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

Dan A. Black
Seth G. Sanders
Evan J. Taylor
Lowell J. Taylor

Online Appendix

A. Selection of States

Our strategy for selection of Deep South states was quite straightforward. First, we limited our sample to states that formed the Confederacy. Second, we wanted states with large concentrations of African Americans, so from those states we further restricted attention to states in which births cohorts of 1916 to 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. See Table A-1. Third, we wished to exclude states that bordered non-Confederate states, because we did not want to include short-distance migration across state borders as part of the Great Migration. 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. As it turns out, this final restriction did not change our selection of states.

Notice from Table A-1 that we disqualify North Carolina and Florida. Of the two, North Carolina appears to be most similar to our sample, as it has a fairly large black representation (over 25%) and a high out-migration rate (nearly 41%).

Table A-1. Description of Deep South States and Other Southern States

	Proportion Black	Proportion of Blacks Moving out of South
A. Deep South States		
Alabama	0.306	0.448
Georgia	0.329	0.363
Louisiana	0.336	0.289
Mississippi	0.444	0.500
South Carolina	0.414	0.455
B. Other Southern States		
Arkansas	0.196	0.624
Florida	0.287	0.278
North Carolina	0.257	0.409
Tennessee	0.144	0.474
Texas	0.134	0.275
Virginia	0.251	0.424

Source: Authors' calculations, 1970 Census, birth cohorts of 1916 through 1932 inclusive.

B. Data Sources and Construction

1. Data from the U.S. Census

We make extensive use of decennial U.S. Census data, from 1880 through 2000. As indicated in the notes in the paper's tables, Census data were taken from the Integrated Public Use Microdata Series (IPUMS). See Ruggles, *et al.* (2010).

2. Duke SSA/Medicare Data,

The Duke SSA/Medicare data are 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). The data are for the period 1976-2001. Because enrollment requires proof of age, we believe age validity of the records is likely high compared with other data sources for the U.S. elderly population. In addition to date of birth, race, sex, and age, the data include zip code of the place of residence in old age.

2.1. Estimated Coverage

One important concern is the extent to which the SSA/Medicare data covers the cohorts we study, black men and women born in the Deep South between 1916 and 1932. We provide some evidence about coverage by comparing counts from these data and corresponding estimated counts from the 1990 and 2000 PUMS of the United States Census. For African Americans in birth cohorts 1916 to 1925, we calculate the number of people recorded in Medicare Part B files of the SSA/Medicare data as of April 1, 1990, and we compare this count to the corresponding estimate from 1990 Census PUMS. For birth cohorts of 1926 to 1932, we undertake an analogous exercise using the 2000 Census PUMS to estimate the population and calculate from the SSA/Medicare counts as of April 1, 2000. The results are presented in Appendix Table B-1 below. Coverage rates are generally quite good. They are somewhat higher for those residing in the South than elsewhere and are generally higher for men than women.

2.2. Matching Records to Place of Birth

The Duke SSA/Medicare data matched Medicare Part B records to Numerical Identification System participant files (known also as the "Numident"), which is a computerized abstract of SS-5 forms. These are the forms people fill out when applying for a social security number. The Numident file has the person's exact date of birth, race, and a 12-character string that identifies the individual's town of birth and a two-character string that identifies the state of birth if applicable. We developed an algorithm that matches this object to place names recorded in the U.S. Geological Service's Geographic Names Information System (GNIS). The GNIS is the master list of all populated place names in the U.S. both current and historic, and includes geographic features including the longitude and latitude of each place. The list of places was obtained from http://geonames.usgs.gov/domestic/download_data.htm. (We downloaded from the web page in 2011.)

Our goal was to accurately assign each 12-letter string to its most likely match among the populated places and counties in the person's birth state based on the percent of letters that are common between the strings. Each unique 12-letter string that has 10 or more occurrences in the data was hand-checked for accuracy. As shown in Table B-2, this accounts for 87% of the cases in the data. Unfortunately, a reasonably common occurrence among these cases had an empty birthplace field or a response of "unknown." In such instances, individuals are designated as *unmatched* and are excluded from our main specification.

For cases in which there were fewer than 10 occurrences, we used an algorithm to form the match (if any). The algorithm searches for similar spellings, based primarily on the first few letters of the written town. (For example, "Chatannoga TN" would be matched to "Chattanooga, TN.") Strings that had no plausible match were designated *unmatched.* Also designated as *unmatched* were cases that had multiple plausible matches. An example would be Jefferson, GA, as Jefferson is the name of a county and a name of a city in Georgia. Altogether we were successful in matching 76 percent of cases to the city or town of birth.

In addition, there were a fairly large number of individuals who provided county of birth (but not town of birth). This was particularly common in Georgia, with its 169 small counties (see Table B-2). We were able to match over 89 percent of cases to a birthplace at least at the county level. Individuals for whom we have birth county, but not birth town, were excluded from town-level analysis if they were born in counties that contained railways. Individuals born in counties without railways could be included in town-level regressions even when we did not have birth town because we could accurately code them as being born in non-railway towns.

3. Vital Statistics Death Certificate Data

For Table 6, Panel B, we used 1960 and 1961 Detailed File of the Vital Statistics to calculate one-year death rates by state of birth.

4. Data from the Behavioral Risk Factor Surveillance System (BRFSS)

The data used to construct Table 7 are from the Behavioral Risk Factor Surveillance System (www.cdc.gov/bfrss/), accessed in 2013. We restricted analysis to black men and women born in the Deep South between 1916 and 1932 inclusive.

6. Death Certificate Data

We use the Detailed Mortality Files (DMF) of the U.S. Vital Statistics registry for the purpose of assessing cause of death.

5. Railway Locations

The location of railway maps is taken from *Gray's New Trunk Railway Map of the United States* from the Library of Congress and may be electronically accessed from:

http://www.loc.gov/item/gm71000844

In Figure B1, we show the distribution of railways in the Deep South. Any of the places from the GIS that are within 2 miles of a railroad line are coded as being a railroad town.

Table B-1. Estimated Coverage Rates of Medicare Part B by State of Birth and Sex, Birth Cohorts of 1916 through 1932

	Men		Women	
Birth State	**Reside in the South**	**Do not Reside in the South**	**Reside in the South**	**Do not Reside in the South**
Alabama	0.872	0.878	0.810	0.748
Georgia	0.902	0.831	0.803	0.733
Louisiana	0.841	0.802	0.794	0.740
Mississippi	0.988	0.891	0.828	0.816
South Carolina	0.918	0.894	0.845	0.862
Total	0.882	0.865	0.815	0.785

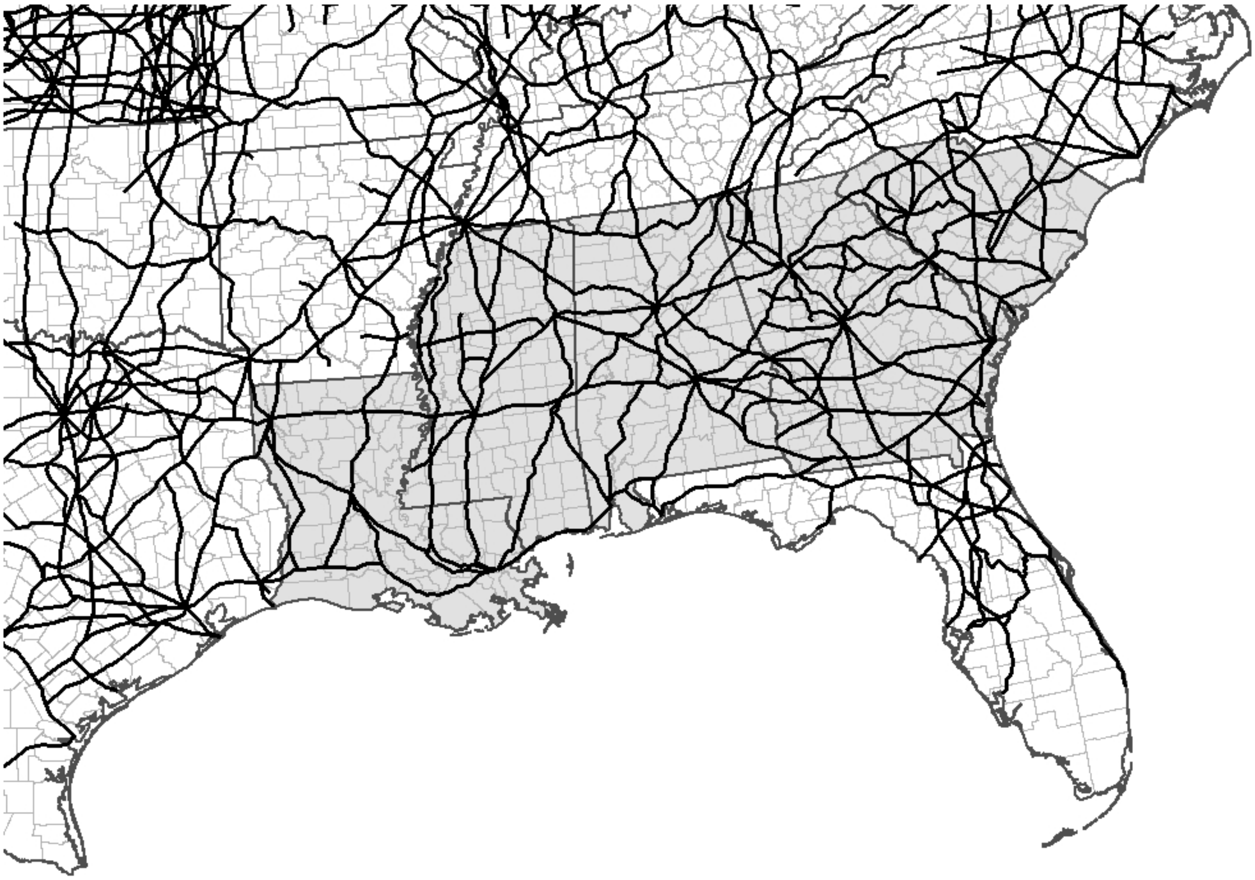
Source: Authors' calculations, Duke/SSA data, and the 1990 and 2000 PUMS of the United States Census. For birth cohorts 1916 to 1925, we compared to the 1990 PUMS. For birth cohorts, 1926 to 1932, we compared to the 2000 PUMS. Duke/SSA counts taken as of April 1 of the Census year.

Table B-2. Match Rates of the Duke/SSA Data to Location of Birth

State	Individuals In Sample	Cases that were Hand-Checked	Cases Matched to Birth Town	Cases Matched to County Only	Percent Matched to Towns
Alabama	212,433	185,511	170,986	22,029	80%
Mississippi	234,871	199,154	178,241	27,044	76%
Louisiana	170,868	148,359	148,936	7,141	87%
Georgia	215,248	190,289	135,045	57,464	63%
S. Carolina	243,866	212,018	180,799	33,767	74%
Total	1,077,296	935,331	814,007	147,445	76%

Source: Author calculations using the Duke/SSA data. All analysis is for black men and women born from 1916 to 1932 inclusive.

Figure B1: Railways in the Deep South



Appendix Reference:

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota, 2010.