

THE IMPACT OF MIGRATION ON INFANT HEALTH:
EVIDENCE FROM THE GREAT MIGRATION

Katherine Eriksson
Gregory T. Niemesh

Abstract: The Great Migration of African Americans from the rural South to the urban North entailed a significant change in the health environment, particularly of infants, during a time when access to medical care and public health infrastructure became increasingly important. We create a new dataset that links individual infant death certificates to parental characteristics to assess the impact of parents' migration to Northern cities on infant mortality. The new dataset allows the paper's key innovation, which is to control for selection into migration and detailed parental characteristics. Conditional on parents' pre-migration observable characteristics and county-of-origin fixed effects, we find that black infants were more likely to die in the North relative to their southern-born counterparts. We find no evidence of the "healthy migrant" effect. Given that infant health has a long-lasting impact on adult outcomes, the results shed light on whether and how the Great Migration contributed to African Americans' secular gains in health and income during the 20th century.

Acknowledgements: Eriksson is an Assistant Professor at California Polytechnic University (kerikss@calpoly.edu). Niemesh is an Assistant Professor at Miami University (niemesgt@miamioh.edu). Josh Koller and Raleigh Pearson provided excellent research assistance. We thank participants at the 2014 Cliometric Society Meetings for helpful comments. This research was conducted with the help of funding from the Economic History Association Arthur H. Cole Grant in Aid.

Section 1. Introduction

Between 1910 and 1970, roughly 6 million African Americans left the South in a mass exodus dubbed the "Great Migration," dramatically altering the geographic distribution of the black population. In 1900, 90 percent of blacks lived in the South, which dropped to 53 percent by 1970. Over the same period, receiving regions experienced an increase in the share of the population that was black: four to 19 percent in the Northeast, six to 20 percent in the Midwest, and one to nine percent in the West (McHugh 1987). Most left the rural south to settle in Northern inner cities.

Southern blacks migrated because they expected the North to provide an improvement in opportunities, although it remains an open question whether migrants gained *on net* along all dimensions. For example, employment opportunities were superior in the North, providing a 60-70 percent gain in real income (Collins and Wanamaker 2014)¹. However, outcomes worsened along a number of non-labor market dimensions, namely, higher incarceration rates (Muller 2012) and lower social standing (Flippen 2013).

The Great Migration's impact on the health of migrants has been relatively unexplored despite the intense interest in the wide and persistent racial disparities in longevity over the course of the 20th century.² It is an open question whether the Great Migration contributed to or hindered black economic progress along the dimension of infant mortality. The urban North and rural South provided radically different health environments for children. Consequently, the impact on the health of the second-generation is potentially large. As the Great Migration is one of the most prominent events of African-American history in the 20th century, its contribution to the racial health gap, for both infant mortality and later life health, is an important aspect, and has not yet received much attention because of data availability.

Herein, we create a novel data set linking individual infant mortality outcomes to parental socio-economic characteristics to analyze the interrelationship between economic variables, infant health, and racial and ethnic disparities. This is a new research program motivated by newly available historical individual-level death certificates.³ We view this preliminary work as

¹ However, Eichenlaub, Tolnay, and Alexander (2010) find migrants had lower *occupational status* relative to those that stayed in their Southern communities.

² The exception is Black *et al.* (2011), which finds that migrating North provides no improvement to longevity, and maybe even a slight reduction, conditional on survival to age 65.

³ Logan and Parman (2011) and Preston *et al.* (1996) link death certificates to census observations, but focus on adult mortality and the intergenerational transfer of health.

illustrative of the usefulness of linked data to revisit important aspects of the evolution of infant health over the course of the early 20th century.

Disparities in socio-economic status explain a significant amount of the variation in infant mortality in aggregate statistics (Collins and Thomasson 2004), although there is less agreement on the underlying mechanisms of the relationships. The Great Migration represented a substantial improvement in the labor market opportunities, access to education, and medical care for those African Americans that made the journey north. Black males experienced an increase in of 60 to 70 percent of real income by migrating to Northern cities (Collins and Wanamaker 2014). This alone should have had a large impact on infant mortality based on the standard estimates of the income-gradient (Finch 2003).

Moreover, the disease environment varied across regions, as did the provision of public health infrastructure to mitigate those diseases. In the absence of modern water and sewage systems, city dwellers faced a heightened risk of water-borne diseases, a leading cause of infant deaths. Southerners, on the other hand, were exposed to parasitic infections of malaria and hookworm, which primarily afflicted children and adults. The early 20th century also experienced a wide variety of new public health measures enacted to reduce infant mortality, the timing and extent of which varied across regions.⁴

To our knowledge, the causal impact of migration on infant mortality has yet to be estimated for this time period and population. The motivation of this paper is to use newly collected microdata on births and deaths to determine the extent to which migration accounts for reductions in black infant mortality during the first half of the 20th century. In the base analysis, we compare the likelihood of death for infants whose parents migrated north to those infants whose parents remained in the South. Our methods and results touch upon four related strands of the literature.

A key theme in the literature analyzing the outcomes of migration is that isolating the causal impact is difficult, but essential. Because health can be positively correlated with innate ability, most studies find evidence of positive selection into migration, termed the "healthy migrant effect", which, unaccounted for, biases the estimates of the health return to migration (Halliday and Kimmit 2008, Jasso *et al.* 2006, Black *et al.* 2011). However, negative selection

⁴ Examples include the construction of modern water and sewage systems, malaria and hookworm eradication, promotion of sanitary privies in rural areas, and the Sheppard-Towner Act (1921) which provided federal funding for educational materials and instruction in maternal and infant nutrition, care, and hygiene.

can occur when the cost of migration is low. Irish-born residents of England are less health, on average, than Irish non-migrants (Delaney *et al.* 2013). This paper's key innovation is to first, control for selection into migration using detailed parental characteristics. Secondly, we include county-of-origin fixed effects to account for location-specific unobservables.

Coming from a multitude of sending locations, immigrants faced diverse health environments, which can be used to explore causal mechanisms for observed health disparities in the receiving locations. The health literature leverages the variation in health risks, behaviors, and constraints across migrants to isolate the importance of underlying causes of health disparities seen today among ethnicities and races (Jasso *et al.* 2004). Bias from selection into migration remains a concern, making the construction of the proper counterfactual crucial. Our focus on migration internal to the U.S. during a time of large regional disparities allows us to construct an improved counterfactual by health outcomes and economic controls from the same datasets for both migrants and stayers. For international migration (Jasso *et al.* 2004), health data from the sending countries is often absent or incomplete.

A third important literature links early-life health conditions to adult health and economic outcomes. Recent work provides an understanding that *in utero* and early childhood health conditions partially explain variability in human capital accumulation, earnings, and life expectancy, among a number of other outcomes of interest.⁵ For example, variation in the infectious disease environment during childhood can explain an important part of variation in adult cognitive function (Case and Paxson 2009, Chay *et al.* 2009), convergence in black-white test scores (Chay *et al.* 2009), and subsequent labor market outcomes (Bleakley 2007). While we do not directly estimate the adult gains for the second generation in this paper, our results have important implications for the understanding of black economic progress.

Finally, our work fits into the literature on the reduction in the black-white mortality gap and black economic progress. The rapid decline in the infant mortality rate from 104 deaths per 1,000 live births in 1910 to 7 deaths in 1998 was one of the signature developments in public health in the 20th century (Wolf 2007, Haines 2006). Over that period, the large black-white infant mortality gap gradually declined, as shown in figure 1. Diminished racial disparities in socioeconomic status accounts for most of the convergence prior to World War II (Collins and Thomasson 2004). By 1940 almost 13% of Blacks born in the South resided in the North, the

⁵ See Almond and Currie (2010) for a recent survey of the literature on the persistent impact of early life conditions.

massive movement of the black population potentially had a significant impact on the rate of convergence in the black-white infant mortality gap, and on black life-expectancy (Margo 1990, p. 109). The question is what the counterfactual black infant mortality rate would have looked like in the absence of the massive gains in wages, education, and access to medical care that occurred because of the Great Migration. Moreover, because of the importance of the fetal and infant health environments on long-term labor market outcomes, the health impacts of the Great Migration might have played an important role for black economic progress years into the future.

Prior research on infant mortality and convergence in mortality rates was limited by the availability of aggregate vital statistics. The main contribution of this paper is to estimate the causal impact of migration on infant mortality rates using a newly collected set of individual level microdata of infant mortality linked to parental characteristics. We combine a count of deaths under one-year of age from infant death certificates with a count of surviving infants on April 1st from the decennial censuses. A wealth of parental characteristics are then added by linking observations to the decennial census. The census provides information on parental literacy, age, homeownership, industry, and occupation, which can be used to assign earnings. For analyses using 1940 data, the census provides educational attainment and income.

To measure infant mortality in this time period, we collect individual-level death certificates from 1919-1940 in both Southern and Northern states.⁶ We restrict our attention to infant deaths in the year prior to a census date. To this dataset we append the full universe of infants in the relevant census who were born to parents born in the South but who are living in either location. Our full dataset contains 19,884 infants, of which 1,509 are infant deaths. We have matched the parents of the infants back to their pre-migration location but work is ongoing to digitize the pre-migration characteristics.

The main estimate we are interested in is the treatment effect of migration. To start, we consider the OLS linear probability regression with an indicator for infant death. Controlling for year fixed effects, infants of southern-born parents living in the North are 2.3 percentage points *more* likely to die in the first year of life. We find no evidence of positive selection into migration based on innate infant health capital. However, migrants experience better birth outcomes relative to northern-born blacks, who are 2.8 percentage points more likely to have a

⁶ Currently, we are working with Tennessee, Illinois, and Ohio, but work is underway to add Missouri, North Carolina, and Alabama.

birth end in death within a year. In fact, in the absence of northern-born blacks' higher levels of education the southern mortality advantage would be even larger.

Section 2: Trends in Infant Mortality and the North-South Gap

African Americans faced radically different health environments in the North and the South. The comparison is essentially of the rural South to the urban North. Aggregate vital statistics drawn from the published volumes of the *Vital Statistics of the United States* show that infant mortality rates in Northern cities were as high or higher than in the South early in the 20th century. As seen in figure 2, the black infant mortality rate (IMR) in the North exceeded that of the South initially. In terms of infant mortality, being born in the rural South was preferable to being born in the urban North. The urban south, however, faced even higher rates of mortality than did the urban North. Over time, Northern mortality declined faster than in the South, so that by 1936 a black infant born in the North had a better chance of surviving the first year of life, at least according to the aggregate statistics.

A basic conceptual framework based on a simple health production function guides our interpretation of the difference in mortality between regions, and any subsequent effect of migration. The likelihood of death is determined by infant health capital, which depends on parental inputs and the health environment. Parental inputs would include time, feeding practices, housing, medical care, and parental health endowments. During this period, limited medical knowledge implied that direct medical care likely did not play a large role. The health environment includes the type and likelihood of infection of pathogens, access to medical care, information, and public health provision. The following section discusses the regional differences in inputs to the health production function that might explain the observed differences in mortality rates, and possible mechanisms for the observed impact of migration. We start with differences in the health environment, and follow with a discussion of the impact of economic variables.

Pathogens

Infant diarrhea caused by a number of underlying gastro-intestinal diseases was the leading cause of death in the early death registration areas. The vital statistics available also showed urban suffering more infant deaths than rural areas: 11.3 percent in Massachusetts cities with a population above 100,000, and 9 percent in cities below 10,000 and rural areas (Federal Security Agency 1947, p. 589). The dense urban populations of the North suffered from poor water and sewage disposal, which led to high rates of water-borne disease. Accordingly, much of the focus of urban public health officials over the course of the early 20th century was to improve water sanitation and sewage systems. Over the first few decades of the 1900s, water chlorination, sewage treatment, and other advances dramatically reduced the death rate from water-borne disease in cities (Cutler and Miller 2005).

A newborn in the rural south may have been exposed to fewer pathogens than a newborn in the Northern cities. The rural South also suffered from water-borne diseases, but mortality was never on the scale faced by the urban North. Low population densities meant that the distance between households insulated rural families to some extent from the spread of germs. However, the typical water contamination in rural areas came from improperly stored sewage leaking into the well on a single farm. Although rural areas faced a lower level of intestinal disease than urban areas, public health officials and reformers believed that efforts should be made to reduce it even further.

The solution was two-fold: build sanitary privies to prevent leaching, and move wells away from privies. The 1920s and 30s saw a concerted effort by the U.S. Public Health Service and local public health departments to install sanitary privies and educate homeowners on their importance for health (Ferrel and Mead 1936). While urban centers in the North relied on public funds, most of the effort in the South was private in nature, being funded out of the pockets of homeowners. This limited the speed with which the rural populace received modern water and sewage. As northern cities improved the water supply, the rural South lost its health advantage.

Milk provided an important vector for the spread of pathogens for infants. At the time, mother's faced three options to feed their infant: breast milk, infant formula, and cow's milk. Breastfeeding reduced the transmission of pathogens, and passed along antibodies from the mother, but was less likely to occur if the mother worked outside of the home. Cow's milk was susceptible to contamination at the site of production, during transportation, and in the home.

Even if the milk reached the home uncontaminated, a common practice was to dilute with potentially unsanitary water or allow the infant to suck the milk from a rag. In a 1911 study, babies that were not breastfed died at rates three to four times higher than those that were (Woodbury 1925). Early public health initiatives were aimed at changing infant feeding practices and improving milk safety to reduce infant diarrhea. Surveys of infant feeding practices during the 1920-30s show that poor feeding practices crossed all economic strata, but that significant differences existed between rural and urban mothers (Wolf 2007). Rural mothers breastfed at higher rates and for longer duration, but were more likely to introduce solid food too soon and of detrimental types (Preston and Haines 1991). The cause of the differences between rural and urban infant feeding are unknown, but may stem from differential labor market opportunities outside the home and the availability and price of safe alternatives to breast milk.

The rural South did suffer from a number of diseases that were absent from the North, such as infections of malaria and hookworm. While infecting older children and adults, there was not necessarily a direct impact on infants (Bleakley 2007). Infections may indirectly affect infants through reducing the health and economic resources available to the parents.

Access to medical care

The medical establishment during this period had little to offer a child or mother suffering from disease. Antibiotics would not be introduced until the mid-1930s, and oral rehydration treatment for diarrhea until the 1960s. While health workers did not have the tools to treat disease, they did understand the germ theory and could provide information for proper care and prevention of infection.

The most relevant and dramatic change in medical care for infants was the movement to the hospital of childbirth. The importance of physicians for the birth of child increased dramatically during the first half of the 20th century. Midwives had attended around half of all births at the turn of the century, declining to only 15 percent by 1930. However, midwives were still present at the majority of African-American births in the South (Ladd-Taylor 1988). Hospitals were not necessarily superior locations in terms of outcomes. As hospital births grew in share, maternal mortality was stagnant and infant deaths from birth injuries increased 40-50% (Thomasson and Treber 2008).

As such, racial segregation and poor provision of medical services in the South may not have been an important factor for regional infant mortality differences prior to the 1940s. By the 1960s, treatments for the common causes of infant death were developed and it is clear that segregation had negative and long-lasting impacts on the health and economic outcomes of southern-born blacks (Almond, Chay, and Greenstone 2006, Chay, Guryan, and Mazumder 2009).

Income

Economic resources available to black families differed significantly across regions. While not entering the infant health production function directly, the variation in income potentially lead to differences in the demand for parental inputs. Were migrants able or willing to use the 60-70 percent average increase in real income to partially mitigate the increased likelihood of infant death they faced in the North? Evidence exists of an income-infant mortality gradient in data after the 1980s (Finch 2003). The 1900 Decennial Census asked women the proportion of their children surviving, which Preston and Haines (1991) use to show the existence of an occupation-income gradient for infant and childhood mortality. Direct evidence for the period and population we study is absent from the literature because of a lack of data. One of our goals with this project is to fill this gap, and potentially trace out changes in the importance of income in explaining variation in infant death rates.

Section 3: Data Description

The analysis in this paper uses a micro-sample of live and deceased infants. We pool observations from state death certificates with live children less than one year of age in each of the respective censuses. We focus on census years 1920, 1930, and 1940. To create a sample of deceased infants, we collect death certificates from Tennessee, Illinois, and Ohio and restrict our attention to black infants born in the year prior to the census year and who died before April 1st of the census year. To generate a comparable sample of non-deceased infants, we use all infants under one year of age in the respective census. We take census data from the full indexes from FamilySearch.org.

Each death certificate includes both parents' names as well as information about the father's place of birth. We use the father's name and birthplace to conduct our first match from

the death certificate to the father's observation in the census. We restrict to observations within one year of the census in order to minimize migration after infant death.⁷ To match fathers to the census, we match by first name, last name and birthplace. Unfortunately, father's age is not available in the death certificates, which decreases our match rates due to double matches. The matching procedure is outlined below, except that for this match we do not iterate based on age but rather match directly based on name and birthplace.

Once we have matched the death certificates to the census, we have a sample of live and deceased infants for each census year. We identify the father of the infant within the census indices. Currently, we require fathers to be household heads and the infant to be identified as the son of the household head.⁸ We match the fathers of the infants backwards to a previous census wave either 10 or 20 years earlier. Many fathers are successfully matched to both waves so we end up with pre-migration childhood household information but also pre-migration occupation information. When we have identified the individual in a previous census index, we digitize by hand the characteristics that are not available in the indexes: occupation, home ownership, literacy, school attendance (if applicable).

To match individuals, we follow the procedure pioneered by Ferrie (1996) and used in Abramitzky, Boustan, and Eriksson (2012). The procedure is as follows:

- (1) We begin by standardizing the first and last names of men in the later year to address orthographic differences between phonetically equivalent names using the NYSIIS algorithm (Attack and Bateman 1992). We also recode any common nicknames to standard first names (e.g. Will becomes William).
- (2) We match observations backwards from the later year to the earlier year using an iterative procedure. We start by looking for a match by first name, last name, race, state of birth, and exact birth year. There are three possibilities:
 - (a) if we find a *unique* match, we stop and consider the observation "matched";
 - (b) if we find multiple matches for the same birth year, the observation is thrown out;
 - (c) if we do not find a match at this first step, we try matching within a one-year band (older and younger) and then with a two-year band around the reported birth year; we

⁷ We plan to expand our sample by looking at older children who may have died as early as 5 years before the census year. This larger sample size comes at the price of potentially lower match rates as well as selective migration post-infant-death.

⁸ Work is ongoing to include the remaining live and deceased infants. We worry that the sample of infants with a father as a household head is different than those with a father who is not household head.

only accept unique matches. If none of these attempts produces a match, the observation is discarded as unmatched.

Our matching procedure generates a final sample of 164 deceased infants and 6,332 non-deceased infants with fathers born in Tennessee but who are living in Tennessee, Illinois, or Ohio in the outcome year. We can successfully match 9.4 percent of deceased infants and 32 of non-deceased infants.⁹

Individuals can fail to match due to (i) non-unique name-birthplace-race combinations; (ii) multiple matches in the earlier year; (iii) misreporting of age; and (iv) complete misspellings of the name. Note that mortality cannot account for any failure to match due to starting with the later year and matching backwards. Other papers using this method (Abramitzky, Boustan and Eriksson 2012) find that non-unique combinations in 1940 account for 40% of match failures, while misreported ages account for about 10% of failure to match. Finally, misspellings account for about 38% of failures.¹⁰

Based on the same previous work, we expect that matched individuals come from slightly higher socio-economic status. Generally, matched individuals have less common names, which is potentially correlated with better economic outcomes. For our estimates to be unbiased within our sample, we worry about this matching bias only if it is differential across states and differential across deceased and non-deceased infants. In our current sample, match rates are higher in northern states and for live births, leading us to underestimate the mortality differential between the north and south.

Section 4: Black Infant Mortality in the North and South

In this section, we tackle the question of which region provided a better all-encompassing environment for infant health. In addition, this exercise provides a useful test of our new dataset in comparison to the published vital statistics.

First, we compare the constructed birth counts to those found in the published volumes of the *Vital Statistics of the United States*. To form the denominator for the constructed mortality rates, we add the count of observations found in the full census born in each state and birth year to the deaths of infants and non-infants who died during the intervening years. The birth year is

⁹ Part of the reason that these are so different is that the deceased infants have to be matched twice. We are planning to investigate the extent to which this causes bias in our estimates.

¹⁰ We expect these numbers to be similar in our dataset but plan to test this after establishing our final sample.

determined by the reported age at last birthday relative to the census day: January 1st for 1920, April 1st for 1930 and 1940. Table 1 reports the results from this comparison for African-American children.

In general, we find that the census undercounts the number of 0-1 year olds relative to the *Vital Stats*. This holds for both black and white children in all states. However, the census reports more children than the *Vital Stats* for ages two to five. We include birth year fixed effects in all regressions to account for the pattern common to all states and census periods to assuage concerns about the undercounting for 0-1 year olds.

For the older children, it is unclear whether the census or the *Vital Stats* provides a more accurate measure of births. *Vital Stats* undercounted births in the South because of incomplete birth registration, and thus overestimating the southern mortality rate. An evaluation in 1939 showed that African-American births went unregistered at particularly high rates, especially in the South and rural areas (Shapiro 1950). Only 80 percent of 1939 black births were registered in Tennessee, compared to close to 97 percent of black births in northern cities.

We use two procedures for most of the estimates in the following sections: (1) unweighted, (2) weighted so that total births will match the published births from the *Vital Stats* for a specific state and year of birth. Because the *Vital Stats* undercounts births relatively more in the South than in Northern cities, the unweighted constructed measure tends to increase mortality rates and effects in the North. However, this is a potential improvement as census counts might be closer to the true number of births.

The aggregate statistics from the published volumes of the *Vital Statistics of the United States* clearly show a rapidly changing infant health environment for blacks in both regions. Early in the 20th century, infant mortality rates were lower in the southern states, mainly driven by the large proportion of African Americans living in rural areas. By the early 1930s, that relationship reversed, with lower rates in the northern urban areas.

One concern with this comparison is that the underlying characteristics of the parents that contribute to the health of the infant differ between the regions. As shown in table 2, in our sample for 1940, new fathers in the north had higher levels of education and earnings. Our data partially ameliorate the problem by allowing us to control for observable characteristics of the parents. As a first step, we estimate equation (1) with the full sample of data.

$$(1) \quad D_i = \gamma_0 + \gamma_1 N_i + X_i \gamma_2 + \varepsilon_i$$

where D_i is an indicator equal to one if the child died, and N_i is an indicator for a northern birth. The coefficient of interest is γ_1 , which provides the unconditional difference in likelihood of infant death between the northern and southern states in our sample. Without including controls X , estimates of γ_1 are weighted differences of North-South birth year x state cells. However, comparing the results to those reported in the *Vital Stats* can be useful.

Table 3 directly compares the results when weighting observations to sum to the published birth counts (columns 5-8), and when not using weights (columns 1-4). The first column of each panel pools observations across years. Over the entire period, northern births were 1.6 p.p more likely to end in death compared to the southern births. Weighting observations to match the VSUS births decreases the northern mortality disadvantage. This is exactly as expected if black births go unregistered at a relatively higher rate in Tennessee.

The subsequent columns report estimates of γ_1 separately for each year. From 1920 to 1930 the mortality advantage of the South increases slightly, subsequently falling during the 1940s. These findings broadly match the diminishing southern mortality advantage found in the published rates of the *Vital Stats*.

Previous historical analyses of mortality have been hindered by the necessity of using aggregate mortality rates, for example, at the state-level or city-level. Our new individual-level data set allows us to improve on previous methods by accounting for individual level observables. For example, we can use the census information to determine the contribution income to the variation in regional infant mortality rates. Accordingly, we estimate equation (1), which includes controls for the observable characteristics of the parents, X_i : income, literacy or education, and homeownership.

The results from estimating equation (1) including SES controls are reported in column 2 of table 3(b).¹¹ Controlling for a set of economic variables we believed *a priori* to be important factors for infant mortality did not significantly change our estimate of the southern mortality advantage in 1940. Births in the North remain 5 p.p. more likely to end in death within a year. The interpretation of this preliminary result would be that the large regional differences in nominal and real income, education, and homeownership did not account for any meaningful part of the regional difference in infant mortality.

¹¹ The 1940 census data is from the 100% digitized sample housed at the NBER. We plan to expand the analysis to include the controls for previous census years, which need to be digitized from the original census manuscripts.

Surprisingly, the coefficient estimate on income is positive and statistically significant, implying that higher incomes are correlated with higher rates of infant death. This result is at odds with much of the modern and historical work on infant mortality (Preston and Haines 1991, Finch 2003). Income is the log wage and salary income reported by the father, unadjusted for regional differences in the cost-of-living. The self-employed did not report income on the 1940 census, which is important for the southern farm owners and operators. We coded observations without a positive income as 0 and included an indicator for missing. In a future revision, we will assign farm and self-employed incomes based on a comparison to wage workers in the same occupation in 1960 following the method in Collins and Wannamaker (2014).

Section 5: Contribution of migration to infant health gains.

Given the high economic returns to migrating North, and the potential positive selection of migrants, health outcomes might be expected to improve with migration.¹² As infants are potentially more susceptible to the differences in health environment—Northern cities were less sanitary but had better health care as the 20th century continued – infant health provides an important indicator for the overall changes in health from migration, and the cost of obtaining the economic returns.

The basic conceptual framework is to compare infant outcomes of blacks that migrated to the outcomes of those that stayed in the South. One innovation of our work is to include parental controls for income, education, and literacy to assess the mechanisms through which migration may have affected health.

A. Naive Estimates

We estimate the impact of migration on infant mortality using the regression with OLS:

$$(2) \quad D_{i,t} = \alpha + \beta_1 M_i + X_i \beta_2 + u_i,$$

where $D_{i,t}$ is an indicator of infant mortality for individual i , M_i is an indicator of migration (=1 if residing in the North post-migration period). We estimate the equation limiting our sample to fathers born in Tennessee, and in this exercise pool data across time-periods; we also estimate

¹² Surprisingly, the migrants themselves experienced elevated mortality rates compared to non-migrants (Black *et al.* 2011).

the equation separately for three time periods. The coefficient of interest is β_1 , which measures the impact of migration on infant mortality in percentage points.

As seen in column 1 of table 4, the estimate excluding any controls, X , is consistent with migrants facing higher rates of infant mortality. Using the estimates without weighting to VSUS births suggests migrants to the northern cities experienced a 2.3 p.p. increase in infant mortality on average over the entire period compared than those fathers that remained in Tennessee. Moreover, we see some significant changes in the size of this correlation over time, mirroring the (slight) increase and subsequent decreases in the overall southern mortality advantage from table 3. Again, weighting to VSUS birth counts slightly reduces the negative effect of a northern birth. Changes in the composition of migrants, the relative disease environments, or usefulness of parental income and education in health production potentially explain the evolution of this correlation.

B. Evidence of and Accounting for Selection

The main concern with the naive result is that the error term contains individual and local characteristics that are correlated with both the migration decision and infant health, introducing bias into the estimate of β_1 . We take a number of steps to account for the possibility of this selection bias.

A first concern is that migrants differentially came from households with higher levels of income, wealth, or education as is shown in table 5.¹³ As migration is costly, sons from black households with greater resources were more likely to migrate. The migrant's health, and thus the health outcomes of future offspring may be positively correlated with pre-migration economic resources. To address this issue, we control for a rich set of parental background characteristics from our matched-sample of Tennessee-born fathers, and re-estimate equation (2).

X_i is a set of background characteristics of the parents measured in the pre-migration period. Individual level controls include an industrial earnings score based on Lebergott (1964), homeownership, and literacy.

Column (1) of table 6 limits the sample to observations for which we have pre-migration earnings scores. Any changes in the coefficient after controlling for observable pre-migration characteristics are not because of differences in fathers included in the sample.

¹³ Collins and Wannamaker (2014) and Black et al. (2011) also find that migrants are positively selected on a number of socioeconomic variables.

In the next two columns we explore the possibility of selection on observables and unobservables using our matched-sample. In column 2, we include X in the regression to control for observable pre-migration individual and household characteristics of the father.¹⁴ Controlling for earnings score, literacy, and homeownership does not significantly change the coefficient estimate from the naive regression, suggesting that selection on observables makes up a small part of the observed differences in infant mortality between migrants and non-migrants.

The final column adds county-of-origin fixed effects to account for the share of the raw difference in mortality between migrants and non-migrants by location-specific unobservables. We find the estimate is slightly lower, but the change represents only a small portion of the raw migration effect. We do not find strong evidence of selection bias based on the results from our matched-sample.

Section 6: Comparing Native and Migrant Outcomes in the North

In this section we explore how migrant birth outcomes compared to blacks born and residing in the North. Contemporary data suggests that immigrant mothers experience better birth outcomes than do the native-born U.S. population, with a gradual equalization of outcomes as time in the U.S. passes (Hummer *et al.* 2007, Antecol and Bedard 2006). However, evidence from the 1900 decennial census schedules suggests that mortality was higher for children born to immigrants, compared to those of native-born parents, with substantial variability across sending nations (Preston and Haines 1991).

We limit the sample to include all births in northern states, IL and OH, regardless of parent's birthplace. We estimate equation

$$(3) \quad D_i = \lambda_0 + \lambda_1 M_i + X_i \lambda_2 + \varepsilon_i$$

where D_i is a death indicator and M_i indicates that the parent migrated from the South. The coefficient of interest, λ_1 , provides the raw difference in mortality outcomes for migrants and non-migrants. The base results reported in the first column of table 7 are consistent with migrants having better health outcomes than non-migrants in the North for the full period. Northern-born parents were 2.2 p.p. more likely to experience a child's death in the first year. Columns 2

¹⁴ Controls include an industry earnings score following the procedure of Collins and Wannamaker (2014) based on Lebergott (1964), and discussed in the data section. We define literacy as able to both read and write. If the father was found as an adult in a prior census, then that individual's information was used. Otherwise, the head's information is used from the household in which the father is found as a child in a prior census.

through 6 re-estimate equation 3 for each year separately. Our results suggest that over time migrant outcomes were deteriorating relative to non-migrants.

Our data allow us to explore the potential mechanisms for the difference in outcomes between migrants and non-migrants. Migrants and non-migrants in the North differ along a number of dimensions important for infant health. Panel B of table 1 shows summary statistics for our sample of new fathers in 1940. Migrants received higher earnings than northern-born fathers, but had lower levels of education and lower rates of homeownership. The economics literature focused on comparing earnings finds a similar result (Weiss and Williamson 1972, 1975, Smith and Welch 1989).

In a regression framework, we want to observe the change in λ_1 as we add controls for these observable characteristics. The purpose is to determine if the migrant advantage in infant health disappears when controlling for observable characteristics believed to enter the health production function. Moving from column 5 to column 6, we see that the southern-born mortality advantage widens after controlling for income, homeownership status, and educational attainment. In the absence of the northern-born blacks' higher levels of education, the gap would have been even wider. We interpret these results as evidence that southern-born migrants held more health capital or had better knowledge of healthy behaviors that translated into better infant outcomes compared to northern-born blacks. A result similar to the "Hispanic paradox" found in recent U.S. data, that is generally explained by the "healthy migrant" effect (Hummer *et al.* 2007).

Section 7: Discussion

The movement North of African Americans during the early 20th century was associated with large increases in infant mortality, despite the large increases in income. We do not find evidence of the "healthy migrant" effect using infant mortality as an indicator of the parental health. More importantly, if healthier fathers were more likely to migrate north, they were unable to transfer it to their children's initial health capital stock to fully account for the negative health influences experienced in northern cities.

In future versions, we plan to use the rich information we have on parental characteristics before and after migration to further explore the implications of our findings:

1. Through which channels did the migration decision affect mortality rates? For instance, post-migration incomes were on average 90 percent higher (Collins and Wannamaker 2014). How much higher would mortality have been in the absence of the income gains? Rural southerners not only migrated to northern cities, but to urban areas in the South. To what extent is this an urban-rural migration story? Finally, different channels would work through different causes-of-death. For example, neonatal death is associated with mother's health and prenatal causes, whereas the majority of non-neonatal deaths are from infection. The published *Vital Stats* show that neonatal deaths made up a larger share of total infant deaths in the North, 56 percent to 43 percent. We hope to leverage the detailed birth and death date and the cause-of-death information listed on the death certificates.
2. Our result that moving North *increased* infant mortality implies that the black-white infant mortality gap and life-expectancy gap would have declined even faster in the absence of the Great Migration. The choices of when and where to migrate depended on the job prospects of blacks once they got to the North. The migration may have occurred decades earlier if large numbers of international immigrants did not flood the urban job centers (Collins 1997). Much of the total movement of blacks out of the South occurred after 1940, with the highest rates occurring during the 1940s and 50s, potentially related to the industrialization in the north, mobilization, and the war effort (Boustan 2009). What would African-American health progress have been under the plausible counterfactuals of earlier, faster, or no Great Migration?

References

- Abramitzky, Ran, Leah Boustan, and Katherine Eriksson. 2012. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration". *American Economic Review*. 102(5): 1832-1956.
- Atack, Jeremy, Fred Bateman, and Mary Eschelbach Gregson. 1992. "'Matchmaker, Matchmaker, Make Me a Match': A General Personal Computer-Based Matching Program for Historical Research." *Historical Methods* 25(2): 53–65.
- Almond, Douglas, Kenneth Y. Chay and Michael Greenstone, 2006, "Civil Rights, the War on Poverty, and Black-White Convergence in Infant Mortality in the Rural South and Mississippi" (December 31, 2006). MIT Department of Economics Working Paper No. 07-04
- Almond, Douglas and Janet Currie, 2011. "Human Capital Development before Age Five," *Handbook of Labor Economics*, O. Ashenfelter & D. Card (ed.), edition 1, volume 4, number 5, Elsevier.
- Antecol, H. and K. Bedard. 2006. Unhealthy Assimilation: do Immigrants Converge to American Weights? *Demography*, 43 (2), May 2006, 337-360.
- Black, Dan A., Magdalena Muszynska, Seth G. Sanders, Evan J. Taylor, and Lowell J. Taylor, 2011. "The Great Migration and African-American Mortality: Evidence from Mississippi," Working Paper.
- Bleakley, Hoyt, 2007. "Disease and Development: Evidence from Hookworm Eradication in the American South," *Quarterly Journal of Economics*, vol. 122(1), pages 73-117, February.
- Boustan, Leah, 2009. "Competition in the Promised Land: Black migration and Racial Wage Convergence in the North, 1940-1970," *The Journal of Economic History*, vol. 69(3), pages 756-783, September.
- Case, Anne and Christina Paxson, 2009. "Early Life Health and Cognitive Function in Old Age," *American Economic Review: Papers and Proceedings*, vol. 99(2), pages 104-109, May.
- Chay, Kenneth Y., Jonathan Guryan and Bhashkar Mazumder, 2009. "Birth Cohort and the Black-White Achievement Gap: The Roles of Access and Health Soon After Birth," NBER Working Papers 15078, National Bureau of Economic Research, Inc.
- Collins, William J., 1997. "When the Tide Turned: Immigration and the Delay of the Great Black Migration," *The Journal of Economic History*, vol. 57(3), pages 607-632, September.
- Collins, William J. and Melissa A. Thomasson, 2004. "The Declining Contribution of Socioeconomic Disparities to the Racial Gap in Infant Mortality Rates, 1920–1970," *Southern Economic Journal*, Southern Economic Association, vol. 70(4), pages 746-776, April.

Collins, William J. and Marianne H. Wanamaker, 2014. "Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data," *American Economic Journal: Applied Economics*, vol. 6(1), pages 220-252, January.

Cutler, David and Grant Miller. 2005. "The Role of Public Health Improvements in Health Advances: The Twentieth Century United States". *Demography*. 41(1): pp1-22.

Delaney, Liam, Alan Fernihough and James P. Smith, 2013. "Exporting Poor Health: The Irish in England," *Demography*, vol. 50(6), pages 2013-2035.

Eichenlaub, Suzanne C., Stewart E. Tolnay and J. Trent Alexander, 2010. "Moving Out but Not Up: Economic Outcomes in the Great Migration," *American Sociological Review*, vol. 75(1), pages 101-225, February.

Federal Security Agency, 1947. *Vital Statistics Rates in the United States 1900-1940*. Washington, DC: Government Printing Office.

Ferrell, John and Pauline Mead, "History of county health organizations in the united States," *Public Health Bulletin*, 222: U.S. Government Printing Office, 1936.

Ferrie, Joseph. (1996) "A New Sample of Males Linked from the Public Use Micro Sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census Manuscript Schedules." *Historical Methods* 29: 141–56.

Finch, Brian K., 2003. "Early Origins of the Gradient: The Relationship Between Socioeconomic Status and Infant Mortality in the United States," *Demography*, vol. 40(4), pages 675-699, November.

Flippen, Chenoa, 2013. "Relative Deprivation and Internal Migration in the United States: A comparison of Black and White Men," *American Journal of Sociology*, vol. 118(5), pages 1161-1198, March.

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 (2006).

Halliday, Timothy J. and Michael C. Kimmit, 2008. "Selective Migration and Health in the USA, 1984-93," *Population Studies*, vol. 62(3), pages 321-334.

Hummer, R., D. Powers, S. Pullum, G. Gossman, and W. Frisbie. 2007. "Paradox found (again): infant mortality among the Mexican-origin population in the United States". *Demography* 44(3): 441-57

Jasso, Guillermina, Douglas S. Massey, Mark R. Rosenzweig, and James P. Smith, 2004.

"Immigrant Health-Selectivity and Acculturation." Pp. 227-266 in N. B. Anderson, R. A. Bulatao, and B. Cohen (eds.), *Critical Perspectives on Racial and Ethnic Differences in Health in Late Life*. Washington, DC: National Academy Press.

Ladd-Taylor, Molly. 1988. "'Grannies' and 'Spinsters': Midwife Education under the Sheppard-Towner Act." *Journal of Social History* 22 (2), 255-275.

Lebergott, Stanley. 1964. *Manpower in Economic Growth: the American Record Since 1800*. New York, NY: McGraw-Hill.

Logan, Trevon, and John M. Parman. 2011. "Race, Socioeconomic Status, and Mortality in the 20th Century: Evidence from the Carolinas." PSC Research Report No. 11-739. May 2011.

Margo, Robert A., 1990. *Race and Schooling in the South, 1880-1950*. Chicago: University of Chicago Press.

Margo, Robert A. 1996. "Wages." In *Historical Statistics of the United States, Millennial Edition*, edited by Susan B. Carter et al., pp. 2-254-2-300. New York: Cambridge University Press.

McHugh, Kevin, 1987. "Black Migration Reversal in the United States," *Geographical Review*, vol. 77(2), pages 171-182, April.

Muller, Christopher, 2012. "Northward Migration and the Rise of Racial Disparity in American Incarceration, 1880-1950," *American Journal of Sociology*, vol 118(2), pages 281-326, September.

Pamuk E, Makuc D, Heck K, Reuben C, Lochner K. Socioeconomic Status and Health Chartbook. Health, United States, 1998. Hyattsville, Maryland: National Center for Health Statistics. 1998.

Preston, Samuel H. and Michael R. Haines, 1991. "Fatal Years: Child Mortality in Late Nineteenth-Century America," NBER Books, National Bureau of Economic Research, Inc, number pres91-1.

Preston, S.H., I.T. Elo, I. Rosenwaik, and M. Hill (1996). "African American Mortality at Older Ages: Results from a Matching Study," *Demography* 35: 1-21.

Shapiro, S., 1950. "Development of Birth Registration and Birth Statistics in the United States," *Population Studies*, vol. 4(1), pages 86-111, June.

Smith, James P. and Finis R. Welch, 1989. "Black Economic Progress After Myrdal," *Journal of Economic Literature*, v ol. 27, pages 519-564, June.

Thomasson, Melissa A. and Treber, Jaret, 2008. "From home to hospital: The evolution of childbirth in the United States, 1928-1940," *Explorations in Economic History*, Elsevier, vol. 45(1), pages 76-99.

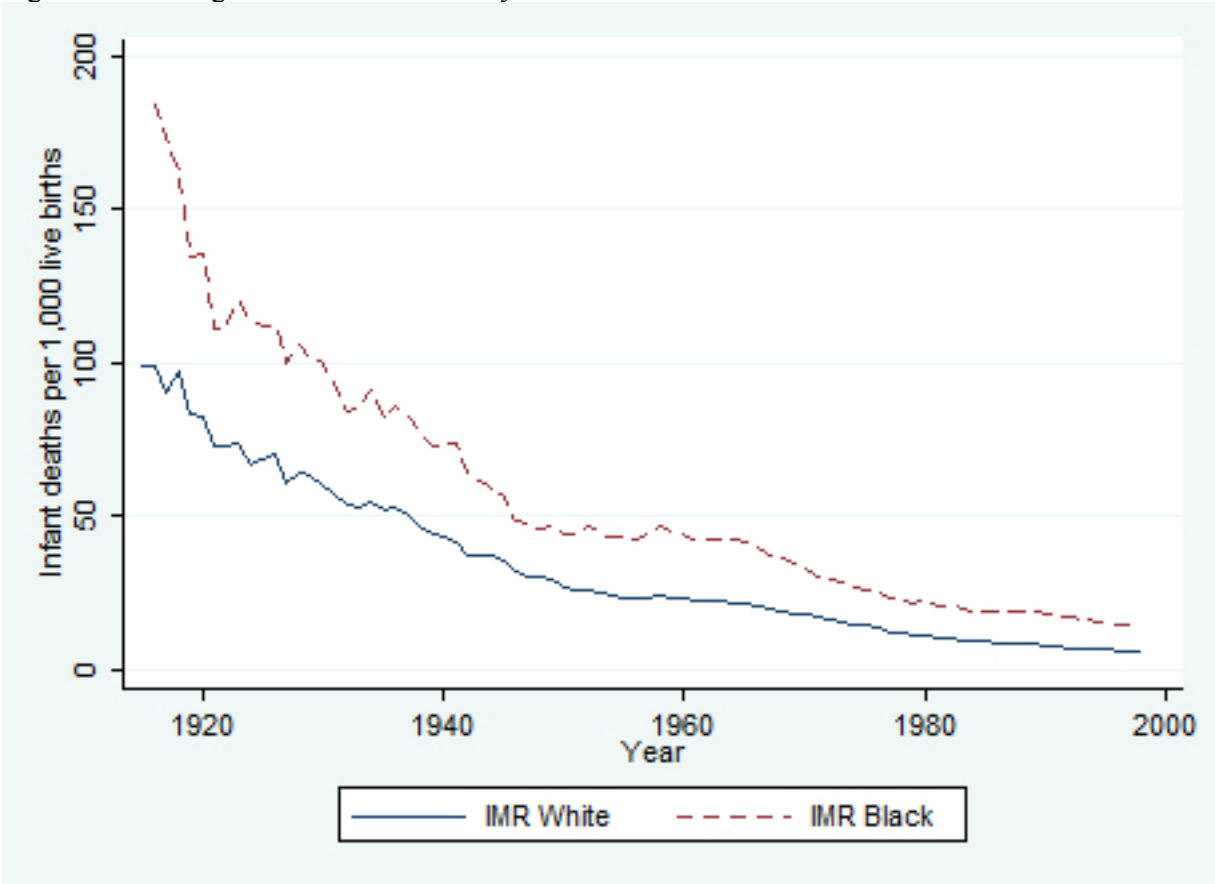
Weiss, Leonard and Jeffrey G. Williamson, 1972. "Black Education, Earnings and Inter-regional Migration: Some New Evidence," *The American Economic Review*, vol. 62(3), pages 372-383, June.

Weiss, Leonard and Jeffrey G. Williamson, 1975. "Black Education, Earnings and Inter-regional Migration: Even Newer Evidence," *The American Economic Review*, vol. 65(), pages 241-244, March.

Wolf, Jacqueline, "Saving Babies and Mothers: Pioneering Efforts to Decrease Infant and Maternal Mortality," in *Silent Victories: The History and Practice of Public Health in Twentieth-Century America*, John Ward and Christian Warren ed., (New York, NY: Oxford University Press, 2007).

Woodbury, R. M. 1995. *Causal Factors in Infant Mortality*. U.S. Children's Bureau, No 142. Washington, DC.

Figure 1: Convergence of infant mortality rates for blacks and whites



Notes: For 1915-1932, data are for the current Birth Registration Area only. Source: Haines (2006).

Figure 2: Urban-Rural-North-South gap in black infant mortality rates.

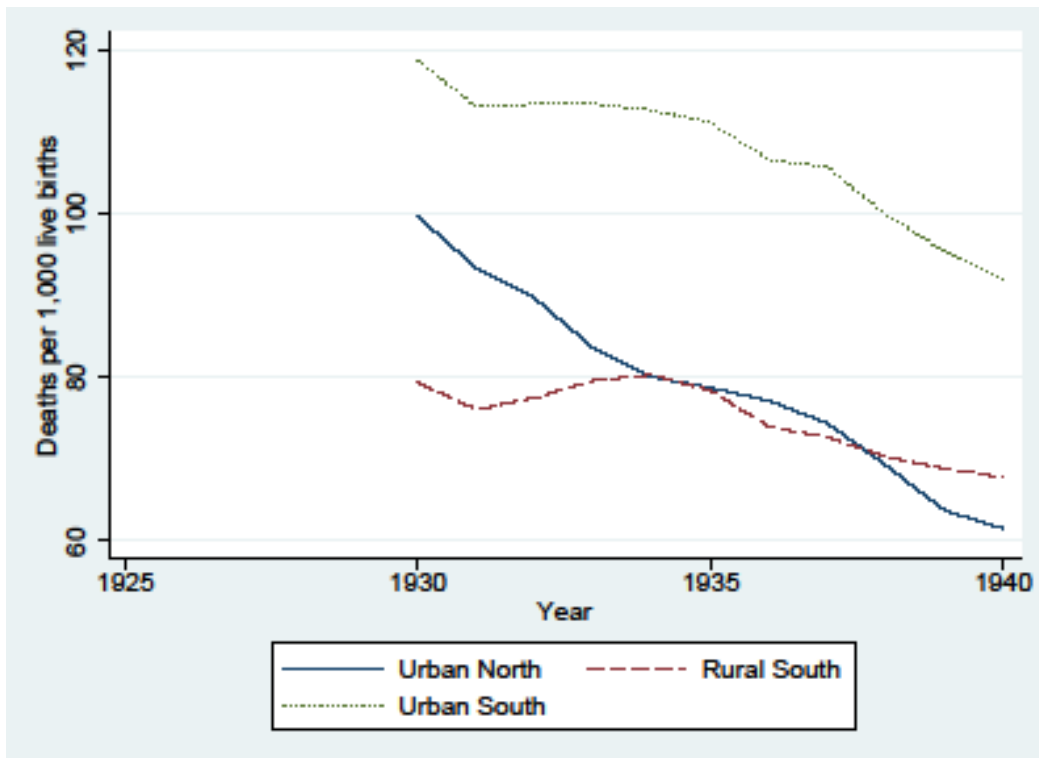
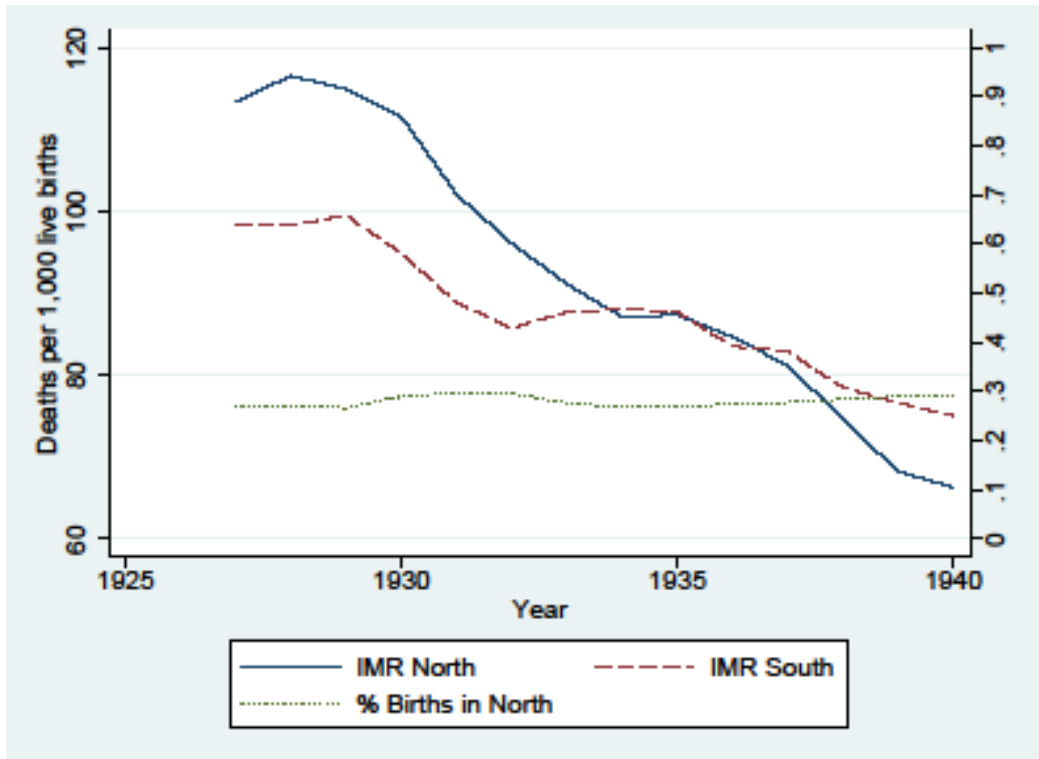


Table 1: Comparison of Constructed Births to *Vital Statistics of the United States*

	Year	VSUS Births	Constructed Births			Ratio of VSUS Births to Constructed Births	
			Census Count	Deaths	Total Births		
Illinois	1940	6,841	5,190	309	5,499	1.2	
	1939	6,473	5,071	500	5,571	1.2	
	1938	6,299	5,770	455	6,225	1.0	
	1937	5,860	5,677	528	6,205	0.9	
	1936	5,593	5,685	646	6,331	0.9	
	1935	5,764	5,710	686	6,396	0.9	
	1930	5,601	4,860	420	5,280	1.1	
	1929	5,548	4,716	791	5,507	1.0	
	1928	5,474	5,160	864	6,024	0.9	
	1927	5,757	5,313	1,057	6,370	0.9	
	Ohio	1940	6,120	4,728	215	4,943	1.2
		1939	5,854	4,724	335	5,059	1.2
		1938	5,972	5,339	450	5,789	1.0
		1937	5,553	5,093	505	5,598	1.0
1936		5,336	5,284	560	5,844	0.9	
1935		5,225	5,249	516	5,765	0.9	
1930		5,514	5,298	756	6,054	0.9	
1929		5,377	5,079	894	5,973	0.9	
1928		5,451	5,715	839	6,554	0.8	
1927		5,687	6,037	881	6,918	0.8	
Tennessee		1940	8,525	7,748	315	8,063	1.1
		1939	8,087	7,650	621	8,271	1.0
		1938	8,049	9,078	776	9,854	0.8
		1937	8,079	8,780	553	9,333	0.9
	1936	7,789	9,233	814	10,047	0.8	
	1935	8,333	9,012	872	9,884	0.8	
	1930	8,125	7,812	839	8,651	0.9	
	1929	8,148	7,702	1,055	8,757	0.9	
	1928	8,096	9,007	1,179	10,186	0.8	
	1927	9,060	9,085	1,304	10,389	0.9	

Notes: Constructed births consist of count of census observations from the closest decennial census for each birth year and deaths in the intervening years allocated to birth year (both infant and non-infant deaths).

Sources: VSUS births comes from the published volumes of the *Vital Statistics of the United States*. Census counts and death data are from the digitized indexes provided by FamilySearch.org.

Table 2: Difference in economic and social resources of new parents in 1940

	<i>Panel A: Location of child's birth</i>		<i>Panel B: Northern children by parental birthplace</i>	
	<i>North</i>	<i>South</i>	Northern-born parents	Southern-born parents
Years of Schooling	7.8 (3.2)	5.6 (2.9)	9.056 (2.821)	7.385 (3.219)
Income	582 (500)	283 (373)	575.12 (544.31)	584.15 (486.95)
Income if >0	731 (453)	450 (383)	706.08 (520.89)	739.94 (431.96)
Home Ownership	0.11 (0.32)	0.13 (0.34)	0.173 (0.379)	0.093 (0.291)

Notes: Observations are from the full 100 percent 1940 census sample, and include the information of fathers of one year olds or infants that died in the previous year that we were able to match to the 1940 full census. The sample was restricted to fathers that were also household heads. Income is wage and salary income. The 1940 income measure did not include any self-employment income. Panel A includes births in IL, OH, and TN and splits the sample by the location of the *child's* birth. Panel B includes only births in IL and OH and splits the sample by location of *parent's* birth.

Table 3: North-South differential in black infant mortality

	<i>Weighted by VSUS death counts</i>				<i>Weighted by VSUS birth and death counts</i>			
	<i>Full</i>	<i>1920</i>	<i>1930</i>	<i>1940</i>	<i>Full</i>	<i>1920</i>	<i>1930</i>	<i>1940</i>
	<i>eq. 1</i>	<i>eq. 1</i>	<i>eq. 1</i>	<i>eq. 1</i>	<i>eq. 1</i>	<i>eq. 1</i>	<i>eq. 1</i>	<i>eq. 1</i>
NORTH	0.016*** (0.001)	0.024*** (0.003)	0.027*** (0.001)	0.0002 (0.001)	0.0095*** (0.001)	0.014*** (0.003)	0.024*** (0.002)	-0.008*** (0.002)
Constant	0.11*** (0.003)	0.10*** (0.003)	0.12*** (0.002)	0.07*** (0.002)	0.12*** (0.003)	0.12*** (0.003)	0.13*** (0.003)	0.08*** (0.002)
Birth year fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Observations	333,991	79,526	131,722	122,743	333,991	79,526	131,722	122,743
R-squared	0.019	0.016	0.012	0.008	0.024	0.020	0.014	0.012

Notes: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The dependent variable is equal to 1 if the child died as an infant and 0 otherwise. The coefficients can be interpreted as a p.p. impact on infant mortality. The sample includes African-American births and infant deaths in Illinois, Ohio, and Tennessee to fathers born in any state. Cols 1 and 4 pool observations from all years, whereas all other columns limit the sample to a single decade. The last column uses a matched sample to the 1940 census to include father's income, homeownership status, and educational attainment as a control. Col 5 limits to the 1940 for sample to observations that can be matched to the census. Cols 5 and 6 weight observations by the inverse of the match rate to the census in the appropriate state and D_i cell.

Table 3(b): North-South differential in black infant mortality

	<i>1940</i>	<i>1940</i>
	<i>eq. 1</i>	<i>eq. 1</i>
North	0.0515*** (0.008)	0.052*** (0.009)
1940 Log income		0.012** (0.006)
1940 Homeownership		0.014 (0.014)
1940 Educational Attainment		-0.0048*** (0.002)
Constant	0.045*** (0.004)	0.0061 (0.036)
Year fixed effects	N	N
Observations	12,147	12,147
R-squared	0.0096	0.0155

Notes: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ The dependent variable is equal to 1 if the child died as an infant and 0 otherwise. The coefficients can be interpreted as a p.p. impact on infant mortality. The sample includes African-American births and infant deaths in Illinois, Ohio, and Tennessee to fathers born anywhere. The last column uses a matched sample to the 1940 census to include father's income, homeownership status, and educational attainment as a control. This table limits the sample to those observations from 1939-1940 that can be matched to the 1940 census. Both columns weight observations by the inverse of the match rate to the census in the appropriate state and D_i cell. Neither column uses weighting schemes of table 4(a).

Table 4: Correlation between migration to the North and Infant Mortality for fathers born in Tennessee

	<i>Weighted by VSUS death counts</i>				<i>Weighted by VSUS birth and death counts</i>			
MIGRANT	0.023*** (0.003)	0.028*** (0.007)	0.039*** (0.004)	0.0002 (0.003)	0.016*** (0.003)	0.017** (0.008)	0.037*** (0.005)	-0.008 (0.003)
Year indicator								
1916	0.023*** (0.00529)	0.023*** (0.005)			0.030*** (0.006)	0.030*** (0.006)		
1917	0.003 (0.0050)	0.003 (0.005)			-0.0006 (0.006)	-0.0006 (0.006)		
1918	-0.004 (0.005)	-0.004 (0.005)			0.0009 (0.006)	0.0006 (0.006)		
1919	-0.0006 (0.006)	-0.0008 (0.006)			-0.015** (0.006)	-0.015** (0.006)		
1920	-0.097*** (0.004)	-0.097*** (0.004)			-0.116*** (0.005)	-0.116*** (0.005)		
1925	-0.030*** (0.005)				-0.032*** (0.005)			
1926	-0.046*** (0.005)		-0.016*** (0.004)		-0.051*** (0.005)		-0.018*** (0.005)	
1927	-0.044*** (0.005)		-0.013*** (0.004)		-0.051*** (0.005)		-0.018*** (0.005)	
1928	-0.042*** (0.005)		-0.012*** (0.004)		-0.043*** (0.005)		-0.010** (0.005)	
1929	-0.052*** (0.005)		-0.021*** (0.004)		-0.065*** (0.005)		-0.032*** (0.005)	
1930	-0.107*** (0.004)		-0.077*** (0.003)		-0.125*** (0.004)		-0.092*** (0.004)	
1935	-0.048*** (0.005)				-0.054*** (0.006)			
1936	-0.067*** (0.00452)			-0.018*** (0.004)	-0.071*** (0.005)			-0.016 (0.005)
1937	-0.049*** (0.006)			-0.0005 (0.005)	-0.058*** (0.006)			-0.003 (0.006)
1938	-0.066*** (0.005)			-0.018*** (0.004)	-0.075*** (0.005)			-0.020 (0.005)
1939	-0.063*** (0.005)			-0.014*** (0.005)	-0.081*** (0.005)			-0.026 (0.005)
1940	-0.115*** (0.004)			-0.066*** (0.004)	-0.134*** (0.004)			-0.079 (0.004)
Constant	0.123*** (0.004)	0.122*** (0.004)	0.090*** (0.003)	0.077*** (0.003)	0.142*** (0.004)	0.142*** (0.004)	0.106*** (0.003)	0.090* (0.004)
N	135,812	42,561	51,237	42,014	135,812	42,561	51,237	42,014
R-squared	0.018	0.014	0.011	0.008	0.022	0.019	0.013	0.011

Table 5: Summary statistics of pre-migration characteristics in matched sample

	<i>Full</i>	<i>Non-migrants</i>	<i>Migrants</i>
Literacy	0.64 (0.48)	0.63 (0.48)	0.75 (0.43)
Income	833 (414)	810 (401)	1079 (467)
Owned home	0.25 (0.43)	0.25 (0.43)	0.25 (0.43)
Obs	1605	888	717

Notes: Mean values with standard errors below in parentheses. The matched sample consists of outcome year migrants and non-migrants matched back to a pre-migration census. All observations are African Americans born in Tennessee. Income is in 1950 dollars as is assigned following Collins and Wannamaker (2014) based on industry earnings from Lebergott (1964) reported in Margo (1996).

Table 6: Impact of migration on infant mortality accounting for selection

	<i>Weighted by VSUS death counts</i>			<i>Weighted by VSUS birth and death counts</i>		
Migrant	0.034*** (0.009)	0.035*** (0.010)	0.031*** (0.01)	0.029*** (0.008)	0.030*** (0.009)	0.027*** (0.010)
1930 indicator	-0.013*** (0.004)	-0.012*** (0.004)	-0.011** (0.005)	-0.013*** (0.004)	-0.011*** (0.004)	-0.010** (0.005)
1940 indicator	0.005 (0.005)	0.007 (0.005)	0.010* (0.006)	0.003 (0.005)	0.005 (0.004)	0.009 (0.006)
Log income		-0.002 (0.004)	-0.003 (0.007)		-0.002 (0.004)	-0.003 (0.007)
Literacy		-0.006 (0.004)	-0.006 (0.005)		-0.006 (0.004)	-0.006 (0.005)
Homeownership		0.009* (0.005)	0.009 (0.006)		0.009* (0.005)	0.009 (0.006)
Constant	0.020*** (0.003)	0.033 (0.030)	0.019 (0.050)	0.020*** (0.003)	0.032 (0.029)	0.018 (0.049)
Observations	1,605	1,605	1,605	1,605	1,605	1,605
R-squared	0.008	0.009	0.068	0.006	0.007	0.066
County Fixed Effects	NO	NO	YES	NO	NO	YES

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 The dependent variable is equal to 1 if the child died as an infant and 0 otherwise. The coefficients can be interpreted as a p.p. impact on infant mortality. The sample includes African-American births and infant deaths in Illinois, Ohio, and Tennessee to fathers born in Tennessee. The matched sample includes all black births and deaths in Ohio and Illinois, and deaths only in TN, for which we can match the parents back to a prior census wave, and a 10 percent sample of births in TN. Results are weighted to account for the sampling and differential match rates to previous census waves. The full sample includes *all* TN births.

Table 7: "Migrant health paradox"? Infant mortality outcomes for migrant parents relative to northern-born parents

	<i>Full</i>	<i>1920</i>	<i>1930</i>	<i>1940</i>	<i>1940 Matched to Census</i>	<i>1940 Matched to Census</i>
Northern born parent	0.022*** (0.002)	0.033*** (0.006)	0.018*** (0.003)	0.020*** (0.002)	0.030* (0.018)	0.040** (0.019)
1940 Log income						0.0098 (0.0124)
1940 Homeownership						0.0095 (0.023)
1940 Educational Attainment						-0.007*** (0.002)
Constant	0.088*** (0.007)	0.095*** (0.008)	0.167*** (0.004)	0.083*** (0.003)	0.089*** (0.008)	0.0095 (0.023)
Birth year fixed effects	Y	Y	Y	Y	N	N
Observations	168,601	27,598	72,901	68,102	6,660	6,660
R-squared	0.026	0.028	0.012	0.010	0.0068	0.0018

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is equal to 1 if the child died as an infant and 0 otherwise. The coefficients can be interpreted as a p.p. impact on infant mortality. The sample includes African-American births and infant deaths in Illinois and Ohio to fathers born anywhere. Col 1 pools observations from all years, whereas columns 2-6 are for individual census periods: 1920 (1915-1920), 1930 (1925-1930), 1940 (1935-1940). The last column uses a matched sample to the 1940 census to include father's income, homeownership status, and educational attainment as a control. This matched sample uses data solely from 1939 and 1940. Col 5 limits to the 1940 for sample to observations that can be matched to the census. Cols 5 and 6 weight observations by the inverse of the match rate to the census in the appropriate state and D_i cell.