THE IMPACT OF MIGRATION ON INFANT HEALTH: EVIDENCE FROM THE GREAT MIGRATION[1]

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 a number of key innovations. First, we construct infant mortality rates specific to migrants and also for a period (1915-1920) prior to the registration of births in many of the states. Second, the microdata allow us to control for the selection into migration and assess a number of potential mechanisms for the migrant health effect. 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 do not find any evidence of migrant selection. 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: Department of Economics, University of California, Davis and NBER (kaeriksson@ucdavis.edu). Niemesh: Department of Economics, Miami University (niemesgt@miamioh.edu). We are particularly grateful to Jeremy Atack, William J. Collins, Gordon Hansen, Adriana Lleras-Muney, Seth Sanders, Marianne Wanamaker, and Sven Wilson. We would also like to thank participants at the 2014 annual meetings of the Cliometric Society, Economic History Association, Social Science History Association, the 2015 meetings of the Population Association of America and All-CA Labor Group, and the 2016 ASSA Meetings. Seminar participants at California Polytechnic State University, Northwestern University, Vanderbilt University, Wake Forest University, UC-Berkeley, UC-Davis, UC-Santa Barbara, and the University of Stellenbosch provided many helpful comments. Josh Koller, Alex McCorkle, and Raleigh Pearson provided excellent research assistance. This research was conducted with the help of funding from the Economic History Association Arthur H. Cole Grant in Aid.

I. 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).[2] However, outcomes worsened along a number of non-labor market dimensions, namely, higher incarceration rates (Muller 2012; Eriksson 2015) 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.[3] 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.

[2] However, Eichenlaub, Tolnay, and Alexander (2010) find migrants had lower *occupational status* relative to those that stayed in their Southern communities.
[3] 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.

In this paper, 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 part of a new research program motivated by newly available historical individual-level death certificates.[4]

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.[5]

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.

---

[4] 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.

[5] 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.

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 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.[6] 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

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

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 percent of Blacks born in the South resided in the North, the 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.

To measure infant mortality in this time period, we collect individual-level death certificates from 1916-1940 in both Southern and Northern states.[7] We restrict our attention to infant deaths in the five years prior to a census date. To this dataset we append the full universe of infants in the relevant

[7] Currently, the sample includes Tennessee, North Carolina, South Carolina, Illinois, and Ohio.

census who were born to parents born in the South but who are living in either location. Our full dataset contains 19,091 infants, of which 1,452 are infant deaths.

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 7.2 percentage points *more* likely to die in the first year of life in the 1915-1920 period. Northern mortality rates fall rapidly relative to the south, and by 1940 the southern mortality advantage has disappeared. We find no evidence of positive selection into migration based on innate infant health capital.

II. Trends in Infant Mortality and the North-South Gap

African Americans faced radically different health environments in the North and the South. 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.

However, infant mortality rates in the published *VSUS* are systematically biased downward in ways that can dramatically change inferences about the causes and consequences of infant mortality. For the first-half of the 20th century, completeness of the birth registration systems varied by state, race, urban status, and education of the parents. Shapiro (1950) describes the results from the first national study to determine the proportion of births registered with state vital statistics offices. Enumerators for the 1940 Decennial Census of Population were required to fill out special infant cards in conjunction with their typical enumeration duties with the census.[8] These infant cards were then compared to the official birth registration certificates in the state vital statistics offices.

The geographic variation in birth registration completeness (for whites and nonwhite combined) can be seen in figure 3. Registrations were more complete in northern states (96.9 percent in Illinois and 95.2 percent in Ohio) compared to the southern states (86.1 percent in North Carolina, 80.4 percent in Tennessee, and 77.5 percent in South Carolina). Black births were even more likely to go unregistered (82.5 percent for nonwhites and 94.0 percent for whites). However, the difference across regions remains when looking solely at black births. The northern cities had close to complete registrations, even for nonwhites (98 percent). Births in hospitals were much more likely to be registered and blacks in northern cities delivered in hospitals. Black mothers in the South, however, were much more likely to deliver outside of a hospital, especially in rural farm areas and smaller cities where getting to a hospital was more difficult. The test suggests that black births were systematically under-registered at higher rates in the southern states relative to the northern states, biasing upward southern mortality rates. Because we want to compare mortality trends across the regions, we take this as our prime motivation for revising mortality estimates from 1920-1940.

A. Revising infant mortality estimates using microdata

State vital statistic offices, in partnership with online genealogy websites, recently published online their collections of death certificate images and indexes. At the same time, digitized versions of complete Decennial Censuses are newly available to researchers from a number of sources.[9] We combine these two sources of data to revise infant mortality estimates for three northern states (Illinois,

[8] The 1940 enumerators were instructed to have a heightened "awareness" to record infants. Thus, we might expect the age-specific component of the census undercount to be less than in prior years (Shapiro 1950).
[9] Census indices and death certificate collections were kindly provided by FamilySearch.org.

Pennsylvania, and Ohio) and three southern states (Tennessee, North Carolina, and South Carolina) for the following years: 1915-1920, 1925-1930, and 1935-1940.[10]

Our procedure amounts to adjusting birth counts in the South to account for under-registration. As an example, we can estimate the cohort infant mortality rate for year t in state s using the following formula:

$$(1) \qquad IMR_{s,t} = \frac{D_{s,b=t,d<365\ days}}{D_{s,b=t,d<365\ days} + D_{s,b=t,d=non-infant} + Census_{s,b=t}}$$

where $D_{b=t,d<365\ days}$ is the number of infant deaths (less than 365 days) of those born in year t in state s, $D_{s,b=t,d=non-infant}$, similary defined, but for non-infant deaths between 1 year and 5 years old, and $Census_{s,b=t}$ is the number of live children enumerated in the decennial census born in year t. In practice, we construct five-year averages for the end of each decade to reduce the year-to-year variability. Infant deaths at the state level (by race and urban status) are drawn from the published *VSUS*, as the aggregate statistics are constructed from the underlying death certificates. Non-infant deaths come from the death indixes.

The complete census indexes provide a method to address a number of potential biases from the migration of young children out of their state of birth. Importantly, $Census_{s,b=t}$ includes living children born in state s regardless of residence. The potential migration of children outside of their state of birth do not bias our estimates of births downward, as long as they remain alive to age five. In the absence of a complete death index for the entire country, we do not observe children who died after migration to a state not included in our sample. We adjust the birth estimates for this potential bias by using the proportion of children living outside their state of birth from the census by applying the state of birth mortality rate to the expected proportion of those that migrated. Finally, we include these deaths in the denominator only, and not in the numerator as infant deaths. Most infant deaths occur in the first 30 days

[10] Data availability limits the number of states included in the study. The matching procedure conducted later to account for selection into migration requires individual death certificate data, for which is available for these six states only.

of life, or even the first day[12]. It seems plausible that families did not travel out of state with unhealthy young infants. In any case, this adjustment does not make a discernible difference to the estimated rates.

Two potential biases remain as concerns: under-enumeration in the census, and under-registration of death. It is well known that the decennial censuses of the early 20th century undercounted young children (Greville 1947). Our estimates of the number of births will be biased downward relative to the true value to the extant under enumeration occurred. Whether our adjustments are an improvement over the published aggregate mortality rates depends on the amount of under-registration of births relative to the amount of under-enumeration in the census. Under-registration was severe in the southern states, and this is exactly where we see a large change in mortality rates from our adjustment, on the order of a 20 percent increase in the number of births in the case of South Carolina. Whereas registration was nearly complete in the northern states, the adjusted births slightly undershoot the registered births. The results confirm our expectations. Our adjustment provides improved estimates of infant mortality where under-registration was more severe than under-enumeration. In practice, we use adjusted births for southern states and registered births in northern states to calculate mortality rates.

Bias from under-registration of deaths is much more difficult to deal with. When a parent decides to not register a death (and likely the birth), no record of the event exists. No direct means to assess the size of under-registration exists (Greville 1947). We believe the extent of death under-registration is less than that of births. Parents had an added incentive to register a death, whereas they did not for a birth. Burial with the family in a cemetery or churchyard required a death certificate. In the absence of a direct assessment of the issue, our adjusted mortality estimates provide an upper bound and the published rates provide a lower bound of the North-South difference in infant mortality in the presence of death under-registration. Moreover, we later provide evidence that our adjusted rates are consistent with mortality rates from sources that would not suffer from under-registration of deaths.

[12] Neonatal deaths make up 50 percent of all deaths in the entire sample. Deaths within 48 hours of birth make up 25 percent of all deaths in the entire sample.

The revised estimates for 5-year average infant mortality rates by state of birth are shown in panel A of table 1 and in figure 4. First, northern black infant mortality was 62 percent (7.0 p.p.) higher than the northern rate during the early part of the 20th century (18.3 p.p. vs 11.3 p.p.). Significant variation exists across states within regions: a 27 percent difference between South and North Carolina (12.6 vs. 9.9 p.p.), and a 15 percent difference between Pennsylvania and Ohio (18.3 p.p. vs. 15.9 p.p.).

The southern mortality advantage remained during the late 1920s, although it was smaller in both absolute and relative terms. Northern rates fell more rapidly during the 1920s, reducing the 64 percent regional gap to only 24 percent by the end of the decade. The absolute decline in black infant mortality in the North continued the rapid decline during the 1930s that it experienced in the 1920s - 35 percent and 32 percent respectively. Southern rates declined by similar small amounts during both periods - 12 percent during the 1920s and 19 percent during the 1930s. By 1940, northern rates converged with and fell below southern rates, completely removing the southern mortality advantage.

We want to emphasize that our microdata method allows us to construct race specific mortality rates for these states in the 1915-1920 period, and 1925-1930 in South Carolina, whereas previously none existed in the *VSUS*. Moreover, it leads to a different set of conclusions about the pace and timing of regional convergence, and cross-region comparisons. Figure 4 plots the revised estimates relative to those found in the *VSUS*.[13] Similar to the previously known published mortality rates, we capture the secular downward trend within regions and the convergence across regions. However, the revised estimates suggest slower declines in the South relative to those in the *VSUS*. The adjustment procedure reduces Southern rates relatively more in the earlier period compared to later decades, reducing the rate of observed declines. Again, black infant mortality was initially larger in the North during the late 1910s, but declined more rapidly than in the South. The adjustment postpones full convergence from the late 1920s to the mid- to late-1930s.

B. Migrant-specific infant mortality rates

[13] Table A1 in the appendix reports the full comparison in more detail.

Migrant-specific mortality rates are the most relevant to our purpose: estimating the difference in infant death between migrant and non-migrant southern-born fathers. Migrant fathers may be systematically different from non-migrant fathers in ways that reflect the underlying causes of infant mortality, such as socioeconomic status. In other words, northern-born infants of southern-born fathers might fact different risks than infants of northern-born fathers. One innovation of the paper is to use the combined death certificate and census index microdata to construct infant mortality rates specific to migrant fathers living in the North using the father's reported place of birth.[15] Similarly, we construct rates for southern-born infants specific to non-migrant father's place of birth (i.e. the sample of Tennessee born infants is restricted to having Tennessee born fathers).

In general, we follow the same procedure outlined above for constructing the state-based revised rates. We scale the total number of infant deaths by the proportion of the sample that reports each birthplace of the father (NC, SC, and TN).[17] We scale the number of births, whether from the published *VSUS* or revised, by the proportion of NC, SC, and TN fathers in the census index and death records.

The migrant-specific infant mortality rates are our preferred estimates for the unconditional difference in likelihood of death for an infant born in the South versus the North. Rates are reported in panel B of table 1, along with the difference to the revised state-based rates from panel A. We find that migrant-specific rates are noticeably higher in the North and lower in the South. Migrant-specific infant mortality rates range from 2.6 to 12 percent *more* than state-average rates in the North, whereas they are 10 to 15 percent *less* than the state-average rates in the South. Importantly for our purposes, the migrant-specific rates widen the North-South gap in infant mortality relative to the state-based revised rates.

C. Migration and infant mortality

[15] We limit the sample based on father's place of birth because in the following section we match fathers back to their childhood homes based on name, year of birth, and place of birth. Mothers cannot be matched to pre-migration childhood homes because of name changes at marriage. In theory, one could construct mortality rates for any specific group for which both the death certificates and census index contain identifying information.

[17] Father's birthplace is missing for some death certificates. Indexers chose not to digitize father's birthplace seemingly at random. However, any systematic omission of father's birthplace will lead to a potential bias in the migrant-specific estimates (e.g. if migrant deaths cluster by date of death and place of residence on a microfilm roll).

The choice of mortality rate influences the potential mortality impacts of South-North migration using the unconditional mean death rate. Table 2 reports the north-south difference in infant mortality rates for each period and each estimation method. We interpret these estimates as the naive unconditional treatment effect of migration North on infant mortality. In summary, the vital statistic estimates of the north-south mortality gap drastically underestimates the true southern mortality advantage because of the systematic underreporting of black births in the south. Our preferred estimates, which we use in the remainder of the paper, are the migrant-specific rates that isolate the sample for which we are most interested: births to fathers born in the three southern states.[18]

III. Contribution of migration to infant health gains and accounting for migrant selection

The unconditional comparison of infant mortality rates clearly show a southern health advantage that declines over time and is eventually surpassed by North by at least the late 1930s.[19] Given the high economic returns to migrating North, we move towards evaluating whether the potential positive selection of migrants biases the unconditional comparison of the previous section. 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. 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.

[18] One remaining concern with our estimates is under-registration of *deaths*. We are working to address this but note that (a) under-registration of deaths was no more common than under-registration of births so an upper bound on this problem would lead us back to the VSUS rates; (b) using the 1910 census and the question about children ever born and children surviving, we see a similar gap between North and South and Urban and Rural areas within the South. This is not perfect because it is a measure of child mortality, not infant mortality, but the two are highly correlated.

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

A. *Census Matching Procedure*

Our goal is to create a sample of live and dead children in each outcome census year, with their fathers matched to a pre-migration census. We do this with two separate matches. First, we match parents of deceased infants from the death certificates to the relevant census manuscripts. Then, we pool the fathers of these infants with the full set of fathers of live infants. Finally, we match this full set of fathers back to a previous census wave; this allows us to control for selection into migration based on childhood characteristics of the fathers. Both matches are described in more detail below. We use the outcome census years of 1920, 1930, and 1940 and are interested in children who were born in the previous five years and who either survived or did not survive the first year of life.

Each death certificate includes both parents' names and states of birth. We use this information to conduct the first match to the census. We restrict to deaths within the five years before the census year and look for both parents living in any state in the census year. For example, even though a household might have had an infant who died in Illinois in 1935, we will still try to find the parents if they move to Indiana or back to the South by 1940. To maximize the number of parents who are matched to the census, we use a procedure similar to Feigenbaum (2015).[20] It proceeds as follows:

1. Calculate Jaro-Winkler string distance scores between all individuals born in the correct state; restrict to possible father matches with a Jaro-Winkler score of 0.8.
2. Calculate a Jaro-Winkler score for all spouses of the possible father matches. Drop observations who are married to a woman with a score of less than 0.8 or with an incorrect birth state.
3. If a household contains an exact match on name for both parents and this household is unique, we consider the death matched to this household.
4. If there is no exact match, but there is one unique match with both parents' scores greater than 0.8, we consider this the match.

[20] Unfortunately, father's age is not available in the death certificates, which decreases our match rates due to multiple potential matches.

5. If there is no candidate for the mother, we match to a father who is not living with his spouse but who is married and has the correct name and birthplace.[21]
6. In the case of multiple matches, we prioritize the observation where the parents are living in the state in which the child died.
7. If there are multiple candidates in the previous steps, we are unable to match the observation.

Our procedure results in a match rate of 39%, with this consistent across states and census years.

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. We match the father of each infant backwards to a previous census wave either 10 or 20 years earlier. The goal is to find the father as a child or young adult, still living in his childhood household, so men who are older than 26 are matched over a 20 year horizon while men younger than 26 are matched over 10 years. When we have identified the individual in a previous census index, we digitize by hand the characteristics of the childhood household that are not available in the indexes: occupation of the father, home ownership, and literacy of parents.

To match individuals across censuses, 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 (Atack 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;

[21] We could do something similar with potential mothers who are not living with their spouse; however, the goal of the exercise is to identify households in which we can match the *father* back to a pre-migration census so we do not look for mothers living alone.

(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 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 1,452 deceased infants and 17,673 non-deceased infants with fathers born in Tennessee, South Carolina or North Carolina but who were born in these states, Illinois, Pennsylvania, or Ohio in the five years previous to the outcome year. We can successfully match 24 percent of deceased infants and 28 percent of non-deceased infants. This results in an overall match rate of 10 percent over the two matches for the deceased infant observations.

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 in the raw data. Therefore, we create a sub-sample which produces the correct mortality weights. We do this by sampling from the births so that the mortality rate in the sample matches the rates presented in Table 1 in each year/state cell.

B. Evidence of and Accounting for Selection

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

(2) $$D_{i,t} = \alpha + \beta_t * M_{i,t} + X_i\theta + u_i,$$

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) interacted with time-period indicators. The coefficients of interest are the β_t's, which measure the impact of migration on infant mortality in percentage points for each period separately. All regressions include census year indicators.

Results are reported in table 3, with base results in column 1. Each observation is weighted to recover the state-level migrant specific mortality rate for the appropriate decade from table 1. Unsurprisingly, the same regional differences and trends are observed as in the table 1. A large southern mortality advantage exists during the 1915-1919 period, which gradually disappears as northern mortality rates converge with those in the South. The base estimate is used as a comparison across specifications as we add in individual and local level controls.

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_t. We take a number of steps to account for the possibility of this selection bias. In the next three columns of table 3, we explore the possibility of selection on observables and unobservables using our matched sample.

Migrants differentially came from households with higher levels of income, wealth, or education as is shown in table 4.[22] 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 southern-born fathers. [23] Controlling for occupational status, literacy, and homeownership does not significantly change the coefficient estimate from the base specification, suggesting that selection on observables makes up a small part of the observed differences in infant mortality between migrants and non-migrants.

Column (3) 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 to be slightly larger, but the change represents only a small portion of the raw migration effect. The county-of-origin fixed effects pick up some of the rural-urban migration patterns within and across regions that

---

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

[23] Controls include a set of indicators for occupational status (Owner operator farmer, tenant farmer, farm laborer, laborer, and an all other category). We define literacy as able to both read and write. We use the head's information from the household in which the father is found as a child in a prior census.

are correlated with the urban mortality penalty. Column (4) combines both the pre-migration controls and county-of-origin fixed effects. Column (5) includes the pre-migration controls interacted with an outcome year indicator to allow for differential effects across years. We do not find strong evidence of selection bias based on the results from our matched-sample. The most stringent specification, column (5), finds similar trends within and across region as the population level estimates of migrant-specific mortality rates from table 1.

IV. Mechanisms and Pathways

In this section, we explore the large number of potential pathways and mechanisms through which the migration impact might have worked. 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, 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 a discussion of the impact of socioeconomic variables and conclude with environmental factors.

A. Socioeconomic and individual characteristics as potential mechanisms

Fertility

Migrant mothers' fertility decisions differ from non-migrant in a way that potentially changes the distribution of infant health. First, we compare the period total fertility rates of southern-born migrant and non-migrant women. The period total fertility rate is calculated by adding across one-year age groups the proportion of women reporting an infant in the household. Table 5 makes this comparison for each census year. Non-migrant southern-born women are more likely than migrant women to have a child in each of the census years: 1.8 times in 1920, 1.5 times in 1930, and 1.46 times in 1940.[24] Figure 5 plots the age-specific fertility rates for 1940 by migrant status, and suggests migrant women are less likely than non-migrants to have a child at every age.[25] Table 7 reports results from cross-sectional regressions of fertility and marital status on regional-migrant status allowing us to control for differences in the age distribution across regions. Results are consistent with migrant women being 11 p.p. less likely to be to have any children in the househould (mean of 78.6 percent), having 1.3 less children in completed fertility (mean of 3.8), and 0.4 to 1.1 years older at the time of their first marriage. Results for the likelihood of reporting ever married are not consistent across census years. Migrants are 1.6 p.p. less likely in 1940, 0.96 p.p more likely in 1930, and the 1920 result is noisy. In sum, all the results point toward a substantial difference in fertility and marriage behavior between the movers and stayers.

The datasets available are not ideal. First, we do not know at what age black females migrated north, or how much return migration occurred. Second, selection into the migration stream might bias any of these estimates away from the true treatment effect. Ideally, we could address both issues with a matched sample for mothers similar to our matched sample for fathers. Unfortunately, matching married women to childhood homes in a previous census is impossible. The majority of women during this period adopt their spouse's surname at the time of marriage. We show the fertility results, nonetheless, as the correlations are informative for potential pathways of the migration effect on infant mortality. However, the direction of the total effect is ambiguous. Later marriage implies older mothers at the time of birth, which would be expected to *increase* mortality in the North. In contrast, smaller family size reduces parity and would be expected to *decrease* mortality in the North. In any case, regressions that control for

[24] The data for this comparison is not ideal as infant deaths are not included in the fertility rate. However, the relatively larger infant mortality rates in the North are not large enough to make up the gap with southern fertility rates.
[25] We choose to report the figure for 1940 because regional differences in black infant mortality had largely disappeared by this time. Figures for 1930 and 1920 show a similar relationship.

age of mother and number of siblings at the time of birth do not reduce the North-South difference in black infant mortality (results unreported).

B.) Environment and Urban Penalty

By further slicing the data into a number of subsamples information on the causes and mechanisms of the negative migrant health impact can potentially be gained. First, we explore whether the urban-rural distribution of births drives the large North-South difference in black infant mortality. Northern blacks lived primarily in urban areas (93 percent urban), whereas Southern blacks were more rural (21 percent urban). Moreover, a large urban penalty for infants, especially black infants, remained well into the 20th century.

Column (2) of table 7 includes indicators for urban status for the non-migrants that remained in the South. Coefficients should be interpreted relative to the black mortality rate in the rural South for that given year. In 1920, migration to the North was slightly "worse" than residing in a southern city. Children born to northern migrants in 1920 faced an 11 percentage point (118 percent) higher mortality rate than southern rural infants. Those born in the urban south faced a 6.1 percentage point (66 percent) higher mortality rate. In 1930, the north-rural south difference was cut in half to 4.5 p.p. (58 percent of the rural rate), and the southern urban penalty also declined by half to 3 p.p. (37 percent of the rural rate). Northern rates continued to converge with southern rural rates during the 1930s, and at a faster pace than the urban South. By 1940, rates in the urban North and urban South were statistically indistinguishable. But, the rural South maintained its mortality advantage.

Black infant health in the North improved dramatically over the course of the early 20th century, whereas improvements in the urban South came slowly. To further explore the different trends across regions and urban status, we split the full sample into populations that are more or less susceptible to certain pathogens and causes of infant death.

Post-neonatal deaths are typically caused by infections diseases, such as diarrhea and pneumonia, contracted after birth and may be indicative of the infectious disease environment. Neonatal deaths are more commonly caused by non-communicable factors: preterm birth, asphyxia, and congenital defects. In the final columns (3) and (4) of table 7, we estimate equation (2) with an indicator for neonatal death and an indicator for post-neonatal death as the dependent variable. The results for post-neonatal death are conditional on survival to at least 30 days. Both neonatal and post-neonatal rates were initially higher in the North, and converged with the South. However, neonatal mortality made up an increasingly large proportion of the total North-South difference in black infant mortality: 55 percent in 1930, 62 percent in 1930, and 111 percent in 1940. Post-neonatal rates converged faster than neonatal rates. Overall, declines in both rates contributed to the disappearance of the Southern mortality advantage. Males fetuses are more susceptible to health insults than are female fetuses. While the North-South difference in male infant mortality is relatively larger than for females, rates for both sexes converged rapidly with southern rates. Taken together, we interpret the results from table 3 and table 7 suggesting that the initial southern mortality advantage was primarily driven by an urban penalty for black infants. The balanced decline in rates across all subsamples suggest a broad improvement in black infant health.

Multiple factors could explain the "urban mortality penalty" in this time period. 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 areas 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). By our period much of the build out of water and sewage systems had already been completed, with no remaining discernible impact on adult mortality rates in cities.

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.

VI. The contribution of the Great Migration to black-white infant mortality convergence

The motivating question behind this paper is what would black infant mortality have looked like if the Great Migration had never happened. One would have initially expected that the migration was a positive factor for black infant mortality, but as we have seen, for individuals born in the North, the probability of dying in the first year was substantially than in the South. In this section, we estimate the contribution of the Great Migration to black infant mortality overall by constructing a counterfactual estimate of overall black infant mortality if all children were born in their father's state of birth.

We start with 1920 where migration increased infant mortality. We start by constructing state-specific black infant mortality rates for each state. Then we use the 1920 Census index to extract all black children 5 years of age or younger. The actual overall black infant mortality rate is a weighted average of state-level rates with the weights the proportion of black births in that state. We re-weight this overall rate by placing the births *back* in the state of birth of the father. For example, if 10% of black births nationally were to fathers born in Alabama, but half of these babies were born in Illinois, we would put all 10% back in Alabama and assign them the Alabama infant mortality rate in this counterfactual experiment. This gives us two numbers in 1920: the actual black infant mortality rate was 10.7 percent and the counterfactual rate was 10.4 percent. Therefore, we can argue that in the absence of the Great Migration, black infant mortality rates would have been 0.3 percentage points lower in 1920 than they actually were.

We repeat this calculation in 1930 and find no gap between the counterfactual and actual rates. Repeating the same in 1940, the difference is -0.16 percentage points (7.38pp vs 7.54pp).

Relative to the black-white infant mortality gap in these years, these numbers are significant. In 1920, black infant mortality rates were 5 percentage points higher than those of whites. This gap would have been 4.7 percentage points in the absence of the Great Migration of blacks, a 6 percent difference. Here, the counterfactual is based on what if *blacks* hadn't moved. By 1940, the gap was 3.1 percentage points and would have been 3.26 percentage points in the absence of migration, a 5.3 percent difference.

Finally, we move forward to the second half of the Great Migration which concluded around 1970. Using NCHS Vital Statistics Birth Data for 1970, we follow the same procedure to assign infants back to their mother's birth state (father's birth place is not available in these data). We find that mortality rates would have been 0.11 pp *higher* (8.2 percent of the 1.31pp gap between black and white infants) in the absence of these mothers' migration to states outside of the South. We do note that this is not necessarily a clean estimate of the impact of the Great Migration since a lot of these mothers are second-generation migrants, but it gives us an idea of North-South differences in this time period.

Overall, the Great Migration appears to have had a mixed impact on black infant mortality. In the first part of the migration, black infant mortality increased since cities in the North had more negative health environments, but by 1930 this gap disappeared and switched directions by 1940. Noteably, however, the Great Migration was not large enough to substantially affect the large pattern of black-white infant mortality convergence.

VII. Conclusion

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 that there was selection into migration based on health status. Importantly, even 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.

We find the largest effect of migration in 1920, but by 1940 the North and South had converged. Mortality rates in the North were 11 percentage points higher in 1920 and this gap fell to 4.5 by 1930 and zero in 1940. Heightened risk in the North of both neonatal and post-neonatal explain the total regional difference. A broad-based regional convergence in rate underscores the fact that a single explaination Even within the South, large differences in mortality rates exist between urban and rural residents. In fact, this difference is almost as large as the differences between North and South, meaning that the negative effects of the Great Migration on mortality were likely primarily due to residing in a city.

Overall, the first part of the Great Migration contributed negatively to black-white convergence—in the absence of this large migration, black infant mortality rates would have fallen *faster* than they in fact did. Our calculations in 1970 suggest that the second wave of the Great Migration was one contributing factor to black-white infant gap convergence, although it did not explain a large part of the black-white infant mortality gap.

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.

Eriksson, Katherine, 2016. “Moving North and Into Jail? The Great Migration and Black Incarceration”. Manuscript.

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. Rosenwaike, 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.

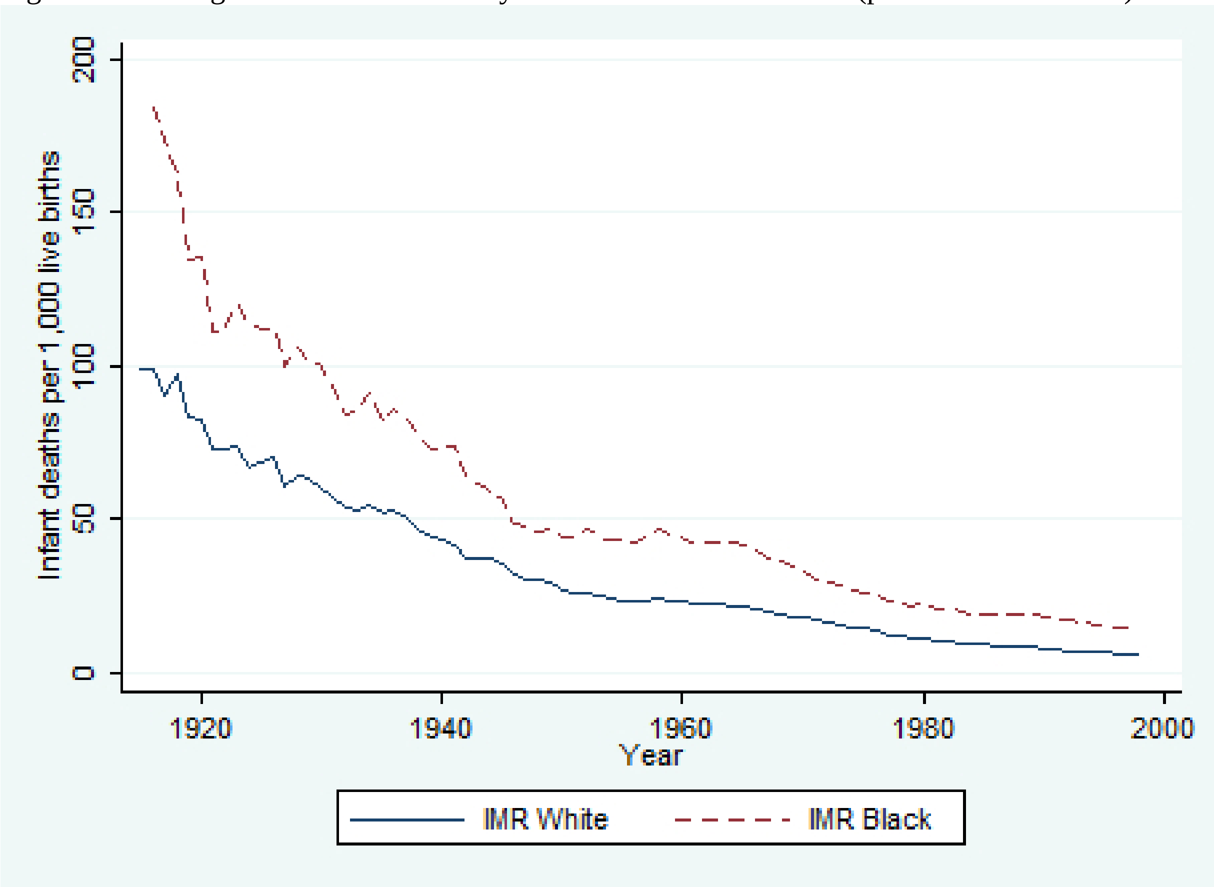
Weiss, Leonard and Jeffrey G. Williamson, 1972. “Black Education, Earnings and Inter-regional Migration: Some New Evidence,” *The American Economic Review*, vol. 62(3), pages 372-383, June.

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).

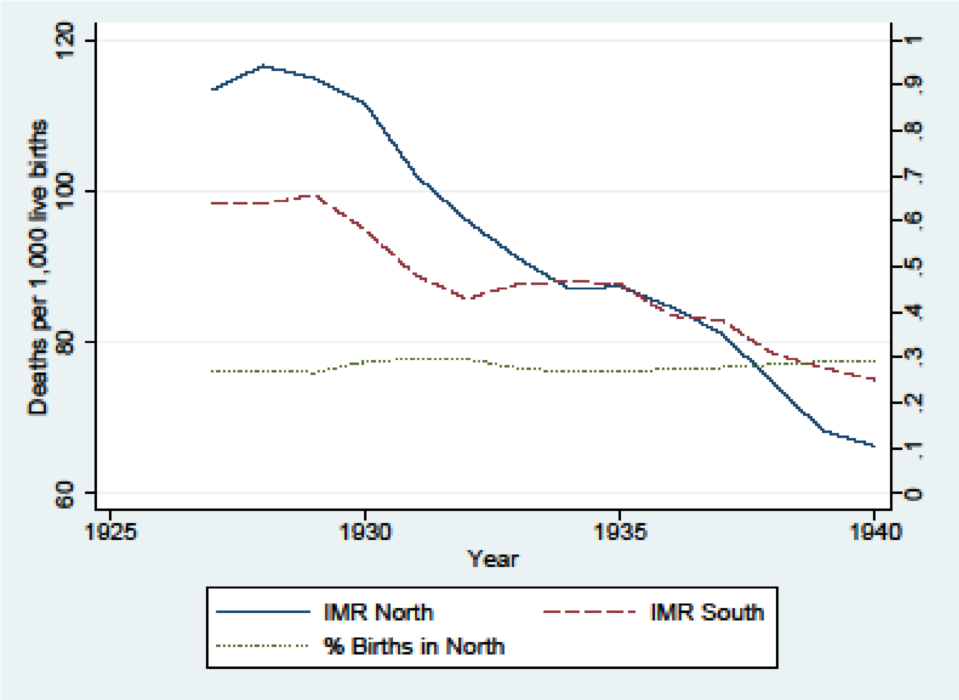
Woodbury, R. M. 1925. *Causal Factors in Infant Mortality*. U.S. Children’s Bureau, No 142. Washington, DC.

Figure 1: Convergence of infant mortality rates for blacks and whites (published *VSUS* data)



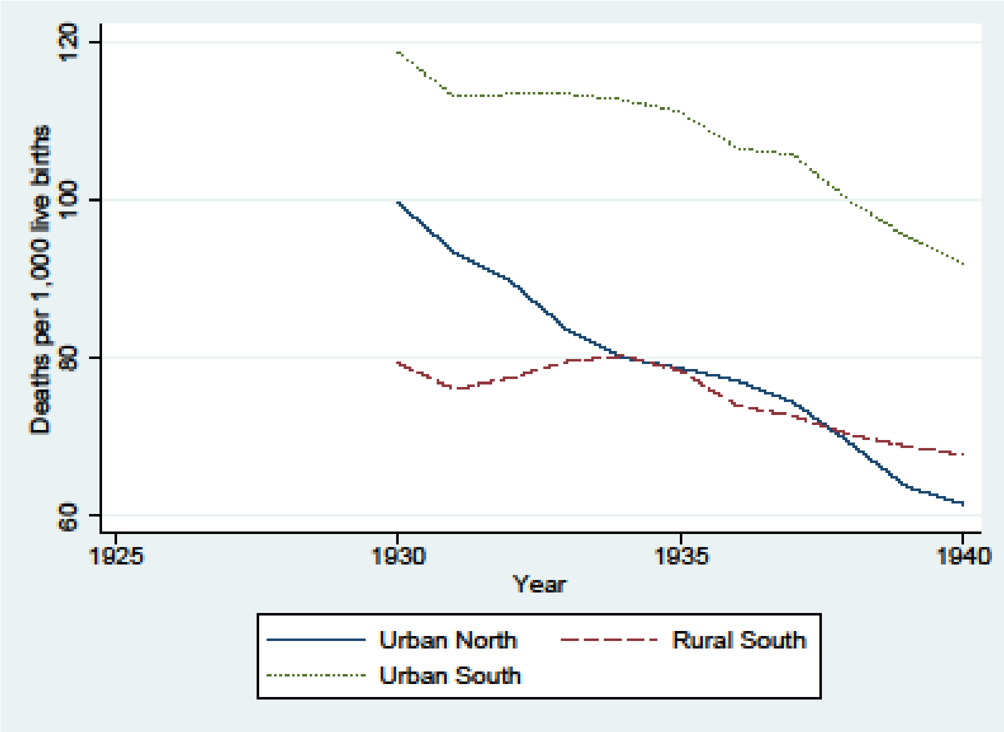
Notes: For 1915-1932, data are for the current Birth Registration Area only. Source: Haines (2006).

Figure 2a: North-South gap in published black infant mortality rates from *VSUS*



Notes: Includes all states reporting infant mortality rates in the published *VSUS*.
Source: Published volumes of aggregate statistics from the *Vital Statistics of the United States* 1927-1940.

Figure 2b: Infant mortality rates for urban status and regions



Notes: Includes all states reporting infant mortality rates in the published *VSUS*. Rural north is excluded as the black population made up a negligible portion of the population.
Source: Published volumes of aggregate statistics from the *Vital Statistics of the United States* 1927-1940.

Figure 3: Undercounting of births in the South: completeness of the birth registration area (white and nonwhite births)

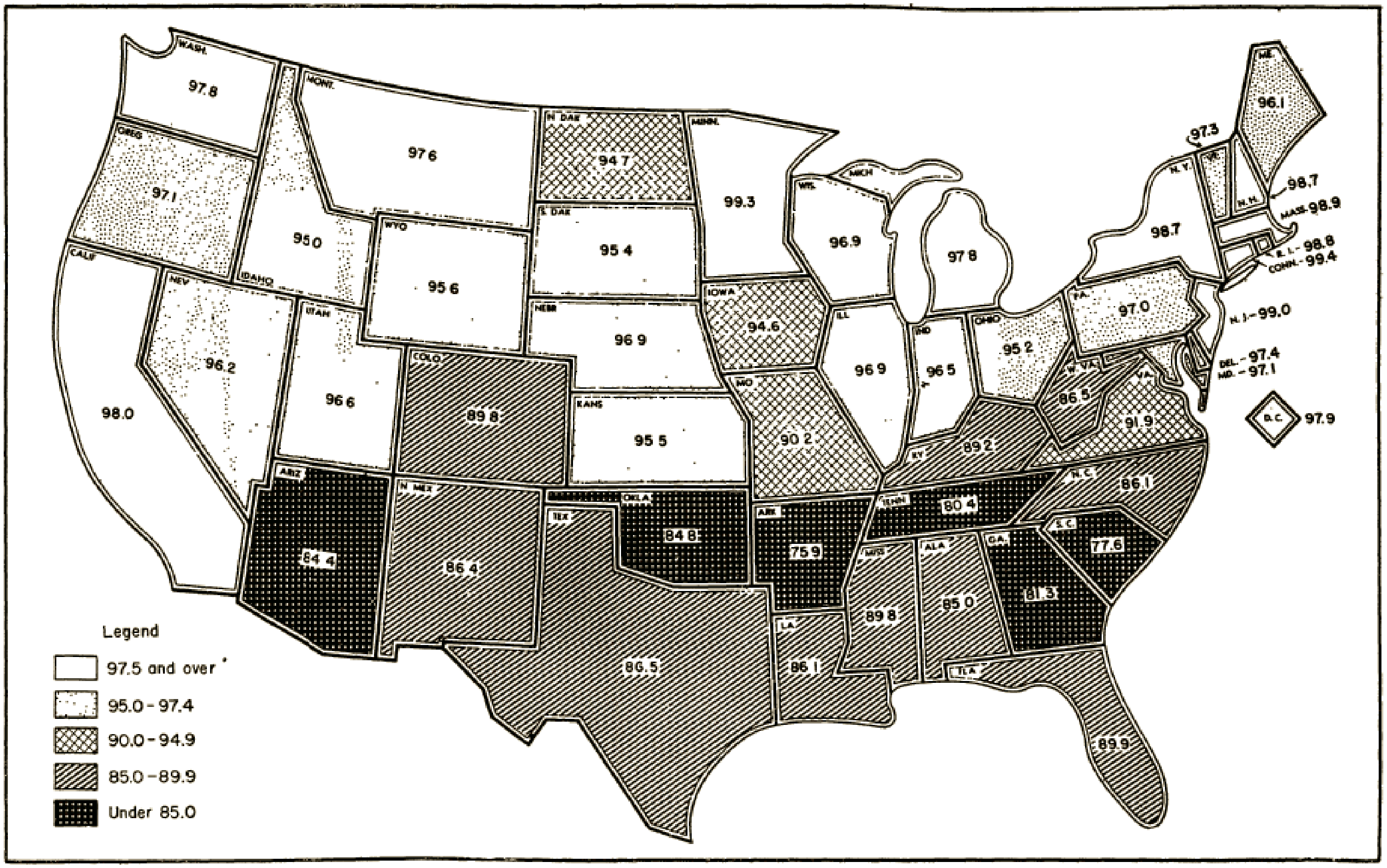
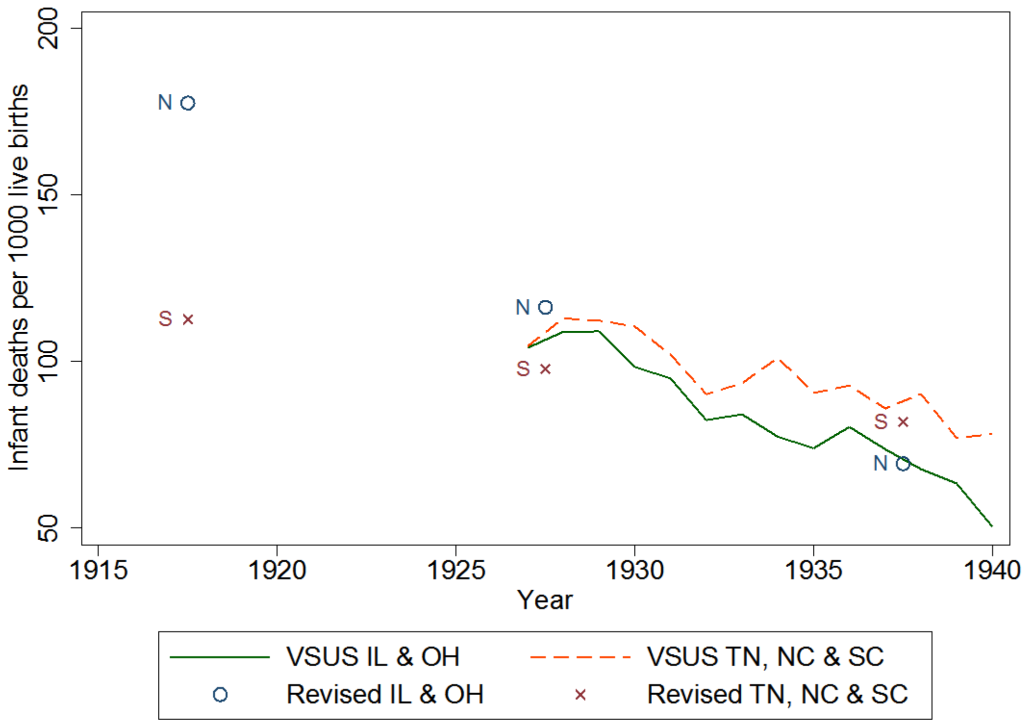


Fig. 2. Percent completeness of birth registration: each State, 1 December 1939 to 31 March 1940.

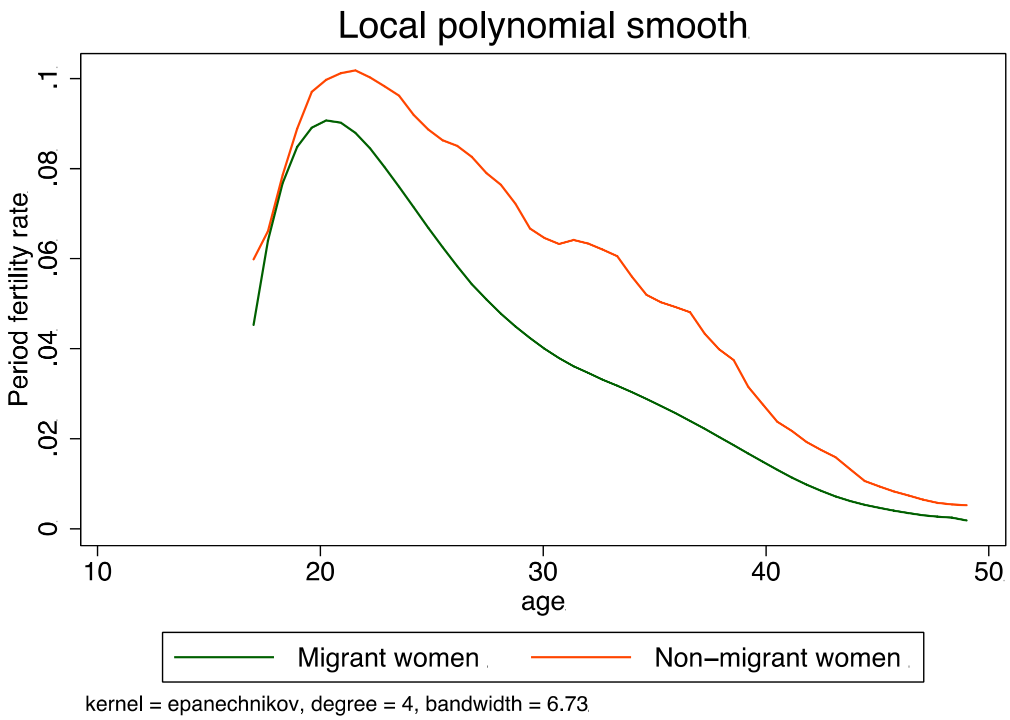
Source: Shapiro (1950). The U.S. Census Bureau conducted an independent check on the completeness of the birth registration system in early 1940 by comparing special infant cards from the March 1940 Decennial Census and official birth certificates in state vital statistics offices.

Figure 4: Revised infant mortality rates compared to published rates from *VSUS*



Notes: "N" denotes the northern states in our sample, whereas "S" denotes the southern states in our sample. Revised rates are 5-year averages calculated from the full census index and death certificate microdata as outlined in section 2. Markers are placed at the midpoint of the 5-year period to simplify the comparison to the published rates. The dashed and solid lines show the annual infant mortality rates from the published *Vital Statistics of the United States* 1927-1940.

Figure 5: Migrant and non-migrant period fertility by age of southern-born African-American women in 1940.



Notes: Units are percentage points. The sample includes southern-born black women aged 16 and over living in the South, Northeast, and Midwest census regions from the IPUMS 1940 census full count. A fourth-degree polynomial smoothing is applied over age groups for an indicator if an infant is reported as present in the household. Figures for 1930 and 1920 show a similar relationship between migrant and non-migrant women.

Table 1: Revised state and new migrant specific black infant mortality rates

Panel A: State of birth infant mortality rate (p.p.) - 5-year average

	1915-1919	1925-1929	1935-1939
Northern states	18.3	12.3	7.8
Southern states	11.3	9.9	8.0
Illinois	16.2	10.5	6.3
Ohio	15.9	11.6	7.2
Pennsylvania	18.3	11.8	7.8
North Carolina	9.9	9.9	8.3
South Carolina	12.6	9.8	8.0
Tennessee	10.5	9.5	6.9

Panel B: Migrant specific infant mortality rate (p.p.) - 5-year average

	1915-1919		1925-1929		1935-1939	
	Rate	Δ total	Rate	Δ total	Rate	Δ total
Northern states	20.5	2.2	13.4	1.1	8.0	0.2
Southern states	9.9	-1.4	8.5	-1.5.	7.2	-0.8
Illinois	14.4	-1.8	10.9	0.4	6.5	0.2
Ohio	16.5	0.6	10.8	-0.8	7.0	-0.2
Pennsylvania	25.1	6.8	15.5	3.7	9.0	1.2
North Carolina	8.8	-1.1	8.1	-1.8	7.7	-0.6
South Carolina	11.4	-1.2	9.1	-0.7	7.1	-0.9
Tennessee	8.8	-1.7	7.6	-1.9	6.0	-0.9

Notes: All entries in the table are measured in terms of p.p. (alternatively deaths per 100 live births). Rates are averaged over five years of data. Panel A reports revised infant mortality rates for all black births in each state for births in each five-year range. Panel B limits births and deaths to specific child state-of-birth/father state-of-birth pairs. For children born in Illinois, Ohio, and Pennsylvania, fathers can be born in Tennessee, North Carolina, or South Carolina. This is our migrant father sample. Children born in each of the southern states are required to have fathers born in the same state. This is our non-migrant sample. The total change column is the difference between from the state based rates in panel A and the migrant specific rates in panel B. See section 2 in the main text or the data appendix for a description of how the revised rates are constructed. Regional averages are weighted means using the revised counts of black births as weights.
Sources: *Vital Statistics of the United States* (1915-1940), indices of the 1920-1940 Decennial Census of Population microdata and collected death certificate indices provided by FamilySearch.org.

Table 2: Unconditional effect of migration on infant mortality by estimation method (north - south in p.p. and as percent of southern rate)

	1915-1919		1925-1929		1935-1939	
	p.p.	*%*	*p.p.*	*%*	*p.p.*	*%*
VSUS	n.a.	n.a.	-1.8	-14	0.7	10
Revised state-based rates	7.0	62	2.4	24	-0.2	-3
Migrant-specific rates	10.6	107	4.9	58	0.8	11

Notes: Summary of results from table 1 and table A1.

Table 3: Evidence of selection bias for the treatment effect of migration on infant mortality

	Infant Death	Infant Death	Infant Death	Infant Death	Infant Death
Migrant*1920	0.1087***	0.1079***	0.1099***	0.1098***	0.1097***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Migrant*1930	0.0443***	0.0434***	0.0471***	0.0467***	0.0451***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Migrant*1940	0.0047	0.0041	0.0075	0.0077	0.0112
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
1930 Indicator	-0.0145*	-0.0152**	-0.0135*	-0.0143*	0.0020
	(0.007)	(0.007)	(0.007)	(0.007)	(0.023)
1940 Indicator	-0.0267***	-0.0286***	-0.0264***	-0.0283***	-0.0564***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.021)
Pre-migration controls					
Farmer (Owner)		-0.0141		-0.0137	
		(0.011)		(0.012)	
Farmer (Tenant)		-0.0028		-0.0004	
		(0.008)		(0.008)	
Farm Laborer		-0.0043		-0.0043	
		(0.012)		(0.012)	
Laborer		0.0120		0.0106	
		(0.010)		(0.010)	
Owns Home		0.0034		0.0046	
		(0.010)		(0.010)	
Head Literate		0.0046		0.0055	
		(0.005)		(0.005)	
Constant	0.0986***	0.0989***	0.0977***	0.0960***	0.1005***
	(0.006)	(0.009)	(0.006)	(0.009)	(0.018)
N	19,091	19,091	19,091	19,091	19,091
R-squared	0.010	0.010	0.026	0.027	0.028
County FE	N	N	Y	Y	Y
Pre-migration controls interacted with year indicator	N	N	N	N	Y

Notes: Robust standard errors in brackets. *** 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 sample includes African-American births and infant deaths in Illinois, Ohio, Pennsylvania, Tennessee, North Carolina, and South Carolina to fathers born in the three southern states. Matched observations are weighted to recover the observed mortality rate in each state. The sample includes observations in the outcome year with fathers that could be matched to a state-of-birth childhood home in a prior census. Pre-migration occupation indicators are relative to the omitted category of "all other occupations." A tenant farmer is an observation that reports occupation as farmer and rents a farm. A head of household is literate if reporting that they can both read and write (in 1900-1920) and if reporting that they are literate in 1930.

Table 4: Summary statistics of pre-migration characteristics in matched sample

	Non-migrants (N= 11,098)	Migrants (N= 7,993)	p-value of difference
Farmer (owner operator)	0.17	0.15	0.034
Tenant farmer	0.50	0.43	0.000
Farm laborer	0.07	0.07	0.126
Laborer	0.13	0.13	0.874
Other occupation	0.14	0.22	0.000
Owner-occupied housing	0.24	0.26	0.000
Literate head	0.53	0.57	0.000

Notes: The sample includes African-American births and infant deaths in Illinois, Ohio, Pennsylvania, Tennessee, North Carolina, and South Carolina to fathers born in the three southern states. Observations are weighted by the number of black births in each state and period within a region. The sample includes observations in the outcome year with fathers that could be matched to a state-of-birth childhood home in a prior census. A tenant farmer is an observation that reports occupation as farmer and rents a farm. A head of household is literate if reporting that they can both read and write (in 1900-1920) and if reporting that they are literate in 1930.

Table 5: Period total fertility rate of southern-born migrant and non-migrant women

	1920	1930	1940
Non-migrant	3.7	2.3	1.9
Migrant	1.7	1.5	1.3

Notes: The sample includes southern-born black women aged 16 and over living in the South, Northeast, and Midwest census regions from the IPUMS 1940 full count, 1930 5%, and 1920 1% samples. The period total fertility rate is calculated by adding across one-year age groups the proportion of women reporting an infant in the household.

Table 6: Fertility and marital outcomes of southern-born migrant and non-migrant African-American women.

	(1)	(2)	(3)	(4)
	Children ever born	P(> 0 kids)	Age at marriage	P(Ever Married)
	(#)	*(p.p.)*	(*years)*	*(p.p.)*
Migrant*1940	-1.27***	-11.1***	0.43***	-1.63***
	(0.02)	(0.34)	(0.032)	(0.047)
Migrant*1930			1.10***	0.96***
			(0.046)	(0.22)
Migrant*1920				0.39
				(0.67)
Controls				
State of birth	Y	Y	Y	Y
Age	Y	Y	Y	Y
Census year	N	N	Y	Y
Constant	3.76***	78.6***	20.92***	85.2***
	(0.03)	(0.35)	(0.05)	(0.22)
IPUMS sample	1940-100%	1940-100%	1940-100% 1930-5%	1940-100% 1930-5% 1920-1%
Observations	88,919	131,031	227,102	3,906,222
R-squared	0.094	0.052	0.070	0.260

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are in parentheses. The sample includes southern-born black women aged 16 and over living in the South, Northeast, and Midwest census regions. All regressions include controls for state of birth and age. Census year indicators are included when sample consists of multiple census years. The variable of interest (Migrant*Year) is an indicator for a migrant mother (i.e. a southern-born female living in the Midwest or Northeast census regions at the time of the decennial census). Each column represent a regression with a separate dependent variable: 1) The number of children ever born, 2) an indicator for having at least one child, 3) age at first marriage, 4) an indicator for ever being married.

Table 7: Impacts on subsamples to explore potential mechanisms

	Base	Urban Residence Status	Neonatal Death	Post-neonatal Death	Male	Female
Migrant*1920	0.1097***	0.1171***	0.0609***	0.0576***	0.1218***	0.0926***
	(0.019)	(0.019)	(0.015)	(0.016)	(0.028)	(0.026)
Migrant*1930	0.0451***	0.0516***	0.0280***	0.0206***	0.0561***	0.0325***
	(0.008)	(0.009)	(0.006)	(0.007)	(0.012)	(0.011)
Migrant*1940	0.0112	0.0192**	0.0125**	-0.0011	0.0151	0.0048
	(0.008)	(0.008)	(0.006)	(0.006)	(0.012)	(0.011)
Urban*South*1920		0.0608***				
		(0.017)				
Urban*South*1930		0.0297**				
		(0.012)				
Urban*South*1940		0.0263**				
		(0.011)				
Constant	0.1005***	0.0934***	0.0828***	0.0297**	0.0775***	0.1227***
	(0.018)	(0.018)	(0.016)	(0.013)	(0.026)	(0.026)
N	19,091	19,091	19,091	18,339	9,652	9,436
R-squared	0.028	0.029	0.024	0.024	0.040	0.044
County FE	Y	Y	Y	Y	Y	Y
Pre-migration controls interacted with year indicator	Y	Y	Y	Y	Y	Y

Notes: Robust standard errors in brackets. *** 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 sample includes African-American births and infant deaths in Illinois, Ohio, Pennsylvania, Tennessee, North Carolina, and South Carolina to fathers born in the three southern states. Urban is defined as living in a city with more than 2500 people in the outcome year. Matched observations are weighted to recover the observed mortality rate in each state. The sample includes observations in the outcome year with fathers that could be matched to a state-of-birth childhood home in a prior census. Pre-migration occupation indicators are relative to the omitted category of "all other occupations." A tenant farmer is an observation that reports occupation as farmer and rents a farm. A head of household is literate if reporting that they can both read and write (in 1900-1920) and if reporting that they are literate in 1930.

Table A1: Comparison of revised black infant mortality rates to published *Vital Statistics of United States*

State of birth	Years	*VSUS* rate	Revised rate	Diff (Revised - VSUS)	
Illinois	1935-1940	6.34	6.39	0.04	=
Illinois	1927-1930	10.50	10.68	0.18	+
Illinois	1915-1920	n.a.	18.22	n.a	
Ohio	1935-1940	7.23	7.44	0.21	+
Ohio	1927-1930	12.15	11.04	-1.10	-
Ohio	1915-1920	n.a.	15.89	n.a	
Tennessee	1935-1940	8.21	6.88	-1.33	-
Tennessee	1927-1930	11.33	9.53	-1.80	-
Tennessee	1915-1920	n.a.	11.64	n.a.	
North Carolina	1935-1940	8.30	8.29	-0.01	=
North Carolina	1927-1930	10.49	9.75	-0.74	-
North Carolina	1915-1920	n.a.	10.52	n.a.	
South Carolina	1935-1940	9.03	7.96	-1.07	-
South Carolina	1925-1930	n.a.	9.85	n.a.	
South Carolina	1915-1920	n.a.	11.82	n.a.	

Notes: All entries in the table are measured in terms of p.p. (alternatively deaths per 100 live births). Infant mortality rates are for all black births in each state. VSUS did not report births prior to 1927 for IL, OH, NC, and TN, and prior to 1932 for SC. Because of this a comparison is not available for some years and is denoted by "n.a.". Revised and *VSUS* rates are six-year averages in 1940 and 4-year averages in 1930 to simplify the comparison. Revised rates for 1920 are six-year averages. See section 2 in the main text or the data appendix for a description of how the revised rates are constructed.
Sources: *Vital Statistics of the United States* (1915-1940), indices of the 1920-1940 Decennial Census of Population microdata and collected death certificate indices provided by FamilySearch.org.