APPLIED ECONOMICS WORKSHOP

Business 33610
Spring Quarter 2010

Seth Sanders
Duke University

(Dan Black, Magdalena Muszynska, Lowell Taylor, and NORC)

"The Great Migration and African-American Mortality: Evidence from Mississippi"

Wednesday, June 2, 2010
1:20 to 2:50pm
Location: HC 3B

For any other information regarding the Applied Economics Workshop, please contact Tamara Lingo (AEW Administrator) at 773-702-2474, tammy.lingo@ChicagoBooth.edu, or stop by HC448.
The Great Migration and African-American Mortality:
Evidence from Mississippi

by

Dan Black
University of Chicago and NORC
danblack@uchicago.edu

Magdalena Muszynska
Duke University
magdalena.muszynska@duke.edu

Seth Sanders
Duke University
seth.sanders@duke.edu

Lowell Taylor
Carnegie Mellon University
lt20@andrew.cmu.edu

April 25, 2010

We are grateful for support from the University of Kentucky Poverty Research Center small grants program contract 3048106739-10-250. We have benefited from excellent research assistance from Ryan Brown and Evan Taylor. All errors are our own.
Abstract

Two inextricably linked phenomena lie at the heart of African American social history in the twentieth century: The first is “black-white economic convergence” that accompanied the decline in discriminatory barriers and narrowing of the black-white gap in human capital. The second is “the great migration”—the movement of millions of African Americans from the South to the North, Midwest, and West. For many years the observation of a declining wage gap between blacks and whites has been widely viewed as a “success,” a clear manifestation of advancement toward equity across races in our society. And the migration of African Americans—seen as a flight from poverty and oppression to relative opportunity and freedom—is understood as an important component of this success. However, welfare is more than wages, and convincing analyses must carefully consider the interaction between income and other measures of well-being. To our knowledge, this study is the first that addresses the role of the Great Migration on longevity in the African American community. Using a unique dataset that links precise place of birth to age at death we find that migration north played no role in lowering mortality among African Americans. This result holds up to careful adjustment for selective migration using the proximity of place of birth to North-South railroad lines which we document greatly increased the probability of migration north. This suggest that if wealth does raise health, that the increases in wages of African Americans did not materially increase wealth accumulation. We suggest a model where the migration itself endogenously raised the cost of living in the north, benefiting Northern land owners but not Southern migrants. An alternative explanation is that much of lifetime health is determined prior to adulthood and in adulthood wealth has limited ability to raise health and increase longevity.
I. Introduction

Two inextricably linked phenomena lie at the heart of African American social history in the twentieth century: The first is “black-white economic convergence” that accompanied the decline in discriminatory barriers and narrowing of the black-white gap in human capital. The second is “the great migration”—the movement of millions of African Americans from the South to the North, Midwest, and West. For many years the observation of a declining wage gap between blacks and whites has been widely viewed as a “success,” a clear manifestation of advancement toward equity across races in our society. And the migration of African Americans—seen as a flight from poverty and oppression to relative opportunity and freedom—is understood as an important component of this success.

This general sense notwithstanding, there are good reasons to be concerned about the validity, or at least precision, of these perceptions. First, inferences about declining racial inequality in earnings are conceptually difficult when people are making voluntary location choices to places with differing labor- and housing-market conditions; from the perspective of social welfare and policy there must be a clear distinction drawn between improvements in welfare that result from individual and family decisions made from a fixed set of choices (like choosing a place to live and work) and the expansion of the choice set via social policy (such as improvements in educational opportunities). Second, welfare is more than wages, and convincing analyses must carefully consider the interaction between income and other measures of well-being, including health and longevity.

The issue of the welfare consequences of the great migration, however defined, has many dimensions. Perhaps the most prominent related literature in economics centers on the black-white convergence of wages during the middle of the 20th century. Of primary interest here is
the effort to decompose the causes of the convergence. In general, explanations have focused on
the role of education, school quality, civil rights legislation including desegregation, affirmative
action, and changing location.\footnote{To mention just a few contributions: Smith and Welch (1989) give a
definitive overview of the broad issues, Butler and Heckman (1977) and Donohue and Heckman (1991) provide
important analyses of the role of civil rights legislation, Card and Krueger (1992) analyze the role of school
quality, Neal and Johnson (1996) and Neal (2006) focus attention on the black-white gap in human
capital as measured by objective measures, and Altonji and Blank (1999) discuss challenges in the empirical
evaluation of market disparities. The large empirical literature is complemented by an important theoretical
literature on discrimination, e.g., Arrow (1971), Becker (1971) and Loury (2002).} While there is extensive
debate about many details, our reading suggests that there is a general consensus, expressed in Smith and Welch (1989),
that as an empirical matter, both the narrowing of the education gap \textit{and} the migration of African
Americans (from generally low-wage to higher-wage labor markets) contributed substantially to
a significant narrowing of racial wage disparity in the U.S. over the last half of the twentieth
century. There is general agreement that civil rights legislation also contributed to decreased
black-white disparity, though there is less agreement about the magnitude of effects and precise
causal paths and about.

Of course, wages—even real wages—represent only a partial story with respect to any
understanding of a larger sense of social “convergence” among blacks and whites. Other
dimensions matter also, including welfare as measured by health and longevity, and the literature
documents that African Americans fare worse than their white counterparts on these dimensions.
Harper, \textit{et al.} (2007), as one example, note that while life expectancy at birth converged for
blacks and whites during the period from 1900 to 1940, that gap remained substantial, and failed
to decline consistently after the 1960s. Levine, \textit{et al.} (2001) find that from 1979 to 1998 the
“black:white ratio of age-adjusted, gender-specific mortality increased for all but one of nine
causes of death that accounted for 83.4\% of all US mortality in 1998.”
While the proximate medical causes for the black-white health gap are reasonably well known, the underlying mechanisms are not. There is ample evidence suggesting that persons of lower socioeconomic status have reduced life expectancies, but some evidence indicates that economic disparities are not the sole source of the black-white gap in mortality. For example, Sorlie, et al. (1992) find that increased income lowers mortality rates for everyone, but that blacks have higher mortality than whites at every level of income. Guralnik, et al. (1993) finds that educational attainment may have a stronger effect than race per se on life expectancy. A very small number of papers focus on childhood and early-life conditions specifically as predictors of late-life racial disparity in mortality. One particularly important piece of research, for our purposes, is Preston, Mill, and Drevenstedt (1998), who were able to show, even with relatively small samples (582 older African Americans) that “children who were exposed to the most unhealthy childhood environments were far less likely to reach age 85 than those living in more favorable environments.” They also showed that the correlation between mortality risks at young ages and mortality risks at older ages is positively correlated for this population.²

More generally, though, little work explores racial differences in mortality in a way that allows consideration of the role of place of birth. And there appear to be no studies assessing the role of the Great Migration in shaping mortality outcomes for African Americans in the U.S. Our view is that this deficit in the literature is largely a function of data limitations. The principal

² Another particularly relevant paper is Warner and Hayward’s (2006) analysis of the racial gap in mortality based on National Longitudinal Survey of Older Men (NLS). Relevant research on the origins of black-white disparities comes also from detailed historical work such as that done by Costa and her colleague, e.g., Costa (2004) and Costa, Helmen, and Wilson (2007). Barker’s work in the medical literature (e.g., his 1990 and 1995 papers) is well-known for drawing a connection between nutrition in utero and in infancy to late-life outcomes. Similarly, in the demography literature, Bengtsson and Lindström (2000, 2003) study the impact of exposure to childhood disease for subsequent mortality. A large literature studies issues surrounding a broad set of fetal-origins and early-childhood exposure hypotheses; contributions include Lundberg (1993), Elo and Preston (1996), Leon, et al. (1998), Preston, et al. (1998), and Vaupel, et al. (1998). Among the many analyses from the economic history literature are the seminal work of Fogel (1993) and Almond’s (2006) study of the fetal-origins impact of the 1918 influenza pandemic.
ways to study mortality are (1) to combine data from vital statistics detailed mortality files (death certificates), (2) to use historic panel survey data or more recently (3) to use administrative data that records data of death for large samples of individuals. There is a large literature from each of these methods and each has benefits and shortcomings. One shortcoming that all have shared is that information on place of birth is limited. In all cases, place of birth is recorded at best as the “State of Birth” and at times is recorded as “Region of Birth” or for confidentiality reasons not available at all.

As we explain in detail below, most of our work relies on Administrative data from the Medicare Part B program which accurately records date of death for the purpose of terminating benefits. While this data does not itself contain place of birth information, as a federal administrative dataset it contains social security numbers. Fortunately the Numident file from the Social Security Administration contains “town or county of birth.” With special permission from the Center for Medicare and Medicaid studies and the Social Security Administration and confidentiality protection, these two files were matched allowing the first broad study of migration and mortality. Knowing the specific town of birth allows a much more complete analysis of issues of selective migration. The benefit of using Medicare data is that almost the entire population of the U.S. is covered and our dataset includes more than 70 million observations on date of death. The cost is that Medicare is a program where eligibility starts at age 65; our analysis is therefore limited to mortality for men and women living to age 65. Having established the relationship between migration and mortality in this dataset we turn to coarser data from the vital statistics detailed mortality files and the U.S. Census. Here we show that the central result in our analysis of our administrative data source holds.
The remainder of our paper is as follows. In section 2 we lay out our data sources and explain the matching and coding of place of birth. This section gives extensive detail on our match rate and how we develop weights to account for observations that could not be matched to a place of birth. In section 3 we lay out our basic findings, comparing mortality of men and women born between 1905 and 1925 in Mississippi who migrated north vs. those that remained in the South. Our central finding from two separate data sources is that the survival rates are indistinguishable. We interpret this as migration failing to have the beneficial effects on health as it had on wages. However, a reasonable alternative is that particularly low skilled African Americans migrated north; then the finding of equal rates of survival should be interpreted with some care. If low skilled African American migrants achieved the same level of survival as the more skilled that remained in the South this should be viewed as an improvement in health and survival. Section 4 addresses this issue. We show that being born on a railroad line decreased the chance of living in Mississippi at old age by 10%. There is little evidence that men and women born in a Mississippi town with a rail line differed systematically at birth than those born off the rail line (and sometimes just a few miles away). Our additional analysis looks at the “reduce form” and assesses whether being born on a rail line increases survival; it does not. Section 5 is a conclusion that concentrates on possible explanations of this result.

II. Data Sources


Our primary data source is the Duke Social Security/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. The data are complete for the period
1976-2001. There are over 70 million records in the dataset, covering 95% 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, exact date of death, and, importantly detailed place of birth information. Specifically, the data include either town and State of birth or town, county and State of birth for all U.S.-born respondents. To our knowledge, this is the only data source that provides detailed place of birth and detailed place of current residence in a very large sample. It is therefore an ideal data set for answering these questions: Which “sending communities” in the South sent people to which “receiving communities” outside the South (which we will often refer to as simply the “North”). The date of death allows us to study mortality among blacks and whites, migrants and non-migrants. A further advantage of this dataset is that death and population counts are based on the same data source.3

b. Coding Place of Birth

While the data set has enormous potential, there is a technical hurdle to overcome before analyses using place of birth can begin. SSA provided 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. The 12-character text field of place of birth, however, poses important technical issues which we resolved in order to code the longitude and latitude coordinates (as well as by Census FIPS codes for Place and County) of each place of birth.

---

3 We discuss the combined use of Vital Statistics mortality detailed data with U.S. Census data below as an alternative method of estimating age-specific mortality.
In order to accomplish this we develop an algorithm that matches the 12-character text field to the U.S. Geological Service’s list of Place Names recorded in the USGS 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. Observations were classified according to the strength of their match between the write-in place of birth on the SSA Numident file and the GNIS list. A “Perfect Unique Match” was a place that was spelled exactly the same in the two files and enough information was available that the location within Mississippi was unique. The remainder of the observations was viewed as having “No Perfect Unique Match” and our algorithm then attempted to match the write-in responses to the “best” match on the GNIS. The matching algorithm works as follows: We compare every letter of the text-field “place of birth” to every letter in each name from the GNIS list to determine how many letters were the same. Order of letters in the two words was disregarded. If there was a single name on the GNIS list that shared the maximum number of letters with the place name on the first list and the first three letters in the two words were identical, we considered it an “Algorithm Match.” The above algorithm divided the remaining names into two categories “Algorithm Matches” and “Algorithm Non-Matches.”

In Table 1 we present the most aggregated results of our matching procedure. The first two columns refer to the number and percentage of places matched and the next two columns the number and percentage of people matched. The place of birth entries which matched exactly to a unique GNIS place name and county (or county name only) formed the “Perfect Unique Match” category and the rest were considered “No Perfect Unique Match.” While only 25.6% of places

---

4 In rare cases, this “Algorithm Match” could be a town name that was not unique; for example there are 6 places named “Antioch, MS.” When this occurred we allocated the person to the place with the largest population if the largest town had 10 times the population of the next largest town of the same name according to our “Perfect Unique Matches.” Otherwise we considered the place of birth as an “Algorithm Non-Match.”
were perfectly and uniquely matched to a GNIS place name, 70.6% of the population were perfectly and uniquely matched. The bulk of the population has a write-in response that is correctly spelled and unique. Of the 29.4% of people that were “No Perfect Unique Matches,” our algorithm could match an additional 17.4% of people (a match rate of approximately 60%). Our automated matching therefore matched 88% of people to their place of birth.5

This computer matching of “No Perfect Unique Matches” is subject to both Type I and Type II error. The algorithm could accept a place on the GNIS list as the write-in place of birth when an alternative town is in fact correct. Additionally, the algorithm could find no matching town when in fact human judgment would find an obvious match. In order to glean as accurate a match as possible and in order to assess our likely error rates we next hand checked both our “Algorithm Matches” for false positive matches and “Algorithm Non Matches” for potential false negative matches.

For both “Algorithm Matches” and “Algorithm Non-Matches” all place of birth entries with a frequency of 5 or more observations were manually checked; this included 87% of all “Algorithm Matches” and these observations make up 49% of the “Algorithm Non-Matched” population. We then randomly sub-sampled spellings of places of birth that occurred 1 to 4 times in for both the “Algorithm Matches” and “Algorithm Non-Matches.” Specifically 200 of the “Algorithm Matches” (1 hour of time) were drawn for each of the 4 groups remaining (4 occurrences, 3 occurrences, 2 occurrences, and 1 occurrence) and hand-checked; 150 for each of the 4 groups remaining of the “Algorithm Non-Matches” were drawn for hand checking (again, 1 hour of time).

5 Of course our match rate for places is much lower matching only 46.4% of all places. This reflects our inability to match by computer algorithm many variants of unusual spellings. Keep in mind that while there is only one way to spell a town correctly, there are an infinite number of ways to spell the town name incorrectly.
This hand checking allowed us to do two things. First, we calculate the false positive rate for “Algorithm Matches” and the false negative rate for “Algorithm Non-Matched.” Second, we formed weights for the random sub-sample of hand checked observations equal to the inverse of the ratio of 150 to the total observations in its specific sub-sample; all observations not hand checked were given a weight of 0. For example, if an observation was part of the subsample of “Algorithm Non-Matched” and the person wrote in a town of birth spelled in a way that occurred twice in this subsample, we constructed a weight of 3,359/150 for each person as there were 3,359 occurrences where a specific spelling occurred exactly twice among “Algorithm Non-Matched” and we randomly sampled 150.

We tried to codify the hand-checking process. First, we compared the place name entry to like spelled GNIS place names, looking for a possible match where the difference is of one or two missing, transposed, or incorrect letters. Second, we tried to find a phonetic match for the entry name, as many of the unmatched place names are simply spelled as they sound. In these cases it is important to be aware of the different interpretations there can be for the spelling of places from different lingual backgrounds. Places with French, German, Italian, Spanish, and Native American names are spelled consistently incorrectly in the same manner, so a place’s particular misspelling or phonetic spelling can give a clue as to the lingual origin of its correct match. Third, we used Google. The Google search can lead to new interpretations of the misspelled place through the “Did you mean?” adjusted spelling suggestions. Also, there are times that the unmatched place name you are working with is just an antiquated/obsolete name for a current place, which Google can find. Finally if these three methods failed we looked for a phonetically similar word which is different in sound due to absence, transposition, or error from
only one or two letters. At this point if a reasonable match had still not been identified, the entry was considered unmatched.

We place each “hand-check outcome” into one of six categories: match-confident, match-some ambiguity, match-county only, no match-multiple possible matches, no match-no possible matches, and unknown. “Match-Confident” pertains to observations for which it would not be reasonable to believe that any other GNIS place and county than the one that has been manually assigned was intended by the respondent. “Match-Some Ambiguity” pertains to an observation, unique geo-site pair which appears exclusively correct but is not as obviously accurate as the former match definition. For example, if the checker finds a unique match using one of the first three procedures listed above, for the most part, this will be listed as a “Match-Confident”, but once the checker needs to start combing search techniques, as mentioned in step four, more subjectivity is used by the checker and the match most likely will be listed as having some ambiguity.

Lack of confidence in a match can come from two sources; picking between several possibilities, or trying to determine if the observation, unique geo-site pair is strong enough to be considered a match. The latter problem relies on some subjectivity from the checker but when a decision cannot reasonably be made between a set of possible unique GNIS place name and county choices, the hand-check outcome is considered “No Match-Multiple Possible Matches.” One example of an entry that would fall into this category would be an observation like “Boltan.” This place name may correspond to either the geo-site place name “Bolton” or “Bolten” and since there is no clear exclusively correct choice the designation “No Match-Multiple Possible Matches” is appropriate. Another example of ambiguity would be the case where the checker can match an observation to a GNIS place like “Greenwood”, but there are
Greenwoods in multiple counties and no way to distinguish to which county this entry refers; this again would be appropriately designated “No Match-Multiple Possible Matches”. “No Match-No Possible Match” is used when all hand-checking procedures have been exhausted and no reasonable match has been found and “Unknown” is used when it is clear that the response was meant to signify that the respondent did not know where they were born.

Table 2 shows our estimates of the match rate and error rate in a sample under various sample construction rules. Table 2 shows that if we were to check all places that were “Algorithm Matches” that occurred 5 or more times we would have a match rate of 88% with an error rate of 0.12%. If we then hand checked all places that were “Algorithm Non-Matches” that occurred 10 or more times we would increase our match rate to 91.74%; checking all places that were “Algorithm Non-Matches” that occurred 5 or more times increases the match rate to 92.44% but increases the amount of time by almost 8 hours. Table 2 shows that that we could achieve an estimated match rate of almost 96% if we checked all “Algorithm Non-Matches;” however this would 264 hours. Instead we sample and weight as we discuss above all observations whose spelling of place of birth occurred fewer than 5 times. We only use hand checked data in our analysis. We omit “Algorithm Matches” that our hand check shows does not match and include people that we are able to hand matches in our “Algorithm Non-Match” sample. Our sample is representative of almost 96% of all observations with an error rate of zero by definition. In our analysis below we only include individuals where there was no ambiguity in the match.

c. Vital Statistics and Census Data

We also use the 1990 Detailed Mortality File (DMF) of the U.S. Vital Statistics registry. This file contains all deaths in the U.S. and includes State of Death and State of Birth. From this
file we can calculate the number of deaths at each age by state of birth for African-Americans. In order to estimate age-specific death rate we divide through by an estimate of the number of African-Americans alive in 1990 by state of birth. This data comes from the 5% IPUMS file of the 1990 Decennial Census. In addition we use the IPUMS Decennial Census files for 1920-1990 to trace out the age of migration for the 1905-1925 cohorts.

### III. Results on Mortality and Migration

In order to aid interpretation of the effect of migration on mortality, we first glean some evidence on the age pattern of migration. From the 1920-1990 Census we examine Black men and women born between 1905 and 1925 in Mississippi. In 1920, these individuals are aged 0 to 15; in 1930 they are aged 5 to 25, etc. For each census year, we code whether the individual is living in Mississippi, in one of the other 10 Confederate States or outside the 11 Confederate States.\(^6\) Then at each age we calculate the fraction of individuals born in Mississippi that were living in Mississippi, in the Confederacy but outside of Mississippi and outside the Confederacy (the “North”). Individuals who were born in Mississippi and are living outside the Confederacy represent migrants to the North; those living in the Confederacy but outside Mississippi are migrants within the South.

Figure 1 plots the fraction of African Americans who migrated to the North and to other areas of the South by age. Two clear patterns emerge. First, by age 60 the total migration rate out of Mississippi was approximately 65%; nearly 50% of African-Americans born between in Mississippi 1905 and 1925 migrated to the North; and additional 15% migrated to other States in the South. This is a massive rate of migration out of Mississippi. Second, the age-migration

\(^6\) The 11 former Confederate States were VA, NC, SC, GA, FL, AL, MS, LA, TN, TX, and AR. KY and MO were officially neutral but were represented by stars in the Confederate flag when the secessionist parts of these states joined the Confederacy in 1861. As shorthand, we refer to the 11 Confederate States as the “South” and the states outside the Confederacy as the “North.”
pattern is quite different for migrants to the North than it is for migrants to other parts of the south. Much of migration within the South occurred before age 20 most likely indicating movements of individuals as children with their parents. But the age pattern for migration to the North is quite different. Migration prior to age 18 is low but there is a steep escalation in migration to the North between ages 18 and 40; thereafter there is only a modest increase in the fraction of African Americans living in the North an indication of lower rates of migration at older ages. This suggests that most migrants to the North migrated after being raised in the South and migrated during their prime working age years. It is very likely the African American migrants to the North spent most of their childhood and adolescent years in the South a similarity they share with non-migrants.

As we discussed, the evidence in the literature suggests that movement from the South to the North greatly increased the earnings of African Americans. Using the 1960 Decennial Census we document the extent of earnings differences between African-American men by place of residence. Table 3 presents the average earnings of African-American men born in Mississippi between 1905 and 1925 who lived in Mississippi, lived in one of the other 10 Confederate states or lived outside of the 11 confederate states. Wage and Salary income is reported in 2010 dollars. Column 1 reports the average level of wage and salary earnings by location and column 2 reports the same among men who have an age an education represented in the northern population and column 3 the same after controlling for single year of age (age 35-55) and level of education of men on the common support. With or without these controls the earnings differences are evident. Black men in Mississippi in 1960 were earning $14,765 annually while migrants to the north were earning nearly $28,687. What is of note is that migrants within the south earn $28,286 nearly the same level of earnings as migrant to the north. This suggests that selective migration
could play a major role if one simply compares northern migrants to non-migrants (most of who remain in Mississippi). Clearly, all migrants had a substantial earnings advantage over non-migrants. Human capital plays virtually no role in explaining the difference in earnings for men who stay in Mississippi vs. those who migrant North. It does play some role in explaining the equivalence between Northern migrants and Southern Migrants – part of the reason Southern Migrants earnings are has high as Northern Migrants is that they in fact are more educated.

We now ask, did migration (and higher earnings) result in lower mortality. We start our analysis with data from our Medicare Part B/SSA Match. We start by simply dividing our data on African Americans born in Mississippi between 1905 and 1925 into those that after age 65 were living in the North and those remaining in the South. Since we observe date of death and date of birth we calculate the age at death for each individual or their age at which our data ends and the observation is censored. We then plot the Kaplan-Meier Survival function for men living in the North vs. living in Mississippi. Figure 2 displays these results.

Figure 2 is startling in the absence of any meaningful difference in the survival rate between migrants and non-migrants; the survival functions effectively lie on top of one and other. At no age do the survival functions differ at any reasonable level of significance in this non-parametric analysis.

Table 4 presents estimates from a Cox proportional hazard model. Table 4 reports the percentage change in the hazard rate with a unit change in a covariate. Column 1 of Table 4 includes only a dummy variable for living in the North and so gives the average difference percent difference in the hazard rate for migrants to the North vs. non-migrants. While the pooling of estimates across ages gives more power for inference, the main message is this – the hazard rate of death was only 2% higher in the North, an effect that is not meaningfully different
but it is statistically significant at the 5% level. In Column 2 we control for several factors that might proxy for the level of depravation as a child including the fraction of the town of birth that was African-American, the size of the town of birth and whether the man spelled the name of his town of birth correctly. While being from a larger town or from a town with a higher share of African-American’s raises the hazard rate of death, the central result remains.

While the Medicare Data has several strengths, it also has obvious shortcomings. The primary one is that mortality prior to age 65 can not be calculated. As discussed above, the alternative to using the single Medicare data source is to use the combination of vital statistics data and Census data to estimate Age specific mortality rates. For 1990 we have a 5% sample of the U.S. population Census and the complete counts of deaths from the Detailed Mortality File of the Vital Statistics. Both of these data sources include state of birth. Focusing on African-American men born in Mississippi we calculate the Age Specific Mortality Rate (ASMR) by place of residence (North vs. South). Specifically our ASMR is calculated as the number of deaths in at each age 50-80 in 1990 over the population alive at that age as of April 1, 1990. We conduct this exercise for two reasons. First it allows us to assess whether our finding is robust to data sources; second it allows us to partially examine mortality at some ages prior to age 65. Since we use a single cross-section, the age-mortality curve represents both changes in mortality with age and with cohort. And we also note that only those aged 65-80 are strictly comparable to our earlier analysis of the 1905-1925 cohorts.

With these caveats, Figure 3 suggests that the patterns we found on survival in the Medicare data are replicated in Vital Statistics and Census data. There is little evidence of large differences in Age specific mortality from age 65 to 80. Inference is somewhat hindered in that our 5% sample from the census leaves our estimate of the at risk population estimated
imprecisely especially at older ages. Again, at no age can we reject that age specific mortality rates differ between migrants to the North and Black Mississippian men who remained in the South. In addition, there is no evidence that ASMR differed prior to age 65. From aged 50-64, Figure 3 suggests that ASMR for migrants and non-migrants were virtually indistinguishable.

In other work we have emphasized that migration may have raised earnings but not wealth and the link with health is usually thought to flow through wealth. But there are at least two other plausible reasons migration might not raise health. One, emphasized by Robert Fogel is that much of health is determined by early life conditions, especially early life nutrition. The age at migration North suggests that this may have been similar for migrants and non-migrants. This theory then would suggest that even if wealth was higher among migrants, that early childhood exposure to deprivation could not be overcome by the additional wealth from migrating to the North, at least in terms of mortality rates. A second explanation revolves around selection. The finding of equal mortality between migrants and non-migrants could be seen as migration in fact improved health of migrants; if the poorest and potentially unhealthiest individuals left Mississippi for the North, the fact that they achieved the same mortality rates as their richer non-migrants could be interpreted as a positive impact of migration. This story is plausible given the history of migration. Historians believe that at least part of the migration North was a function of the mechanization of cotton production; cotton field hands were among the least educated and poorest residents of Mississippi in the beginning of the 20th Century. We now turn to assessing selective migration.

IV. Selective Migration and Mortality

Many historians, economists and sociologists have noted that an important aspect of the Great Migration was that it was “vertical;” African American’s in Chicago disproportionately
were born in the Mississippi Delta region while African American migrants to Washington, DC, Philadelphia and New York tended to be born in Virginia, North Carolina and South Carolina. Historians believe that the railroads played a major role in determining the flow of migrants from specific areas in the south to specific northern cities. The prevailing historical theory is that people tended to get on the nearest train stop and there was a strong tendency to relocate at the terminus of rail lines. Thus there were migratory streams from Mississippi to Chicago, using the *Illinois Central Railroad* line, and from Virginia and the Carolinas to Washington DC, Philadelphia and New York, along the *Pennsylvania Railroad* line. The formation of African American neighborhoods around the terminus has been documented for several cities. Historians believe that because much migration was driven by the search for improved economic prospects, many migrants left the South in their late teens or early 20s a finding that we confirmed above.

The brief history of the expansion of railway construction in the U.S. is helpful. In general, eminent domain was used to allow railroad companies to minimize the cost of connecting large cities. To do this, land was acquired in as straight a path as possible, given geographic restrictions, to connect major cities. For example, what determined whether a town in Mississippi was on a rail line was whether it laid on a straight line between Memphis, TN and Jackson, MS, between Jackson, MS and New Orleans, LA or between other larger cities. Figure 4A presents one relevant train map for Mississippi, a map of the Illinois Central railroad line in 1892. Figure 4B presents a second train map which while less detailed maps all of the trunk rail lines in Mississippi in 1898.

In our SSA/Medicare part B data, we have no information on individuals at the time of migration (we do not even observe the date of migration), but we can nonetheless address this issue in an indirect fashion using the historians insight. Consider, two entirely comparable black
individuals—comparable along “observable” dimensions like schooling, literacy, and education, and also along “unobservable” dimensions, such as ambition and health. The argument is that if one such individual is born near a town with a train stop on a line heading north, he is more likely to migrate than the individual born far from the train line. Thus distance to the train line of the town of birth is an “instrument” that affects the probability of migration, but otherwise has no direct effect on late-life mortality.⁷

Using the detailed place of birth information, we classify whether the town of birth was in fact on a railroad line. We do this by overlaying Figure 4B on a modern GIS map of Mississippi. 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 traced in Figure 4B. 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. In addition, there is some measurement error for the exact path of the rail line that we trace from our map. If our strategy is to be successful, of course, the historians’ observations about the importance of railroads to migration must be empirically correct: being located near transportation lines must be a statistically meaningful determinant of out-migration from the South. Table 5 presents evidence that being born on a railroad line had powerful effects on migration North. We record for each place of birth (a) the fraction of persons living in Mississippi, (b) the fraction living outside Mississippi but within the South (defined here as the 11 former Confederate States), and (c) the fraction living outside the South (defined the same way) at the time the person was first observed in the Medicare Part B data between 1976-2001. We then simply estimate the fraction of people born on a rail line and away from a rail line that lived in each location.

---
⁷ An instrument would have no direct effect on late-life mortality. As discussed below a crude measure of human capital suggests that human capital was no different for individuals born in railway v. non-railway towns.
Perhaps the most astonishing fact in Table 5 is this: among blacks born in towns that had train stops, only 32.8% remained in Mississippi at old age. The rest migrated to the North or elsewhere in the South. The percentage of blacks who remained in Mississippi among those born in towns not on the rail line was 35.5%, a proportion that is 10% higher than for blacks born on the rail line. Statistics in Table 5 also show that living on the rail line did not just increase migration generally, but increased migration to the North specifically. Living on a train line not only decreased the chances of living in Mississippi it decreased the chances of migrating to other areas of the South by 2.5 percentage points (or 25%).

In Table 6 we treat these locations as dependent variables in a series of linear probability regressions; these give the “difference estimator” of the impact on migration of being born in a town along the train line. Columns (1), (4) and (7) do not control for any covariates and replicate the mean differences in the fraction in each location for individuals born on a train line and born off a train line. Columns (2), (5) and (8) present estimated impacts on location in old age controlling for the fraction of the place of birth’s population that was black and a measure of the total size of the town of birth. Blacks are overrepresented in some parts of the state, and this control is meant to mitigate any correlation between migration from certain parts of the State (e.g. the Delta region) where blacks are overrepresented. The factors are in fact correlated with migration to the north – being born in a town with a high concentration of Blacks increases migration north while being from a large town reduces it. However, while these factors reduce slightly the effect of being born on a railroad line on migration north the effect is still strong and significant.

Finally, columns (3), (6) and (9) add some measures of the individual. Specifically we control for gender and a crude measure of literacy. We code whether the individual spelled the
name of his or her town correctly according to the USGS data base. Our prior, given historian’s accounts is that misspelling should increase migration to the north. Our analysis suggests that in fact this is the case, spelling your town of birth correctly increases the probability of remaining in Mississippi and reduces the probability of moving north. It also appears that African American women were less likely to migrate north than African American men. However, controlling for these factors, the effect of living on a railroad on the probability of migrating north remains large and statistically significant.

Since controlling for whether the individual spelled the name of their town correctly has little impact on the effect of living in a railroad town on migration north, this suggests that an ability to spell may be largely uncorrelated with being born on a rail line. In fact, while 75.74% of individuals living off the train line spell their town of birth correctly, a comparable 74.66% of individuals living on the train line spell the name of their town of birth correctly. This 1.08 percentage point difference is not significant at the 10% level (t-stat 1.16). We take this as some evidence that human capital levels may have been largely balanced for men and women born on the rail line vs. those born off of the line. This is important for our strategy to address selective migration. If human capital is balanced, then comparing the survival functions for individuals born off and on the line is an indirect test of the effect of migration on mortality that is insensitive to selective migration. If being on the train line increase migration but was not otherwise associated with health, then a higher rate of survival for individuals born on the train line must come from their higher rate of migration.

Figure 5 plots the Kaplan-Meier survival function for men by their town of birth. Just as in Figure 2, there appears to be no effect of living on a rail line on survival. This indicates indirectly that migration must not have changed survival as if migration had increased survival
we would expect to see survival higher for men born on the rail line. In fact there is no evidence of any difference in survival rates.

Returning to Table 4 which presents estimates from a Cox proportional hazard model, we now estimate the effect of being born on the rail line. Recall Table 4 reports the percentage change in the hazard rate with a unit change in a covariate. To the degree that being in the North appeared to have a small positive effect on mortality, our analysis that attempts to correct for selective migration indicates that this effect was likely due to selective migration. Column 3 of Table 4 suggests that men who were born in a town on the rail line did not have a statistically significant difference in the hazard rate of death than those born off of a railroad line. This lack of significance is not due to the relative hazard rates being estimated imprecisely leading us to a strong conclusion of little if any effects of migration on mortality.

V. Conclusion

Our analysis suggests that migration of African American men out of the South to Northern cities had no impact on survival. This result is somewhat surprising given the voluminous evidence that migration north increased the economic opportunities of African Americans. This suggests that either the economic opportunities did not translate into better health or that we have overstated the welfare gains from migration. Either of these could be true. Eichenbaum, Tolney and Alexander (2010) recently revisited the question of the economic benefit of migration and conclude that the economic benefits may have been smaller than previous estimated. And there is evidence that housing prices African Americans paid upon arriving in the South was substantially higher and forced up by the great migration (Cite from Fishback). This may suggest the reason mortality did not decline for migrants – economic opportunities in fact were not markedly better in their new home.
An alternative is that early life poverty cast a long shadow on the entire African-American population of the cotton south regardless of their later life achievements. This view is consistent with techno-physical revolution that is laid out in Robert Fogel’s book *The Escape from Hunger and Pre-mature Death: 1700-2100*. Here Fogel emphasizes the role of nutrition for giving the body the constitution to withstand insults from pathogens. Fogel’s view is that much of the biological capital is affected by early life nutrition. This is likely to have been similar for migrants and non-migrants. An alternative explanation then is that the poverty and poor nutrition of African Americans growing up in Jim Crow Mississippi could not be reversed by the opportunities that the north offered migrants. While we can not distinguish these two hypotheses empirically with our data it does appear that the lack of improvement is real and not a function of differential selection of migrants.
Literature Cited


Figure 1:
Migration within the South and to the North by Age
African American Men and Women

Source: Author’s Calculation for Individuals in the 1920-1990 Decennial Census
Born in Mississippi between 1905 and 1925
Figure 2:
Kaplan-Meier Survival Rates
Blacks Born in Mississippi between 1905 and 1925

Kaplan-Meier survival estimates

Source: Authors’ calculations from the matched SSA/Medicare Part B files. Analysis was conducted on data where exact place of birth could be determined and was confirmed. Weights were used to represent the entire SSA/Medicare Part B data file. KM analysis on the entire file was nearly identical. Survival plots are for men born in Mississippi who remained and those born in Mississippi who migrated to the North.
Figure 3.
Age Specific Death Rates by Age in 1990
Blacks Born in Mississippi between 1905 and 1925

Source: Authors’ calculations from detail mortality files (DMF) of the Vital Statistics Registry of the U.S., and the 1990 PUMS. Age specific mortality rates were calculated as the number of deaths during 1990 divided by the number of men surviving until 1990 for each age. The numerator was constructed by counting the number of deaths of men born in Mississippi in 1990 at each age 50 to 80 using the DMF’s mid-year population; our population estimate is as of the Census, April 1, 1990 rather than a mid-year estimate. We plot ASMRs where the estimated population of men exceeds 600 (30 Observations in the Census). ASMRs are for men born in Mississippi who remained and those born in Mississippi who migrated to the North.
Figure 4A:
Illinois Central Railroad Stops, June 30, 1892

Figure 4B:
All Railroad Lines in Mississippi, 1898

Source: Gray's new trunk railway map of the United States, Dom. of Canada and portion of Mexico, 1898. Library of Congress, Call Number G3701.P3 1898 .G7 RR 67
Figure 5
Kaplan-Meier Survival Rates
Blacks Born in Mississippi between 1905 and 1925

Kaplan-Meier survival estimates

Source: Authors’ calculations from the matched SSA/Medicare Part B files. Analysis was conducted on data where exact place of birth could be determined and was confirmed. Weights were used to represent the entire SSA/Medicare Part B data file. Survival plots are for men born in on a rail line (railroad=1) and those born away from a rail line (railroad=0).
<table>
<thead>
<tr>
<th></th>
<th>Occurrences of Place Names by Place</th>
<th>Occurrences of Place Name by People</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>%</td>
</tr>
<tr>
<td>A</td>
<td>Total</td>
<td>73,986</td>
</tr>
<tr>
<td>B</td>
<td>Perfect Unique Matches¹</td>
<td>18,913</td>
</tr>
<tr>
<td>C</td>
<td>No Perfect Unique Match²</td>
<td>55,073</td>
</tr>
<tr>
<td></td>
<td>C1 Algorithm Matches</td>
<td>15,384</td>
</tr>
<tr>
<td></td>
<td>C2 Algorithm Non-Matches</td>
<td>39,688</td>
</tr>
<tr>
<td></td>
<td>C3 Blank Entries</td>
<td>1</td>
</tr>
</tbody>
</table>

Notes:

¹ Entries which perfectly match a unique GNIS place name and county or county name;
² Entries which do not perfectly match a unique GNIS place name and county or county name
<table>
<thead>
<tr>
<th>Hand-Coding Algorithm Matches Down</th>
<th>Estimated¹ Match Rate²</th>
<th>Error Rate³</th>
<th>Non-Matched Hand-Coding Hours Necessary</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. 5 Occurrences</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Matched Hand-Coding:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hand-Coding Non-Matches Down to 10</td>
<td>87.91%</td>
<td>0.12%</td>
<td>3.26 hrs.</td>
</tr>
<tr>
<td>B.</td>
<td>91.71%</td>
<td>0.12%</td>
<td>5.70 hrs.</td>
</tr>
<tr>
<td>C. Hand-Coding Down to 5 Occurrences</td>
<td>92.42%</td>
<td>0.12%</td>
<td>13.58 hrs.</td>
</tr>
<tr>
<td>D. Hand-Coding Down to 2 Occurrences</td>
<td>93.54%</td>
<td>0.11%</td>
<td>50.25 hrs.</td>
</tr>
<tr>
<td>E. Hand-Coding All Occurrences</td>
<td>95.93%</td>
<td>0.11%</td>
<td>264.59 hrs.</td>
</tr>
</tbody>
</table>

Notes:
¹ Estimated based on four 200 count samples of entries with 1, 2, 3, and 4 occurrences which were algorithm-matched to a unique GNIS name (Total number of algorithm-matched observations with 1, 2, 3, and 4 occurrence is 18,343) and four 150 count samples of entries with 1, 2, 3, and 4 occurrences which were algorithm-nonmatches (Total number of algorithm-nonmatched observations with 1, 2, 3, and 4 occurrence is 45,755).
² Based on 73,986 unique place names from a population of 817,687, match rate is the ratio of the accepted algorithm matches population to the total population.
³ Error rate is the ratio of the incorrect algorithm matches population that cannot be reasonably hand matched to the accepted matched population.
Table 3:  
Wage and Salary Earnings (in 2010 dollars) 
Black Men Born in Mississippi, 1905-1925 Birth Cohorts

<table>
<thead>
<tr>
<th></th>
<th>Observations</th>
<th>Observations</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Common Support</td>
<td>Earnings</td>
<td>Earnings</td>
<td>Earnings</td>
</tr>
<tr>
<td>Mississippi</td>
<td>1936</td>
<td>1920</td>
<td>$14,765***</td>
<td>$14,660***</td>
<td>$14,583***</td>
</tr>
<tr>
<td>Other Confederate States</td>
<td>976</td>
<td>966</td>
<td>$28,286***</td>
<td>$28,007***</td>
<td>$25,250***</td>
</tr>
<tr>
<td>Outside Confederate States</td>
<td>1184</td>
<td>1184</td>
<td>$28,687***</td>
<td>$28,687***</td>
<td>$28,687***</td>
</tr>
<tr>
<td>Observations</td>
<td>4096</td>
<td>4070</td>
<td>0.53</td>
<td>0.53</td>
<td>0.55</td>
</tr>
<tr>
<td>Common Support</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

Notes: *** p<0.01, ** p<0.05, * p<0.1; Standard errors in parentheses.
Table 4:
Marginal Effects from Cox Proportional Hazard Model,
Percentage Change in Hazard Rate with a Unit Change in Covariates,
Black Men Born in Mississippi, 1905-1925 Birth Cohorts

<table>
<thead>
<tr>
<th></th>
<th>North v. South</th>
<th>On Line v. Off</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>North (=1 in North)</td>
<td>0.022+</td>
<td>0.021+</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Log Population</td>
<td>0.009*</td>
<td>0.009*</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Fraction Black</td>
<td>0.061*</td>
<td>0.071*</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Spelled Birth Place</td>
<td>0.008</td>
<td>0.006</td>
</tr>
<tr>
<td>Correctly</td>
<td>(0.011)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Railroad (=1 On Line)</td>
<td>-0.000</td>
<td>-0.012</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Observations</td>
<td>74,541</td>
<td>74,541</td>
</tr>
</tbody>
</table>

Notes:
Marginal Effects are calculated as the average change in the log of the hazard rate with a 1 unit change in each covariate of interest. Standard Errors in parentheses (+sig. at 0.05, *sig. at 0.01).
### Table 5:
Location at Old Age for Blacks Born in Mississippi, 1905-1925 Birth Cohorts

<table>
<thead>
<tr>
<th></th>
<th>Born in Town with Railroad Stop</th>
<th>Born in Town with No Railroad Stop</th>
<th>The “Effect” on Location of Being on a Railroad Stop</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Living in Mississippi (MS)</td>
<td>32.8%</td>
<td>35.5%</td>
<td>-2.7% (4.31)</td>
</tr>
<tr>
<td>% Living Outside the South</td>
<td>59.1%</td>
<td>53.9%</td>
<td>+5.2% (8.06)</td>
</tr>
<tr>
<td>% Living in South, Outside MS</td>
<td>8.1%</td>
<td>10.6%</td>
<td>-2.5% (6.26)</td>
</tr>
<tr>
<td>Sample Size</td>
<td>107,029</td>
<td>57,161</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Authors’ calculations using the Duke Social Security/Medicare Dataset. The sample includes 164,190 African Americans born between 1905 and 1925 in Mississippi.
Table 6:
Linear Probability Regressions: Impact of Birth Location
on Location at Old Age for Blacks Born in Mississippi, 1905-1925 Birth Cohorts

<table>
<thead>
<tr>
<th></th>
<th>Living in Mississippi</th>
<th>Living Outside the South</th>
<th>Living in South, Outside MS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>On Train Line</td>
<td>-0.027*</td>
<td>-0.021*</td>
<td>-0.019*</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Log Population</td>
<td>0.008*</td>
<td>0.005*</td>
<td>-0.007+</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Fraction Black</td>
<td>-0.125*</td>
<td>-0.127*</td>
<td>0.158*</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Sex (Female=0)</td>
<td></td>
<td></td>
<td>-0.018*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.006)</td>
</tr>
<tr>
<td>Correct Spelling</td>
<td></td>
<td></td>
<td>-0.041*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.007)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.355*</td>
<td>0.376*</td>
<td>0.322*</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.022)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.001</td>
<td>0.006</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Notes: Authors’ calculations using the Duke Social Security/Medicare Dataset. Robust t statistics in parentheses (+sig. at 0.05, *sig. at 0.01). “On Train Line” equals to 1 if the Place of Birth is a train stop on a Railroad Line (according to a 1893 map), 0 otherwise. “Fraction Black” and “Log Population of Place of Birth” are calculated as the number of blacks divided by the number of people and the log of the number of people in each Place of Birth in the Medicare Dataset, respectively. For Place of Birth (POB) we code the spelling of the town of birth as follows: Spelling = 1 if spelled correctly, 0 otherwise.