Medical Birth Register: a summary of content and quality. 2003; .

113. Kirkwood BR, Sterne JAC. Essential Medical Statistics. 2nd ed. ed. United Kingdom: Wiley-Blackwell Etextbooks; 2010.

114. Papiernik E, Zeitlin J, Delmas D, et al. Termination of pregnancy among very preterm births and its impact on very preterm mortality: results from ten European population-based cohorts in the MOSAIC study. *BJOG* 2008; 115(3): 361-8.

115. Rosicova K, Madarasova Geckova A, van Dijk JP, Kollarova J, Rosic M, Groothoff JW. Regional socioeconomic indicators and ethnicity as predictors of regional infant mortality rate in Slovakia. *International Journal of Public Health* 2011; 56(5): 523-31.

116. European Comission. National strategy for Roma integration. 2016; Available at: http://ec.europa.eu/justice/discrimination/roma-integration/romania/national-strategy/national_en.htm, 2017.

117. NCD Risk Factor Collaboration (NCD-RisC). A century of trends in adult human height. *Elife* 2016; 5: 10.7554/eLife.13410.

118. Farrant BM, Stanley FJ, Hardelid P, Shepherd CC. Stillbirth and neonatal death rates across time: the influence of pregnancy terminations and birth defects in a Western Australian population-based cohort study. *BMC Pregnancy Childbirth* 2016; 16: 112,016-0904-1.

119. Harron K, Gilbert R, Cromwell D, Oddie S, Guttmann A, van der Meulen J. International comparison of emergency hospital use for infants: data linkage cohort study in Canada and England. *BMJ Qual Saf* 2017; .

120. Anderson JG, Baer RJ, Partridge JC, et al. Survival and Major Morbidity of Extremely Preterm Infants: A Population-Based Study. *Pediatrics* 2016; .

121. Thygesen LC, Ersboll AK. When the entire population is the sample: strengths and limitations in register-based epidemiology. *Eur J Epidemiol* 2014; 29(8): 551-8.

122. Maret-Ouda J, Tao W, Wahlin K, Lagergren J. Nordic registry-based cohort studies: Possibilities and pitfalls when combining Nordic registry data. *Scand J Public Health* 2017; 45(17_suppl): 14-9.

123. Gissler M, Louhiala P, Hemminki E. Nordic Medical Birth Registers in epidemiological research. *Eur J Epidemiol* 1997; 13(2): 169-75.

124. Administrative Data Liaison Service. NHS Scotland. 2017; Available at: http://www.adls.ac.uk/nhs-scotland/, 2017.

125. DECIPHer. Wales Electronic Cohort for Children. ; Available at: http://decipher.uk.net/research-page/wales-electronic-cohort-children/#, 2017.

126. Mental Health and Community Care Team, HSCIC. Maternity and Children's Data Sets (MCDS) FAQs. 2015; Available at: http://content.digital.nhs.uk/media/1674/Maternity--Childrens-Data-Set-FAQs/pdf/Maternity_and_Children's_Data_Sets_FAQs.pdf, 2017.

127. NHS Digital. Maternity Services Monthly Statistics. England, June 2017, Experimental Statistics. November 2017; Available at: http://digital.nhs.uk/media/33523/Maternity-Services-Monthly-Statistics-England-June-2017-Experimental-statistics-Executive-Summary/default/msms-jun17-exp-rep. Accessed October, 2017.

128. Office for National Statistics. Births Statistics: Metadata. July 2015; .

129. Ghosh RE, Ashworth DC, Hansell AL, Garwood K, Elliott P, Toledano MB. Routinely collected English birth data sets: comparisons and recommendations for reproductive epidemiology. *Arch Dis Child Fetal Neonatal Ed* 2016; .

130. Paul Arrowsmith, Health and Social Care Information Centre. Change to NHS Numbers for Babies (NN4B) Standard – Proposal for Change. 2014; Available at: https://groups.ic.nhs.uk/SCCIDsupport/dashboard/SCCISecretariat/2014-05-28/ISB%201555%20NHS%20Number%20for%20Babies%20Proposal%20for%20Change.pdf. Accessed 06/15, 2016.

131. Moser K, Macfarlane A, Chow YH, Hilder L, Dattani N. Introducing new data on gestation-specific infant mortality among babies born in 2005 in England and Wales. *Health Stat Q* 2007; (35)(35): 13-27.

132. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol* 2017; .

133. NHS Health Research Authority. Confidentiality Advisory Group. 21 Nov 2017; Available at: https://www.hra.nhs.uk/approvals-amendments/what-approvals-do-i-need/confidentiality-advisory-group/, November 2017.

134. NHS Digital. Hospital Episode Statistics. ; Available at: http://content.digital.nhs.uk/hes. Accessed 08/02, 2017.

135. Department of Health. *A Simple Guide to Payment by Results* . 2011; Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213150/PbR-Simple-Guide-FINAL.pdf. Accessed 08/11, 2017.

136. NHS Digital. HES Data Dictionary: Admitted Patient Care. 20.12.2016; Available at: http://content.digital.nhs.uk/media/19425/APC-DD-Final-Doc2/pdf/DD-APC-V7.pdf. Accessed 03/05, 2017.

137. Murray J, Saxena S, Modi N, Majeed A, Aylin P, Bottle A, Medicines for Neonates Investigator Group. Quality of routine hospital birth records and the feasibility of their use for creating birth cohorts. *J Public Health (Oxf)* 2013; 35(2): 298-307.

138. Knight H, Cromwell D, van der Meulen J, Gurol-Urganci I, Richmond D, Mahmood T, et al. Patterns of maternity care in English NHS hospitals 2011/12. UK: Royal College of Obstetricians and Gynaecologists, London; 2013.

139. WHO ICD Implementation Database (WHOFIC). Morbidity. Results by country. 2014; Available at: http://apps.who.int/gho/data/node.whofic.S03Q01_02?lang=en, 2017.

140. NHS Digital. Introduction to Healthcare Resource Group. ; Available at: http://content.digital.nhs.uk/hrg. Accessed 08/14, 2017.

141. NHS Digital. Hospital Episode Statistics (HES) Analysis Guide. 2.3rd ed. NHS Digital; June 2015.

142. Office for National Statistics. Mortality Statistics: Childhood, infant and perinatal, England and Wales (Series DH3), 2006. 2008; 39.

143. NHS Digital. A Guide to Linked Mortality Data from Hospital Episode Statistics and the Office for National Statistics . June 2015; .

144. Office for National Statistics. Impact of registration delays on mortality statistics. 2011; Available at: http://webarchive.nationalarchives.gov.uk/20160106020016/http://www.ons.gov.uk/ons/guide-method/user-guidance/health-and-life-events/impact-of-registration-delays-on-mortality-statistics/impact-of-registration-delays-on-mortality-statistics-article.pdf, 2016.

145. NHS Digital. Linked HES-ONS mortality data. ; Available at: http://content.digital.nhs.uk/article/2677/Linked-HES-ONS-mortality-data. Accessed 11/04, 2015.

146. NHS Information Standards Board. Data Standards: NHS Numbers for Babies. 2003; Available at: http://webarchive.nationalarchives.gov.uk/+/http://www.isb.nhs.uk/library/dscn/dscn2002/612002.pdf. Accessed 06/15, 2016.

147. Hipisley-Cox J. Validity and completeness of the NHS Number in primary and secondary care Electronic data in England 1991-2013 . 2015; Available at: http://eprints.nottingham.ac.uk/3153/3/QResearchReport2013.pdf.

148. HES Development Team. Methodology for creation of the HES Patient ID (HESID). 2.0th ed. NHS Digital; 2014.

149. Hagger-Johnson G, Harron K, Gonzalez-Izquierdo A, et al. Identifying Possible False Matches in Anonymized Hospital Administrative Data without Patient Identifiers. *Health Serv Res* 2015; 50(4): 1162-78.

150. Medical Research Council and NHS Research Authority. Do I need NHS REC approval? 2015; Available at: http://hra-decisiontools.org.uk/ethics/. Accessed 06/05, 2017.

151. Heath and Social Care Information Centre. A Guide to Linked Mortality Data from Hospital Episode Statistics and the Office for National Statistics . 2013; .

152. Office for National Statistics. Mortality Statistics: Childhood, infant and perinatal, England and Wales (Series DH3) 1999-2007. ; Available at: http://webarchive.nationalarchives.gov.uk/20160105160709/http:/www.ons.gov.uk/ons/rel/vsob1/mortality-statistics--childhood--infant-and-perinatal--england-and-wales--series-dh3-/index.html, 15/06/2016.

153. Office for National Statistics. Child mortality statistics: Childhood, infant and perinatal 2008-2013. ; Available at: http://webarchive.nationalarchives.gov.uk/20160105160709/http://ons.gov.uk/ons/rel/vsob1/child-mortality-statistics--childhood--infant-and-perinatal/index.html, 16/06/2016.

154. Harron K, Gilbert R, Cromwell D, van der Meulen J. Linking Data for Mothers and Babies in De-Identified Electronic Health Data. *PLoS One* 2016; 11(10): e0164667.

155. Sandall J, Murrells T, Dodwell M, et al. The efficient use of the maternity workforce and the implications for safety and quality in maternity care: a population-based, cross-sectional study. *Health Services and Delivery Research* 2014; 2(38).

156. Pan H, Cole T. LMSgrowth, a Microsoft Excel add-in to access growth references based on the LMS method. 2012; .

157. Oken E, Kleinman KP, Rich-Edwards J, Gillman MW. A nearly continuous measure of birth weight for gestational age using a United States national reference. *BMC Pediatr* 2003; 3: 6.

158. Dobbins TA, Sullivan EA, Roberts CL, Simpson JM. Australian national birthweight percentiles by sex and gestational age, 1998-2007. *Med J Aust* 2012; 197(5): 291-4.

159. Roberts CL, Lancaster PA. Australian national birthweight percentiles by gestational age. *Med J Aust* 1999; 170(3): 114-8.

160. HES Data Quality Team. HES 2012-13 Inpatient Data Quality Note. 05/11/2013; .

161. Office for National Statistics. Mortality Statistics: Childhood, infant and perinatal, England and Wales (Series DH3). 2003-2007; .

162. Office for National Statistics. Birth statistics, England and Wales (Series FM1) 1998-2008. 2016; Available at: http://webarchive.nationalarchives.gov.uk/20160108035551/http://www.ons.gov.uk/ons/rel/vsob1/birth-statistics--england-and-wales--series-fm1-/index.html, 2016.

163. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009; 338: b2393.

164. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011; 30(4): 377-99.

165. White IR, Royston P. Imputing missing covariate values for the Cox model. *Stat Med* 2009; 28(15): 1982-98.

166. Collett D, 1952-, Collett D. Modelling survival data in medical research / D. Collett. London: London : Chapman & Hall; 1994.

167. Dattani N, Datta-Nemdharry P, Macfarlane A. Linking maternity data for England, 2005-06: methods and data quality. *Health Stat Q* 2011; (49):53-79. doi(49): 53-79.

168. Dattani N, Datta-Nemdharry P, Macfarlane A. Linking maternity data for England 2007: methods and data quality. *Health Stat Q* 2012; (53)(53): 4-21.

169. Harron K. email communication. 2016; .

170. Cortina Borja M. email communication. 2017; .

171. Moser K, Hilder L. Assessing quality of NHS Numbers for Babies data and providing gestational age statistics. *Health Stat Q* 2008; (37)(37): 15-23.

172. Jay M. email communication. 2017; .

173. Lewis K. email communication. 2017; .

174. Wijlaars L, Zylbersztejn A, Verfürden M, Gilbert R, Hardelid P. Phenotyping congenital anomalies in England and Scotland: a comparison of three coding clusters using retrospective hospital data. *IJPDS* 2017; 1(1): 351 Proceedings of the IPDLN Conference (August 2016).

175. Office for National Statistics. Personal Demographics Service data. 2017; Available at: https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/datasourceoverviews/personaldemographicsservicedata, November 2017.

176. Imperial College London. Neonatal Data Analysis Unit. Neonatal data. ; Available at: http://www.imperial.ac.uk/neonatal-data-analysis-unit/neonatal-data/, 2017.

177. Moser K. Gestation-specific infant mortality by social and biological factors among babies born in Enland and Wales in 2006. *Health Stat Q* 2009; (42)(42): 78-87.

178. Hilder L, Moser K, Dattani N, Macfarlane A. Pilot linkage of NHS Numbers for Babies data with birth registrations. *Health Stat Q* 2007; (33)(33): 25-33.

179. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol* 2009; 24(11): 659-67.

180. Johansson LA. Dödsorsaksstatistik Historik, produktionsmetoder och tillförlitlighet Komplement till rapporten Dödsorsaker 2008. 2010; .

181. Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health* 2011; 11: 450,2458-11-450.

182. Cnattingius S, Ericson A, Gunnarskog J, Kallen B. A quality study of a medical birth registry. *Scand J Soc Med* 1990; 18(2): 143-8.

183. Socialstyrelsen (The National Board of Health and Welfare). Diagnoskoder (ICD-10). 2017; Available at: http://www.socialstyrelsen.se/klassificeringochkoder/diagnoskodericd-10, 2017.

184. Socialstyrelsen (The National Board of Health and Welfare). In English – the National Patient Register. ; Available at: http://www.socialstyrelsen.se/register/halsodataregister/patientregistret/inenglish, 2017.

185. Stig K, Paulsson Lütz I. Financing of Health Care in the Nordic Countries. Copenhagen, Denmark: Nordic Medico Statistical Committee; 2013.

186. Brooke HL, Talback M, Hornblad J, et al. The Swedish cause of death register. *Eur J Epidemiol* 2017; 32(9): 765-73.

187. Johansson LA. Targeting Non-obvious Errors in Death Certificates [dissertation]. Uppsala University; 2008.

188. Statistics Sweden. Longitudinal integration database for health insurance and labour market studies (LISA by Swedish acronym). ; Available at: http://www.scb.se/lisa-en, 15/02/2017.

189. Haglund B. Birthweight distributions by gestational age: comparison of LMP-based and ultrasound-based estimates of gestational age using data from the Swedish Birth Registry. *Paediatr Perinat Epidemiol* 2007; 21 Suppl 2: 72-8.

190. Gissler M, Mohangoo AD, Blondel B, et al. Perinatal health monitoring in Europe: results from the EURO-PERISTAT project. *Inform Health Soc Care* 2010; 35(2): 64-79.

191. Lindström K. Long-term Consequences of Preterm Birth: Swedish National Cohort Studies [dissertation]. Stockholm, Sweden: Department of Clinical Sciences and Education, Karolinska Instituet; 2011.

192. Feinstein JS. The relationship between socioeconomic status and health: a review of the literature. *Milbank Q* 1993; 71(2): 279-322.

193. Davies SC. Annual Report of the Chief Medical Officer 2012, Our Children Deserve Better: Prevention Pays. London: Department of Health; 2013.

194. Hardelid P, Davey J, Dattani N, Gilbert R, Working Group of the Research and Policy Directorate of the Royal College of Paediatrics and Child Health. Child deaths due to injury in the four UK countries: a time trends study from 1980 to 2010. *PLoS One* 2013; 8(7): e68323.

195. Global Burden of Metabolic Risk Factors for Chronic Diseases Collaboration (BMI Mediated Effects), Lu Y, Hajifathalian K, Ezzati M, Woodward M, Rimm EB, Danaei G. Metabolic mediators of the effects of body-mass index, overweight, and obesity on coronary heart disease and stroke: a pooled analysis of 97 prospective cohorts with 1.8 million participants. *Lancet* 2014; 383(9921): 970-83.

196. Elovainio M, Hakulinen C, Pulkki-Raback L, et al. Contribution of risk factors to excess mortality in isolated and lonely individuals: an analysis of data from the UK Biobank cohort study. *Lancet Public Health* 2017; 2(6): e260-6.

197. Alm B, Mollborg P, Erdes L, Pettersson R, Aberg N, Norvenius G, Wennergren G. SIDS risk factors and factors associated with prone sleeping in Sweden. *Arch Dis Child* 2006; 91(11): 915-9.

198. Feudtner C, Feinstein JA, Zhong W, Hall M, Dai D. Pediatric complex chronic conditions classification system version 2: updated for ICD-10 and complex medical technology dependence and transplantation. *BMC Pediatr* 2014; 14: 199,2431-14-199.

199. Schoenfeld D. Partial Residuals for The Proportional Hazards Regression Model. *Biometrika* 1982; 69(1): 239-41.

200. Liew Z, Olsen J, Cui X, Ritz B, Arah OA. Bias from conditioning on live birth in pregnancy cohorts : an illustration based on neurodevelopment in children after prenatal exposure to organic pollutants. *Int J Epidemiol* 2015; 44(1): 345-54.

201. McAndrew F, Thompson J, Fellows L, Large A, Speed M, Renfrew M. Infant Feeding Survey 2010. 2012; .

202. Kramer MS. Determinants of low birth weight: methodological assessment and meta-analysis. *Bull World Health Organ* 1987; 65(5): 663-737.

203. Iams JD, Romero R, Culhane JF, Goldenberg RL. Primary, secondary, and tertiary interventions to reduce the morbidity and mortality of preterm birth. *Lancet* 2008; 371(9607): 164-75.

204. Chamberlain C, O'Mara-Eves A, Porter J, Coleman T, Perlen SM, Thomas J, McKenzie JE. Psychosocial interventions for supporting women to stop smoking in pregnancy. *Cochrane Database Syst Rev* 2017; 2: CD001055.

205. Gill PJ, Goldacre MJ, Mant D, Heneghan C, Thomson A, Seagroatt V, Harnden A. Increase in emergency admissions to hospital for children aged under 15 in England, 1999–2010: national database analysis. *Arch Dis Child* 2013; 98(5): 328-34. doi:10.1136/archdischild-2012-302383.

206. Keeble E, Kossarova L. Focus on: Emergency hospital care for children and young people. What has changed in the past 10 years? Research report. The Health Foundation and Nuffield Trust; 2017.

207. Johansson LA, Westerling R. Comparing hospital discharge records with death certificates: can the differences be explained? *J Epidemiol Community Health* 2002; 56(4): 301-8.

208. Johansson LA, Westerling R. Comparing Swedish hospital discharge records with death certificates: implications for mortality statistics. *Int J Epidemiol* 2000; 29(3): 495-502.

209. Limerick SR, Bacon CJ. Terminology used by pathologists in reporting on sudden infant deaths. *J Clin Pathol* 2004; 57(3): 309-11.

210. Grey R, Headley J, Oakley L, Kurinczuk JJ, Brocklehurst P, Hollowell J. Inequalities in infant mortality project briefing paper 3.Towards an understanding of variations in infant mortality rates between different ethnic groups in England and Wales. 2010; .

211. Chou D, Daelmans B, Jolivet RR, Kinney M, Say L, Every Newborn Action Plan (ENAP) and Ending Preventable Maternal Mortality (EPMM) working groups. Ending preventable maternal and newborn mortality and stillbirths. *BMJ* 2015; 351: h4255.

212. Bauld L. The impact of smokefree legislation in England: evidence review. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216319/dh_124959.pdf ed. ; March 2011.

213. NHS Digital. Statistics on Women's Smoking Status at Time of Delivery. England: April 2016 to March 2017. June 2017; Available at: https://digital.nhs.uk/media/31132/Statistics-on-Women-s-Smoking-Status-at-Time-of-Delivery-England-Quarter-4-2016-17-Report/default/stat-wome-smok-time-deli-eng-q4-16-17-rep, 2017.

214. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. *Cochrane Database Syst Rev* 2016; 4: CD006611.

215. Rice VH, Hartmann-Boyce J, Stead LF. Nursing interventions for smoking cessation. *Cochrane Database Syst Rev* 2013; (8):CD001188. doi(8): CD001188.

216. Tobacco Policy Team, Healthier Lives Division, Public and International Health Directorate. Tobacco Packaging Guidance Guidance for retailers, manufacturers and distributors of tobacco products, enforcement agencies and the public on changes to tobacco packaging from 20 May 2016. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/596139/Tobacco_Packaging_Guidance.pdf ed. Department of Health; 2016.

217. McNeill A, Gravely S, Hitchman SC, Bauld L, Hammond D, Hartmann-Boyce J. Tobacco packaging design for reducing tobacco use. *Cochrane Database Syst Rev* 2017; 4: CD011244.

218. Opray N, Grivell RM, Deussen AR, Dodd JM. Directed preconception health programs and interventions for improving pregnancy outcomes for women who are overweight or obese. *Cochrane Database Syst Rev* 2015; (7):CD010932.

219. Kramer MS, McDonald SW. Aerobic exercise for women during pregnancy. *Cochrane Database Syst Rev* 2006; (3):CD000180.

220. Thangaratinam S, Rogozinska E, Jolly K, et al. Interventions to reduce or prevent obesity in pregnant women: a systematic review. *Health Technol Assess* 2012; 16(31): iii,iv, 1-191.

221. Public Health England. Health matters: obesity and the food environment. 31 March 2017; available at https://www.gov.uk/government/publications/health-matters-obesity-and-the-food-environment/health-matters-obesity-and-the-food-environment--2#national-policies-to-tackle-obesity.

222. Hadley A, Ingham R, Chandra-Mouli V. Implementing the United Kingdom's ten-year teenage pregnancy strategy for England (1999-2010): How was this done and what did it achieve? *Reprod Health* 2016; 13(1): 139,016-0255-4.

223. Penman-Aguilar A, Carter M, Snead MC, Kourtis AP. Socioeconomic Disadvantage as a Social Determinant of Teen Childbearing in the U.S. *Public Health Rep* 2013; 128(Suppl 1): 5-22.

224. Gray R, Bonellie SR, Chalmers J, Greer I, Jarvis S, Kurinczuk JJ, Williams C. Contribution of smoking during pregnancy to inequalities in stillbirth and infant death in Scotland 1994-2003: retrospective population based study using hospital maternity records. *BMJ* 2009; 339: b3754.

225. Cromwell DA, Knight HE, Gurol-Urganci I. Parity derived for pregnant women using historical administrative hospital data: accuracy varied among patient groups. *J Clin Epidemiol* 2014; 67(5): 578-85.

226. Ludvigsson JF, Almqvist C, Bonamy AK, et al. Registers of the Swedish total population and their use in medical research. *Eur J Epidemiol* 2016; 31(2): 125-36.