HHS Public Access
Author manuscript
Am Econ Rev. Author manuscript; available in PMC 2015 September 03.

Published in final edited form as:
Am Econ Rev. 2015 February ; 105(2): 477–503. doi:10.1257/aer.20120642.

The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South

Dan A. Black,
University of Chicago, 1155 East 60th Street, Chicago IL 60637, and NORC
(danblack@uchicago.edu)

Seth G. Sanders,
Duke University, 419 Chapel Drive, Box 90097 Durham NC 27708, and NORC
(seth.sanders@duke.edu)

Evan J. Taylor, and
University of Michigan, 238 Lorch Hall, 611 Tappan Ave., Ann Arbor MI 48109
(evanjt@umich.edu)

Lowell J. Taylor
Carnegie Mellon University, 5000 Forbes Ave., Pitsburgh PA 15232, NORC, and NBER
(lt20@andrew.cmu.edu)

Abstract

The Great Migration—the massive migration of African Americans out of the rural South to largely urban locations in the North, Midwest, and West—was a landmark event in U.S. history. Our paper shows that this migration increased mortality of African Americans born in the early twentieth century South. This inference comes from an analysis that uses proximity of birthplace to railroad lines as an instrument for migration.

The Great Migration—the early twentieth-century migration of Southern-born African Americans to locations with better social and economic opportunities—played a pivotal role in the lives of millions of individuals. At the onset of the Great Migration and as it was underway, there were suggestions that migration out of the South might be an important avenue for improving the health of African Americans (Wright, 1906; Myrdal, 1948), but also concerns about adverse health conditions in urban areas to which blacks were migrating (Bureau of the Census, 1918). To our knowledge, though, no research has evaluated the long-term impact of the Great Migration on health outcomes of migrating individuals. Our paper tackles this issue, focusing on mortality among African Americans born in the "Deep South" states of South Carolina, Georgia, Alabama, Mississippi, and Louisiana.

Figure 1 gives a sense of magnitude and timing of African American migration out of the Deep South, providing out-migration rates from the Deep South by ages 40–49 for birth cohorts 1891 through 1955. Migration rates were approximately 20 percent for the earliest

Correspondence to: Lowell J. Taylor.
The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

of these cohorts, and migration subsequently increased substantially; for each birth cohort 1914 through 1943, the migration rate was at least 40 percent. As we describe below, due to data availability, we study birth cohorts 1916 through 1932, which are among the high-migration cohorts.

Economic historians refer to a variety of “push” and “pull” factors were at play in shaping the scale and timing of the Great Migration. To give a few examples, Collins (1997) and Carrington, Detragiache and Vishwanath (1996) emphasize increasing demand for labor in the North post World War I, which was in part the consequence of restrictions on foreign immigration; Alston and Ferrie (1993) discuss the breakdown of “paternalistic” labor practices due to the mechanization of cotton production; Lange, Olmstead, and Rhode (2008) assign a key role to damage to agriculture from the boll weevil; and Boustan (2010) demonstrates that county-level out-migration of African Americans from the South was influenced over time by variation in a variety of agricultural conditions.

A key challenge for drawing credible inferences about the impact of this migration on mortality is that migrants are a select group; these individuals pay a possibly high cost—direct migration costs, and also often a loss of present-day economic security and diminished contact with community and family—in the hope that life will be better elsewhere. Migrants thus plausibly have relatively high aspirations and motivation, traits associated with human capital investment generally, including investments in health. As Norman, Boyle, and Rees (2005) note, at least since Farr (1864), scholars have understood that migrants may differ systematically from non-migrants in terms of characteristics related to health. Any effort to estimate the causal impact of the Great Migration on mortality must account for this issue.[1] In our work below, we employ two research strategies for drawing inference about the impact of migration on mortality:

Our first and most innovative strategy is to use the proximity of birthplace to railroads as an instrument for migration. Historians have emphasized the outsized role played by railroads in facilitating the Great Migration and shaping migratory pathways. For example, railroads appear to have contributed to a “vertical” pattern of much of the Great Migration. It seems that African Americans from the South often began travel from the nearest train stop, and then often settled near the terminus of the same rail lines. Such tendencies may have contributed to the migratory streams from the Carolinas to cities on the eastern seaboard (Washington D.C., Philadelphia and New York) via the *Southern Railway* and *Pennsylvania Railroad*, and from Mississippi to Chicago via the *Illinois Central Railroad*. Below we show that proximity of birthplace to a railroad is a powerful predictor of migration for Southern-born African Americans. Under the assumption that being born in a railroad town otherwise has no impact on late-life health, proximity of birthplace to a rail line can be used as an instrumental variable (IV) to estimate the impact of migration on late-life mortality. Of course we are concerned about the validity of the exclusion restriction; our biggest concern is that railroads may have brought prosperity to towns, and this in turn improved prospects

[1]Many papers examine this selection process, which demographers often call the “healthy migrant hypothesis.” See, e.g., Halliday and Kimmitt (2008), for evidence that the healthy tend to have higher geographic mobility.

for lifetime health. We return to this issue in some detail below, arguing that under plausible conditions we can at least credibly bound estimates.

This first empirical approach requires data that provide a great deal of detail on place of birth—far more detail than in the primary sources that are generally used for the study of mortality (e.g., Census records, vital statistics, or historical panel survey data). We use administrative data from the Medicare Part B program, which record birth and death dates, matched to files of the Social Security Administration which record "town or county of birth." We were able to construct records of more than one million individuals for our research—black men and women born in the Deep South, 1916–1932. These data are used in our primary analyses, which focus on mortality rates among the older population (those aged 65 and older), who are included in Medicare data.

To evaluate the impact of migration on mortality at younger ages, we turn to a second approach, based on Heckman and Robb's (1985) "repeated cross-section" evaluation strategy. The idea is quite simple: if moving North has an impact on mortality, then cross-cohort variation in migration rates should be correlated with cross-cohort age-specific mortality. This analysis, which uses data from the 1960 U.S. Census and corresponding death records from Vital Statistics, covers many more birth cohorts (1885–1940) and ages of death (ages 20 to 65) than does our analysis with SSA/Medicare data.

The remainder of our paper proceeds as follows: In Section I we present a simple model of selective migration, and discuss strategies for identification of a causal impact of migration on mortality. Section II has a description of data sources. In Section III we present our basic findings about migration patterns of African Americans out of the Deep South, documenting that (i) migratory flows are influenced by rail lines, (ii) most migration is undertaken when individuals are relatively young (prior to age 30), and (iii) migrants tend to have higher education than non-migrants. Section IV then provides our analysis of mortality. We find that survival rates declined among older African Americans as a consequence of migrating North. By examining cause of death, we get some sense about mechanisms involved.

I. A Model of Human Capital Investment and Migration

To fix basic ideas and set the stage for empirical analysis to come, we provide a simple model that generates selective migration. Suppose human capital is a function of endowed latent ability, α, and schooling, E, chosen by the individual (or by parents): $f(\alpha, E)$, a strictly concave function with $f_\alpha(\alpha, E) > 0$, $f_E(\alpha, E) > 0$, and $f_{E\alpha}(\alpha, E) > 0$ (so the marginal value of education is higher for high-ability people). Then if w is the market return to human capital and c is the cost per unit of education, individuals maximize lifetime utility, given by

$$U(E)=wf(\alpha, E) - cE. \quad (1)$$

Individuals choose the level of education that solves $wf_E(\alpha, E^*) = c$. Standard arguments lead to sensible comparative statics: $\partial E^*/\partial w > 0$, $\partial E^*/\partial c < 0$, and $\partial E^*/\partial \alpha > 0$.

With this basic model of human capital accumulation in mind, consider a decision to migrate, which, like schooling, constitutes a form of investment.[2] In particular, consider an individual living in a Southern state who anticipates earning a higher return on human capital in the North: $w_N > w_S$. To obtain the higher return, he incurs a migration cost *m*. Thus the individual compares utility in the North to utility in the absence of migration, i.e.,

$$U_N^* = w_N f(\alpha, E_N^*) - cE_N^* - m \quad (2)$$

and

$$U_S^* = w_S f(\alpha, E_S^*) - cE_S^*, \quad (3)$$

where educational attainment is chosen optimally in each case (which for any #x003B1; gives $E_N^* > E_S^*$).

This model predicts "selective migration." Let $\hat{\alpha}$ be the level of ability such that an individual is indifferent between migrating North and remaining in the South, i.e., the value of α for which (2) and (3) are equal. Then individuals with latent ability lower than $\hat{\alpha}$ will remain in the South, while those with higher ability migrate. This selection complicates attempts to empirically evaluate the effect of migration on individual outcomes. Observed differences in characteristics between migrants and those who remain in the South-in terms of education, income, health, etc.—could be the consequence of living in the North rather than the South, but could also be due to unobserved differences in traits of migrants and non-migrants. Within this framework, we are interested in estimating an effect that might reasonably be thought of as "the causal impact on longevity of migrating North" for African Americans born in the South. In so doing, we want to allow for the possibility that longevity, *y*, is positively related to innate ability or human capital more generally, both of which are unobservable.[3]

The key to identifying causal effects of migration is to exploit variation in migration costs. This follows from an intuitively sensible comparative static,

$$\frac{\partial \hat{\alpha}}{\partial m} = \frac{1}{w_N f_\alpha(\hat{\alpha}, E_N^*) - w_S f_\alpha(\hat{\alpha}, E_S^*)} > 0; \quad (4)$$

an increase in the migration cost *m* shifts upward the migration threshold.[4]

To see the how variation in migration costs helps with identification, suppose there are two types of location in the South: *railway towns*, which have a relatively low migration cost, and *non-railway towns*, which have a higher migration cost. Let α_0 be the ability threshold for migration North for individuals born in railway towns and let $\alpha_1 > \alpha_0$ be the

[2]In our conception, both schooling and migration are decisions that occur early in life, prior to labor market participation. Of course, some migration (and schooling, for that matter) occurs at older ages. However, as we document below, most migration by African Americans out of the Deep South was indeed by younger individuals.

[3]At most we can only observe educational attainment. Across countries there is strong negative relationship between education and mortality (as shown in Preston, 1975, and many papers that followed), and the same is true within countries. In terms of our model, this correlation could be the consequence of a positive relationship between our latent characteristic (α) and longevity, or because education improves prospects for longevity, or both. Cutler and Lleras-Muney (2010) give a good discussion of the issues involved, and provide links to the literature. Work by Black, *et al*. (2013) shows why identification of the effect of education on mortality is difficult.

[4]To see this result, set (2) and (3) to be equal and use the implicit function theorem.

corresponding threshold for those born in non-railway towns. We can divide individuals into three groups: First, those whose latent ability level is below α_0 are "never movers," *set N*, who remain in the South regardless of which type of town they were born in. Second, those whose latent ability is greater than α_1 are "always movers," *set A*. Finally, individuals for whom $\alpha_1 > \alpha > \alpha_0$ are "compliers," *set C*. This group is so named because conceptually they are people who "comply" with the proposed instrument-moving North if born in a railway town and remaining in the South if born in a non-railway town.

Now let $M = 1$ designate an individual's decision to migrate and $M = 0$ otherwise. Let $y_{M=1}$ be longevity for an individual if he or she migrates, while $y_{M=0}$ is longevity if the individual does not migrate—noting that one of these is observed, while the other is an unknown counterfactual. We can estimate $E(y_{M=1}|A)$ and $E(y_{M=0}|N)$, respectively, by means $\bar{y}_{M=1,z=0}$ and $\bar{y}_{M=0,z=1}$, where $Z = 1$ indicates birth a railroad town and $Z = 0$ otherwise. Of course we can never estimate $E(y_{M=1}|N)$.or $E(y_{M=0}|A)$ Remarkably, simple algebra shows that we can recover both $E(y_{M=1}|C)$ and $E(y_{M=0}|C)$, under the following assumptions: Birthplace, $Z \in \{0,1\}$, induces some individuals to migrate who otherwise would not have migrated (as predicted by our model), and it is statistically independent of $(y_{M=0}, y_{M=1})$.[5]

Given this, we can evaluate two predicted inequalities from our theory:

$$E(y_{M=1}|C) < E(y_{M=1}|A) \text{ and } E(y_{M=0}|N) < E(y_{M=0}|C). \quad (5)$$

More importantly, we can find an estimator whose probability limit is $E(y_{M=1} - y_{M=0}|C)$. Following steps like those in footnote 5, that estimator is found to be a Wald estimator,

$$\frac{\overline{y}_{Z=1} - \overline{y}_{Z=0}}{\overline{M}_{Z=1} - \overline{M}_{Z=0}}. \quad (6)$$

In setting up estimation, we have closely followed Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996)—proposing to estimate a "local average treatment effect" (to use their expression), where the term "local" emphasizes the fact that the estimate pertains for a particular subset of the population, and term "treatment effect" refers to the impact of migration.

Three points are clarified by our theoretical set-up. First, our estimate applies for the middle-ability group only; impacts might differ for higher- and lower-ability individuals. Second, the estimated effect includes the impact of behavioral responses made in anticipation of migration (e.g., increased educational attainment). Third, if migrants are positively selected into migration, the LATE estimate will be smaller than the corresponding OLS coefficient.

One final observation follows from our set-up: Suppose we observe two cohorts with differing rates of migration, which in (2) and (3) could be due to some combination of

---

[5]Notice, for instance, that once we assume independence, we can use Bayes rule to write out $E(y_{M=0}|C) = \{[Pr(N) + Pr(C)]/Pr(C)\} E(y_{M=0}|N \cup C) - \{Pr(N)/Pr(C)\} E(y_{M=0}|N)$. Each element on the right-hand side of this latter equation corresponds to an easily-estimated moment: the term $[Pr(N) + Pr(C)]$ is estimated by the mean migration rate for individuals born in non-railway towns ($\bar{M}_{z=0}$); the term $E(y_{M=0}|N \cup C)$ has expectation $E(\bar{y}_{M=0,z=0})$, i.e., mean longevity among non-migrants from non-railroad towns; and so forth.

differing migration costs, differing returns to remaining in the South, and differing returns in the North. Then, regardless of the source of differing migration rates, if migration increases mortality— which, we show below, seems to be the case—cohort level mortality should be higher cohorts that have higher migration rate. We use this idea to motivate our second approach to evaluating the impact of migration on mortality idea (drawing on Heckman and Robb, 1985).

II. Data

Our ability to study the impact of the Great Migration on mortality hinges on access to unique data sources.

A. The Duke SSA/Medicare Dataset

Our primary data source is the Duke SSA/Medicare Dataset. These data consist of the Master Beneficiary Records from the Supplementary Medical Insurance Program (Medicare Part B) merged by Social Security Number to records from the Numerical Identification Files (NUMIDENT) of the Social Security Administration (SSA).[6] The data are for the period 1976–2001. There are over 70 million records in the data, covering a very high proportion of the population aged 65 years and older. Because enrollment requires proof of age, the age validity of the records is high compared with other data sources for the U.S. elderly population. In addition to race, sex, and age, information includes entitlement status (primary versus auxiliary beneficiary), zip code of the place of residence in old age, exact date of death (for the deceased), and, importantly, detailed place of birth information. Specifically, the data include either town and state of birth or county and state of birth for U.S.-born respondents.

To our knowledge, this is the only data source that provides detailed place of birth and detailed place of residence in older age in a large sample of individuals born in the early 20th century United States. The data are valuable for answering this key question: Which “sending communities” in the South sent people to which “receiving communities” outside the South.[7] A further advantage of these data is that death and population counts are based on the same data source.

Before the SSA/Medicare data could be used for our purposes, there was a technical hurdle to overcome concerning location of birth. The SSA provides a 12-character text field for the place of birth as well as a two-character abbreviation for the state of birth. The state of birth abbreviations follow the Postal Service abbreviations and pose only minor issues to convert to Census state FIPS codes. However, the research strategy outlined above requires that we establish birthplace at a detailed level, so that we can determine precise longitude and latitude coordinates, and then determine proximity to railway lines using appropriate historical records. In order to establish the birthplace from the 12-character text field, we developed an algorithm that matches this object to place names recorded in the U.S.

[6]The merge was made possible by the Center for Medicare and Medicaid Studies and the Social Security Administration (and with extensive confidentiality protection). We are grateful to James Vaupel for his role in making these data available to us.

[7]We observe residence only in older age, so some individuals classified as non-migrants could be individuals who migrated and then returned to the South in old age. Analysis below suggests this is a limited phenomenon (see footnote 12).

Geological Service's Geographic Names Information System (GNIS). The GNIS is the master list of all place names in the U.S., both current and historic, and includes geographic features, including the longitude and latitude of each place. Our algorithm essentially classifies places according to the strength of their match between the write-in place of birth on the SSA NUMIDENT file and the GNIS list. We were able to match places at very high rates, and, we believe, with modest error. We find that the data have quite high coverage rates (typically 0.80 or above) for the 1916–1932 cohorts, but much lower rates for earlier cohorts. Thus we restrict attention to only the 1916–1932 cohorts. (A data appendix available on line provides additional details.)

B. Vital Statistics and Census Data

In our analysis below we also use the Detailed Mortality Files (DMF) of the U.S. Vital Statistics registry. These files contain all deaths in the U.S. and for 1960 include state of death and state of birth. For our analysis of 1960 mortality, we are interested in estimating age-specific death *rates* by state of birth for black men and women by birth cohort and state of birth, so we need to form estimates of the number of African Americans alive in specific years by state and birth cohort. Data to form these estimates come from 1960 Integrated Public Use Samples (IPUMS) of the Decennial U.S. Censuses. We also make use the of IPUMS Decennial Census files for 1910–1990 for various other descriptive exercises below.

III. Patterns of Migration and Labor Market Outcomes of Migrants

As a first step in our empirical analysis, we present evidence concerning the migration flows from Deep South states. In selecting states for analysis, we first limited our sample to states that were in the Confederacy. Second, we focused on states with large concentrations of African Americans, and so further restricted attention to states in which births cohorts 1916 through 1932 were at least 30 percent black (as measured in the 1970 U.S. Census). This left us with five States—Alabama, Georgia, Louisiana, Mississippi, and South Carolina. In addition, we wished exclude states that bordered non-Confederate states, because we did not want to include short-distance migration across state borders as part of our analysis of the Great Migration.[8] As it turns out, this final restriction did not change our selection of states.

A. Migratory Flows

Table 1 uses data from the Duke SSA/Medicare dataset and the 1970 U.S. Census to show location of adult residence for black individuals in birth cohorts 1916–1932 for each of the five Deep South states. In this Table, "South" is defined to be the Deep South plus the other six Confederate states, and those living outside the Confederacy are said to be in the "North."[9] Notice that for the SSA/Medicare data we are looking at residence at age 65 or older, while with the Census data we are looking at ages approximately 38 to 69. The striking feature in this table is the high proportions of black Americans residing outside the

[8]For example, we would not want to include Virginia, because conceptually a move from a birthplace in Northern Virginia to nearby Washington DC is quite different than the long-distance moves that were most common during the Great Migration.

[9]The 11 former Confederate states are Virginia, North Carolina, South Carolina, Georgia, Florida, Alabama, Mississippi, Louisiana, Tennessee, Texas, and Arkansas. (Kentucky and Missouri were officially neutral, though they were represented by stars in the Confederate flag when secessionist parts of these states joined the Confederacy in 1861.) We refer to non-Southern states as the "North" merely for convenience.

South when they are middle aged. The proportion migrating out of the South ranges from approximately 0.30 (migrating out of Louisiana) to approximately 0.50 (migrating out of Mississippi).

Many social scientists have noted “vertical” migratory patterns the emerged during the Great Migration, and have suggested that railroads played a role in shaping these streams.[10] Table 1 provides further documentation of such patterns, using both the Duke SSA/Medicare and Census data. Conditional on residing in the North, individuals born in South Carolina and Georgia tended to move to Eastern cities such as Washington, Philadelphia, and New York, with a substantial proportion of Georgians moving also to Detroit. Migrants from Alabama and Mississippi tended to migrate to such Midwestern cities as Detroit and Chicago. Finally, while there was significant migration from Louisiana to Chicago, a strikingly high faction of these migrants moved west to cities in California. Most migration was to urban locations.[1] Most African Americans who remained in the South remained in their home state.

A potential key role for the railways in shaping migration can be illustrated with maps that we developed in the process of building the data used in our regressions below. Consider, for example, the famous migration of African Americans born in Mississippi to Chicago via the *Illinois Central Railroad*. Figure 2 reproduces a map of this line from the turn of the century. Then Panel A of Figure 3 shows the fraction of black individuals within each of Mississippi’s counties who resided in Chicago at age 65 or older toward the end of the 20th century. There is a striking concentration of migration to Chicago from counties along the *Illinois Central*. Similarly, as shown in Panel B or Figure 3, the *Mobile and Ohio*—an important line with a southern terminus in Mobile and a northern terminus in St. Louis—played a key role facilitating migration to St. Louis by blacks born in eastern Mississippi counties along that rail line.

Because of similarities in migratory patterns, in some of our analyses below we divide our data into three groupings: South Carolina and Georgia, which tended to have migration up the East coast; Alabama and Mississippi, which tended to have migration to the Midwest; and Louisiana, which had concentrated migration to cities in the West.

To get a sense of the typical ages at which migration occurred for African Americans born 1916–1932 in the Deep South, we use Census data from 1920 through 1990 to calculate at each age the following three proportions: (1) those who were still living in their birth state, (2) those who were living outside their birth state but still in the South (i.e., other states of the former Confederacy), and (3) those who were living in the North. Figure 4 plots the proportions who had migrated to the North and to other areas of the South by the indicated age. Like Table 1, Figure 4 shows that total lifetime migration out of the Deep South was extremely high. The figure also shows clear trends in the age of migration: Nearly all of the migration within the South occurred by age 25. Migration to the North is quite low prior to age 18—about the same as migration to other Southern states—but there is a steep

[10]Boustan (2010) discusses migratory patterns and gives references to the extant literature. In early work, Wright (1906) noted the emergence of these migratory streams (as discussed in Trotter, 1991).

[11]In contrast, individuals in our sample who remain in their home States often reside in non-metropolitan areas: 0.45 in South Carolina, 0.42 in Georgia, 0.33 in Alabama, 0.77 in Mississippi, and 0.28 in Louisiana.

escalation in migration to the North between ages 18 and 37. Thereafter we observe only a small increase in the fraction of African Americans living in the North—an indication of low migration rates at older ages.[12]

Migration to the North therefore occurs primarily in the early prime years of labor market participation. Because most African American migrants to the North spent their childhood and adolescent years in the South, they presumably are generally similar to non-migrants with respect to the quality of available schooling and exposure to early-life health conditions.

B. Income and Education among Migrants

There is a large literature that emphasizes the role of migration out of the South as an important means whereby income increased for African Americans by the mid-twentieth century.[13] In Table 2 we provide evidence consistent with this idea. Specifically, using the 1970 Decennial Census, we estimate regressions in which men's earnings are specified to be a function of birth cohort (entered as indicator variables), and an indicator variable for moving North. We use two measures of earnings: wage and salary income and total personal income, both reported in 2010 dollars. For both earnings measures used, and for each of the three groups of States we analyze, black men who migrate have much higher income than those who remain in the South. The last two columns also show that among both men and women, those migrating North had higher levels or schooling than non-migrants.

In sum, then, available evidence establishes three important facts about migration relevant to the analysis of mortality that follows: First, patterns of migration differed quite markedly across States, with much of the migration following the "vertical" paths noted by historians. Railroads played a major role in determining these flows. Second, most migration occurred among the relatively young, though typically after childhood and adolescence. Third, migrating individuals had more education and higher earnings than those who remained behind—facts broadly consistent with the idea that migrants were positively selected.

IV. Migration and Mortality

We turn now to our key empirical analysis of the impact of migration on mortality.

A. Migration and Mortality at Older Ages

We use the Duke SSA/Medicare dataset to construct survival measures among those aged 65 and older, which we compare for migrant and non-migrants, using birth in a railway town as an instrument for migration. As for the location of the railways themselves, a brief history of the expansion of railways in the U.S. is helpful. In general, eminent domain was used to

---

[12]As for return migration at older age, we conduct an analysis using the 2000 Census, which asks the location of residence 5 years earlier (i.e., 1995). Our analysis was restricted to black men and women born in the Deep South who were aged 60 and older in 2000. We estimate transition probabilities conditional on residence in 1995. Among those still living in the Deep South in 1995, virtually all (over 99%) remained in the Deep South in 2000. As for individuals living in the North, 97.4% remain in the North and only 1.9% return to the Deep South, with 0.7% moving elsewhere in the South (e.g., Florida). In short, it appears that mobility rates are quite low at older ages.

[13]Among many contributions are Smith and Welch (1989), Maloney (1994), and Margo (1995). Collins and Wanamaker (2012) provide evidence of positive selection into migration. There are difficulties in interpretation given price differentials across locations (see Black, *et al*., 2013).

allow railroad companies to minimize the cost of connecting large cities and other important transportation hubs. Land was generally acquired in reasonably straight paths between these locations, given geographic restrictions. For example, the primary line of the *Illinois Central* followed a fairly straight path from Cairo, a city located in southern Illinois at the confluence of the Mississippi and Ohio Rivers, though central Mississippi, including the state's largest city, Jackson, and on to New Orleans. In consequence, many African Americans living in central rural Mississippi lived in or near towns on the railway while many others were located in comparable towns far from the rail line.

Using the detailed birthplace information, we classify whether the town of birth was on a railroad line. We do this by overlaying maps of all rail lines at the turn of the twentieth century on a modern GIS map of the U.S. We consider a town to have a rail stop if the longitude and latitude of a town is within two miles of the train lines; we allow for a two mile radius as the longitude and latitude reflects the city center and the train stop could be away from the city center (and there may be small measurement error in the exact path of the rail line).[14]

Table 3 provides our key regression results. Recall that individuals enter our sample at age 65 (with the exception of a small number who enter because of disability, and whom we also include in our sample if they survived to 65). Our outcome measures are survival to age 70 or to age 75. We specify these survival variables be a function of where individuals live in old age— North or South. We also include an indicator variable for each of the 34 gender by cohort cells in the data, and indicator variables for state of birth. Column (1) presents initial baseline OLS results, which show that among individuals who survive to age 65, migration North is not associated with survival to age 70 or to age 75.[15]

For a fairly substantial proportion of our sample we are able to match on state and county of birth, but not on the town of birth (usually because the respondent only supplied the county). To implement our IV strategy we need proximity of birthplace to the railway. So we estimate a second OLS regression, reported in column (2), in which we include only those cases in which we are able to match on the town of birth. The estimated coefficients are extremely close in the two regressions. In the second regression, and in the IV exercise that follows, we use clustered standard errors—clustering on birthplace.

Columns (3) and (4) report results from our IV estimation. The first stage, reported in column (3), shows that close proximity of one's birthplace to a railroad line has a large impact on migration North, increasing migration by approximately 6 percentage points. This finding provides strong empirical evidence (for the first time, as far as we know) backing the widely held view that railroads played a key role in facilitating the Great Migration. Estimated railroad coefficients have t statistics that exceed 40 in our regressions, so marginal F statistics exceed 1600.

---

[14]Additional details are available in an appendix online.

[15]We have access to data through 2002. For the regression in which dependent variable is survival to age 70, we use birth cohorts, 1916–1932. For survival to age 75, we can only use birth cohorts, 1916–1927. Hence sample size is smaller in the second regression.

Coefficients on "Living in the North," reported in column (4), are estimated LATEs of migration. If there is positive selection into migration, we expect this effect to be lower than the OLS estimate, i.e., to be negative given that OLS estimates are close to zero. We find this to be the case. The estimated LATE on survival to age 70 is approximately −0.06, and the estimated LATE on survival to age 75 is −0.10. Given baseline survival probabilities of 0.87 and 0.71, respectively, our estimated effects are substantial. To get a sense of the magnitudes involved, we conducted back-of-the-envelope calculations for the impact on life expectancy at age 65. These show that migration out of the Deep South reduced life expectancy by at least 1.5 years.[16]

We conducted four additional analyses that shed light on our main results.

First, we conducted our analysis for survival to age 70 using two divisions of the data—cohorts born 1916 to 1923 and cohorts born 1924 to 1932. Not surprisingly, proximity to railroad lines has a larger effect on migration in the earlier period than in the later period, when use of bus and automobile transportation became more widespread. But the first stage is highly significant in both cases, and the estimated LATE is virtually identical over the two periods; it is −0.059 (s.e. 0.018) in the earlier cohorts and −0.058 (s.e. 0.028) for the later cohorts.

Second, we tried an alternative specification in which in the first stage the instrument is a continuous measure of the distance of an individual's birthplace from the nearest rail town (the log of one plus this distance, measured in miles). In this case the estimated LATE is −0.060 (s.e. 0.018) for survival to age 70 and −0.108 (s.e. 0.022) for survival to age 75.[17]

Third, we conducted our analysis omitting large cities. Specifically, in our sample there are no non-railroad towns with more than 12,500 observations. In contrast, there are several railroad towns with more than 12,500 observations. So we conduct our analysis for a subsample that includes the smaller cities only (i.e., cities fewer than 12,500 observations).[18] Standard errors increased with this analysis, but again there was virtually no change in the magnitude of the coefficients.

---

[16]Using our data we find that life expectancy at age 65 conditional on *not* surviving to age 75 is 4.8. Life tables from 2001 indicate that life expectancy at age 75 for blacks was approximately 10, so at age 65 life expectancy conditional on surviving to 75 was approximately 20. Given our estimate that the LATE of migration on survival to age 75 is −0.10, around a mean survival rate of approximately 0.70, migration is estimated to reduce the probability of survival to age 75 from approximately 0.75 to 0.65. Let E_N be life expectancy at age 75 for those who move North and E_S be the life expectancy for those remaining in the South. Given all this, the effect of migration on expected extra years (at age 65) is $[0.65\times(E_N+10) + 0.35\times4.8] - [0.75\times(E_S+10) + 0.25\times4.8]$. If for those over age 75 migration has *no* effect on mortality, $E_S=E_N=10$, and we have an estimated impact of *−1.52 expected extra years*. If we were to suppose instead that at age 75 migrants continue to have higher mortality, this number increases in absolute value. For instance, if migrants have life expectancy 1 year less than non-migrants at age 75 (E_N=9.5 compared to E_S=10.5), our calculation is *−2.22 expected extra years*.

[17]As yet another alternative, we tried a cruder location variable—the fraction of people within one's *county* who are born in a railway town—which then allows us to include people for whom we have county of birth but not birth town. As noted above, in baseline regressions we exclude about one quarter of our sample because we lack exact town of birth. When we use a county-based strategy, our sample size increases to 960,552 for the regressions in Panel A and to 673,356 for Panel B. For these samples, OLS estimates are nearly identical to those in column (1); the 2SLS estimate (with standard error) for "survival to age 70" is −0.096 (0.032); and the 2SLS estimate for "survival to age 75" is −0.155 (0.046). Of course this is a different LATE—one generated using an identification strategy that corresponds less well to the theory we presented above. Still, qualitative inferences are similar.

[18]Our sample of individuals born in smaller towns only eliminates those born in New Orleans, LA; Atlanta, GA; Birmingham, AL; Savannah, GA; Shreveport, LA; Charleston, SC; Montgomery, AL; Mobile, AL; Spartanburg, SC; Macon, GA; Columbia, SC; Greenville, SC; Augusta, GA; Jackson, MS; Sumter, SC; Columbus, GA; Baton Rouge, LA; Meridian, MS; Tuscaloosa, AL; and Selma, AL.

Fourth, we repeated our key analyses separately for three state groupings: South Carolina and Georgia; Alabama and Mississippi; and Louisiana. The basic message was qualitatively similar in each case, though standard errors were sufficiently large that not all 2SLS estimates were statistically significant at the 0.05 level.

In Table 4 we analyze men and women separately. In general, survival rates for women are higher than for men, and this is true, we find, in our sample of older African Americans. The estimated effects of migration on survival are negative for both men and women.

Finally, we consider the distinctive predictions, given in (5), that follow from our model of positive selection into migration. We expect, first of all, that among individuals who remain in the South, those born in railway towns (set N, the "never movers") should have lower longevity than individuals born in non-railway towns who would have migrated had they been born in railway towns (set C, the "compliers"), i.e., $E(y_{M=0}|N) < E(y_{M=0}|C)$. We find that this inequality holds for men (0.626 < 0.680) and for women (0.766 < 0.819). Second, among individuals who migrate to the North, those who migrate only because of birthplace was in railway towns (set C, the "compliers") should have lower longevity than those who migrate regardless of birthplace (set A, the "always movers"), i.e., $E(y_{M=1}|C) < E(y_{M=1}|A)$. Again, this inequality holds for both men (0.613 < 0.635) and women (0.711 < 0.774).[19]

B. Identification: Issues with the Exclusion Restriction

The validity of our inferences hinges on the assumption that birth in a railroad town has no direct impact on mortality at older ages. To see the problems that follow if this assumption is violated, suppose for individual *i* born in town *j* we have

$$y_{ij}=\gamma+\beta M_i+\delta_j+\varepsilon_{ij}, \quad (9)$$

where now δ_j is an effect specific to the birth town (normalized to be mean zero). Our key identifying assumption is that $u_{ij} = \delta_j + \varepsilon_{ij}$ is uncorrelated with our instrument, $Z \cup \{0,1\}$, the indicator for whether *j* is a railroad town. If, instead, town-specific effects differ for railway and non-railway towns, our IV estimator is inconsistent if the mean value of δ_j differs for railway and non-railway towns (*i. e.*,$\bar{\delta}_{z=1} - \bar{\delta}_{z=0} \neq 0$).

A potential violation of the exclusion restriction is that the arrival of a railroad brings prosperity to a town. A large literature in economics examines the role of railways for spurring local economic development.[20] In turn, this prosperity might bring better nutrition and child healthcare, thereby having a positive impact on older-age longevity.[21] If so, $\bar{\delta}_{z=1}$ −

[19]Wald estimators (see (6) above) can be calculated directly from these means; they are 0.613 – 0.680 = −0.067 for men, and 0.711 – 0.819 = −0.108 for women. These differ slightly from the estimates presented in Table 4 because those estimates are from regressions that also include cohort and State indicator variables.

[20]To give two recent examples, Atack and Margo (2011) show that the arrival of rail transportation increased farmland value in the American Midwest in the mid nineteenth century, and Donaldson's (forthcoming) work suggests that railroads increased local real income levels in nineteenth and early twentieth century India.

[21]The idea that early-life economic conditions play a crucial role for shaping long-run health has been developed in an important literature, e.g., Barker (1990), and Fogel (2004). Importantly, for our study, Preston, Hill, and Drevenstedt (1998) provide evidence that among African Americans, an unhealthy childhood environment was associated with a reduced probability of survival at every age, up to age 85.

$\bar{\delta}_{z=0} > 0$, and in this case it is easily shown that our IV estimator forms an upper bound for β; we understate the detrimental impact of migration.

One way to provide evidence on this matter is to notice that while lifetime migration out of the South among our cohorts was very high for African Americans (approximately 43 percent), it was quite low for corresponding whites (less than 9 percent). Thus, while the older black population living in the South is a highly select group, the older white population living in the South is not. With that in mind, consider regressions for *whites* born in the Deep South, 1916–1932, in which a survival to age 70 (conditional on survival to 65) is taken to be a function of cohort by gender fixed effects, state effects, and an indicator variable for birth in a railroad town. Using 1,530,557 observations we estimate a very small positive "railroad town effect" of 0.0012, with a clustered s.e. of 0.00068 (statistically significant at the 0.10 level).[22] If the beneficial impact of railway-town birth is as large for blacks as for whites, then we can show that the LATE of migration for survival to age 70 decreases from −0.058 to −0.079.[23]

Of course, local value created by railroads may have excluded much of the black population (Margo, 1984). For example, even where rail transportation increased value to local farmland, incomes of black agricultural workers may have been largely unchanged.[24] Moreover, any long-term health benefits that accompanied increased local income were likely attenuated for black families due to discrimination in healthcare (see, e.g., Almond, Chay, and Greenstone, 2006, and Jayachandran, Lleras-Muney, and Smith, 2010). Given these factors, birth in a railway town may have had a smaller beneficial health impact for Southern-born blacks than whites.

While we do not have evidence that directly compares contemporaneous health outcomes of individuals born in Southern railroad towns and non-railroad towns in the early twentieth century, Table 5 does provide evidence for a limited number of socio-economic outcomes at the county level that are plausibly correlated with health. Here we first construct county-level measures of the socio-economic variables from the U.S. Census. Then we use these as dependent variables in regressions that include as an explanatory variable the fraction of people within the county who were born in railway towns (calculated from the SSA/ Medicare data). Importantly, we notice that in 1910—prior to the start of the migration that we study—literacy and school attendance in the black population were moderately higher in counties with high railway concentration. The same is true, but to a lesser extent, in 1920 and 1930. For interest sake, we provide comparable statistics for whites.

In sum, our examination of existing evidence gives us little reason to believe that in the early twentieth century South railroad towns likely generated particularly poor health outcomes for either black or white children.

---

[22]We tried this exercise also for survival to age 75, estimated with 1,052,479 observations. Here the "railroad town effect" is similarly very small, 0.0018, with a clustered s.e. of 0.0013 (not statistically significant).

[23]This follows from observing that in model (9), $\text{plim}(b_{IV}) = \beta + \text{plim}\{(\bar{\delta}_{z=1} - \bar{\delta}_{z=0})/(\bar{M}_{z=1} - \bar{M}_{z=0})\}$. The denominator of the adjustment term is calculated using statistics from column (3) of Table 3.

[24]Many black agricultural workers in the South were in sharecropping arrangements. As is clear from discussions such as Braverman and Stiglitz (1982), an increase in farmgate crop prices (surely plausible when rail transportation arrives) needn't increase sharecrop income.

C. Migration and Mortality at Younger Ages

To this point our analysis focuses on mortality post age 65. Ideally we would like to conduct similarly detailed evaluations of mortality at ages younger than age 65. While this is not possible with available data, we can undertake an alternative approach, adapting Heckman and Robb's (1985) repeated cross-section research design. We apply this idea using *mortality* rates or *log mortality* rates rather than survival rates as above. Our regression specifies the mortality rate, measured at the cohort (*c*) by birth state (*s*) level, as a function of cohort and birth state effects and the proportion of the state-cohort group that migrated North, say P_{cs},

$$m_{CS}=\mu_S+\mu_C+\gamma P_{CS}+\varepsilon_{CS}. \quad (12)$$

Suppose there is year-to-year cross-state variation in migration as a consequence of shifts in the economic returns to remaining in one's home state and/or moving North. Then marginal changes in P_{cs} will affect observed state-cohort mortality if moving North has a causal impact on migrating individuals. Indeed, $\hat{\gamma}$ estimates a LATE, assessing the impact of migration on mortality among "marginal individuals"—conceptually those who migrate in circumstances that generate relatively high cohort-state migration rates but not in circumstances that result in marginally lower migration rates.

To illustrate the general idea, we start by applying this method to our SSA/Medicare data. We construct 85 cells (for 17 cohorts by 5 states) and for each calculate mortality from age 65 to 70. Given results reported in Panel A of Table 4, we might expect our estimates of γ to be in the neighborhood of 0.07 for men and 0.05 for women (i.e., an increase in P_{cs} from 0 to 1 would increase cohort-level mortality by 0.07 for men and 0.05 for women). Our exercise is woefully under-powered. Still, results reported in Table 6 are somewhat encouraging; we estimate positive coefficients, and in the case of men the estimates actually line up quite well.

Next we try this same strategy for individuals younger than aged 65 and younger, using data from 1960. Here we follow standard practice from the demography literature by using *log* morality rates, which is probably more appropriate given that we are evaluating mortality over a wide range of ages. We conduct our analysis for 1960 because in that year we can construct base population estimates using Census data, which have age, gender, race, state of current residence, and birth state. Death counts come from the Detailed Mortality File of the Vital Statistics from 1960 and 1961, which also record these data elements.[25] We construct 12-month log mortality rates for the period April 1960 through March 1961, for blacks born in the Deep South, for cohort × gender × state of birth cells, from birth cohorts 1885 through 1940, i.e., for individuals aged 20 to 65 in 1960.

Results, reported in Panel B of Table 6, are consistent with our analysis for older ages; migration increases mortality among African Americans. Taken at face value, our estimates suggest that migrating North increases age-specific mortality rates by about half for women and somewhat less than half for men. These inferences are qualitatively similar to those

[25]Unfortunately, the Vital Statistics death data do not record state of birth for other years of interest to us, e.g., 1950 and 1970.

shown in Panel A of Table 4.[26] Having noted that, we see the identification strategy used for analysis of mortality at older ages to be quite compelling, while the identification for impacts at younger ages requires assumptions that are possibly less tenable (see Heckman and Robb, 1985), leading us to view that evidence as suggestive in nature.

D. Potential Causal Mechanisms

There are a number of plausible, possibly overlapping, pathways whereby migration may have been damaging to health of African Americans. These might include, for example, increased stress due to dislocation from families and communities, higher exposure to environmental hazards or poorer sanitary conditions, and changes in behaviors related to smoking, alcohol consumption, and diet. Unfortunately, there is little previous research that evaluates the health consequences of migration. One exception is Gibson, *et al*. (forthcoming). In a study that compares successful and unsuccessful Tongan applicants to a New Zealand migration lottery, these authors find that migration increased individuals' blood pressure and hypertension, and they provide evidence that these effects may have stemmed from higher stress levels or increased dietary sodium.[27]

While we do not have health records or medical expenditures in our SSA/Medicare data, we can make headway by appealing to other data sources. We focus on two behavioral factors: smoking and alcohol use. Smoking increases mortality due to many disease processes, including chronic obstructive pulmonary disease (COPD), stroke, heart disease, and various forms of cancer. For example, among Americans aged 55 and older, the risk of death due to COPD is more than 20 times higher for smokers than for those who have never smoked (Thun, *et al*., 2013). As for alcohol consumption, heavy drinking contributes to poor health via a wide range of diseases, and is by far the most important factor in mortality due to cirrhosis of the liver (Rehm, *et al*., 2010).

Evidence in Table 7 provides reason to believe that some of the increased mortality due to migration may be related to increased smoking and drinking. Panel A shows rates of self-reported smoking and alcohol use from the 1984 though 2010 waves of the Behavioral Risk Factor Surveillance System (BRFSS) for black men and women born in the Deep South, 1916–1932. Notice that smoking is much higher among men than among women, and drinking is moderately higher among men than women. More important, from our perspective, among both men and women, smoking and alcohol use is lowest among non-migrants (those in their birth State at older age), higher among those who migrated elsewhere in the South, and higher yet for those who migrated North. These behavioral patterns likely contributed to outcomes shown in Panel B. According to indications on death certificates, COPD is much higher among men than women, and cirrhosis is moderately higher among men than women. Our key observation is that for both genders, COPD and cirrhosis are more common among migrants than non-migrants.

[26]For instance, we estimate that from age 65 to 70, the effect of migration on women's mortality is 0.05, from a base of 0.10, and the effect on men's mortality is 0.07, from a base of 0.18.

[27]Gibson, *et al*. (forthcoming) do *not* argue, though, that migration therefore reduces overall well-being among Tongans, because migration increases income (McKenzie, Gibson, and Stillman, 2010), and may also have contributed to improved mental health (Stillman, McKenzie, and Gibson, 2009).

Further evidence about differences in the cause of mortality between Southern-born African American migrants and non-migrants (birth cohorts 1916 through 1932) is provided in Table 8. Here we study on the *primary* cause of death, focusing on the 16 most commonly listed causes.[28] For each leading cause of death we construct a dissimilarity index equal to the ratio of two constructs: (i) the proportion of deaths among individuals in the North for which that cause is the primary cause of death, and (ii) the corresponding proportion for individuals who remained in the South. We notice that among both men and women there is a striking over-representation in the North of deaths with the primary cause, chronic liver disease and cirrhosis—a drinking-rated condition. As for prominent smoking-related conditions, we also observe that cancer of respiratory and intrathoracic organs is significantly higher in the North among both men and women, while COPD as is over-represented in the North for women but not men.[29] While a complete analysis of competing risks in this context is beyond the scope of our current paper, we view the evidence as consistent with the possibility that increased drinking and smoking among migrants to the North contributed to differential disease processes from non-migrants.

V. Conclusion

At the very outset of the Great Migration, Wright (1906) noted a commonly-held concern that migration to the urban North held health risks for Southern-born African Americans (he cites risks from tuberculosis and pneumonia, in particular).[30] But he expressed an optimistic view of “many positive evidences of a healthful effect” of migration. By the mid-twentieth century, Myrdal (1948) similarly suggested that migration out of the South might lead to improved health among African Americans. Our paper uses an extensive assemblage of data collected over the ensuing decades to evaluate this issue—providing, we believe, the first attempt to establish a link between the Great Migration and mortality. We find support for the “healthy migrant hypothesis”—the positive selection of individuals into migration. We also find that, contrary to hopes expressed by Wright and Myrdal, migration out the South *reduced* longevity.

Given the central role played by the Great Migration in shaping social and economic advancement of African Americans in the twentieth century, our inference about its adverse impact on mortality might be surprising. However, while the Great Migration was surely a means for improving of economic opportunities among African Americans—resulting in higher wages and better job prospects among migrants, as documented by Smith and Welch

[28]Importantly, an individual can have COPD or cirrhosis listed as a present condition (as analyzed in Table 7), but have some other more-common primary cause of death, such as cardiovascular disease or cancer of respiratory and intrathoracic organs.

[29]For each cause of death we formed 95 percent confidence intervals, using a bootstrap procedure (with 1000 replications). For chronic liver disease and cirrhosis, the index confidence intervals were 1.38–1.65 for men and 1.27–1.56 for women. For cancer of respiratory and intrathoracic organs, the corresponding index confidence intervals were 1.03–1.08 for men and 1.29–1.38 for women.

[30]Along these same lines, Higgs (1977) points to a U.S. Census (1918) report in which Census officials noted that most of the black population of the time lived outside the largely urban “registration area” for which mortality data were available, but then speculated that “it is highly probable that mortality is much lower in this rural element than it is in the population of the registration area, which is largely urban and largely a migrant population.” These demographers express a concern that black migrants from the South were “subjected to conditions similar in some respects to those encountered by the foreign immigrant, and the difficulties of adjustment to these conditions may be reflected in the higher mortalities from such causes as tuberculosis and pneumonia” (p. 314). Cutler and Miller (2005) show that from 1900 to 1936 there was a sharp drop in percentage of deaths due to infectious diseases in major U.S. cities, and argue more generally that the “urban mortality penalty” largely disappeared during this period, thanks to clean water technologies.

(1989), Margo (1995), and Maloney (1995), among others—the economic and historical literature also emphasizes that African Americans often faced daunting circumstances in the North, including high costs in discriminatory housing markets and uneven employment prospects. Real economic gains to moving North may have been modest or non-existent for many African Americans (Eichenlaub, Tolnay, and Alexander, 2010),[31] thus attenuating improved health prospects associated with increasing prosperity. In any event, any beneficial health benefits due to economic and social improvement were apparently swamped by other forces, such as changes in behavioral patterns that were detrimental to long-term health, including higher propensities to smoke and consume alcohol. Any overall welfare evaluation centering on migration would be complex.

Our findings suggest a new layer of complication for the vast literature that evaluates links between health and location, education, income, and race, in the United States. We mention two examples: First, our work provides insight into one mechanism that generates local variation in mortality of the sort described in Geronimus, Bound, and Colen's (2011) study of excess mortality in high-poverty pockets in the U.S. Given the selective nature of migration we find, high-mortality high-poverty localities can be engendered by high levels of selective outmigration. Migration patterns might thus be part of the explanation for the low life expectancy found in many counties in the rural South (documented in Kulkarni, *et al*., 2011).

Second, our results indicate why it is very difficult to sort out the impact of improved educational opportunities among African Americans for lifetime health. As discussed in Cutler and Lleras-Muney (2010), there are many good reasons to believe that better education should lead to improved health. But among black children in the early twentieth-century South, as Aaronson and Mazumder (2011) show, better education led to an increased propensity to migrate to the North. Our work here, in turn, show that this migration led to higher mortality. In on-going research, we are hoping to sort through these complex issues.

Finally, our research may be relevant for evaluating current trends in developing countries, many of which are experiencing extraordinary levels of migration from rural areas to urban centers. Our assessment of the long-term consequences of the Great Migration suggests that dislocation due to migration might have substantial costs in terms of individual health.

Acknowledgments

We are gratefully acknowledge support from Eunice Kennedy Shriver National Institute of Child Health and Human Development (RO1 HD062747). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development or the National Institutes of Health. We are grateful also for support from the University of Kentucky Poverty Research Center. We have benefited from excellent research assistance from Ryan Brown, and from comments by seminar participants at Peking University, UC-Berkeley, UCLA, UC-Santa Barbara, Carnegie Mellon, Stanford, Northwestern, Chicago, Kentucky, Amsterdam, the Tindenberg Institute, Wisconsin, Atlanta Federal Reserve, Chicago Federal Reserve, and conferences sponsored by NBER and IZA. Any errors are our own.

[31] Black, *et al*. (2013) express concerns about comparing earnings across local markets with differing prices. Also, Foote, Whatley, and Wright (2003) note that even in jobs in which in whites and blacks were similarly paid in the North, black workers were sometimes disproportionately assigned particularly unpleasant and dangerous work.

References

Alston, Lee J.; Ferrie, Joseph P. Paternalism in Agricultural Labor Contracts in the U.S. South: Implications for the Growth of the Welfare State. American Economic Review. 1993; 83(40):852–876.

Aaronson, Daniel; Mazumder, Bhashkar. The Impact of Rosenwald Schools on Black Achievement. Journal of Political Economy. 2011; 119(5):821–888.

Almond, Douglas; Chay, Kenneth; Greenstone, Michael. Civil Rights, the War on Poverty, and Black-White Convergence in Infant Mortality. Massachusetts Institute of Technology; 2006. manuscript

Angrist, Joshua; Imbens, Guido W.; Rubins, Donald B. Identification of Causal Effects using Instrumental Variables. Journal of the American Statistical Assoc. 1996; 91:444–472.

Atack, Jeremy; Margo, Robert A. The Impact of Access to Rail Transportation on Agricultural Improvement. Journal of Transport and Land Use. 2011; 4(2):5–18.

Barker DJP. The Fetal and Infant Origins of Adult Disease. British Medical Journal. 1990

Black, Dan A.; Hsu, Yu-Chieh; Taylor, Lowell J. The Effect of Early-Life Education on Later-Life Mortality. Carnegie Mellon University; 2013. manuscript

Black, Dan A.; Kolesnikova, Natalia; Sanders, Seth; Taylor, Lowell. Racial Wage Disparity when Wages and Prices Differ Across Location. Carnegie Mellon; 2013. manuscript

Boustan, Leah Platt. Competition in the Promised Land: Black Migration and Racial Wage Convergence in the North, 1940–1970. Journal of Economic History. 2009; 69(3):756–783.

Boustan, Leah Platt. Was Postwar Subsurbanization 'White Flight'? Evidence from the Black Migration. Quarterly Journal of Economics. 2010:417–443.

Boustan, Leah Platt; Fishback, Price V.; Kantor, Shawn. The Effect of Internal Migration on Local Labor Markets: American Cities during the Great Depression. Journal of Labor Economics. 2010; 28(4):719–745.

Braverman, Avishay; Stiglitz, Joseph E. Sharecropping and the Interlinking of Agrarian Markets. American Economic Review. 1982; 72(4):695–715.

Carrington, William; Detragiache, Enrica; Vishwanath, Tara. Migration with Endogenous Moving Costs. American Economic Review. 1996; 86(4):909–930.

Collins, William J. When the Tide Turned: Immigration and the Delay of the Great Black Migration. Journal of Economic History. 1997; 57:607–632.

Collins, William; Wanamaker, Marianne. Selection and Income Gains in the Great Migration of African Americans: New Evidence from Linked Census Data. manuscript.

Cutler, David; Lleras-Muney, Adriana. Understanding Differences in Health Behaviors by Education. Journal of Health Economics. 2010; 29(1):1–28. [PubMed: 19963292]

Cutler, David; Miller, Grant. The Role of Public Health Improvements in Health Advances: The Twentieth Century United States. Demography. 2005; 42(1):1–22. [PubMed: 15782893]

Donaldson, Dave. Massachusetts Institute of Technology; 2012. Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. manuscript

Eichenlaub, Suzanne; Tolnay, Stuart; Alexander, JTrent. Moving Out But Not Up: Economic Outcomes in the Great Migration. American Sociological Review. 2010; 75(1):101–125.

Farr, W. Supplement to the 25th Annual Report of the Registrar General. London: HMSO; 1864.

Fogel, Robert W. The Escape from Hunger and Premature Death, 1700–2100. Cambridge: Cambridge University Press; 2004.

Foote, Christopher L.; Whatley, Warren C.; Wright, Gavin. Arbitraging a Discriminatory Labor Market: Black Workers at the Ford Motor Company, 1918–1947. Journal of Labor Economics. 2003; 21(3):493–532.

Geronimus, Arline T.; Bound, John; Colen, Cynthia G. Excess Black Mortality in the United States and in Selected Black and White High-Poverty Areas, 1980–2000. American Journal of Public Health. 2011; 101(4):720–729. [PubMed: 21389293]

Gibson, John; Stillman, S.; McKenzie, D.; Rohoura, H. Natural Experiment Evidence on the Effect of Migration on Blood Pressure and Hypertension. Health Economics. forthcoming

Halliday, Timothy; Kimmitt, Michael. Selective Migration and Health in the USA, 1984-93. Population Studies. 2008; 62(3):321–334. [PubMed: 18937145]

Heckman, James J.; Robb, Richard. Alternative Methods of Evaluating the Impact of Interventions: An Overview. Journal of Econometrics. 1985; 30:239–267.

Higgs. Competition and Coercion: Blacks in the American Economy 1865–1914. Cambridge: Cambridge University Press; 1977.

Imbens, Guido W.; Angrist, Joshua. Identification and the Local Average Treatment Effects. Econometrica. 1994; 62(2):467–476.

Jayachandran, Seema; Lleras-Muney, Adriana; Smith, Kimberly V. Modern Medicine and the Twentieth Century Decline in Mortality: Evidence on the Impact of Sulfa Drugs. American Economic Journal: Applied Economics. 2010; 2:118–146.

Kulkarni, Sandeep C., et al. Falling Behind: Life Expectancy in U.S. Counties from 2000 to 2007 in an International Context. Population Health Metrics. 2011; 9(16)

Lange, Fabian; Olmstead, Alan L.; Rhode, Paul W. The Impact of the Boll Weevil, 1892–1932. Journal of Economic History. 2009; 69(3):685–718.

McCord, Colin; Freeman, Harold P. Excess Mortality in Harlem. New England Journal of Medicine. 1990; 322(3):173–177. [PubMed: 2294438]

McKenzie, David; Gibson, John; Stillman, Steven. How Important is Selection? Experimental vs. Non-Experimental Measures of Income Gains from Migration. Journal of the European Economic Association. 2010; 8(4):913–945.

Margo, Robert A. Accumulation of Property by Southern Blacks Before World War I: Comment and Further Evidence. American Economic Review. 1984; 74(4):768–776.

Margo, Robert A. Explaining Black-White Wage Convergence, 1940–1950. Industrial and Labor Relations Review. 1995; 48(3):470–481.

Maloney, Thomas N. Wage Compression and Wage Inequality Between Black and White Males in the United States. Journal of Economic History. 1994; 54(2):358–381.

Myrdal, Gunnar. Social Trends in American and Strategic Approaches to the Negro Problem. Phylon. 1948; 9(3):196–214.

Norman P, Boyle P, Rees P. Selective Migration, Health and Deprivation: A Longitudinal Analysis. Social Science and Medicine. 2005; 60(12):2755–2771. [PubMed: 15820585]

Preston, Samuel H. The Changing Relation between Mortality and Level of Economic Development. Population Studies. 1975; 29(2):231–248. [PubMed: 11630494]

Preston SH, Hill M, Drevenstedt G. Childhood Conditions that Predict Survival to Advanced Ages Among African Americans. Social Science Medicine. 1998; 47(9):1231–1246. [PubMed: 9783866]

Rehm, Jürgen, et al. The Relation Between Different Dimensions of Alcohol Consumption and Burden of Disease – An Overview. Addiction. 2010; 105(5):817–843. [PubMed: 20331573]

Smith, James P.; Welch, Finis. Black Economic Progress After Myrdal. Journal of Economic Literature. 1989; 27(2):519–564.

Stillman, Steven; McKenzie, David; Gibson, John. Migration and Mental Health: Evidence from a Natural Experiment. Journal of Health Economics. 2009; 28(3):677–687. [PubMed: 19349087]

Thun, Michael J.; Carter, Brian D.; Feskanich, Diane; Freedman, Neal D.; Prentice, Ross; Lopez, Alan D.; Hartge, Patricia; Gapstur, Susan M. 50-Year Trends in Smoking-Related Mortality in the United States. The New England Journal of Medicine. 2013; 368(4):351–364. [PubMed: 23343064]

Trotter, Joe W. Black Migration in Historical Perspective: A Review of the Literature. In: Trotter, editor. The Great Migration in Historical Perspective. Bloomington: In. Univ. Press; 1991.

U.S. Bureau of the Census. Negro Population. 1918:1790–1915.

Wright, Richard R. The Migration of Negroes to the North. Annals of the American Academy of Political and Social Science. 1906; 27:97–116.

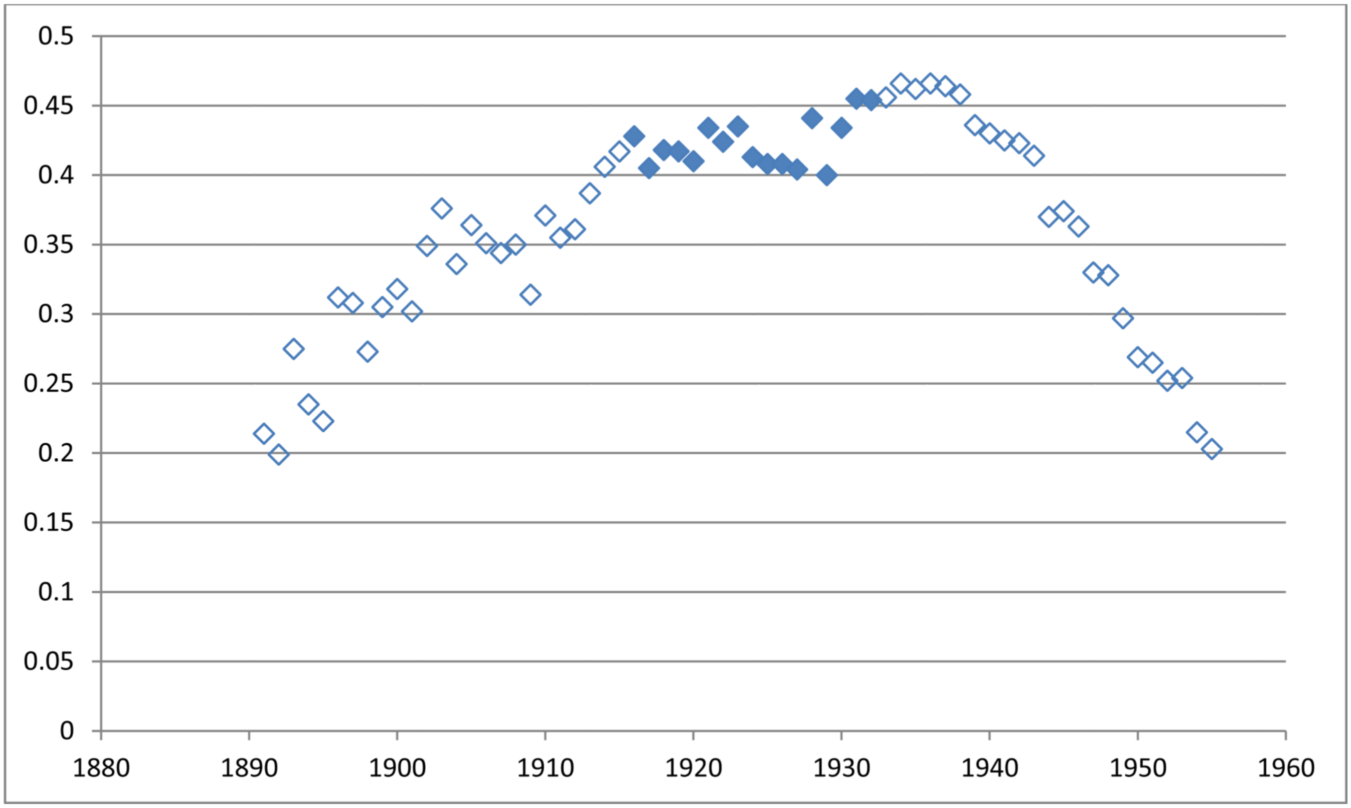


Figure 1. Proportion of African Americans Migrating out of the Deep South as of their 40s, Birth Cohorts 1891–1955

Source: Authors' calculations using 1940–2000 U.S. Census data for black men and women born in the Deep South. For birth cohorts 1891–1899, we measure those living in the North in 1940; for birth cohorts 1900–1909 we measure those living in the North in 1950; and so forth. Thus we record migration as of ages approximately 40–49. The dark-colored data points are for our primary cohorts of study, 1916–1932.

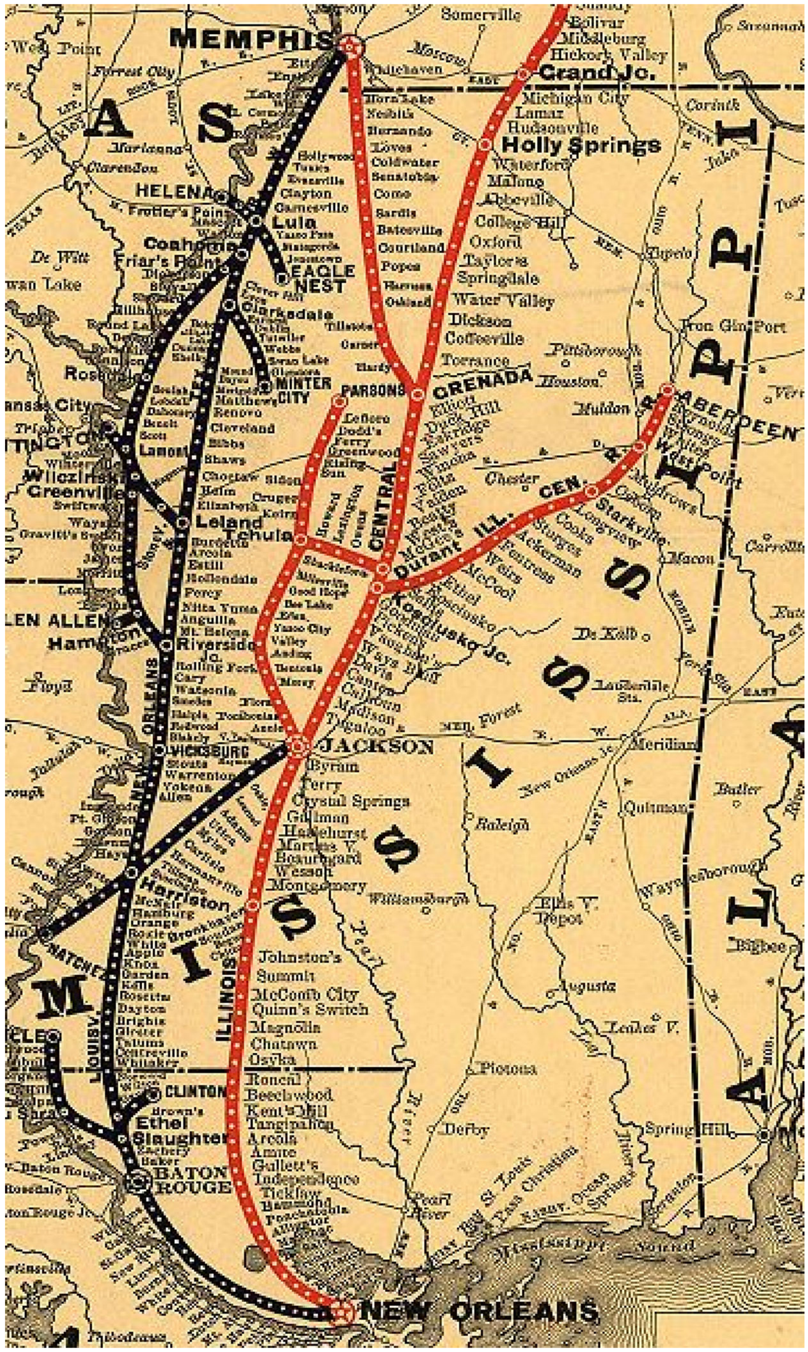
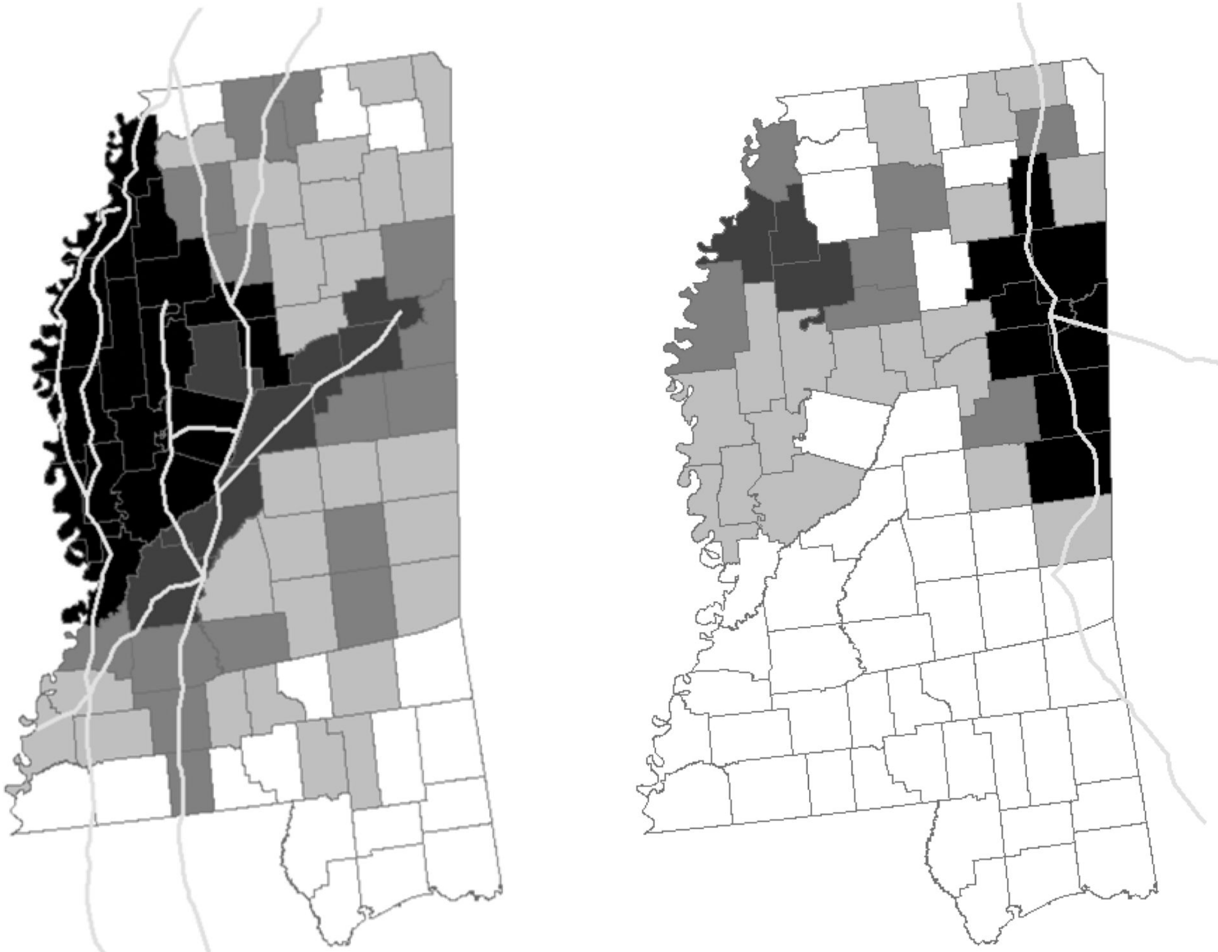


Figure 2. Illinois Central Railroad Stops in Mississippi, 1892

Source: Rand McNally & Co., Engravers, Chicago. Library of Congress Call Number G4041.P3 1892 .R3 RR431.



Panel A. Proportion of Black Population in the County Migrating to Chicago

Panel B. Proportion of Black Population in the County Migrating to St. Louis

Figure 3. Migration of Blacks from Mississippi Counties along the *Illinois Central* to Chicago and from Mississippi Counties along the *Mobile and Ohio* to St. Louis

Source: Authors' calculations from the Duke-SSA data, birth cohorts 1916 through 1932. In Panel A the proportion of the black population in a county migrating to Chicago is indicated by shading, from light to dark: <0.10, 10–0.14, 0.14–0.18, 0.18–0.22, and >0.22. Also in Panel A the *Illinois Central* line is highlighted. In Panel B the proportion of the black population in a county migrating to St. Louis is indicated by shading, from light to dark: <0.04, 0.04–0.06, 0.06–0.08, 0.08–0.10, and >0.10. Also in Panel B the *Mobile and Ohio* line is highlighted.

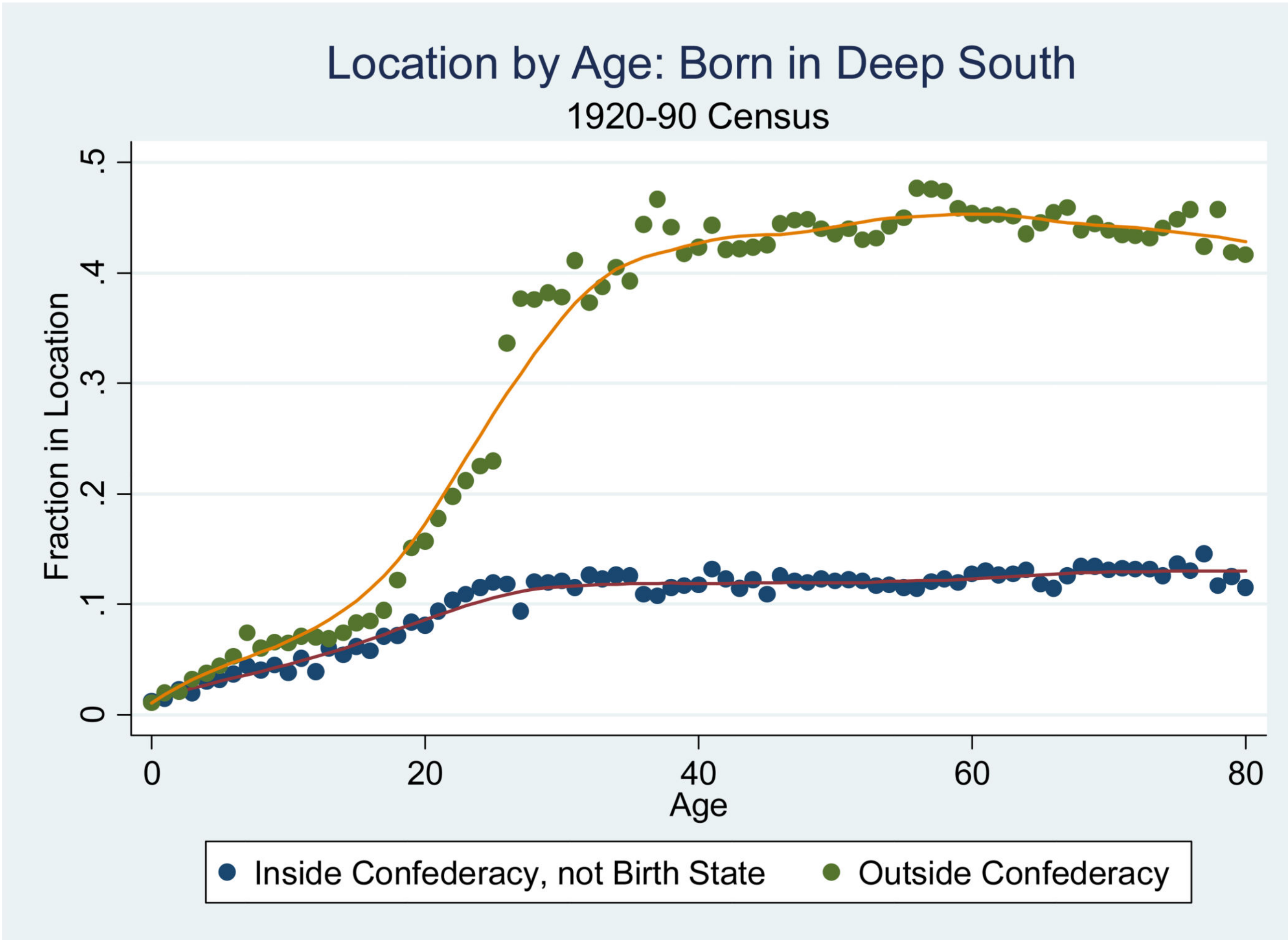


Figure 4. Migration within the South and to the North among African Americans Born 1916–1932

Source: Authors' calculations for blacks in the 1920–1990 Decennial Census born in the Deep South, 1916 through 1932.

Table 1

State of Residence in Adulthood, African Americans Born in the Deep South, 1916–1932

Born in South Carolina	**Proportion**	
	Duke Data	**Census Data**
Reside in South Carolina	0.42	0.43
Reside in rest of South	0.15	0.11
Reside in North	0.43	0.46
Conditional on North, proportion residing in:		
New York City	0.41	0.48
Washington	0.19	0.10
Philadelphia	0.17	0.12
Non-metro area	0.0095	---
Born in Alabama	Proportion	
	Duke Data	Census Data
Reside in Alabama	0.42	0.45
Reside in rest of South	0.13	0.09
Reside in North	0.45	0.45
Conditional on North, proportion residing in:		
Detroit	0.19	0.20
Chicago	0.14	0.15
Cleveland	0.12	0.11
Non-metro area	0.018	---
Born in Louisiana	Proportion	
	Duke Data	Census Data
Reside in Louisiana	0.53	0.59
Reside in rest of South	0.15	0.12
Reside in North	0.32	0.29
Conditional on North, proportion residing in:		
Los Angeles	0.30	0.27
San Francisco	0.19	0.21
Chicago	0.11	0.10
Non-metro area	0.016	---
Born in Georgia	Proportion	

Born in South Carolina	**Proportion**	
	Duke Data	**Census Data**
	Duke Data	Census Data
Reside in Georgia	0.46	0.49
Reside in rest of South	0.19	0.16
Reside in North	0.35	0.35
Conditional on North, proportion residing in:		
New York City	0.23	0.22
Detroit	0.15	0.16
Philadelphia	0.11	0.09
Non-metro area	0.014	---
Born in Mississippi	Proportion	
	Duke Data	Census Data
Reside in Mississippi	0.32	0.37
Reside in rest of South	0.15	0.14
Reside in North	0.53	0.50
Conditional on North, proportion residing in:		
Chicago	0.36	0.34
Detroit	0.11	0.12
St. Louis	0.10	0.10
Non-metro area	0.025	---

Source: Authors' calculations, Duke SSA/Medicare data and 1970 Census, birth cohorts 1916 through 1932 inclusive. We list destination cities that have a proportion of 0.10 or more. In the SSA/Medicare data we have residence at ages 65 and older. (For Census estimates of cities, we use the proportion of urban residents whose metropolitan area is identified.)

Table 2

Earnings (in 2010 Dollars) and Education by Residence in 1970, African Americans Born in the Deep South, 1916–1932

Born in Georgia or South Carolina	**Men's wage and earnings**	**Men's total personal income**	**Men's education**	**Women's education**
Mean for Individuals Residing in the South	26,684	29,675	6.96	7.98
Coefficient on "Residing in the North"	18,214*** (758)	20,150*** (795)	2.05*** (0.11)	1.41*** (0.09)
N	5,084	5,084	5,084	6,208
Born in Alabama or Mississippi	Men's wage and earnings	Men's total personal income	Men's education	Women's education
Mean for Individuals Residing in the South	25,806	29,123	7.22	8.20
Coefficient on "Residing in the North"	20,988*** (762)	22,224*** (841)	1.80*** (0.10)	1.47*** (0.09)
N	5,023	5,023	5,023	6,142
Born in Louisiana	Men's wage and earnings	Men's total personal income	Men's education	Women's education
Mean for Individuals Residing in the South	28,359	31,502	7.24	8.19
Coefficient on "Residing in the North"	17,931*** (1,304)	19,545*** (1,298)	2.19*** (0.18)	1.98*** (0.16)
N	2,143	2,143	2,143	2,513

Source: Authors' calculations, 1970 PUMS, state sample, black men and women born 1916 through 1932 inclusive. Earnings and income are for 1969, but are adjusted 2010 using the CPI. Standard errors are in parentheses

*** significant at the 0.01 level.

Table 3

Impact of Living in the North on Survival to Age 70 or Age 75 Conditional on Survival to Age 65, African Americans Born in the Deep South, 1916–1932

A. Survival to Age 70	**(1) OLS**	**(2) OLS**	**(3) IV Estimation**	
	(Full Sample)	**(Town-Matched)**	**1st Stage**	**2nd Stage**
Mean of the Dependent Variable	0.864	0.866	0.43	0.866
Living in the North	0.0005 (0.0007)	0.0015 (0.0010)	--	-0.058^{***} (0.017)
Born on Railroad Line	--	--	0.056^{***} (0.0013)	--
N	1,077,296	828,179	828,179	828,179

B. Survival to Age 75	(1) OLS	(2) OLS	IV Estimation	
	(Full Sample)	(Town-Matched)	1st Stage	2nd Stage
Mean of the Dependent Variable	0.704	0.708	0.43	0.708
Living in the North	0.0006 (0.0011)	0.0014 (0.0017)	--	-0.102^{***} (0.026)
Born on Railroad Line	--	--	0.062^{***} (0.0015)	--
N	757,790	575,379	575,379	575,379

Source: Authors' calculations using Duke SSA/Medicare data. The sample is African Americans born in South Carolina, Georgia, Alabama, Mississippi, and Louisiana, 1916–1932. The regressions also include gender × cohort indicator variables and state of birth indicator variables. Standard errors, given in parentheses, are clustered by birthplace for regressions (2), (3) and (4)

*** significant at the 0.01 level.

Table 4

Impact of Living in the North on Survival to Age 70 or Age 75 Conditional on Survival to Age 65, Men and Women Born in the Deep South

A. Survival to Age 70	Men			Women		
		(2) IV Estimation			(4) IV Estimation	
	(1) OLS	1st Stage	2nd Stage	(3) OLS	1st Stage	2nd Stage
Mean of the Dep. Variable	0.825	0.44	0.825	0.898	0.42	0.898
Live in North	0.0033** (0.0016)		−0.071** (0.030)	0.0001 (0.0010)		−0.048*** (0.018)
Born on Railroad Line		0.050*** (0.0019)			0.060*** (0.0017)	
N	364,988	364,988	364,988	463,192	463,192	463,192
A. Survival to Age 75	*Men*			*Women*		
		(2) IV Estimation			(4) IV Estimation	
	(1) OLS	1st Stage	2nd Stage	(3) OLS	1st Stage	2nd Stage
Mean of the Dep. Variable	0.631	0.44	0.631	0.768	0.42	0.768
Live in North	0.0050** (0.0020)		−0.070* (0.038)	−0.0016 (0.0021)		−0.123*** (0.030)
Born on Railroad Line		0.057*** (0.0023)			0.067*** (0.0020)	
N	253,254	253,254	253,254	322,125	322,125	322,125

Source: Authors' calculations using Duke SSA/Medicare data. The sample is African Americans born in 1916–1932 for whom town matching was possible. The regressions also include cohort indicator variables and state of birth indicator variables. Standard errors, given in parentheses, are clustered by birthplace

*** significance at the 0.01 level

** significance at the 0.05 level; and

* significance at the 0.10 level. Marginal *F* statistics for the first stage in 2SLS (not shown) are in every case very high (greater than 614).

Table 5

Relationship between County Level Socio-Economic Outcomes and Railroad Concentration in the Deep South

	Literacy Rate		*School Attendance among Children Aged 6–12*		*School Attendance among Children Aged 13–18*	
Year	Black	White	Black	White	Black	White
1910	0.056*** (0.019)	0.051*** (0.012)	0.060* (0.031)	0.033 (0.023)	0.073** (0.033)	−0.003 (0.029)
1920	0.039* (0.021)	0.022** (0.010)	0.048 (0.032)	0.053*** (0.018)	0.067** (0.034)	0.001 (0.026)
1930	0.036* (0.019)	0.012 (0.008)	0.002 (0.030)	0.034** (0.016)	0.049 (0.034)	0.017 (0.025)

Source: Authors' calculations: dependent variables are calculated using the U.S. Census IPUMS and railroad concentration is calculated using Duke SSA/Medicare data as described in footnote 14. These are county-level regressions for the Deep South (including all counties for which data elements are available; $n = 368$ to $n = 418$), and include State fixed effects. Standard errors are in parentheses

*** significance at the 0.01 level

** significance at the 0.05 level; and

* significance at the 0.10 level.

Table 6

The Relationship between Mortality and Proportion of Cohort Migrating North, African Americans Born in the Deep South

A. Mortality Rate, Ages 65–70, Duke/SSA Data	**Men**	**Women**
Proportion of Cohort in North (at Last Residence)	0.059* (0.033)	0.016 (0.023)
N	85	85
B. Log Annual Morality Rate in 1960, Calculated with Census and Vital Statistics Data		
Proportion of Cohort in North (in 1960)	0.54** (0.22)	0.66* (0.35)
N	230	230

Source for Row A: Authors' calculations, Duke SSA/Medicare data. The sample is African Americans born in the Deep South, 1916–1932. Dependent variable is the five-year mortality rate of a State-level cohort. (To retain comparability with results reported above, we use levels). The regression includes cohort effects (1916 is the excluded category), state-of-birth fixed effects, and the fraction of each cohort that migrates North.

Source for Row B: Authors' calculations, 1960 and 1961 Detailed File of the Vital Statistics and 1960 Public Use Micro Samples of the Decennial Census. Dependent variable is the log of the one-year death rate for cohorts born 1895–1940 (i.e., those aged approximately 20–65); the numerator is obtained from Vital Statistics records and the denominator is constructed from Census records. Regressions also contain cohort effects (with year 1895 as the excluded category) and birth state effects (Alabama is the excluded category).

Standard errors are in parentheses

** significant at the 0.05 level

* significant at the 0.01 level.

Table 7

Evidence about a Possible Link between Migration and Mortality Related to Smoking and Alcohol Use, African Americans Born in the Deep South, 1916–1932

A. Proportion who Smoke and Drink Alcohol				
	Proportion of Men		***Proportion of Women***	
Residence in Old Age	**Smoke**	**Drink**	**Smoke**	**Drink**
Birth State (in South)	0.626	0.376	0.266	0.255
Other State in South	0.642	0.397	0.347	0.272
In North	0.658	0.456	0.405	0.320
N	13,945	13,945	31,796	31,796
B. Proportion of Deceased with Conditions Related to Smoking and Alcohol				
	Proportion of Men		*Proportion of Women*	
Residence in Old Age	COPD	Cirrhosis	COPD	Cirrhosis
Birth State (in South)	0.083	0.013	0.041	0.011
Other State in South	0.094	0.017	0.051	0.012
In North	0.098	0.021	0.064	0.016
N	240,222	240,222	209,409	209,409

Source: Authors' calculations. Samples are restricted to black men and women born in the Deep South, 1916–1932. Data for Panel A are from the Behavioral Risk Factor Surveillance System (BRFSS). "Smoke" is a self-report indicating that an individual is a current or former smoker. "Drink" is someone who indicates that he or she drinks any alcohol. Panel B is based on Death Certificate data, for individuals dying between ages 65 and 75. An individual is included in the count if chronic obstructive pulmonary disease (COPD) or cirrhosis of the liver was indicated as a present condition on the death certificate.

Table 8

Dissimilarity of Cause of Death between Migrants and Non-Migrants among African Americans Born in the Deep South

A. Men

Primary Cause of Death	Dissimilarity	Rank
Chronic liver disease and cirrhosis	1.51*	13
Cancer: stomach, colon, pancreas, peritoneum	1.15*	3
Pneumonia and influenza	1.14*	8
Cancer: Leukemia	1.11*	15
Cancer: respiratory and intrathoracic organs	1.06*	2
Cancer: urinary organs	1.05	12
Homicide	1.04	16
Cancer: Other	1.02	5
Other infectious and parasitic diseases	1.00	10
Cancer: genital organs	0.99	4
Nephritis	0.99	9
Diabetes	0.98	7
Chronic obstructive pulmonary diseases	0.97	6
Major cardiovascular diseases	0.96*	1
All other accidents	0.81*	11
Motor vehicle accidents	0.62*	14

B. Women

Primary Cause of Death	Dissimilarity	Rank
Chronic liver disease and cirrhosis	1.41*	14
Cancer: respiratory and intrathoracic organs	1.33*	4
Chronic obstructive pulmonary diseases	1.29*	9
Pneumonia and influenza	1.27*	10
Cancer: stomach, colon, pancreas, peritoneum	1.14*	2
Cancer: breast	1.12*	6
Cancer: urinary organs	1.11*	13
Cancer: Leukemia	1.10	15
Other infectious and parasitic diseases	1.07	11
Cancer: Other	1.01	5
Major cardiovascular diseases	0.95*	1
Cancer: genital organs	0.91*	7
Diabetes	0.87*	3
Nephritis	0.81*	8

A. Men		
All other accidents	0.76*	12
Motor vehicle accidents	0.58*	16

Source: Author' calculations from Death Certificate data. Samples are restricted to black men and women born in the Deep South, 1916–1932, and who died between the ages of 65 and 75.

*Indicates that the 95 percent confidence interval does not include 1 (using a bootstrap procedure with 1000 replications). For additional details, see the appendix online.