Death in the Promised Land: The Great Migration and Black Infant Mortality[1]

Katherine Eriksson
UC-Davis and NBER

Gregory T. Niemesh
Miami University and NBER

December 7, 2016

Abstract: The Great Black Migration from the rural South to the urban North in the first half of the 20th century drastically lowered the health environment of infants. We show that migrating to northern cities increased the likelihood that an infant born to a migrant would die in the first year of life. We find no evidence that differences are driven by migrant selection. Much of this gap is due to residential location in unhealthy neighborhoods within northern cities. Our paper adds to the literature on the costs and benefits of the Great Migration.

[1] We are particularly grateful to Jeremy Atack, Leah Boustan, William J. Collins, Dora Costa, 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 and Australian Cliometric Society 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 emigrated from the South during the Great Black Migration, dramatically altering the geographic distribution of the black population in the United States. In 1900, 90 percent of blacks lived in the South, which dropped to 53 percent by 1970 at the conclusion of the migration. Over the same period, receiving regions experienced an increase in the share of the population that was black: from 4 percent 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 cities.[2]

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 (Grossman, 1991; Lemann, 1991; Wilkerson, 2010). For example, employment opportunities were superior in the North, providing a 60-70 percent gain in real income (Collins and Wanamaker, 2014).[3] However, outcomes worsened along a number of non-labor market dimensions, namely, higher incarceration rates (Muller, 2012; Eriksson, 2016), lower social standing (Flippen, 2013), and reductions in longevity for older adults (Black, et al., 2015).

In this paper, we show that the urban North and rural South provided radically different health environments for children, with large impacts on the health of the second generation. The Great Migration was one of the most important events in 20th century African-American history, and so the contribution of migration to the racial health gap, for both infant mortality and later

[2] The Great Depression and World War II divide the Great Migration into two periods: 1914-1940 and 1940-1970. This paper focuses on the first Great Migration, in which the percent of southern-born African-Americans residing outside the South tripled from 5 percent (450,000 people) in 1910 to 15 percent (1.5 million people) in 1940. Little migration took place in the decade between 1930 and 1940 following the Great Depression (Boustan, 2017)

[3] However, Eichenlaub et al. (2010) find migrants had lower occupational status relative to those that stayed in their Southern communities.

life health, is potentially large, and has not yet received much attention because of data availability.[4]

We are the first to explore the effect of the Great Migration on black infant mortality, and proceed in three steps. First, we compare infant mortality rates between northern and southern states between 1915 and 1940. The underlying rates contain two innovations. First, we use state-level infant mortality adjusted for the severe under-registration of births in the South following the census based method used in Eriksson and Niemesh (2016). Second, using complete death indices for a set of six southern and northern states, we construct mortality rates specific to infants of southern-born parents. We find that infants born in northern states were 8.9 percentage points more likely to die (93% higher than in the South) in 1920, and that this gap was 4.5 percentage points by 1930; by 1940 the two regions were indistinguishable. We show that the early gap is largely, but not entirely, accounted for by the fact that cities were unhealthy, particularly for black infants, even as late as 1940. Thus, the regional convergence between North and South was driven primarily by the relatively more rapid improvement in mortality rates in Northern cities. The broad patterns are consistent either with negative selection (based on health) into migration or with place-based effects.

The second part of the paper rules out the selection channel, leaving place-based effects as the likely mechanism. Potential selection bias is addressed through a matching strategy that controls for pre-migration family characteristics and place-of-origin unobservables similar to that of Collins and Wanamaker (2013). We compile individual-level death certificates from three southern states (North Carolina, South Carolina, and Tennessee) and three northern states (Illinois, Ohio, and Pennsylvania). Observations for live infants come from the decennial census.

[4] The obvious exception is Black et al (2015), who find that migrating North provides a slight reduction in longevity for the migrants themselves, conditional on survival to age 65.

We use these to create an individual-level dataset containing migrant and non-migrant parents of deceased and live infants in each census year, 1920, 1930 and 1940. To control for potential selection into migration, we match the fathers to an earlier census; this allows us to control for observable characteristics which predict migration as well as origin county fixed effects.[5] We do not find evidence for the hypothesis that migrants were selected on the basis of health, a pattern that is consistent with the lack of selection into black migration documented elsewhere (Collins and Wanamaker, 2014; Boustan, 2017).

Finally, as place-based effects remain as the likely mechanism, we turn to exploring black settlement patterns within northern cities as an explanation for the initial southern mortality advantage in 1920, and the subsequent regional convergence by 1940. We use two cities in our sample of states, Chicago and Philadelphia, for which ward level information on health outcomes is available, and a large interdisciplinary literature on racial residential settlement patterns exists.[6] We show that (a) intra-urban disparities in health were large throughout the period; (b) black migrants from the South tended to live in initially less healthy neighborhoods. These two facts together explain the persistent urban penalty for black infants which had mostly been eradicated for white infants by 1920. We hypothesize that discriminatory housing patterns account for initial residence patterns (Cutler and Glaeser, 1997; Shertzer and Walsh, 2016; Vigdor, 1999; Drake and Cayton, 1945). We argue that the high black infant mortality rates in the North were not driven entirely by the stress of migrating or behavioral changes, but also partly by characteristics of the locations themselves.

---

[5] We also consider an instrumental variables strategy to account for selection into migration. However, in our context the distance-to-a-railroad instrument used in Black et al. (2015) is not excludable. While proximity to a railroad predicts migration, in the context of contemporaneous migration, it is unlikely that the exclusion restriction holds: urban areas were more likely to send migrants to the North, but infants also faced high mortality environments. We have run IV regressions using this instrument but do not believe it valid in this situation.

[6] These two cities alone account for 20 percent of migrants from the South as well as 20 percent of migrants from our three states above.

Our paper touches upon three related strands of the literature. A first 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.[7] 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), 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, our results have important implications for the understanding of black economic progress.

Our work also fits into the literature about black-white health disparities. 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 the period 1910 to 1998, the large black-white infant mortality gap gradually declined, as shown in Figure 1. Diminished racial disparities in socioeconomic status account for most of the convergence prior to World War II (Collins and Thomasson, 2004; Boustan and Margo, 2016). 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).

Finally, we add to the literature on health in cities in the early 20th century. While much is known about intra- and inter-city differences in health outcomes in the late 19th and first decade of the early 20th century (Costa and Kahn 2015, Cain et al 2016, Troesken 2004), we are the first

[7] See Almond and Currie (2011) for a recent survey of the literature on the persistent impact of early life conditions.

to connect health outcomes from 1915 to 1940 to residential decisions by race. Our time period comes after most sanitary investments (Alsan and Goldin, 2016; Cutler and Miller, 2005) were completed in northern cities, and thus sheds light on the remaining role of within city variation in health environment for infant outcomes.

The paper proceeds as follows. The next section discusses the historical background of the Great Migration and then of infant mortality in the 20th century. Section 3 shows the broad patterns in infant mortality between North and South and describes how we construct a new set of infant mortality rates. In Section 4, we construct a micro-level dataset and show that selection is not responsible for the main differences. In Section 5, we explore residential decisions within northern cities. Section 6 discusses and concludes.

II. The Great Migration and Historical Infant Mortality

A. The Great Migration

The Great Migration is divided into two periods by the Great Depression and World War II: 1910-1940 and 1940-1970. By 1970, 1.6 million African-Americans lived in the North or West of the country. Only 47 percent of blacks still lived in the South by 1970. This paper will focus solely on the First Great Migration. Figure 2 shows the destinations of migrants leaving the South during both phases. During the first phase, most migrants moved directly north of their birth state: for example, fifty percent of migrants from Mississippi went to Chicago while over half of migrants from Virginia migrated to Philadelphia. Table 1 shows the top ten destination cities of black migrants in 1920 and 1930. In each decade, the top ten cities account for almost 60 percent of migrants. A large number of whites also left the South by 1940 but did not move to the same cities. Panel B of Table 1 shows the top destinations of white migrants—these ten cities only account for twenty percent of white migration. Figure 3 shows the overall number of Southern-born blacks living in three northern states, Ohio, Illinois, and Pennsylvania, by decade from 1850 to 1950.

The Great Migration was caused by a combination of push and pull factors. Within the South, the passing of Jim Crow laws incentivized rural blacks to move North in hope of less discrimination. The boll weevil infestation reduced cotton productivity as the boll weevil swept across the South starting in Texas around 1900 (Lange et al., 2008). In the North, demands for industrial labor during World War I meant that industrial cities needed more labor than was available—large companies sent labor agents to the South to recruit labor.

Collins and Wanamaker (2014) show that blacks who migrated could expect a 60-70 percent increase in real occupational income. However, blacks were more likely to be incarcerated after moving North (Mueller, 2012; Eriksson, 2016) and race riots were common. The new black migrants often clashed with immigrant populations with whom they were competing for jobs. Overcrowding in housing within cities, as well as industrial pollution, potentially had large negative impacts on the health of migrants. Black et al. (2013) show that migrants from the Deep South had decreased longevity, conditional on surviving to age 65, and that this was mainly driven by behavioral changes—smoking and drinking. This paper is the first to look at the health of second generation infants.

B. Historical infant mortality

Infant mortality of both races decreased quickly in the first part of the 20^{th} century. Figure 1 plots infant mortality rates by race for the set of states in the birth registration area. The birth registration area was composed of mostly Northern states until 1920, and most blacks in northern states lived in cities, so the black infant mortality rate is probably higher than the actual national rate. Nonetheless, black infant mortality rates were higher than white rates. The convergence by 1940 is the topic of this paper.

Figure 4 shows the evolution of infant mortality rates for blacks and whites between 1920 and 1940 in the North and South. Not captured in this graph is that infant mortality rates in northern cities fell starting in the late 19th century with the introduction of clean water sources and sewer systems (Cutler and Miller, 2005). For example, Chicago's sewer system was installed in the nineteenth century and water was chlorinated after 1916. Philadelphia's water was chlorinated after 1913. These systems benefited black residents as much or more than white residents (Troesken 2004). Our data picks up after the main sanitary infrastructure was in place. Subsequent decreases in infant mortality were likely caused by cleaner milk (Lee, 2007), the promotion of breast feeding, and public health outreach programs. In Section V, we discuss some specific programs in Chicago and Philadelphia which helped urban infant mortality rates continue to fall through the 1930's.

Southern mortality rates fell as well, but at a slower pace as the North converged with the South. Likely explanations for this decline include malaria and hookworm eradication programs carried out with the aid of southern philanthropists (Bleakley 2007; Barreca et al 2012). In addition, convergence in wages and education between whites and blacks accounts for a large part of the convergence in infant mortality, likely in the South as well as North (Collins and Thomasson 2004).

III. Comparing Infant Mortality in the North and South between 1920 and 1940

Infant mortality rates in the South were systematically biased upwards during this time period due to incomplete birth registration. For the first-half of the 20th century, completeness of the birth registration system 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. 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 nonwhites combined) can be seen in Figure 5. 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). Differences were largely driven by hospital versus home birth (Shapiro, 1950). The completeness 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 use revised mortality estimates from 1920-1940.

The full procedure is described in Eriksson and Niemesh (2016). The procedure amounts to revising birth counts by state, race, and year using the full count census indexes. We then combine these with published death statistics to create revised mortality estimates. For our purposes here, to reduce year to year variability, we pool the five years of data prior to the census date.[8]

We want to estimate the difference in infant death between migrant and non-migrant southern-born fathers, so we use the death indexes from six states—Tennessee, South Carolina, North Carolina, Illinois, Ohio, and Pennsylvania – to create migrant-specific mortality rates for migrants from the three southern states to the three northern states. Northern-born infants of southern-born fathers might face different risks than infants of northern-born fathers due to differences in socioeconomic status or residential location choice. We use the full death certificate indexes from FamilySearch.org to create infant death counts based on father's place of birth.[9] Similarly, we construct rates for southern-born infants specific to non-migrant father's place of birth (e.g. the sample of Tennessee born infants is restricted to having Tennessee born fathers). We follow the same procedure outlined above for constructing the state-based revised rates.

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 2, 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-

---

[8] We account for migration out of states of birth between birth and death using the census indexes. 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).[8] Our estimates of the number of births will be biased downward relative to the true value to the extent under enumeration occurred—for example, in South Carolina, we find a 20 percent increase in births relative to the birth registration statistics. All contemporary sources argued that deaths were more likely to be registered than births due to higher incentives to register the death than the birth. See Eriksson and Niemesh (2016) for more discussion.

[9] We create rates based on father's place of birth because in the following section we match fathers back to their childhood census 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.

average rates in the North, whereas they are 10 to 15 percent lower 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.

The original revised estimates for 5-year average infant mortality rates by state of birth are shown in Figure 2. First, northern black infant mortality was 62 percent (5.7 p.p.) higher than the southern rate during the early part of the 20th century (16.6 p.p. vs 10.9 p.p.). Importantly, revised rates show convergence between the North and South to be later than the published statistics indicate—after 1935 instead of around 1926.

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.

Figure 6 plots black infant mortality rates in our northern and southern states using the migrant-specific adjusted rates. Table 3 reports these unconditional differences in infant mortality rates between regions. Using migrant-specific rates, the gap is 8.9 percentage points in 1920, falls to 4.5 percentage points in 1930, and about 0.5 percentage points by 1940. We turn in the next two sections to possible explanations for the large but decreasing gap. We consider three types of explanations: selection of migrants, place-specific effects of Northern cities, and behavioral changes.

IV. Accounting for selection using a matched sample

A. Constructing a matched sample

In this section, we turn to estimating the extent of selection into migration from the South. If migrants are positively selected based on health, we would actually be underestimating the true causal effect of moving North. 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., 2004; Black et al., 2015). However, negative selection can occur when the cost of migration is low. Irish-born residents of England are less healthy, on average, than Irish non-migrants (Delaney et al., 2013). In the context of the Great Migration, most literature finds little evidence of selection into migration (Collins and Wanamaker, 2014; Boustan, 2017).

We use an empirical strategy that aims to remove selection bias from the estimates by controlling for a host of pre-migration characteristics. The basic conceptual framework is to compare infant outcomes of black parents that migrated to the outcomes of infants of those parents that stayed in the South conditional on a rich set of pre-migration household characteristics, including county-of-origin fixed effects. We restrict to the six states described above for which we have individual-level data.

We use two main sources of data: complete digitized indexes of death certificates made available from FamilySearch.org and Ancestry.com[10], and the complete digitized decennial censuses for 1900-1940 also from FamilySearch.org and the NBER restricted full count data. The combination of these recently available data sources allow for the construction of a data set

[10] We were able to receive the full death indexes from FamilySearch.org, a volunteer-based genealogical website, for all states except Pennsylvania. We hand-collected the full sample of deceased black infants born to fathers born in one of our southern states in the relevant years from Ancestry.com.

that includes a measure of infant health, post-migration parental characteristics, and pre-migration parental characteristics at the individual level.

Our goal is to create a sample of births, which may or may not end in an infant death, born to southern-born black male migrants and non-migrants with measures of pre-migration characteristics with which to predict the migration decision. As in the discussion of regional rate differences above, five years of data prior to each decennial census are combined to form outcome periods ending in 1920, 1930, and 1940.

We construct a sample of live and deceased infants in each census year 1920-1940 along with their parents observed in the census in the following way:

> (First Match - Death records to census) Match infant deaths from individual vital statistics death certificates to parents in the census index, which provides us with the parental characteristics of both the infants who died (from the death records) and the surviving infants (from the census).
>
> (Second Match - Census to Census) Take the post-migration decision outcomes from the first match and find the father in his pre-migration household in a prior census, providing us a set of controls to remove the bias from the causal estimate of migration.

Details of the match procedure are described in Appendix A.

After both matches, our final sample consists of 9,574 deceased infants and 173,016 births not ending in infant death. We take a 10 percent sample from North Carolina and South Carolina, a 25 percent sample from Tennessee, and a 100 percent sample from the northern states: this results in a final sample of 19,091 infants, 1,452 of which died in infancy. After matching, we look up by hand the characteristics in the pre-migration census: household head occupation, literacy, and home-ownership status.[11]

[11] Full census indexes are now available for 1910, 1920, 1930, and 1940. Because some men are matched to 1900, however, we take a sample to look at individual manuscript pages by hand. When the full count 1900 data is released by IPUMS, we will be able to use pre-migration characteristics for all 182,590 matched individuals.

B. Evidence of and Accounting for Selection

We estimate the impact of migration on infant mortality by estimating the following linear probability model:

$$\textbf{(1)}\ D_{it} = \alpha_c + \beta_1 Mig1920 + \beta_2 Mig1930 + \beta_3 Mig1940 + \gamma_t + X_{it}\theta + u_{it}$$

Where D_{it} is an indicator equal to one if the infant died and zero otherwise. $Mig1920, Mig1930, and\ Mig1940$ are equal to one if the child is born in the north in 1920, 1930, and 1940, respectively. We include year fixed effects. The coefficients of interest are β_1, β_2, and β_3 which measure the impact of migration on infant mortality for each period separately.

Migrants were differentially drawn from households with higher levels of income, wealth, or education as is shown in Table 4.[12] 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 set of parental background characteristics from the pre-migration census: home ownership, literacy, and dummies for a set of occupational categories of the household head.[13] We also control for a set of county fixed effects to control for the idea that healthier counties may send more or less migrants.

Results from Equation (1) are reported in Table 5. Each observation is weighted to recover the state-level migrant specific mortality rate for the appropriate decade from Table 2. In Column (1), we replicate the results from Section III. A southern mortality advantage of 8.8 percentage points exists in 1920. By 1930, the advantage is only 4.1 percentage points, and the

---

[12] Collins and Wanamaker (2014), Collins and Wanamaker (2015), Black et al. (2015) also find that migrants are positively selected on a number of socioeconomic variables.

[13] 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

North is zero by 1940. This 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 estimates of β. We take a number of steps to account for the possibility of this selection bias. In the next three columns of Table 5, we explore the possibility of selection on observables and unobservables using our matched sample. Column (2) adds the household-level pre-migration controls. The coefficient estimate does not significantly change from the initial specification, suggesting that selection on observables does not bias the main result. Column (3) adds county-of-origin fixed effects to account for location-specific unobservables. We find the estimate to be slightly larger, but the change represents only a small portion of the raw migration effect. 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 2.

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. To explore whether the urban-rural distribution of births drives the large North-South difference in black infant mortality, we estimate equation (2) but with additional indicators for urban status in the South interacted with year.

We show these results in Table 6. In Column (2), coefficients should be interpreted relative to the black mortality rate in the rural South for that given year. In 1920, migration to the North increased mortality relative to residing in a southern rural area. Children born to northern migrants in 1920 faced an 9.7 percentage point (118 percent) higher mortality rate than southern rural infants. Those born in the urban south faced a 5.6 percentage point (66 percent) higher mortality rate. In 1930, the north-rural south difference was cut in half to 4.8 p.p. (58 percent of the rural rate), and the southern urban penalty also declined by half to 2.7 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. However, the rural South maintained a slight mortality advantage of 2.4 p.p. relative to the urban South and 1.6 p.p. relative to the urban North.

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, in the rest of Table 6 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 infectious 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. These are likely correlated with the health of the mother.[14]

In the final columns (3) and (4) of Table 6, we estimate equation (2) with an indicator for neonatal death and an indicator for post-neonatal death as the dependent variable. The results for

[14] Costa (2004) argues that higher black prematurity rates in Northern cities were caused by a higher prevalence of syphilis among blacks than whites. We would expect this to fall after penicillin was available in the 1940's and 1950's.

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: 57 percent in 1930, 87 percent in 1930, and 360 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, but it appears that post-neonatal causes of death converged faster, consistent with public health improvements in northern cities. 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 (Columns 5 and 6).

Taken together, we interpret the results from Table 5 and Table 6 to show that the initial southern mortality advantage was primarily driven by an urban penalty for black infants. The balanced decline in rates across all subsamples suggests a broad improvement in black infant health in all locations, but at differential rates. Our results suggest that the negative impact on infant mortality from migration North stems from a rural to urban migration pattern followed by blacks during the Great Migration during a period when an urban penalty still existed for black infants.

In Table 7, we show evidence that the black urban penalty for migrants through 1940 did not mimic the patterns for white migrants but that the rates for northern-born blacks in northern cities were similar to those of southern-born migrants.

C. Fertility differences between North and South

Before turning to looking at city-specific factors driving the North-South infant mortality gap, we consider the role of one behavior which could differ across region, namely fertility. Given that babies born later in the birth order (except for the difference between first- and second- born babies) have a lower chance of surviving the first year of life, if fertility rates are higher in the North this could be an important factor.

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 8 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.[15] Figure 7 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.[16]

Columns (1) and (2) of Table 8 report 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 show that migrant women are 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 (Columns (3) and

---

[15] 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.
[16] 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.

(4)). In sum, all the results point toward a substantial difference in fertility and marriage behavior between the movers and stayers.[17]

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 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). In the next section, we turn to look at two specific cities in the north, Chicago and Philadelphia, for which we have data on health outcomes broken down by ward.

V. Patterns in Infant Mortality in Chicago and Philadelphia

A. Ward-Level Information from Chicago and Philadelphia

In this section, we use Chicago and Philadelphia as case studies to understand why infant mortality was so high in northern cities and how it was changing over time. We chose these cities for two reasons: first, Chicago and Philadelphia described the modal experience of migrants—over 20% of migrants from the South went to one of these two cities and their black infant mortality rates were not substantially different than other large receiving cities; Second, detailed ward-level information on mortality rates as well as GIS ward boundary files are available from the Union Army Historical Ecological Data project.[18]

[17] 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.
[18] See uadata.org/hue

In Chicago, Department of Health reports reported ward-level overall mortality rates per 1,000 population as early as 1880. In Philadelphia, the same reports published ward-level overall mortality rates after 1890 and ward-level infant mortality rates after 1914. Ideally, we would have ward-level infant mortality rates broken down by race, but these were not reported. We are working to geocode addresses from the death certificate index in Chicago so as to create these rates. This will allow us to even look at the enumeration-district level and consider factors such as crowding in housing, distance to coal-polluting manufacturing plants, and accessibility of sewer line.

Two hypotheses drive our analysis. First, areas of cities to which black migrants moved might have been unhealthy *before* they came. This would be consistent with place-based causal factors. If, on the other hand, we see that certain areas become unhealthy *after* the black migrants arrive, this is consistent with behavior differences, overcrowding in housing, or discriminatory provision of health care.

Second, as public health departments reached out more to unhealthy parts of town in the 1920's and 1930's, it's possible that mortality rates converged between healthy and unhealthy parts of town. Costa and Kahn (2015) show convergence in typhoid mortality rates within New York and Philadelphia as clean water interventions are made. Anecdotal evidence suggests that Chicago and Philadelphia's health departments were sending lots of public health nurses and other types of medical care into the unhealthiest wards. If the mortality rates of wards within cities are converging over time, we'd expect this to drive the lower mortality rate by 1920.

We look first at Chicago in 1920 in Figure 8. Panel A shows the overall ward-level mortality rates from 1920 plotted against circles which are proportional to the southern-born black population size in 1920. We see, unsurprisingly, that the black population was living predominantly in wards with high mortality rates. Figure 9 shows the scatter plot of black population versus ward level mortality. There is a positive, statistically significant correlation with a regression coefficient of 1266.22 (se = 312.35).

To test our first hypothesis, we then plot mortality from 1900 with the black population from 1920. We find that places to which black migrants moved were actually among the most healthy places in the city in 1900 (although, still unhealthier than they were in 1920). The neighborhoods in Chicago to which black migrants moved have been nicknamed the "Black Belt" and were actually previously middle-to-upper-class white neighborhoods in 1900 (Hirsch, 1998). Given that public infrastructure was in place by 1900, it is unlikely that these neighborhoods were not connected to water and sewer by 1920, suggesting the explanation for high black mortality is not just "bad neighborhoods". Panel C in Figure 8 shows that density was higher in neighborhoods with high mortality, so part of the story is probably housing density and ease of transmission of communicable diseases. Panel D shows that immigrants and blacks did not live in the same parts of town by 1920, so this experience was not common to both disadvantaged groups.

We turn now to intra-city convergence in mortality rates. Figures 10 and 11 show that infant and total mortality rates *did not* converge between "more black" and "less black" wards between 1915 and 1930 in either Chicago or Philadelphia. We define the "more black" wards according to the population distribution in 1920. As we see in the Figures, the two types of wards track pretty closely and even appear to diverge by 1930. Consistent with a lack of convergence, the overall variance of mortality within cities increases between 1920 and 1930 (not shown).

These patterns go against what contemporary observers were arguing. In a 1925 address to the National Social Work Convention, Forrester B. Washington, president of Philadelphia's

Armstrong Association, described public health nurses working in predominantly black wards in Philadelphia, New York City, and Cincinnati. Yet, he admitted that cities were not doing enough to reduce infant mortality among blacks in northern cities. He also highlighted overcrowding, noting that, in 1924, Philadelphia had built 35,000 new houses, “not a single one was available for (black) tenancy”.

From the above analysis, we conclude that initial high black infant mortality rates in Chicago and Philadelphia likely were a result of behavioral differences (breastfeeding, medical care), overcrowding, and discrimination, not a neighborhood effect. The lack of convergence between more and less black parts of town suggests that, despite health interventions to bring down black infant mortality, the black population was growing fast enough to keep infant mortality rates high.

VI. Discussion and 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 8.9 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 death explains the total regional difference. A broad-based regional convergence in rate underscores the fact that a single explanation cannot account for the patterns. 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.

The motivating question behind this paper is what would black infant mortality have looked like if the Great Migration had never happened? We calculate a rough estimate of 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.

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. We find 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.

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.

Alsan, Marcella and Claudia Goldin. 2015. "Watersheds in Infant Mortality: The Role of Effective Water and Sewerage Infrastructure, 1880-1915." NBER Working Paper Number 21263.

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.

Barreca, Alan, Price Fishback, and Shawn Kantor. 2012. "Agricultural Policy, Migration, and Malaria in the 1930s United States". *Explorations in Economic History* 49: 381-98.

Black, Dan A., Seth G. Sanders, Evan J. Taylor, and Lowell J. Taylor. 2015. "The Impact of the Great Migration on the mortality of African-Americans: Evidence from the Deep South". *American Economic Review* 105(2): 477-503.

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.

Boustan, Leah. 2017. *Competition in the Promised Land: Black Migrants in Northern Cities and Labor Markets*. Princeton University Press, Princeton, NJ.

Boustan, Leah and Robert Margo. 2016. "Racial Differences in Health in the United States: A Long-Run Perspective" in *The Oxford Handbook of Economics and Human Biology*. Ed. John Komlos and Inas R. Kelly. Oxford University Press.

Cain, Louis and Sok Chul Hong. 2009. "Survival in 19th Century Cities: The Larger the City, the Smaller Your Chances". *Explorations in Economic History* 45(4): 450-463.

Cain, Louis, Sok Chul Hong, and Carlos Villareal. 2016. "Intra-Urban Health Disparities: Survival in the Wards of 19th Century American Cities". Manuscript.

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* 70(4): 746-776.

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): 220-252.

Costa, Dora and Matthew Kahn. 2015. "Declining Mortality Inequality within Cities during the Health Transition" *American Economic Review P&P* 105(5): 564-569.

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.

Cutler, David, Edward L. Glaeser & Jacob L. Vigdor, 1999. "The Rise and Decline of the American Ghetto," *Journal of Political Economy*, 107(3): pp 455-506.

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

Drake, St. Clair. 1945. *Black metropolis; a study of Negro life in a northern city*. New York: Harcourt, Brace and Co.

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.

Eriksson, Katherine and Gregory T. Niemesh. 2016. “Revising Infant Mortality Rates for the Early 20th Century United States”. Manuscript.

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

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

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.

Grossman, James, *1989. Land of Hope: Chicago, Black Southerners, and the Great Migration*. Chicago: University of Chicago Press.

Haines, Michael R., 2006. “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.

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.

Hirsch, Arnold R. 1998. *Making the Second Ghetto: Race and Housing in Chicago, 1940-1960.* University of Chicago Press, Chicago, IL.

Hornbeck and Naidu (PUT IN TEXT)

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.

Lange, Fabian, Alan Olmstead and Paul Rhode. 2009. "The Impact of the Boll Weevil, 1892-1932". *Journal of Economic History* 69(3): 685-718.

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

Lee, Kwang-Sun. 2007. "Infant Mortality Declines in the Late 19th and Early 20th Centuries: The Role of Market Milk". *Perspectives in Biology and Medicine* 50(4): 585-602.

Lemann, Nicholas, 1991. *The Promised Land : The Great Black Migration and How It Changed America*. New York: Vintage Books.

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.

Shertzer, Allison and Randall Walsh. 2016. “Racial Sorting and the Emergence of Segregation in American Cities” NBER Working Paper Number 22077.

Smith, James P. and Finis R. Welch, 1989. “Black Economic Progress After Myrdal,” *Journal of Economic Literature*, vol. 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.

Troesken, Werner. 2004. *Water, Race, and Disease*. MIT Press, Boston, MA.

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.

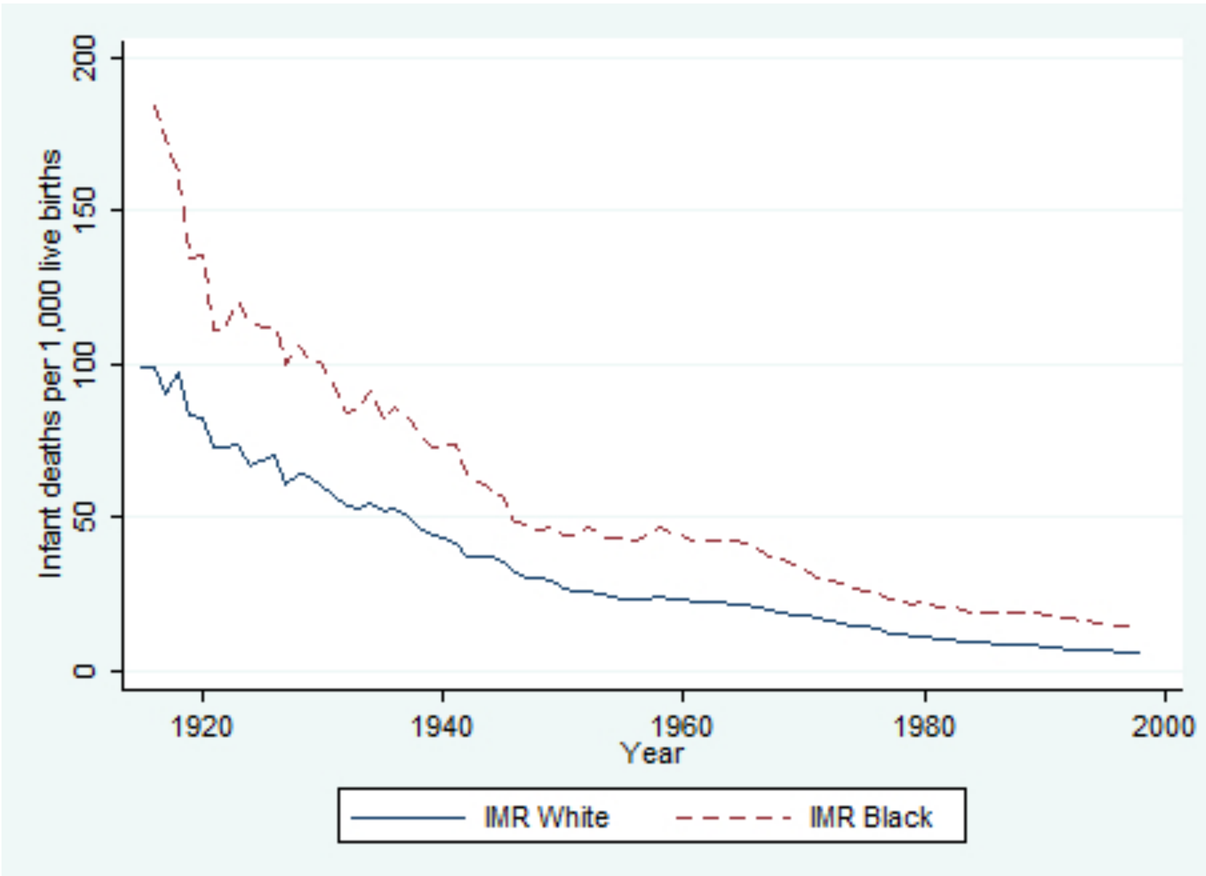
Wilkerson, Isabel. 2010. *The warmth of other suns: the epic story of America's great migration*. New York: Random House.

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.

Figures

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 2: Migration patterns during the First and Second Great Migrations

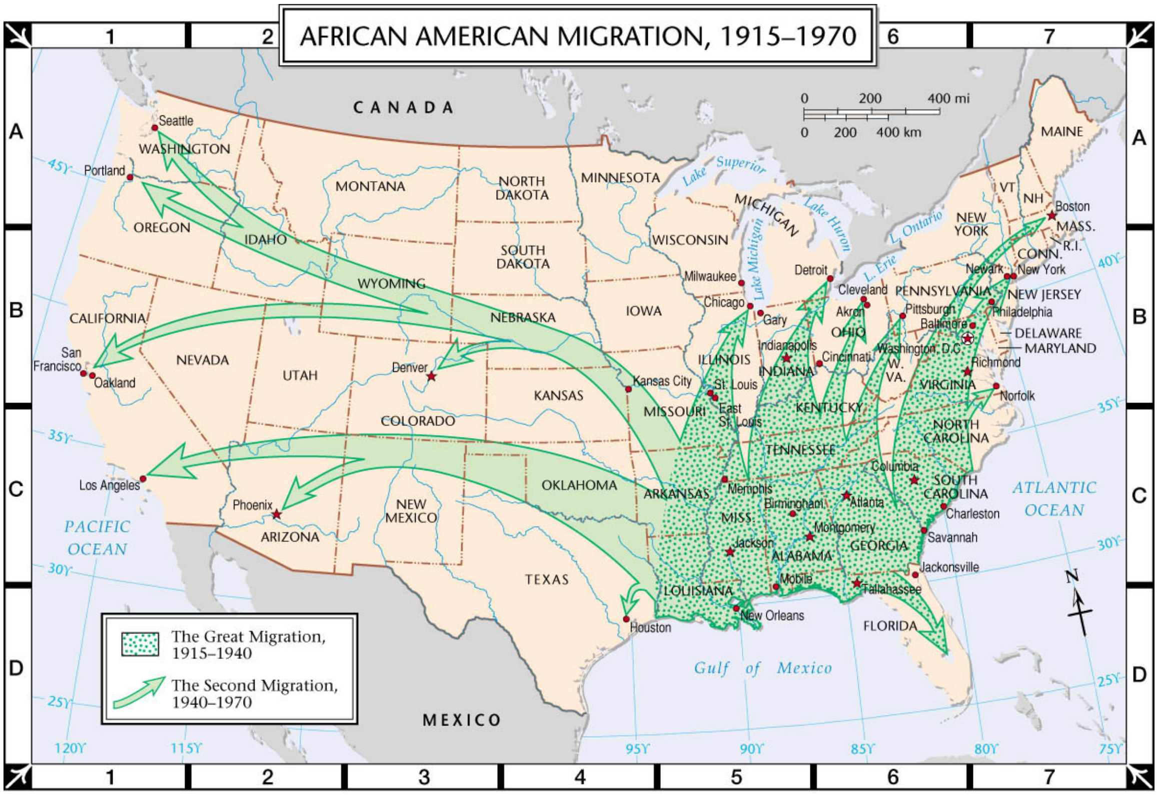
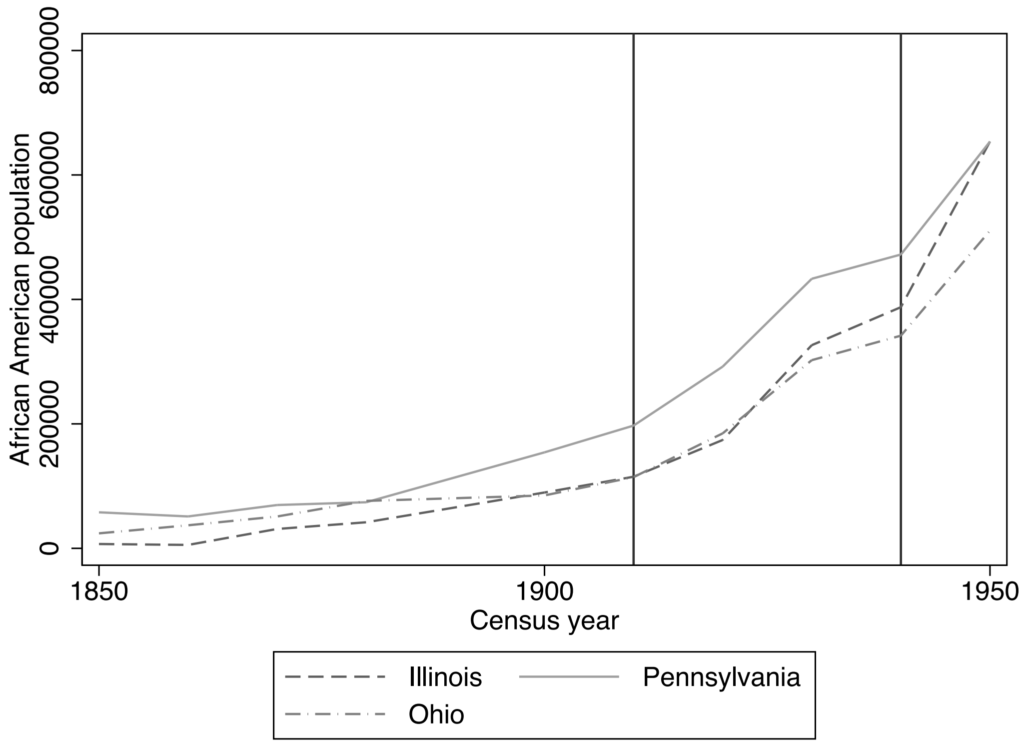


Figure 3: Increase in African-American population in sample of northern states



Source: IPUMS 1850-1950.

Figure 4A: Black Infant Mortality Rates by North and South

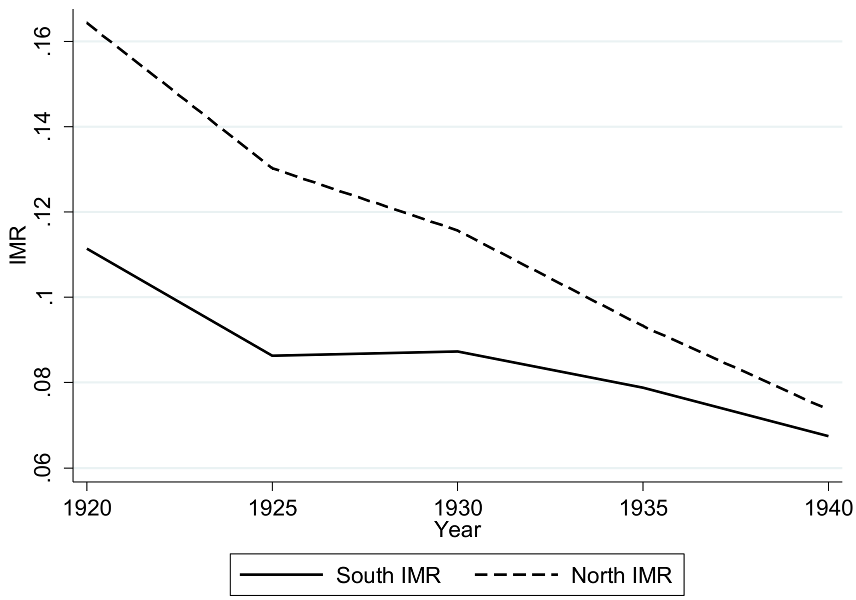
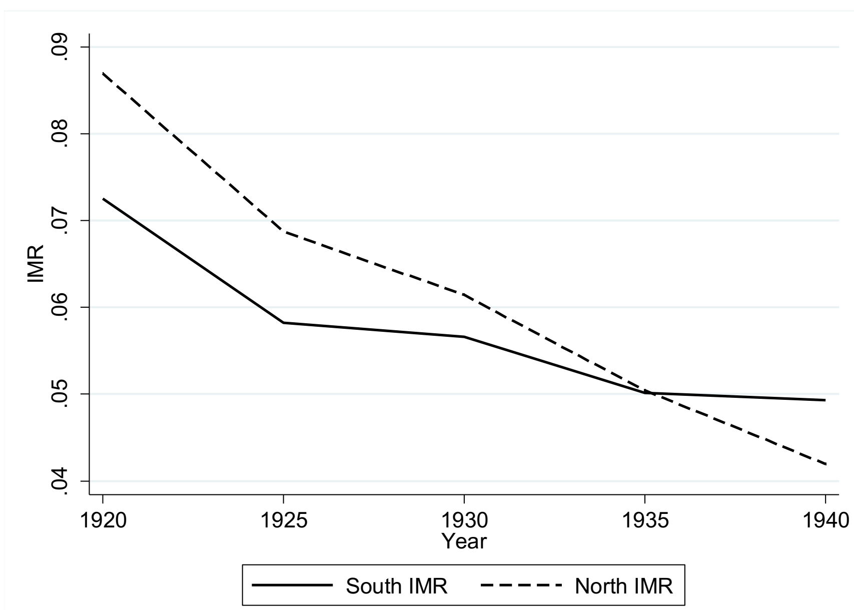


Figure 4B: White Infant Mortality Rates by North and South



Notes: Rates are adjusted using the method in Eriksson and Niemesh (2016). States included if they were members of the Death Registration Area. Southern states: Mid-Atlantic and South Atlantic census regions; Northern States: Northern and North Central census regions. Rates are reported as infant deaths per live birth. Data comes from the published Vital Statistics Death reports as well as the decennial censuses of 1920, 1930, and 1940.

Figure 5: Undercounting of births in the South: completeness of the birth registration area (white and nonwhite births)

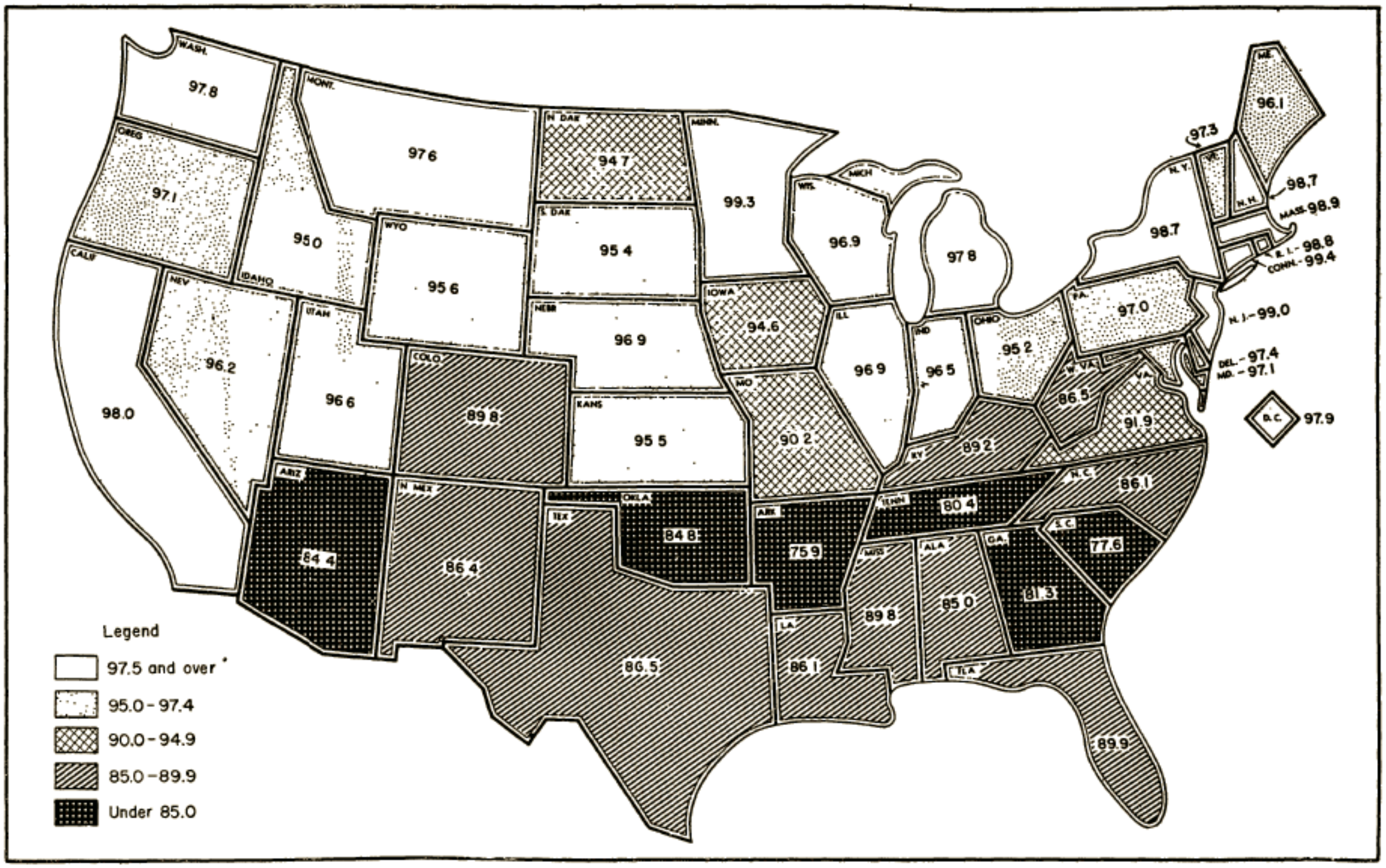
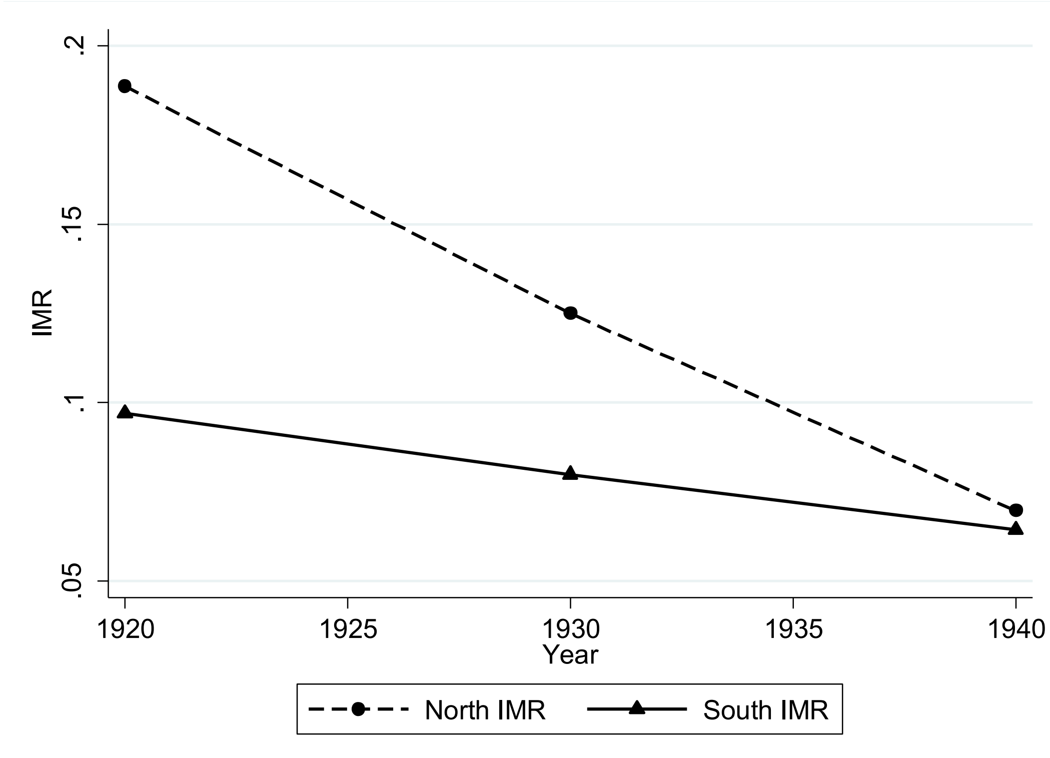


Fig. 2. Percent completeness of birth registration: each State, 1 December 1939 to 31 March 1940.

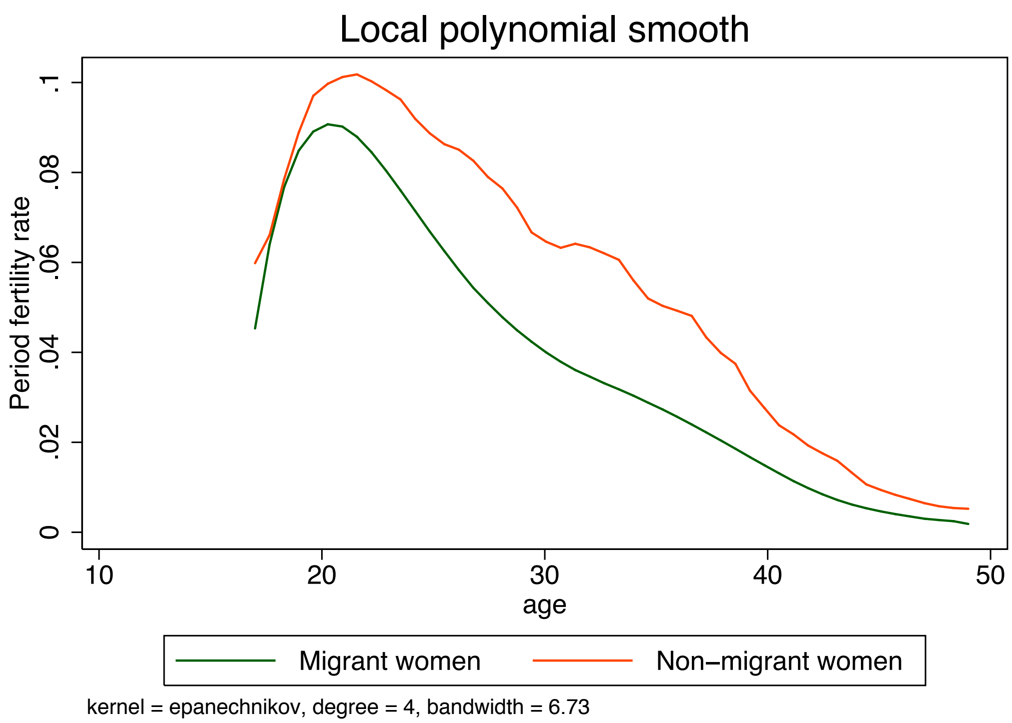
Source: Shapiro (1950). The U.S. Census Bureau conducted an independent check on the completeness of the birth registration system in early 1940 by comparing special infant cards from the March 1940 Decennial Census and official birth certificates in state vital statistics offices

Figure 6: Convergence in Migrant-specific Infant Mortality Rates, by region



Notes: Rates are adjusted using the method in Eriksson and Niemesh (2016). Northern states: Illinois, Ohio, Pennsylvania. Southern states: North Carolina, South Carolina, Tennessee. Data comes from the full count death indexes from the six states, as well as the decennial censuses of 1920, 1930, and 1940.

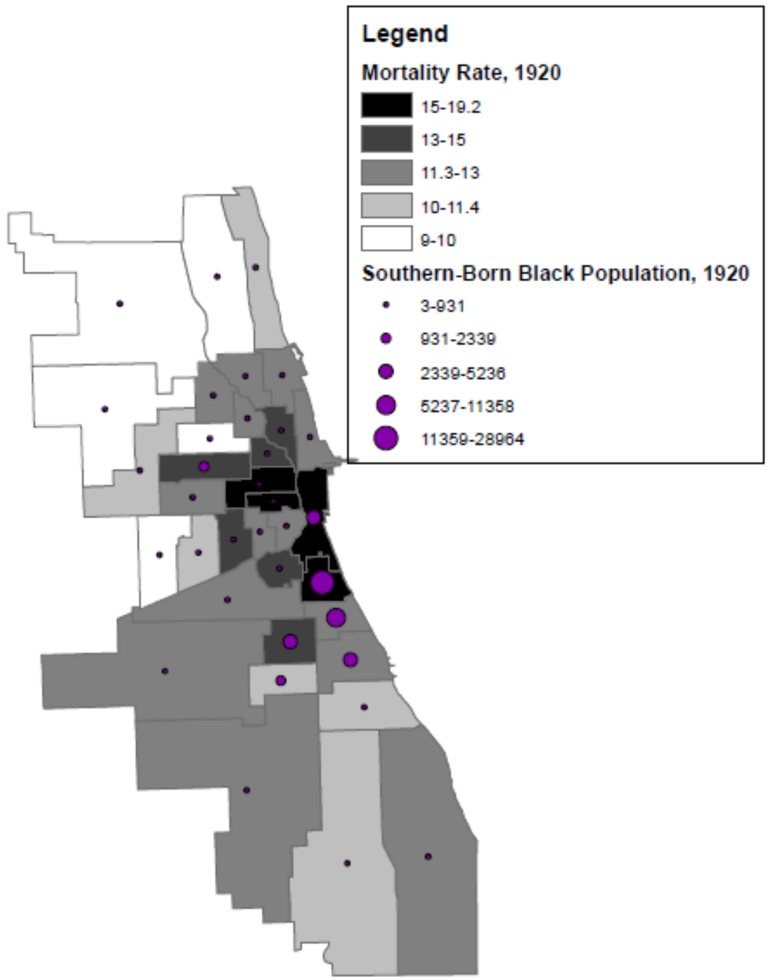
Figure 7: Migrant and non-migrant period fertility by age of southern-born African-American women in 1940.



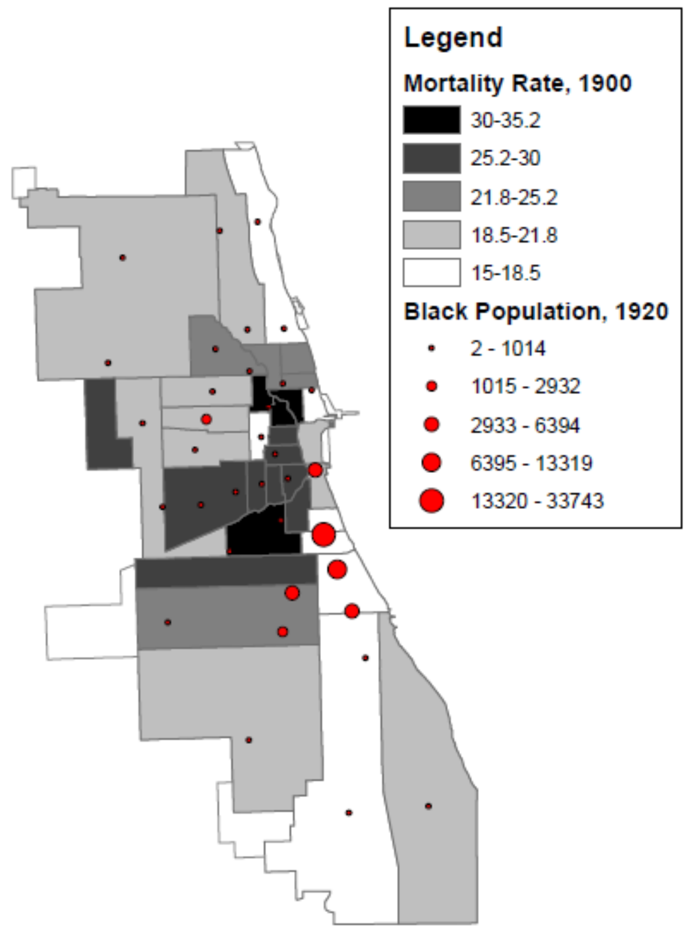
Notes: Units are percentage points. The sample includes southern-born black women aged 16 and over living in the South, Northeast, and Midwest census regions from the IPUMS 1940 census full count. A fourth-degree polynomial smoothing is applied over age groups for an indicator if an infant is reported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

Figure 7: Chicago Mortality Rates, Black Population, and Population Density

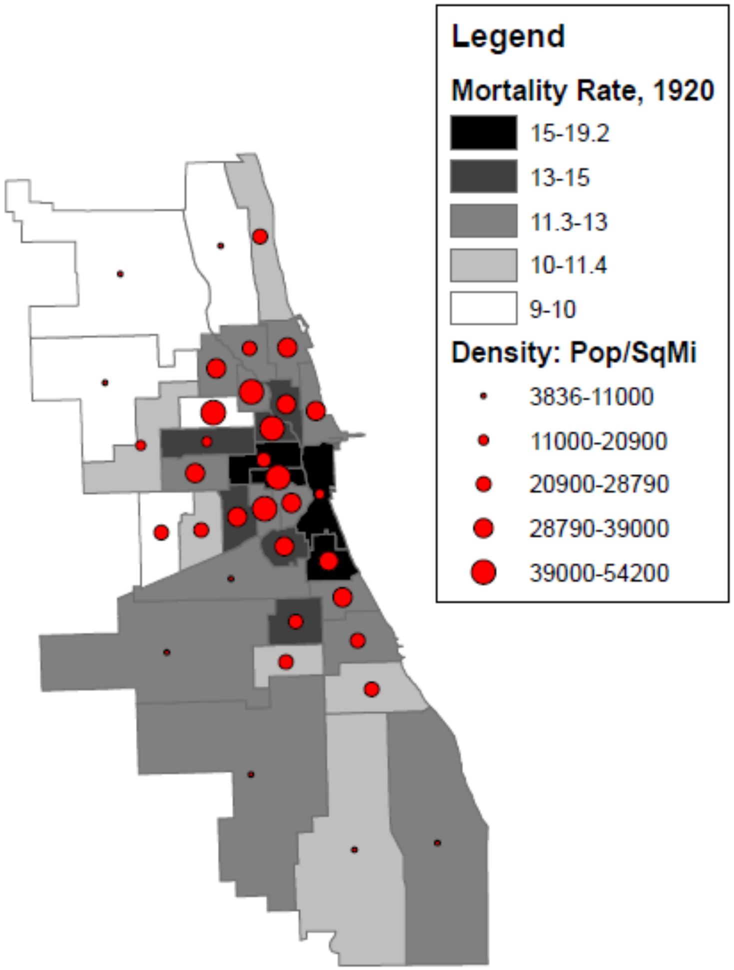
A.



B.



C.


Legend
Mortality Rate, 1920
15-19.2
13-15
11.3-13
10-11.4
9-10
Density: Pop/SqMi
3836-11000
11000-20900
20900-28790
28790-39000
39000-54200

D.

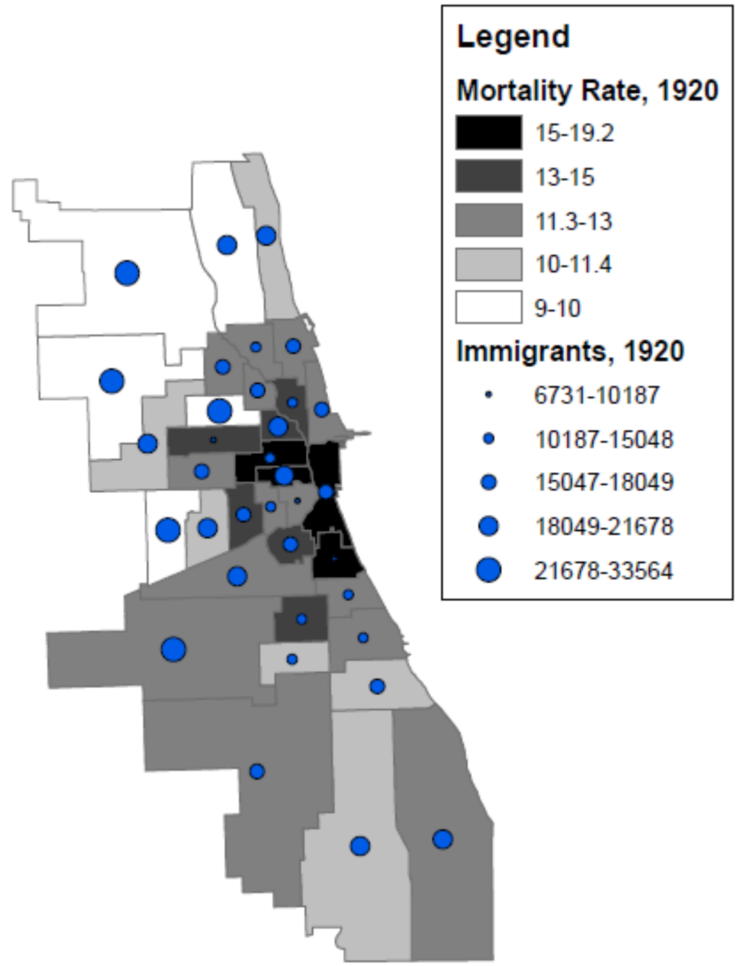
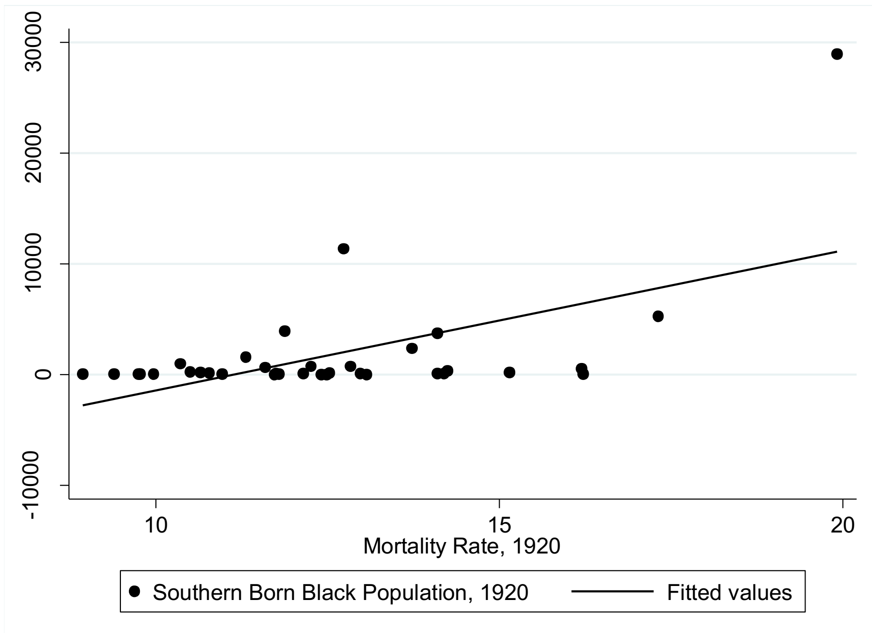

Legend
Mortality Rate, 1920
15-19.2
13-15
11.3-13
10-11.4
9-10
Immigrants, 1920
6731-10187
10187-15048
15047-18049
18049-21678
21678-33564

Figure 8: Correlations between Ward Level Mortality Rates and Characteristics—Chicago, 1920



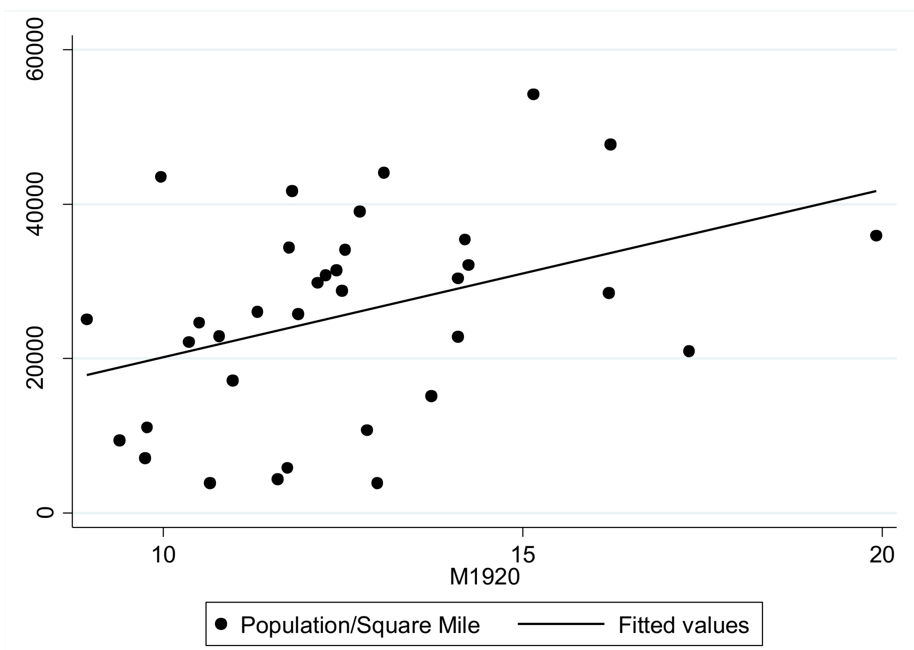
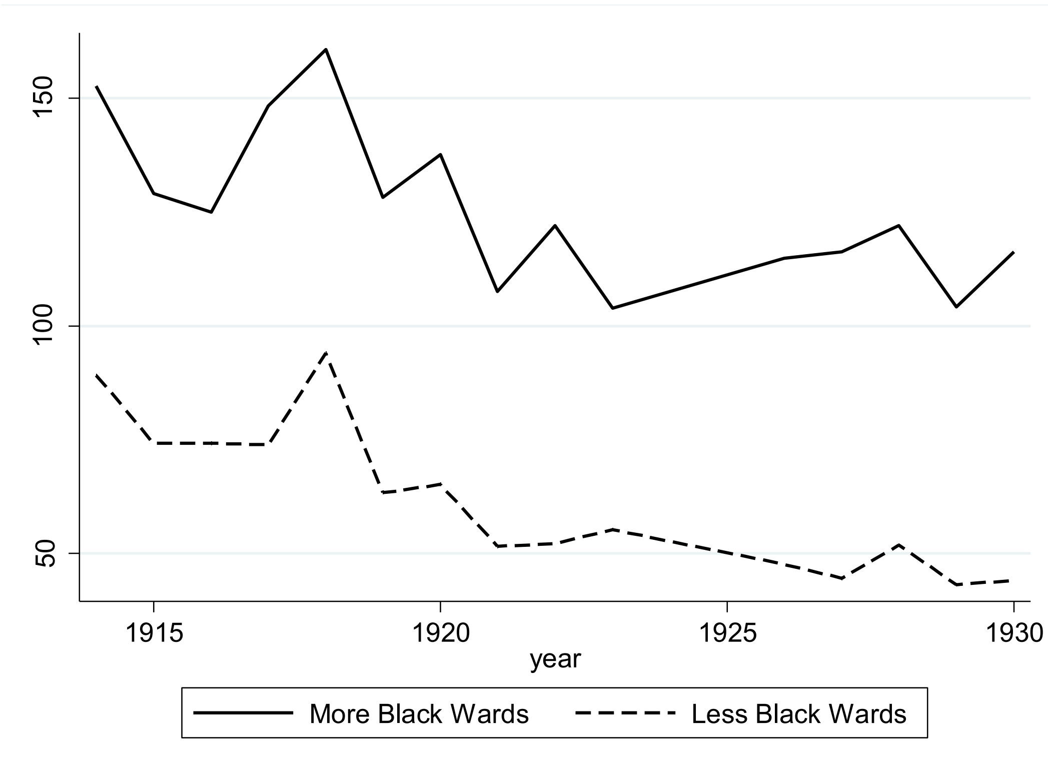
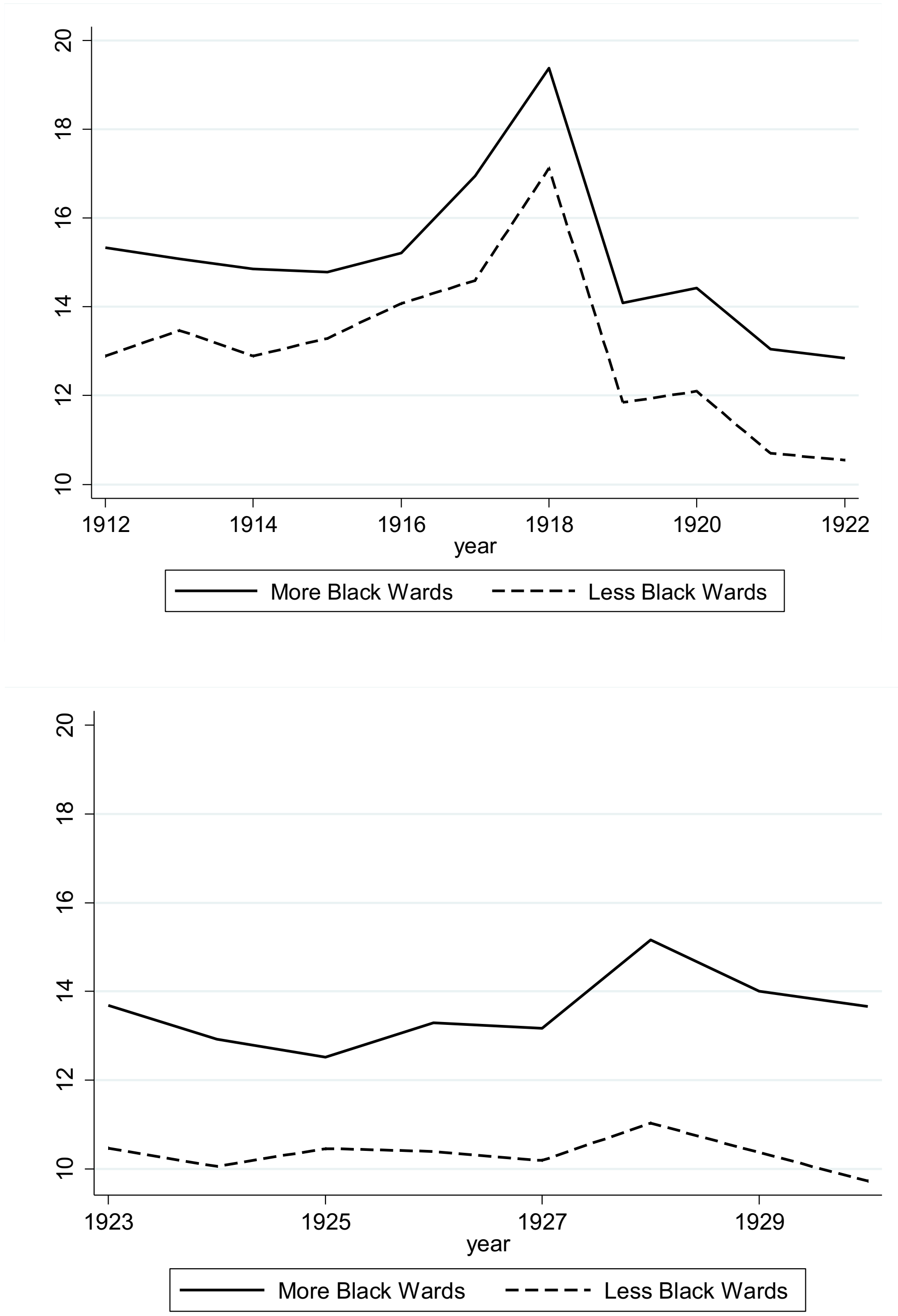


Figure 9: Most Black versus Least Black ward infant mortality rates, Philadelphia 1914-1930



Notes: Data taken from the Union Army Project's Historical Urban Ecological data. Two groups defined by population distribution in 1920 full count data.

Figure 10: Most Black versus Least Black ward infant mortality rates, Chicago 1914-1931



Notes: Data taken from the Union Army Project's Historical Urban Ecological data. Two groups defined by population distribution in 1920 full count data.

Tables

Table 1: Top 10 Destination Cities of Black Great Migrants, 1920

City	% of Black Migrants	City Infant Mortality Rate
Philadelphia, PA	11.6	17.8
Chicago, IL	9.75	--
New York City, NY	9.08	16.4
St Louis, MO	4.98	--
Detroit, MI	4.10	--
Cleveland, OH	3.30	--
Pittsburgh, PA	3.19	17.9
Cincinnati, OH	2.72	19.0
Indianapolis, IN	2.70	13.5
Kansas City, KS	2.44	21.3
Columbus, OH	1.57	14.6
Total	*55.5*	

Notes: Author's calculation using full count census indexes from FamilySearch.org. Sample restricted to southern-born men between ages 18 and 65 who are living outside the South in 1920.

Table 2: 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	16.6	11.4	6.7
Southern states	10.9	9.3	6.2
Illinois	14.6	9.1	4.1
Ohio	15.3	10.6	5.2
Pennsylvania	16.7	10.8	6.2
North Carolina	9.8	9.1	6.1
South Carolina	11.1	9.0	6.2
Tennessee	9.4	8.9	5.4

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	18.8	1.8	12.5	1.1	6.9	0.2
Southern states	9.7	-1.1	8.0	-1.3	6.4	0.2
Illinois	13.4	-1.2	10.4	1.3	5.8	1.7
Ohio	15.8	0.5	10.6	0	6.4	1.2
Pennsylvania	22.5	5.8	14.1	3.3	7.6	1.4
North Carolina	9.3	0.5	7.8	-1.3	6.9	0.8
South Carolina	10.4	-0.7	8.5	-0.5	6.2	0
Tennessee	8.2	-1.2	7.1	-1.8	5.5	0.1

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 3: 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	5.6	53	2.1	23	0.5	8
Migrant-specific rates	9.1	94	4.5	56	0.5	8

Notes: Summary of results from Table 1 and Table A.2.

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

	(1) Infant Death	(2) Infant Death	(3) Infant Death	(4) Infant Death	(5) Infant Death
Migrant*1920	0.0878***	0.0871***	0.0900***	0.0898***	0.0899***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Migrant*1930	0.0412***	0.0405***	0.0439***	0.0437***	0.0420***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Migrant*1940	0.0029	0.0026	0.0061	0.0064	0.0091
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
1930 Indicator	-0.0167**	-0.0174**	-0.0159**	-0.0166**	0.0010
	(0.007)	(0.007)	(0.007)	(0.007)	(0.022)
1940 Indicator	-0.0320***	-0.0337***	-0.0321***	-0.0339***	-0.0566***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.019)
Pre-migration controls					
Farmer (Owner)		-0.0121		-0.0118	
		(0.011)		(0.011)	
Farmer (Tenant)		-0.0009		0.0016	
		(0.007)		(0.008)	
Farm Laborer		-0.0030		-0.0032	
		(0.011)		(0.012)	
Laborer		0.0121		0.0109	
		(0.009)		(0.009)	
Owns Home		0.0032		0.0040	
		(0.010)		(0.010)	
Head Literate		0.0044		0.0052	
		(0.005)		(0.005)	
Constant	0.0961***	0.0950***	0.0953 ***	0.0924***	0.0959***
	(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	0.026	0.027
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; standard errors are clustered at the childhood census county level. 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. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Impacts on subsamples to explore potential mechanisms

	(1) Base	(2) Urban Residence Status	(3) Neonatal Death	(4) Post-neonatal Death	(5) Male	(6) Female
Migrant*1920	0.0878***	0.0973***	0.0503***	0.0461***	0.1002***	0.0752***
	(0.017)	(0.017)	(0.013)	(0.014)	(0.025)	(0.023)
Migrant*1930	0.0412***	0.0480***	0.0260***	0.0190***	0.0527***	0.0299***
	(0.008)	(0.008)	(0.006)	(0.006)	(0.012)	(0.011)
Migrant*1940	0.0029	0.0164**	0.0107*	-0.0015	0.0130	0.0029
	(0.007)	(0.008)	(0.006)	(0.005)	(0.011)	(0.010)
Urban*South*1920		0.0566***				
		(0.016)				
Urban*South*1930		0.0276**				
		(0.012)				
Urban*South*1940		0.0241**				
		(0.010)				
Constant	0.0891***	0.0778***	0.0286**	0.0772***	0.1131***	0.0959***
	(0.017)	(0.015)	(0.012)	(0.024)	(0.024)	(0.017)
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: Standard errors in brackets; standard errors are clustered at the childhood census county level. 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. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Infant Mortality Rates for White and Northern-Born Blacks

	1920	1930	1940
Panel A: White Infant Mortality Rates in North and South			
North	9.1	6.5	4.3
South	6.9	5.9	4.9
Diff: Northern minus Southern	3.2	0.6	-0.6
Panel B: Comparing Northern-born and Southern-born Mortality Rates			
Northern-born Black			
Illinois	13.8	9.9	6.6
Ohio	10.7	10.3	6.9
Diff: NB Black – SB Black			
Illinois	0.4	-0.5	0.8
Ohio	-5.1	-0.3	0.5
Panel C: Chicago-specific Infant Mortality Rates			
SB Black	14.4	10.3	5.4
NB Black	13.6	10.7	7.7

Notes: Rates taken from Eriksson and Niemesh (2016), except for Chicago. Chicago-specific infant mortality rates are calculated from the full count death indexes from Illinois.

Table 8: Fertility and marital outcomes of southern-born migrant and non-migrant African-American women.

	(1)	(2)	(3)	(4)
	Children ever born (#)	P(> 0 kids) *(p.p.)*	Age at marriage (*years)*	P(Ever Married) *(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.

Appendix A: Match Procedure and Comparison of Matched Sample to the Population

First Match - Death records to census: The index of death records lists the race of the child, birthplace of the child, and the names and birthplaces of both parents. We match each death record to the father's household in the decennial census conducted after the child's death. Infant deaths from 1915-1919 are matched to the 1920 census, deaths from 1925-1929 to the 1930 census, and deaths from 1935-1939 to the 1940 census. We allow the father to reside in any state in U.S. regardless of the state in which his child died. 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 within the North to Indiana or return to the South by 1940. To maximize the number of parents who are matched to the census, we use a procedure similar to (Feigenbaum 2016).[19] It proceeds as follows:

1. Blocking on the father's state of birth, calculate the Jaro-Winkler string distance (for surname and given name) between that listed in the death record and all potential matches in the census; restrict to possible father matches with a string distance of at least 0.8 for both surname and given name.
2. For the spouse of each potential match, calculate a Jaro-Winkler string distance between the given name listed in the death record and the name listed in the census. Drop potential father matches with a spouse Jaro-Winkler string distance less than 0.8 or a spouse with an incorrect state of birth.

[19] Unfortunately, father's age is not available in the death certificates, which decreases our match rates because of multiple potential matches when we use a conservative limitation of the sample based on the age of potential fathers. The pool of potential matches in the census is limited to men aged 18 to 50, which accounts for 95 percent of all fathers of black children under the age of 5 enumerated in the census.

3. If a household contains an exact match on name for both parents, this household is unique, and there is no other potential match with both scores above 0.8, 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.
5. If we do not find a match in the previous two steps, we consider married men who are not living with their spouse.[20]
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.

This procedure yields a sample of 2,457 infant deaths to migrants and 39,878 infant deaths to nonmigrants and for an overall match rate of 40 percent, with some consistency across states and census year. Match rates by state and year are shown in Table A.1. Rates are consistently slightly higher in the South than the North, most likely because those who migrate are more geographically mobile and thus less likely to be enumerated successfully in the Census. In Table A.2, we show the breakdown of matches by selection criteria. Most of the individuals who we cannot match do have a possible match in the Census, but the name, spouse name, and each state of birth are not unique enough to find a definitive match. We match approximately 75 percent of our successful matches using both spouses' names and birth states, with the remaining

[20] Note that the goal of the exercise is to identify the father in the outcome year census so that he can be matched to a pre-migration census. However, we could similarly take unmatched observations in step 5 and attempt to match them to a mother not living with a spouse, but this observation would not be included in the empirical strategy that depends on father's state of birth to indicate migrant status.

25 percent of fathers matched because they are not living with their wives in the Census but do report being married.

In Table A.3, we use the limited set of characteristics available in the death indexes to assess whether the matched sample differs from the population of dead infants that we try to match. We do not find much evidence that this is the case. Matched infants die slightly earlier than non-matched but the difference is only 0.5 percent of the mean. We also consider quarter of birth and quarter of death and find no differences. We are more likely to match infant deaths of children born to fathers from North Carolina and slightly less likely to match those born to fathers from Tennessee. We re-weight all regressions to account for differential match rates across parental birthplace.

The first match procedure constructs the sample of births that ended in infant death. From the census indexes, we extract all black children under age five who were born in one of the sample states to southern-born fathers from one of the three southern sample states.[21] We define migrant status based on the birth state of the *child.* Together, these two data sources represent the sample of birth outcomes.

Second Match - Census to Census: The next step involves taking the father for each birth in the sample and finding his pre-migration household in a previous census wave, either 10 or 20 years prior to the outcome year census. The goal is to find the father as a child or young adult, still living in his childhood household. Men who are older than 26 are matched over a 20 year horizon while men younger than 26 are matched over 10 years. To match fathers across

[21] Note that the census enumeration, to the extent that is complete, provides the population of black children meeting the sample restrictions that remain alive to the census date. Children that die after infancy but before the census date, clearly are not enumerated in the census index. However, we have a list of non-infant deaths from the death index for which we complete the match procedure above. These matches are included in the regressions to follow as a birth that did not end in an infant death. What we cannot include in the sample are births that migrated out of the sample states and ended in death prior to the census. For example, a child born in Illinois in 1935 that survived past infancy, emigrated to Mississippi in 1937, and died at age 4 in 1939 in Mississippi, will not be included in the sample.

censuses, we follow the iterative procedure pioneered by Ferrie (1996) and used in Abramitzky et al. (2012). It proceeds 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 et al. 1992). Any common nicknames are recoded to standard first names (e.g. Will becomes William).
2. Observations are matched backwards from the post-migration census year to the pre-migration census 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, yielding 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 only accept unique matches. If none of these attempts produces a match, the observation is discarded as unmatched.

Match rates for the second match are reported in Table A.4. We find lower match rates for black men residing in the North as opposed to the South. The differential match rates arise from the fact that it is easier to find individuals that do not migrate. Table A.5 compares observable characteristics from the matched sample from the second match to the population as a whole. Similar to the findings from Abramitzky et al. (2012), we find that matched individuals come from slightly higher socio-economic status. 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. Specifically, our sample is 1.3 percentage points (about 8 percent) less likely to be living in an urban area in the last census year. Fathers are more likely to be literate and have 0.23 more years of education in the matched sample than the population. Surprisingly, they are slightly less likely to own their home. Finally, they are more likely to be represented in farm occupations in the South, consistent with the more rural status of the matched sample. We reweight results to account for differences in urban/rural status.

Table A.1: Match rates by state and time period, First match

	1915-1919	1925-1929	1935-1939
Illinois	39.50	40.35	31.80
Ohio	36.94	33.57	46.60
Pennsylvania	31.76	34.96	34.24
North Carolina	39.98	40.17	44.02
South Carolina	37.86	39.35	44.15
Tennessee	36.41	37.04	42.74

Notes: Match rates for the first match procedure (death certificates to subsequent census)

Table A.2: Causes of Match Failure, First match

Category	1915-1919	1925-1929	1935-1939
Potential death records to be matched	49,411	33,726	23,702
- Name/state-of-birth combination not found	116	365	68
- Name/state-of-birth/residence combination not unique	30,381	20,231	13,343
Actual matches	18,914	13,130	10,291
Matched with spouse in household	14,391	9,825	8,027
Matched with mother not in household	4,523	3,305	2,264

Table A.3: Comparing the matched sample with the universe of possible matches, First Match

	Population	Difference: (Match - Pop)
Age in days	204.37	-1.241
Quarter of birth		
Q1	0.260	-0.004
Q2	0.257	-0.001
Q3	0.258	0.001
Q4	0.223	0.005*
Quarter of death		
Q1	0.252	0.002
Q2	0.274	0.001
Q3	0.233	-0.001
Q4	0.239	-0.001
Father's birthplace		
North Carolina	0.359	0.011***
South Carolina	0.481	-0.003
Tennessee	0.159	-0.007***
N	106,839	149,249

Notes: Column (1) contains the mean of individual characteristics for the full population of death certificates across all time periods and states. Column (2) reports coefficients and statistical significance for the differences between the matched sample and the population of death certificates ($*p < 0.10, **p < 0.05, ***p < 0.01$).

Table A.4: Match rates by child's state of birth and time period, second match

	1915-1919	1925-1929	1935-1939
Illinois	22.82	25.42	14.65
Ohio	19.32	24.67	15.95
Pennsylvania	27.44	28.75	22.07
North Carolina	27.62	29.40	31.82
South Carolina	26.03	27.81	31.78
Tennessee	29.67	29.81	32.54

Table A.5: Comparing the matched sample to the population, second match

	Population	Difference: (Match - Universe)
Urban	0.192	-0.013***
Age of father	32.31	-0.997***
Father literacy (for 1920 and 1930 only)	0.695	0.035***
Father's years of schooling (for 1940 only)	4.486	0.234***
Owns own home	0.189	-0.005***
Within South:		
Father is owner-occupier farmer	0.028	0.005***
Father is farm laborer	0.300	0.018***
Father is tenant farmer	0.118	0.004***
N	743,967	908,073