FETAL HEALTH STAGNATION: HAVE HEALTH CONDITIONS IN UTERO IMPROVED IN THE US AND WESTERN AND NORTHERN EUROPE OVER THE PAST 150 YEARS?

Dr Eric B. Schneider[1]

Lecturer in Economic History
University of Sussex

e.b.schneider@sussex.ac.uk

5 June 2015

*Work in progress, please do not cite or circulate without permission

Abstract

Many empirical studies have shown that health conditions *in utero* can have long lasting consequences for health across the life course. However, despite this evidence, there is no clear consensus about how fetal health has changed in the very long run. This paper analyses historical birth weights and perinatal mortality rates to construct a coherent picture of how health conditions *in utero* have changed in the western world over the past 150 years. In short, the evidence suggests that fetal health has been relatively stagnant. Birth weights had already reached their current levels in North America and Northern and Western Europe by the late nineteenth century, and they have changed very little in between. Perinatal mortality rates have fallen dramatically since the late 1930s, but this decline was mainly caused by improvements in intrapartum treatments after the introduction of Sulfa drugs and antibiotics. Thus, the health benefits associated with the perinatal mortality decline were concentrated among those at risk and did not influence the population at large. Finding stagnant fetal health during a period when many other indicators of health improved dramatically is provocative and suggests two conclusions: either fetal health did not improve or the indicators used to measure fetal health, indicators still widely used today, may not accurately capture all aspects of health *in utero*. If fetal health has been stagnant, then better conditions *in utero* cannot explain cohort improvements in life expectancy over the twentieth century. If the indicators of fetal health are problematic, then researchers must move beyond birth weight and perinatal mortality to understand how developmental plasticity based on the prenatal environment influences later life health.

[1] The author wishes to thank attendees at the 2015 NBER Cohort Studies Meeting in Boston, at the demography seminar at Umeå University and at the 2014 Beyond Birth Weight Workshop at the University of Sussex for helpful comments on the paper. The usual disclaimer applies.

Introduction

Many empirical studies have shown that health conditions experienced by foetuses *in utero* have significant long-lasting health consequences. Babies exposed to poor conditions are at higher risk for heart disease, stroke and diabetes in later life and have lower lifetime earnings and educational attainment and greater disability than healthier cohorts (Barker, 1997; Almond and Currie, 2011; Godfrey *et al.*, 2007). Studies to date have been tremendously helpful in establishing a causal link between fetal health and later life health, but they have often relied on exogenous shocks to cleanly identify causal links. Thus, there have been relatively few studies that attempt to explain how fetal health has changed over time.

The period between 1860 and the present has been a period of epidemiological transition where many standard indicators of human health have improved dramatically around the world. Crude death rates, child mortality, infant mortality and stillbirth rates have fallen. Life expectancy and average adult height have increased. Western Europe and North America led these trends with the rest of the world following suit in the second half of the twentieth century. The earliest aspects of the mortality decline occurred apart from modern medical science before the germ theory of disease or antibiotics, highlighting the importance of improvements in sanitation and to a lesser extent nutrition in reducing mortality in the nineteenth century (Floud *et al.*, 2011).

Despite these general improvements in health, there is as of yet no consensus on the trajectory of fetal health over the same time period. Woods (2009) and his co-authors (2006) have reconstructed perinatal mortality rates for a number of countries in Western Europe and North America, showing that these rates have declined in the past 150 years. Likewise, several studies have shown that mean birth weights of infants born in hospitals in the nineteenth century were very close to their modern levels (Rosenberg, 1988; Goldin and Margo, 1989; Ward, 1993; Costa, 1998). This paper attempts to collate all of this evidence into a coherent story about how fetal health has changed over time.

In short, the available evidence suggests that fetal health has been stagnant over the past 150 years. Birth weight means and distributions are nearly identical in the past and present, and although there has been some change in mean birth weight over time, these fluctuations are small relative to the cross-national variation in birth weight (Kramer, 1987; Kramer *et al.*, 2002). In addition, declines in perinatal mortality were mainly caused by improvements in intrapartum treatments after the introduction of Sulfa drugs and antibiotics in the late 1930s and 1940s (Løkke, 2012). This result is provocative because it goes against the historical improvements in all other aspects of health and conflicts with observed cohort improvements in life expectancy. Thus, we must also consider whether the proxies used to measure fetal health, proxies which are still used today, are as useful as researchers might like.

The paper first defines and discusses the complications in measuring fetal health. It then presents the historical birth weight evidence and estimates the influence of changes in environmental and demographic factors on birth weight over time. It closes with a detailed analysis of trends in perinatal mortality and a discussion of the consequences of the results.

Measuring Fetal Health

Before discussing changes in fetal health in the very long run, it might be helpful to briefly discuss what fetal health means in this context and some of the challenges and problems with measuring fetal health in general. Plasticity is very strong in the embryonic and fetal period making the developing child extremely sensitive to changes in conditions *in utero*. Poor conditions such as a nutritional shortage, a lack of key micronutrients, the infection of the placenta or a viral infection can stunt prenatal development harming organ functioning and fetal growth among other negative consequences. Recent research suggests that these conditions and the physiological responses of the fetus to the environment *in utero* may have consequences for the health of an individual across their life course (Godfrey *et al.*, 2007). Thus, the purpose in attempting to measure fetal health over the past 150 years is to understand how the prevalence of unhealthy conditions and unhealthy physiological responses has changed over time and influenced cohort health. This purpose shifts the focus of analysis toward understanding the average health and distribution of health outcomes of the population rather than identifying a subset of individuals that might be at risk. It also leads to an emphasis on conditions that would significantly alter fetal development and on the health of surviving infants since these will influence trends in cohort morbidity and mortality.

Given the complexity of prenatal development and the requirements of measuring fetal health, it is very unlikely that any one indicator would be able to perfectly capture fetal health. Thus, we are left with imperfect options from which to choose, especially when pushing measurement into history. Birth weight and length reflect the outcome of fetal growth at one point in time, but they cannot reveal the trajectory of fetal growth before birth. Fetal growth itself is determined by some combination of genetic and epigenetic inheritance as well as dynamic responses to conditions in the womb. Thus, using birth anthropometry, it is impossible to distinguish between an individual born with high inherited growth potential who experiences intrauterine growth restriction and is born at a normal birth weight close to the population mean and an individual of average inherited growth potential who does not experience poor conditions and is born at the same birth weight. Measuring fetal growth directly using ultrasound technology may help ameliorate this problem, but these measurements are not available historically. In addition, fetal growth (and especially weight gain) occurs mostly in the third trimester, so birth weight may not fully capture fetal health in the first and second trimester (Roseboom *et al.*, 2011; Hanson *et al.*, 2015).

Another potential proxy for fetal health is perinatal mortality since poor conditions *in utero* can lead to stillbirths or early neonatal deaths. Perinatal mortality is especially attractive as a historical proxy since perinatal deaths were systematically registered in a number of countries beginning in the nineteenth century (Woods, 2009). However, perinatal deaths were a relatively rare occurrence even in the nineteenth century when 3-6% of total births ended in a stillbirth or neonatal death. Thus, using perinatal mortality as an indicator of population fetal health could be problematic if the factors that led to these extreme outcomes did not reflect the general, population experience of children during the prenatal period.

Clearly, measuring fetal health is difficult and complex, especially when one is limited to indicators available in historical periods. However, the indicators

available historically are more or less the same indicators that many scholars are using to track fetal health today, so it is worthwhile to attempt to understand how fetal health has changed over time using these measures.

Birth Weights, 1840 to the present

Rosenberg (1988), Goldin and Margo (1989), Ward (1993) and Costa (1998) pioneered the study of historical maternity records containing birth weight that have survived for a number of European and North American maternity hospitals. These hospitals served mostly the working class populations in their hinterlands and can provide an indication of general birth weight levels despite issues of selection effects and representativeness. As table 1 shows these authors found that birth weight levels in the mid to late nineteenth century and early twentieth century had already reached modern levels. Indeed if we put faith in the recent INTERGROWTH-21st standards, nearly all of these populations had birth weights at or above the median birth weight for full-term babies (Villar *et al.*, 2014). There was some small improvement among African Americans in the United States and among the Irish, but these improvements are tiny relative to the differences in average birth weights across countries (Kramer 1987): in the 1980s the average birth weight in Pakistan was 2,770 g whereas it was 3,500 g in Norway (table A1).

Table 1: Birth weights of children in historical populations compared with children of like populations born in the 1970s and 1980s.

	Historical			Modern (1970s-80s)		
Country/Place	Years	Mean Birth Weight (g)	LBW (%)	Years	Mean Birth Weight (g)	LBW (%)
North America						
Boston, USA						
Whites	1872-1900	3,409	5.9	1974-77	3,480	3.6
Blacks	1872-1900	3,094	12.3	1974-77	3,230	7.7
All	1872-1900	3,397	6.2	1977	3,299	7.4
New York, USA	1910-1931	3,463	5.5	--	--	--
Philadelphia, USA	1848-1873	3,403	8.1	--	--	--
Montreal, Canada	1851-1905	3,375	5.7	1988	3,303	6.3
Europe						
Vienna, Austria	1865-1930	3,097	12.9	1978	3,320	5.8
Dublin, Ireland	1869-1930	3,282	8.0	1978-79	3,473	4.4
Edinburgh, Scotland	1847-1920	3,132*	15.2	1974	3,062*	7.7
Three Cities, Norway	1860-1920	3,400-3,500	--	--	--	--

Notes: All samples reflect mainly working class populations. LBW means low birth weight, a birth weight under 2,500 grams. *Scottish mean birth weight includes both live births and stillbirths, driving down the mean figure.

Sources: New York – Costa (1998, p. 992); Philadelphia – Goldin and Margo (1989, p. 365); Norway – Rosenberg (1998, p. 281); All Others – Ward (1993, p. 134).

There is not space in this paper to discuss all of the historical maternity hospitals in detail, but a closer study of the maternity hospitals in Boston, MA may assuage doubts that historical maternity hospitals' patients were not representative or that the high birth weights were driven by the selection of women into each hospital. Ward (1993) collected three samples of maternity patient records from nineteenth-

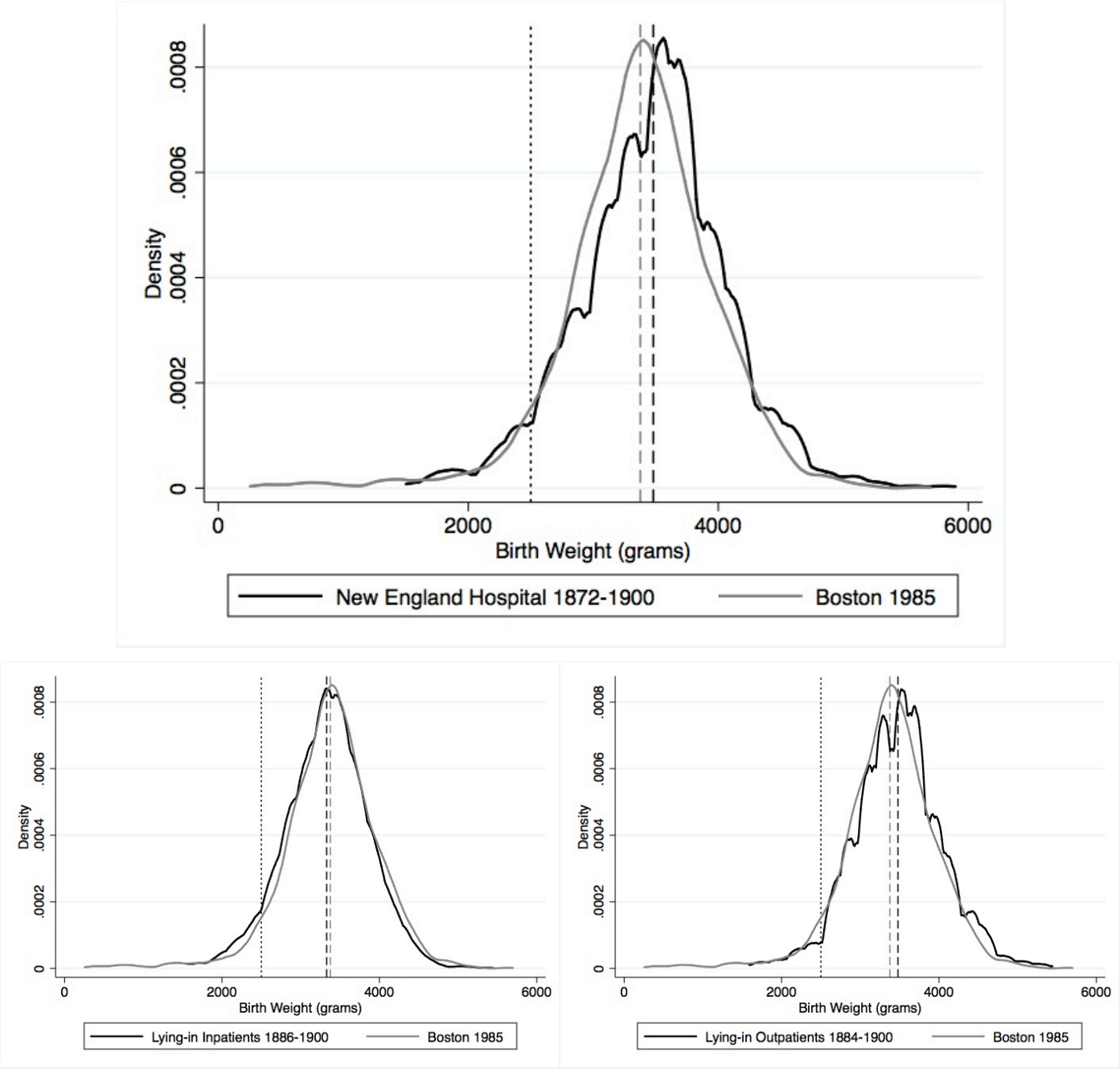
century Boston: the New England Hospital for Women and Children (NEH) maternity records (1872-1900), the Boston Lying-in Hospital inpatient records (1886-1900) and the Boston Lying-in Hospital outpatient records (1884-1900). The NEH and Boston Lying-in inpatient ward provided women a place to give birth and recover afterwards, often for three to four weeks. These hospitals served mainly married and respectable single women. Both required a fee to be paid by the women. The NEH charged $10 per week, which was substantially less than private fees for maternity at the time, but also probably excluded some poorer patients from giving birth there. The occupations of women patients in the NEH suggest that most women were from the upper working class or lower middle class. The Boston Lying-in inpatient ward also charged a $20 fee for women resident in Boston, but it did not turn away impoverished women and 70 per cent of women giving birth in the Lying-in hospital had their fee waived. Thus, the Boston Lying-in inpatient ward seems to have treated a poorer cross-section of the working class along with some complicated pregnancies from higher classes. Both the NEH and Boston Lying-in inpatient ward also tended to serve primiparous women, who made up around 60 per cent of births in each hospital.

The Boston Lying-in outpatient department was something different altogether. It sent medical students to attend births in women's homes around the city of Boston and then follow up with the women by revisiting them at home for several weeks. The outpatient department was a more equal opportunity operation: it did not charge a fee and it served all women whether respectable or not. Ward argues that it was the most representative of the working class of the three sets of Boston patient records. It also overwhelmingly served multiparous women.

Given the differences in populations served and selection mechanisms into the three Boston patient record samples, it seems highly unlikely that the high mean birth weights in each could be driven by the hospitals oversampling from the healthy and/or wealthy residents of Boston. What is even more striking is that the birth weight distributions in these hospitals are very similar to the birth weight distribution of the population of white, singleton births, the most comparable group, in Boston in 1985 (Figure 1). The NEH and Boston Lying-in outpatient samples are slightly above the distribution for Boston in 1985 whereas the Boston Lying-in inpatient sample is slightly below. In addition, the birth weight distribution of children born in the Philadelphia Almshouse in the mid nineteenth century is also very similar to the population distribution of white singleton births in Philadelphia in 1985 (see figure A1).

The evidence presented above suggests that mean birth weights and birth weight distributions were very similar in the nineteenth century and today, but what has happened in between? The twentieth century saw a rapid increase in maternal height and BMI, a fall in the stillbirth rate and the rise and decline again of smoking, all of which may have influenced the birth weight distribution. However, focussing again on the United States where data are most prevalent, there is little evidence that the average birth weight or the distributions of birth weight have changed dramatically across the twentieth century. Costa's studies of births in New York City and Baltimore in the first half of the twentieth century match or exceed mean birth weights of the previous century (Table 1).

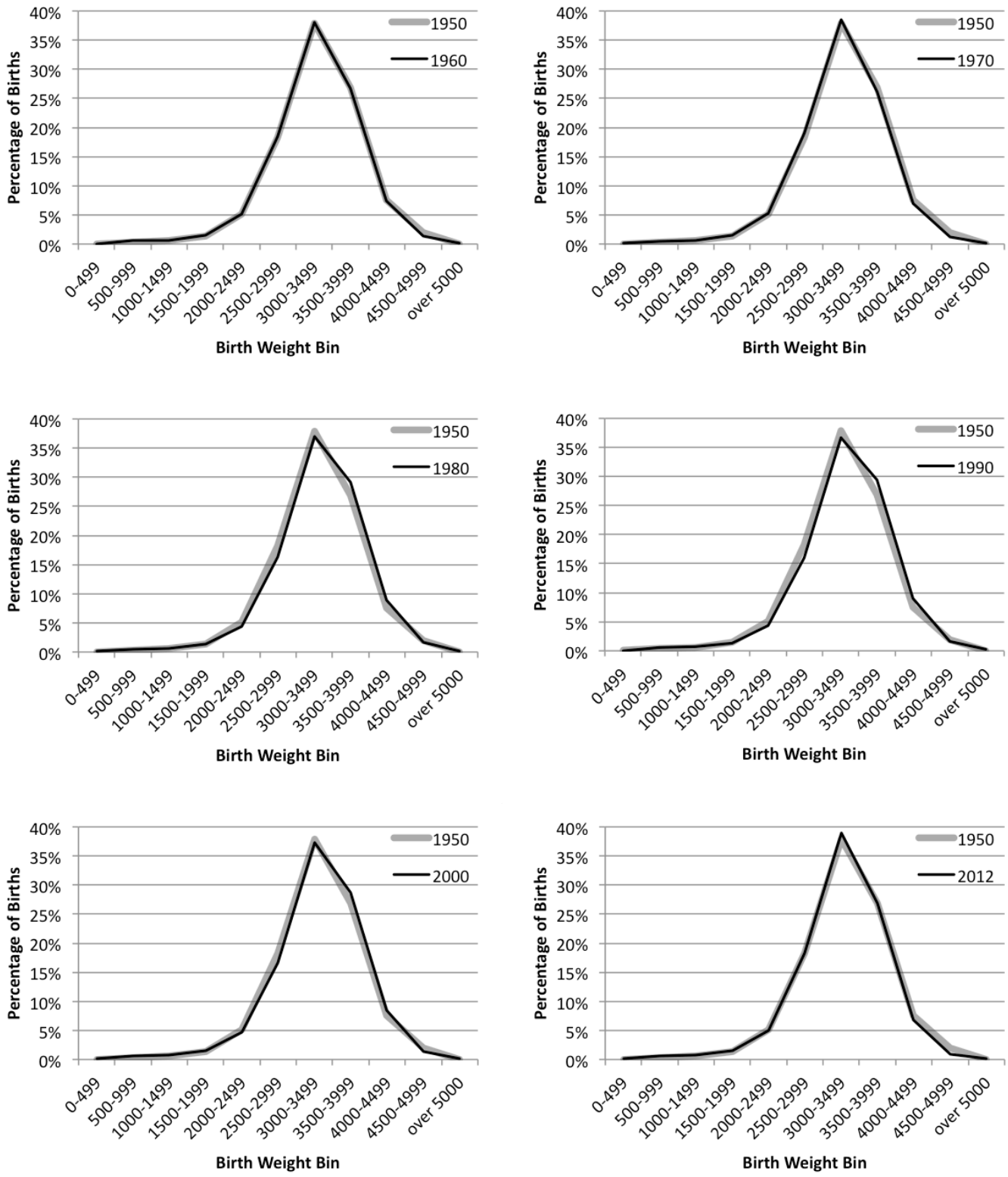
Figure 1: Birth weight distributions in historical hospitals in Boston in the nineteenth century compared to the population distribution of Boston in 1985.



Notes: The dotted vertical line marks the low birth weight cut-off of 2,500 grams. The dashed vertical lines mark the mean of each distribution. The distribution of birth weights in Boston in 1985 only includes white, singleton births to make it most comparable with the historical data.

Sources: Historical birth weight datasets – Ward and Gagné (2012); Birth weights in Boston 1985 – U.S. Department of Health and Human Services, 'Linked bith/infant death data, 1985 birth cohort' (1990).

Figure 2: Stagnation of birth weight distributions in the United States from 1950 to the present.



Notes: Distributions before 1990 are based on a 50 per cent sample of registered births in the 50 states of the United States. From 1990 onward they are based on the population of births.

Sources: Vital Statistics of the United States, 1950-2012.

Figure 2 presents the birth weight distributions reported in the Vital Statistics of the United States for every decade from 1950 to 2012. The median birth weight declined by around 20 grams from 1950 to 1970, but the distributions were nearly identical. Between 1970 and 1980 there was a 60 gram increase in the median birth weight and the distribution shifted slightly upward. The median birth weight increased slightly in 1990 and has fallen by 60 grams since then. The birth weight distribution in 2012 has even shifted downward, closely approximating the distribution in 1950. This evidence for the United States is bolstered by a Dutch study that shows that birth weight by gestational age percentiles calculated for two Amsterdam clinics between 1931 and 1967 are virtually identical near term to birth weight percentiles calculated from the population of births in the Netherlands in 2001 (Visser *et al.*, 2009). It is also supported by long run evidence from Norway (Rosenberg, 1988). Thus, despite substantial improvements in public health and medical technology, it does not appear that the mean or distribution of birth weights has changed substantially in the past 150 years.

Factors influencing birth weight over time

The remarkable consistency of birth weight means and distributions over time is especially stark considering that many factors influencing birth weight have changed dramatically over the past century. Maternal height has increased, smoking prevalence among pregnant women and the population in general increased and then fell, the parity mix of births has changed, and the stillbirth rate has fallen dramatically. Each of these will be discussed in detail in an attempt to understand how each might have affected the birth weight distribution.

Maternal height is positively associated with birth weight because the process of maternal constraint prevents a fetus from out-growing the size of the vaginal canal. Thus, as women have grown larger over the past 150 years, it is possible that they could have given birth to larger babies. Historical and modern estimates of the influence of maternal height on birth weight controlling for a number of confounding factors are between 9 and 12 grams increase in birth weight per extra cm of maternal height (Costa, 1998 and authors' calculations from 2013 US Natality Public Use File). Measuring the secular increase in female height over the past 150 years is somewhat difficult because of the paucity of historical sources, but evidence from nineteenth-century penitentiaries places the mean height of US white women at around 159.5 cm at its lowest point in the nineteenth century (Carson, 2011). This means that the average height of white American women has increased by 3.5 cm to 163 cm for white women giving birth in 2013 (US Natality Public Use File). Thus, if we simply multiplied these two figures, the increase in maternal height in the United States could have accounted for a 31.5 to 42 gram increase in birth weight. However, this likely overestimates the effect because in cross-section 70 per cent of variation in height is driven by genetic or other inherited factors. Thus, part of the relationship between maternal height and birth weight may be reflecting conditions *in utero* but the larger part is reflecting the fact that genetically larger women have larger babies. If it were possible to separate the inherited and environmental influence of maternal height on birth weight, the environmental influence would likely be substantially smaller than the genetic influence, diminishing the influence of the secular increase in maternal height on the birth weight distribution. Having said this, the influence of the

secular increase in maternal height on the birth weight distribution may be stronger in European countries like the Netherlands that have experienced a 14 cm increase in female average height (de Beer, 2010; Schönbeck *et al.*, 2012).

The rise in tobacco smoking prevalence to the middle of the twentieth century and its subsequent decline could have also influenced the birth weight distribution. Fetal nicotine exposure is associated with intrauterine growth restriction and a number of other pregnancy complications. Studies have also shown that exposure to passive tobacco smoke (second-hand smoke) is nearly as harmful as modest cigarette consumption by the mother, leading to a 200 gram reduction in birth weight. Birth weights are 450 grams lower for mothers who smoke more than ten cigarettes per day (Roquer *et al.*, 1995). Thus, the increase and eventual decrease in pregnant women's passive and direct exposure to tobacco smoke could have strongly influenced the birth weight distribution. Cohort studies and surveys have generally placed the peak of smoking prevalence for men in the 1960s with the peak for women following a decade or so thereafter (Birkett, 1997; Kemm, 2001; CDC, 2007; and Lund and Lund, 2014). Smoking prevalence has then declined fairly steadily since the 1960s and 1970s with larger declines in North America than in Europe (Graham, 1996; CDC, 2007).

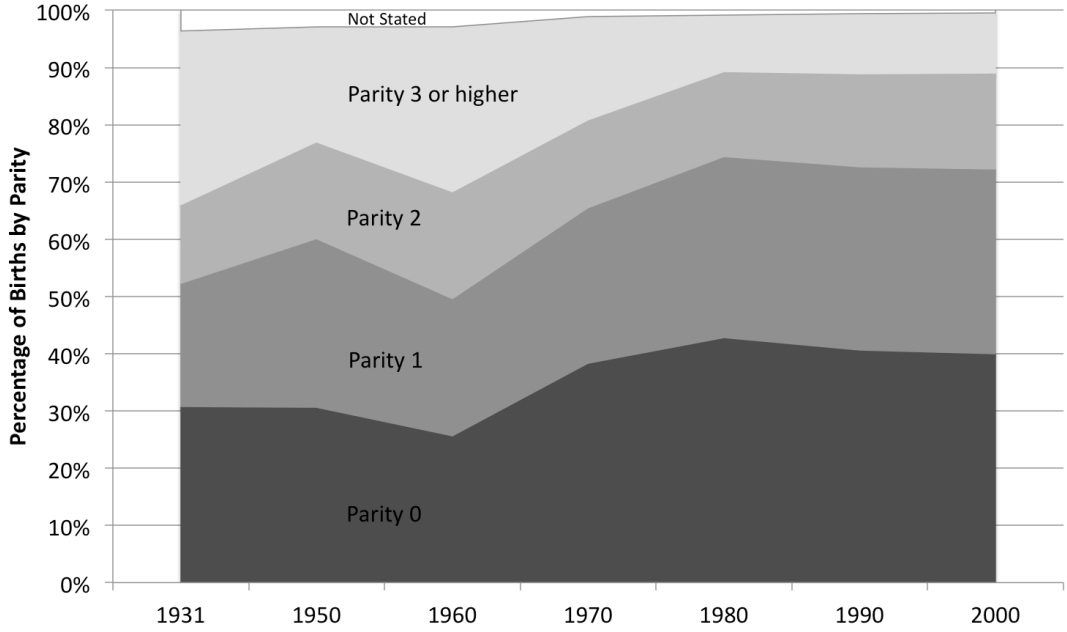
Using Roquer *et al.*'s (1995) estimates of the birth weight penalty for various levels of smoking exposure and the approximate shares of mothers exposed at different points in time, we can estimate how the increase and later decrease in smoking prevalence should have influenced the birth weight distribution. In each case, I have tried to exaggerate the swings in smoking rates in order to present upper bound limits on the effect of smoking on birth weights. Table A2 presents these results in detail. Assuming that women did not smoke in the nineteenth century and only received some passive exposure from men (10 per cent of mothers), we can assume that birth weights were not strongly influenced by nicotine exposure. However, between the nineteenth century and the 1960s and 1970s, nicotine exposure increased dramatically. Assuming that 30 per cent of mothers were never exposed, 30 per cent were exposed passively, 30 per cent were light smokers and 10 per cent were heavy smokers, this increase in nicotine exposure would have decreased the average birth weight by 145 grams, a substantial decline. However, the decline in smoking prevalence over the past 40 or 50 years should have increased birth weight counteracting some of this earlier decline. Thus, I also predicted the mean birth weight using the self-reported smoking prevalence in the 2013 US Natality Public Use File covering 87.4 per cent of births in the United States, which suggests that 5 per cent of mothers were light smokers and 4 per cent were heavy smokers. The US Natality data do not contain information about passive smoke exposure, so I made an educated guess of 10 per cent exposure. Taking a new weighted average of Roquer *et al.*'s figures, we would get a 118 gram increase in birth weight from the decline in smoking over the past 40-50 years. Given the large swings in mean birth weight predicted by the changes in smoking prevalence over the past 150 years, the stagnant birth weight distributions are even more striking.

Birth order, or parity, also influences birth weight since first born children tend to have substantially lower birth weights than higher birth order children. Indeed, the difference in birth weight between first born and higher birth order children seems to have been higher in the past, sometimes exceeding 100 grams (see Figures A2 and

A3). This could influence the birth weight distribution because the parity mixture of births has changed dramatically over the past century as total fertility declined from five or six children per woman to less than 2 children per woman. As women have had fewer children the share of first-born children out of all children born has increased. Figure 3 shows the parity mixture of births from 1931 to 2000. This does not quite capture the full effect of the fertility decline, which began long before 1931, but unfortunately it is very difficult to estimate the parity mixture of births before national vital registration began in the United States in 1931.

In any case, the share of first born children out of all births increased by 17 percentage points between 1931 and the present. Because first-born children have lower birth weights, a higher share of first-born children would lead to a downward shift in the birth weight distribution. It is possible to measure the effect of changes in the parity mix on birth weight by estimating the influence of parity on birth weight in a multivariate regression. Then, one can calculate weighted average birth weights for each time period using the predicted mean values of birth weight for each parity from the regression and the share of births in each parity. Since there appeared to be different historical and modern effects of parity on birth weight, I used both the predicted average birth weights in the Boston Lying-in outpatient clinic and the U.S. 1985 birth cohort. Applying this check to the data, the outpatient clinic birth weights suggest a 25 gram decrease in mean birth weight due to changes in parity mix; however, the decrease is much smaller when using the 1985 birth cohort data at only 10 grams. Thus, it does not appear that changes in the parity mixture of births, at least since 1931, have led to dramatic changes in the birth weight distribution.

Figure 3: Change in share of births by parity in the United States, 1931-2000.

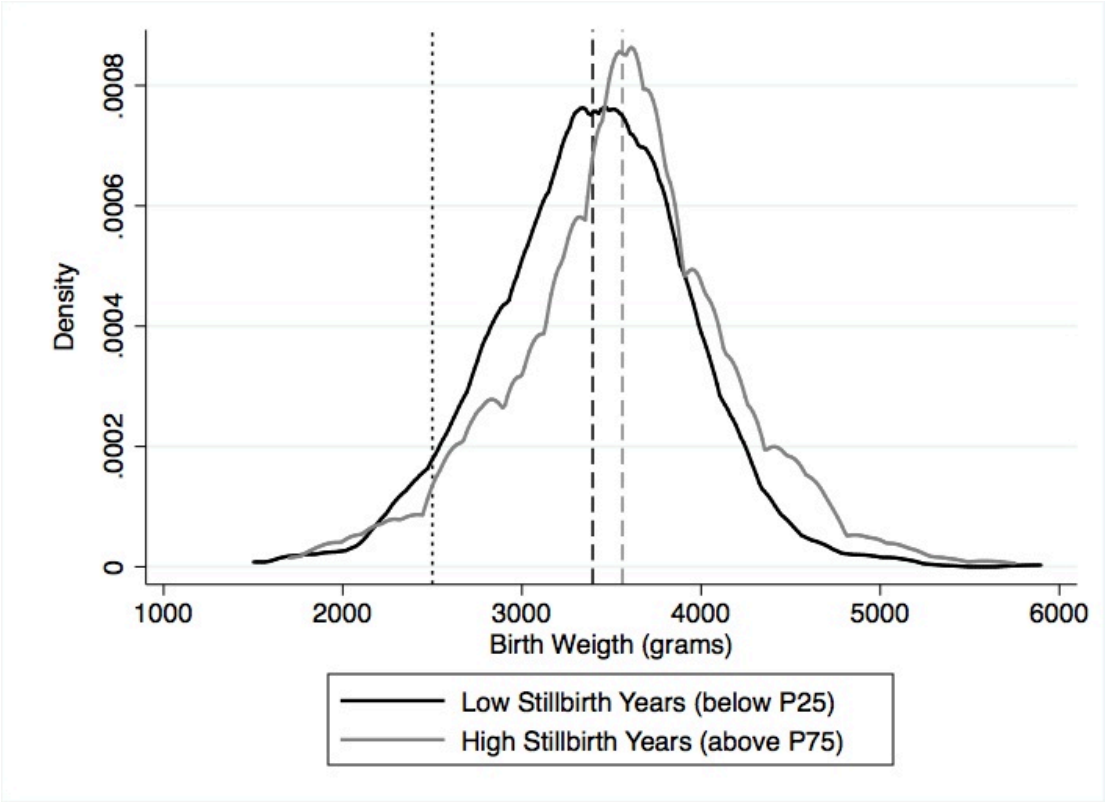


Notes: Data are from the entire birth registration area across the dates covered.

Sources: Vital Statistics of the United States 1931-2000.

Finally, declining stillbirth rates may have influenced the average health of the surviving population of infants and shifted the birth weight distribution. Before the twentieth century, doctors could do very little to prevent stillbirths and neonatal deaths. However, stillbirth rates and neonatal death rates varied substantially year to year because of viral epidemics, the introduction of lead pipes and associated lead poisoning, venereal diseases, etc., all factors exogenous to the medical treatment of the time. Thus, it is possible to test selection and scarring effects on health from these insults. If the selection effect dominates, then we would expect to see a higher average birth weight when the stillbirth rate was high since these diseases would be more likely to kill foetuses that were already small and unhealthy. On the other hand, if the scarring affect dominated, then we would expect to see lower average birth weights when the stillbirth rate was high because poor intrauterine environments weakened all foetuses, not just those who died. Studies conducted in the past thirty years have mainly found that small to moderate increases in birth weight in general have made up for scientific advancements that have improved survival rates for low birth weight, premature infants (Kramer *et al.*, 2002). However, these studies cover a period where stillbirths and neonatal mortality had already fallen substantially, weakening any selection effects. Thus, it would be helpful to test this for the late nineteenth century.

Figure 4: Birth weight distributions of children born in the New England Hospital, Boston in high and low stillbirth rate years, 1872-1900.



Notes: The average stillbirth rate in low stillbirth rate years was 39.96, and the average stillbirth rate in high stillbirth rate years was 51.06.

Sources: Ward and Gagné (2012); Registry Department, City of Boston (1908, pp. 306-307).

Figure 4 provides a first, tentative response. It compares the birth weight distributions of babies born in the New England Hospital in Boston, USA between 1872 and 1900 during high stillbirth and low stillbirth years, years where the stillbirth rate in Boston was either above the 75th or below the 25th percentile for the period. The selection effect seems to dominate. The average birth weight is significantly lower in low stillbirth years than in high stillbirth years with the whole distribution of birth weights shifted to the left in low stillbirth years. This suggests that as stillbirth rates rose, lower birth weight foetuses were more likely to be stillborn than higher birth weight foetuses. Thus, declining stillbirth rates or neonatal mortality rates could have led to a decrease in the average birth weight of surviving foetuses, contrary to the expectation that these two indicators should move in opposite directions. If this speculative evidence for Boston holds across North and Western Europe and in North America, it raises significant questions about whether perinatal mortality can be used as a proxy for fetal health at all. A strong selection effect suggests that perinatal mortality is more a proxy of poor health among the foetuses and infants that die, not poor health across the whole population.

To summarize, there have been a number of factors influencing the birth weight distribution over time. The secular increase in maternal height and the decline in nicotine exposure since the 1960s would have shifted the birth weight distribution upward whereas the increase in smoking prevalence before the 1960s, the shift toward lower parity births, and the decline in the stillbirth rate may have shifted the birth weight distribution downward. The fall in median US birth weight over the past twenty years may have been driven by the rise of induced labor and fraternal twins conceived through *in vitro* fertilization (Zhang *et al.*, 2010). However, the net effect of all of these forces seems to have been a static birth weight distribution over the past 150 years. Thus, the important research question, beyond the scope of this paper, is whether the short and long-run health benefits or consequences of these various shifts are equal. Is birth weight perfectly able to proxy the health costs and benefits of these various exposures, with static birth weight distributions suggesting that there simply has not been improvement in fetal health over the past 150 years? Or for instance, is the health cost of nicotine exposure *in utero* more severe than the health benefit from having a taller, healthier mother? Does the lower birth weight of first-born children reflect poorer health than their higher birth order counterparts and thus lead to poorer health in old age? Using birth weight as a simple proxy for fetal health does not enable us to distinguish between these underlying mechanisms and limits our ability to determine whether clinical and policy interventions have improved or will improve health across the life course.

Causes of Decline in Perinatal Mortality

Although birth weight distributions seem to have been relatively constant over the past century and a half, there have been substantial declines in perinatal mortality during the same time period. The selection effect of stillbirths on birth weight in late nineteenth-century Boston has already raised doubts about whether perinatal mortality accurately proxies the health *in utero* of the general population of births, but it is possible that substantial improvements in the perinatal mortality rates (rather than fluctuations around a stationary trend) could have marked an improvement in fetal health. Perinatal mortality rates were also recorded more systematically and widely in

historical periods than birth weights, making them a potentially attractive proxy for fetal health.

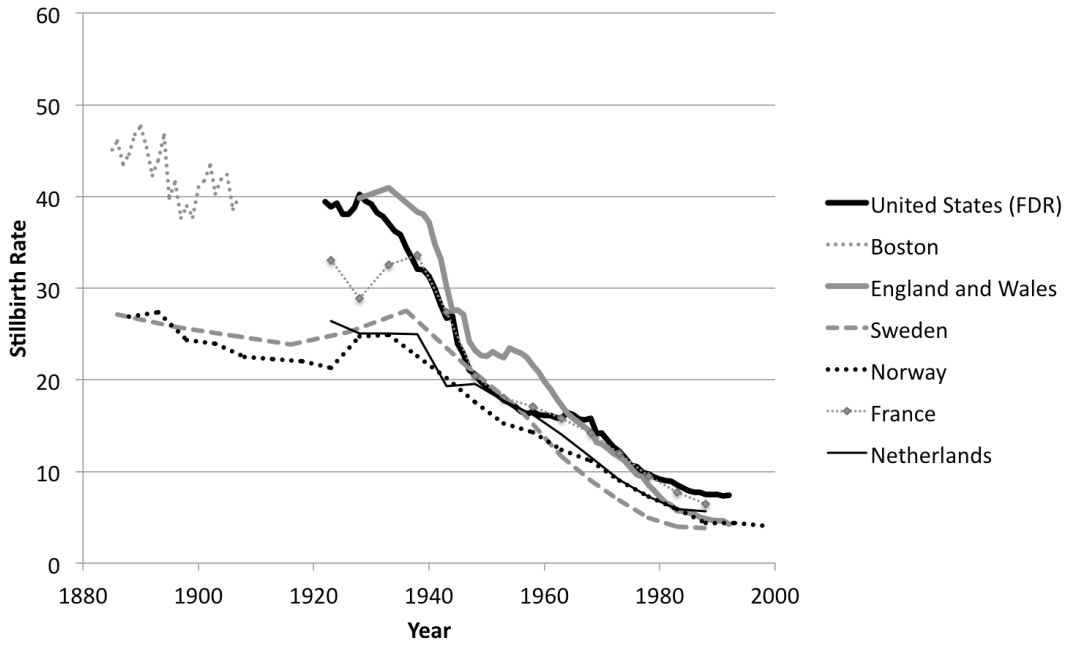
However, stillbirth (SBR) and early neonatal mortality rates (ENMR) are not without their problems. Most importantly from the perspective of measuring fetal health, stillbirths and early neonatal deaths may occur because of infection or intra-uterine growth restriction, factors that indicate poor fetal health, or because of intrapartum complications, which could be unrelated to fetal health especially in a period before the widespread use of Caesarean sections. Thus, it will be important to understand why perinatal mortality declined in order to interpret whether this decline suggests improvement in fetal health or not. In addition, SBRs and ENMRs suffer inconsistencies in registration across countries from differing definitions of stillbirths and cultural practices around baptism of infants. Registration appears to have been fairly good in Scandinavian countries, but more problematic in Catholic countries such as France and Italy (Woods, 2009). Another potential problem arises from the periods for which perinatal mortality data are available. While stillbirth rates were recorded for Scandinavian countries since the mid eighteenth and early nineteenth centuries, these data were not systematically registered in the US, the UK, France and Italy until well into the twentieth century. Woods (2009) analysed historical trends in stillbirths across a wide range of countries accounting for the various registration and other problems with the data. His analysis will form the basis of the discussion below.

Figures 5 and 6 show the trends in SBR and ENMR for countries in Europe and North America from the 1880s to the present. The trends, if not the starting levels, are very similar. SBRs and ENMRs were relatively stagnant from the 1880s until the late 1930s when they began to decline dramatically. Woods (2009) attributed the differences in level between countries in the early twentieth century to obstetric practice in each country. Scandinavian countries and the Netherlands had larger numbers of highly trained doctors and midwives who promoted best practice, and therefore, they had substantially lower stillbirth rates than England, the United States, France or Italy (Woods *et al.*, 2006). However, all of these countries experienced a sharp decline in SBR and ENMR in the late 1930s. The common timing of the decline is extraordinary and rules out any country-specific environmental or health policy changes. Thus, Woods (2009) suggested that the sharp decline could have been caused by the introduction of sulfa drugs and antibiotics from the mid 1930s onward. However, he was not able to determine the precise mechanism through which the adoption antibiotics would have influenced the SBR since most bacterial infections cannot cross the placental barrier.

Løkke (2012) has provided the most extensive evidence to date linking the decline in perinatal mortality to the introduction of antibiotics. Studying reports on causes of death from the National Hospital in Copenhagen, Denmark between 1910 and 1975, she found that antibiotics reduced perinatal mortality through two mechanisms. First, antibiotics allowed physicians to conduct more invasive surgeries to protect the life of the fetus than they had in the past. In the pre-antibiotic era, Danish obstetricians placed the highest value on preventing maternal mortality, so they only carried out Caesarean sections or other invasive surgeries when the mother's life was at risk. However, once Prontosil and Sulfa drugs became available in the mid-1930s followed by Penicillin in the mid-1940s, Danish obstetricians began performing more invasive intrapartum surgeries because they could protect mothers

from puerperal fever. The most common intrapartum complications affected were placental abruption, placenta praevia, eclampsia, contracted pelvis and a prolapsed umbilical cord. Løkke found that reduced mortality of mothers and infants treated for these five complications in the National Hospital accounted for over half of the decline in perinatal mortality in Copenhagen between 1937 and 1957. Thus, the majority of the decline in perinatal mortality appears to have been driven by improvements in intrapartum care rather than any change in health conditions *in utero*.

Figure 5: Stillbirth rates (SBR) in Western Europe and North America declined simultaneously beginning in the late 1930s.



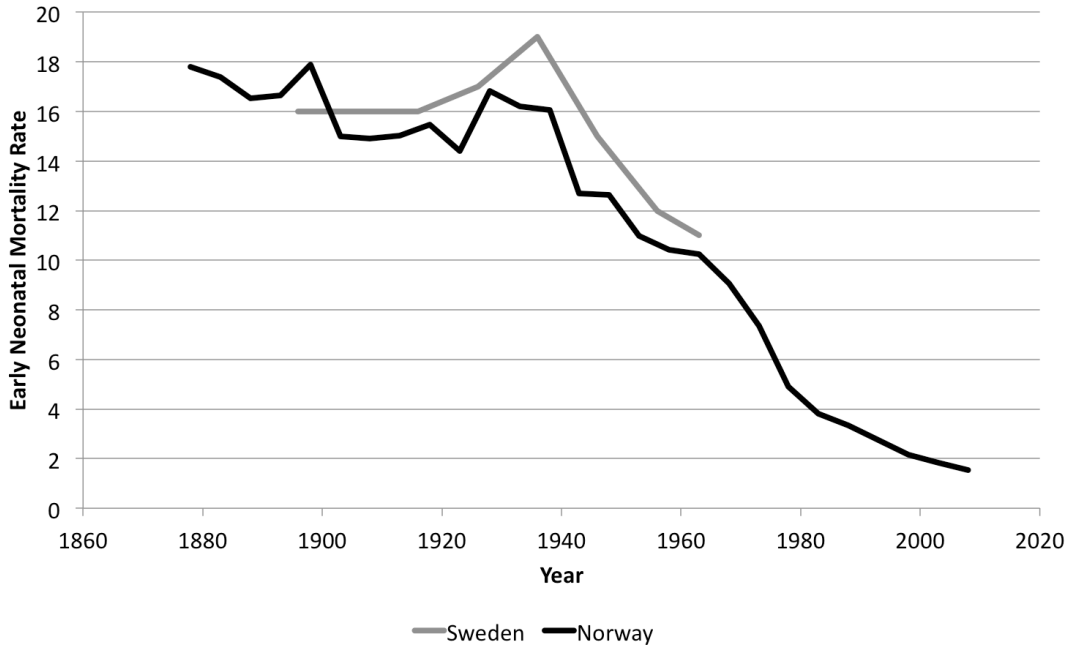
Notes: Data for the United States are fetal death ratios (FDR), stillbirths over live births, rather than stillbirth rates, stillbirths over total births. Data for Norway, the Netherlands and France are quinquennial averages, and data for Sweden are decadal averages until 1961 and quinquennial averages thereafter. Data before 1921 for France and the Netherlands are not reported because an unknown number of early neonatal deaths were recorded as stillbirths before then.

Sources: United States – Haines (2006, table Ab912-927); Boston – Registry Department, City of Boston (1908, pp. 306-307); England and Wales – Office of National Statistics (2014); Sweden – *Historisk Statistik för Sverige* (1967, pp. 108-109) and Macfarlane *et al.* (2000, pp. 664-665); Norway – Statistics Norway (2014, table 05860); France and the Netherlands – Macfarlane *et al.* (2000, pp. 664-665).

However, Løkke's second mechanism does allow for improvements in fetal health. She argues that antibiotics accelerated the cohort improvement in women's health across the twentieth century because they reduced infections during pregnancy. A general reduction in infections during pregnancy likely improved maternal health and thus fetal health on its own, but most important from the perspective of the decline in perinatal mortality was a reduction in women entering the hospital with syphilis. 5.8 per cent of mothers in 1927 entered the hospital with syphilis, but this rate had fallen to 0.1 per cent by 1957. This reduction in the syphilis rate may explain

the decline in perinatal mortality not attributed to the improvements in intrapartum care described above.

Figure 6: Early neonatal mortality rates (ENMR), deaths in the first seven days, also declined sharply at the end of the 1930s.



Notes: Data for Norway are quinquennial averages, and data for Sweden are decadal averages until 1961 and quinquennial averages thereafter.

Sources: Sweden – *Historisk Statistik för Sverige* (1967, pp. 115-116); Norway – Statistics Norway (2014, table 05860).

Thus, if Denmark is representative, it appears that much of the decrease in perinatal mortality beginning in the late 1930s across Northern and Western Europe and North America can be attributed to better interventions in response to intrapartum complications and to the reduction in syphilis prevalence made possible by antibiotics. However, neither of these factors accurately reflect the health of the population in general. The vast majority of children did not experience serious intrapartum complications and syphilis only affected those infected. Thus, most of the decline in perinatal mortality was not caused by a general improvement in fetal health. Rather, it was made possible by preventing deaths among foetuses and neonates most at risk. This conclusion is bolstered by evidence that even since the mid-1970s in the United States, reductions in perinatal mortality have been mainly driven by better neonatal care for at-risk infants and that most nutritional interventions have failed to reduce intrauterine growth restriction or preterm births (Goldenberg and Culhane, 2007). These findings, along with the selection effects of stillbirths shown for late nineteenth-century Boston above, raise serious questions about whether declines in perinatal mortality reflect improvements in fetal health over the past 150 years.

Discussion

The available evidence seems to suggest that fetal health has not improved substantially over the past 150 years in Western countries. Birth weight averages and distributions have remained remarkably similar over time and the decline in perinatal mortality beginning in the late 1930s was driven by factors unrelated to the general health *in utero* of the population. This striking and somewhat shocking finding could point toward two conclusions, both of which have critical implications for contemporary biomedical and social science research. On the one hand, it is possible that fetal health has remained stagnant over the past 150 years despite substantial improvements in living standards, medical technology and health infrastructure. Drastic improvements in nearly every aspect of postnatal health perhaps did not strongly influence prenatal health because foetuses were taking all of the resources they needed from their mothers before the improvements in health. If this is true, then improving fetal health cannot explain cohort improvements in life expectancy and mortality risk age profiles or gains in human capital over the twentieth century (Arora, 2013; Almond and Currie, 2011). These cohort effects may have been driven by the reduction in infant and childhood scarring from diseases and undernutrition (Crimmins and Finch, 2006; Quaranta, 2014; Hatton, 2014).

On the other hand, the evidence may suggest that the indicators used to measure fetal health, indicators that are still widely used today, are not as helpful as researchers might hope. Perhaps fetal health has improved over the past 150 years, but we do not currently have indicators that can capture this improvement. For instance, micronutrient supplementation (iodine and folic acid) along with a reduction in rubella infections among pregnant women after the introduction of the MMR vaccination must have led to an improved level of fetal health even if there were no shift in the birth weight distribution. This would suggest looking for new ways to measure fetal health in the past and future. Birth weight is an especially problematic indicator since it proxies fetal health with so much error. As mentioned above, it provides little information about the growth trajectory *in utero* and cannot distinguish between a high growth potential fetus who has experienced growth restriction and a average growth potential fetus who experienced normal development. In addition, birth weight is most sensitive to poor conditions in the third trimester, but may be unaffected by health shocks earlier in the prenatal period (Roseboom *et al.*, 2011). Thus, from a statistical perspective, regressions using birth weight to predict later life outcomes could suffer from omitted variable bias and/or attenuation bias, seriously compromising the parameter estimates and any causal inference about the relationships tested.

In addition, these historical findings question the standards for birth weight by gestational age developed recently by the INTERGROWTH-21st group based on the pregnancies of healthy, non-obese women from eight different countries (Villar *et al.*, 2014). If birth weights have remained remarkably stable in the West over a period where health improved so dramatically, can we expect to see convergence to the 'optimal' birth weight means and distributions as developing countries experience a similar transition? If not, then why do birth weights vary so dramatically around the world? This historical evidence suggests that it is difficult to define any single pattern of fetal growth, even one derived as precisely as Villar *et al.*'s pattern, as optimal (Hanson *et al.*, 2015).

In order to move past these problems, researchers need to focus on developmental pathways that might influence health rather than on one simple indicator, such as birth weight. For social scientists this might involve studying factors influencing fetal health directly rather than relying on birth weight to capture the net effects. For instance, how does maternal smoking during pregnancy affect the later life health of a fetus? How strong is that effect relative to the effect of birth weight on its own? In addition, it would be interesting to know whether demographic factors such as parity, mother's age and gestational age that influence birth weight would also influence later life health outcomes. Researchers could also incorporate a wider range of proxies for fetal health such as the Ponderal index, birth length, and apgar score in an attempt to better proxy the fetal environment. Medical researchers could focus on epigenetic markers, which could be used as additional proxies for development trajectories *in utero*. Thus, only by studying the developmental process directly will we be able to move beyond birth weight to a greater and more complex understanding of fetal health.

References

Data

Goldin, Claudia, and Robert A. Margo, 'Birthweight Data From the Philadelphia Almshouse Hospital, 1848-1873', ICPSR 20701, Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2008-01-03. http://doi.org/10.3886/ICPSR20701.v1.

Haines, Michael R., 'Fetal death ratio, neonatal mortality rate, infant mortality rate, and maternal mortality rate, by race: 1850–1998', Table Ab912-927 in *Historical Statistics of the United States, Earliest Times to the Present: Millennial Edition,* edited by Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright (New York: Cambridge University Press, 2006). http://dx.doi.org/10.1017/ISBN-9780511132971.Ab912-113710.1017/ISBN-9780511132971.Ab912-1137.

Historisk Statistik för Sverige, Del 1. Befolkning, Andra Upplagan, 1720-1967 (Stockholm, 1967).

Macfarlane, Alison *et al.*, *Birth counts: Statistics of pregnancy and childbirth, Volume 2, tables* (London, 2000).

Martin Joyce A., Brady E. Hamilton, Michelle J. K. Osterman *et al.*, 'Births: Final data for 2012', *National vital statistics reports*, 62, no 9 (2013), available at http://www.cdc.gov/nchs/data/nvsr/nvsr62/nvsr62_09.pdf.

National Center for Health Statistics (2013). Data File Documentations, Natality, 2013 (machine readable data file and documentation), National Center for Health Statistics, Hyattsville, Maryland. Available at http://www.nber.org/data/vital-statistics-natality-data.html.

Office of National Statistics, 'Birth Summary Tables, England and Wales 2013', (2014), electronic dataset, accessed 4/6/2015, http://www.ons.gov.uk/ons/rel/vsob1/birth-summary-tables--england-and-wales/2013/index.html.

Registry Department, City of Boston, *Annual Report of the Registry Department of the City of Boston for the Year 1907* (Boston, 1908).

Statistics Norway, 'Table: 05860 – Perinatal deaths and infant deaths', StatBank Norway, accessed 14/10/2014, https://www.ssb.no/en/befolkning.

U.S. Department of Commerce, Bureau of the Census, *Birth, Stillbirth, and Infant Mortality Statistics for the Birth Registration Area of the United States, 1931: Seventeenth Annual Report* (Washington, 1934).

U.S. Department of Health and Human Services, National Center for Health Statistics, 'Linked birth/infant death data, 1985 birth cohort: United States', ICPSR 3266. Hyattsville, MD: U.S. Dept. of Health and Human Services, National Center for Health Statistics [producer], 1990. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2003. http://doi.org/10.3886/ICPSR03266.v1.

U.S. Department of Health, Education and Welfare, Public Health Service, *Vital Statistics of the United States, 1950: Volume II, Marriage, Divorce, Natality, Fetal Mortality and Infant Mortality Data* (Washington, 1953).

U.S. Department of Health, Education and Welfare, Public Health Service, *Vital Statistics of the United States, 1960: Volume 1 – Natality* (Washington, 1960).

U.S. Department of Health, Education and Welfare, Public Health Service, *Vital Statistics of the United States, 1970: Volume 1 – Natality* (Rockville, Maryland, 1975).

U.S. Department of Health, Education and Welfare, Public Health Service, *Vital Statistics of the United States, 1980: Volume 1 – Natality* (Hyattsville, Maryland, 1984).

U.S. Department of Health, Education and Welfare, Public Health Service, *Vital Statistics of the United States, 1990: Volume 1 – Natality* (Hyattsville, Maryland, 1994).

U.S. Department of Health, Education and Welfare, Public Health Service, *Vital Statistics of the United States, 2000: Volume 1 – Natality*, online publication, http://www.cdc.gov/nchs/data/statab/t001x26.pdf.

Ward, W. Peter and Monique Hélène Gagné, 'Birth weight and economic growth data sets, New England Hospital for Women and Children, Boston, 1872-1900', University of British Columbia Library Data Services (2012), http://hdl.handle.net/2429/44366.

Ward, W. Peter and Monique Hélène Gagné, 'Birth weight and economic growth data sets, Boston Lying-in (inpatient services), 1886-1900', University of British Columbia Library Data Services (2012), http://hdl.handle.net/2429/44368.

Ward, W. Peter and Monique Hélène Gagné, 'Birth weight and economic growth data sets, Boston Lying-in (outpatient services), 1883-1900', University of British Columbia Library Data Services (2012), http://hdl.handle.net/2429/44367.

Articles

Almond, Douglas and Janet Currie, 'Killing me softly: The fetal origins hypothesis', *Journal of Economic Perspectives*, 25, no. 3 (2011), pp. 153-172.

Arora, Suchit, 'Understanding aging during the epidemiological transition', *Research in Economic History*, 29 (2013), pp. 1-69.

Barker, David J. P., 'Maternal nutrition, fetal nutrition, and disease in later life', *Nutrition*, 13, no. 9 (1997), pp. 807-813.

de Beer, Hans, 'Physical stature and biological living standards of girls and young women in the Netherlands, born between 1815 and 1865', *The History of the Family*, 15, no. 1 (2010), pp. 60-75.

Birkett, Nicholas J., 'Trends in smoking by birth cohorts for births between 1940 and 1975: A reconstructed cohort analysis of the 1990 Ontario health survey', *Preventive* Medicine, 26, no. 4 (1997), pp. 534-541.

Carson, Scott Alan, 'Height of female Americans in the 19th century and the antebellum puzzle', *Economics and Human Biology*, 9, no 2 (2011), pp. 157-164.

Center for Disease Control and Prevention, 'Cigarette smoking among adults – United States, 2006', *Morbidity and Mortality Weekly Report*, 56, no. 4 (2007), pp. 1157-1161.

Costa, Dora L., 'Unequal at Birth: A long-term Comparison of Income and Birth Weight', *The Journal of Economic History*, 58, no. 4 (1998), pp. 987-1009.

Crimmins, Eileen M. and Caleb E. Finch, 'Infection, Inflammation, Height, and Longevity', *Proceeding of the National Academy of Sciences*, 103, no. 2 (2006), pp. 498-503.

Floud, Roderick, Robert W. Fogel, Bernard Harris and Sok Chul Hong, *The Changing Body: Health, Nutrition and Human Development in the Western World since 1700* (Cambridge, 2011).

Godfrey, Keith M., Karen A. Lillycrop, Graham C. Burdge, Peter D. Gluckman and Mark A. Hanson, 'Epigenetic Mechanisms and the Mismatch Concept of the Developmental Origins of Health and Disease', *Pediatric Research*, 61, no. 5 (2007), pp. 5R-10R.

Goldenberg, Robert L. and Jennifer F. Culhane, 'Low birth weight in the United States', *American Journal of Clinical Nutrition*, 85, no. 2 (2007), pp. 584S-590S.

Goldin, Claudia and Robert A. Margo, 'The Poor at Birth: Birth Weights and Infant Mortality at Philadelphia's Almshouse Hospital, 1848–1873', *Explorations in Economic History*, 26, no. 3 (1989), pp. 360-379.

Graham, Hilary, 'Smoking prevalence among women in the European community, 1950-1990', *Social Science & Medicine*, 43, no. 2 (1996), pp. 243-254.

Hanson, Mark, Torvid Kiserud, Gerard H. A. Visser, Peter Brocklehurst and Eric Schneider, 'Essay: Optimal fetal growth – a misconception?', *American Journal of Obstetrics and Gynecology* (2015), forthcoming.

Hatton, Timothy J., 'How have Europeans grown so tall?', *Oxford Economic Papers*, 66, no. 2 (2014), pp. 349-372.

Kemm, John R., 'A birth cohort analysis of smoking by adults in Great Britain 1974-1988', *Journal of Public Health Medicine*, 23, no. 4 (2001), pp. 306-311.

Kramer, Michael S., Isabelle Morin, Jong Yang, Robert W. Platt, Robert Usher, Helen McNamara, K. S. Joseph and Shi Wu Wen, 'Why are babies getting bigger? Temporal trends in fetal growth and its determinants', *The Journal of Pediatrics*, 141, no. 4 (2002), pp. 538-542.

Løkke, Anne, 'The antibiotic transformation of Danish obstetrics. The hidden links between the decline in perinatal mortality and maternal mortality in the mid-twentieth century', *Annales de démographie historique*, (2012), pp. 205-224.

Lund, Ingeborg and Karl Erik Lund, 'Lifetime smoking habits among Norwegian men and women born between 1890 and 1994: a cohort analysis using cross-sectional data', *British Medical Journal Open*, 4 (2014), pp. 1-9.

Quaranta, Luciana, 'Early life effects across the life course: The impact of individually defined exogenous measures of disease exposure on mortality by sex in 19th- and 20th-century southern Sweden', *Social Science & Medicine*, 119 (2014), pp. 266-273.

Roquer, J. M., J. Figueras, F. Botet and R Jiménez, 'Influence on fetal growth of exposure to tobacco smoke during pregnancy', *Acta Paediatrica*, 84, no. 2 (1995), pp. 118-121.

Roseboom, Tessa J., Rebecca C. Painter, Annet F. M. van Abeelen, Marjolein V. E. Veenendaal, and Sussane R. de Rooij, 'Hungry in the womb: What are the consequences? Lessons from the Dutch famine', *Maturitas*, 70, no. 2 (2011), pp. 141-145.

Rosenberg, Margit, 'Birth Weights in Three Norwegian Cities, 1860-1984. Secular Trends and Influencing Factors', *Annals of Human Biology*, 15, no. 4 (1988), pp. 275-288.

Schönbeck, Yvonne, Henk Talma, Paula can Dommelen, Boudewijn Bakker, Simore E. Buitendijk, Remy A. HiraSing and Stef van Buuren, 'The world's tallest nation has stopped growing taller: the height of Dutch children from 1955 to 2009', *Pediatric Research*, 73, no. 3 (2012), pp. 371-377.

Villar, José *et al.*, 'International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project', *Lancet*, 384 (2014), pp. 857-868.

Visser, Gerry H. A., Paul H. C. Eilers, Patty M. Elferink-Stinkens, Hans M. W. M. Merkus, and Jan M. Wit, 'New Dutch reference curves for birthweight by gestational age', *Early Human Development*, 85 (2009), pp. 737-744.

Ward, W. Peter, *Birth Weight and Economic Growth: Women's Living Standards in the Industrializing West* (Chicago, 1993).

Woods, Robert I., Anne Løkke and Frans van Poppel, 'Two hundred years of evidence-based perinatal care: Late-fetal mortality in the past', *Archives of Disease in Childhood*, 91, no. 6 (2006), pp. F445-F447.

Woods, Robert I., *Death before birth: Fetal health and mortality in historical perspective* (Oxford, 2009).

Zhang, Xun, K. S. Joseph and Michael S. Kramer, 'Decreased term and postterm birthweight in the United States: impact of labor induction', *American Journal of Obstetrics and Gynecology*, 203, no. 2 (2010), pp. 124e.1-124e.7.

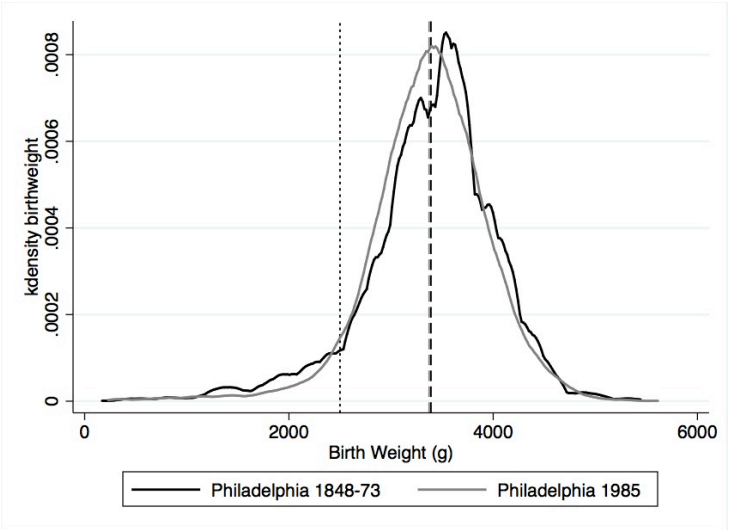
Web Appendix

Table A1: Birth weights of children around the world in the 1980s.

Country	Mean Birth Weight (g)	LBW (%)
Latin America		
Brazil	3,170-3,298	9.0
Chile	3,340	9.0
Colombia	2,912-3,115	10.0
Guatemala	3,050	17.9
Mexico	3,019-3,025	11.7
Africa		
Egypt	3,200-3,285	7.0
Kenya	3,143	12.8
Nigeria	2,880-3,117	18.0
Tunisia	3,210-3,376	7.3
United Republic of Tanzania	2,900-3,151	14.4
Zaire	3,163	15.9
Asia		
China	3,215-3,285	6.0
India	2,493-2,970	30.0
Indonesia	2,760-3,027	14.0
Iran	3,012-3,250	14.0
Iraq	3,540	6.1
Japan	3,200-3,208	5.2
Malaysia	3,027-3,065	10.6
Pakistan	2,770	27.0

Sources: Kramer (1987, p. 665).

Figure A1: Birth weights in the Philadelphia Almshouse (1848-73) compared to the population distribution of births in Philadelphia in 1985.



Notes: The dotted vertical line marks the low birth weight cut-off of 2,500 grams. The dashed vertical lines mark the mean of each distribution. The distribution of birth weights in Phildadelphia in 1985 only includes white, singleton births to make it most comparable with the historical data.

Sources: Philadelphia Almshouse – Goldin and Margo (2008); Birth weights in Boston 1985 – U.S. Department of Health and Human Services, 'Linked bith/infant death data, 1985 birth cohort' (1990).

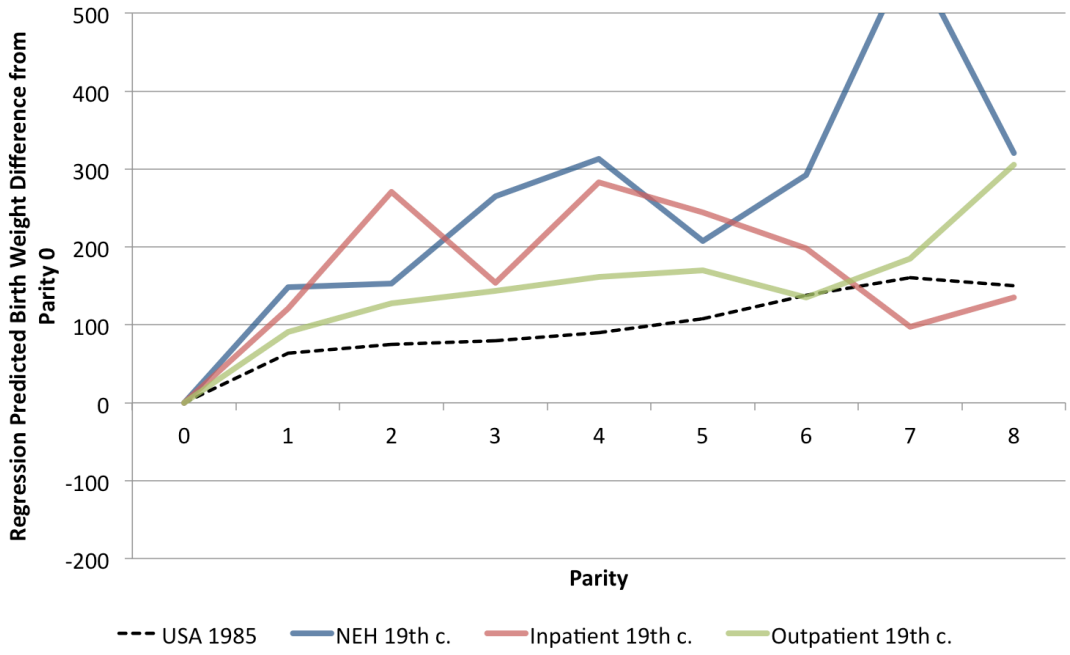
Table A2: Estimates of the effect of tobacco smoke exposure on average birth weight in the United States.

	Smoking Exposure				Predicted Birth Weights	
	None	Passive	Low	High	Mean	Change
Birth Weights (Roquer et al., 1995)	3407	3215	3205	2948		
Share of Mothers Exposed						
Minimum Smoking (c. 1900)	90%	10%	0%	0%	3388	
Maximum Smoking (c. 1970)	30%	30%	30%	10%	3243	-145
2013 USA	82%	10%	5%	4%	3361	118

Notes: Predicted mean birth weights are the weighted average of Roquer *et al.*'s figures based on the assumed shares in each category.

Sources: Roquer *et al.* (1995); National Center for Health Statistics (2013).

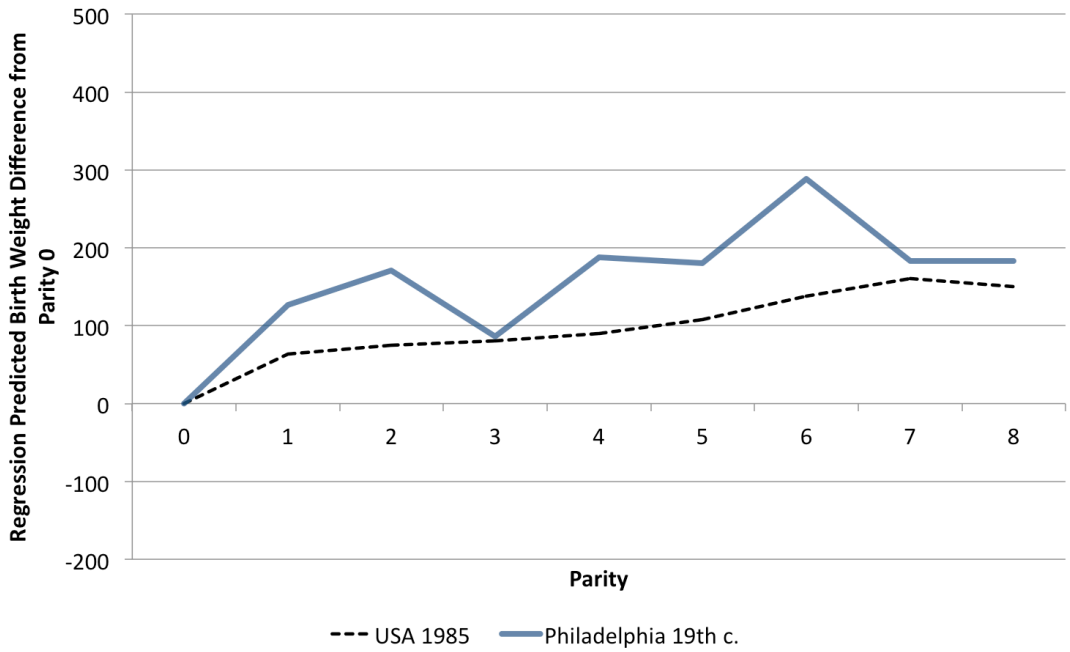
Figure A2: Regression predicted birth weight difference from parity zero for hospitals in nineteenth-century Boston compared to the US figure in 1985.



Notes: The birth weight difference at each parity was estimated using multiple regression analysis with a dummy variable for children of each parity. The regressions also included dummy variable controls for maternal age, sex of the child and gestational age where possible.

Sources: Historical birth weight datasets – Ward and Gagné (2012); Birth weights in Boston 1985 – U.S. Department of Health and Human Services, 'Linked bith/infant death data, 1985 birth cohort' (1990).

Figure A3: Regression predicted birth weight difference from parity zero for the Philadelphia Almshouse (1848-73) compared to the US figure in 1985.



Notes: The birth weight difference at each parity was estimated using multiple regression analysis with a dummy variable for children of each parity. The regressions also included dummy variable controls for maternal age, sex of the child and gestational age where possible.

Sources: Historical birth weight datasets – Philadelphia Almshouse – Goldin and Margo (2008); Birth weights in Boston 1985 – U.S. Department of Health and Human Services, 'Linked bith/infant death data, 1985 birth cohort' (1990).