NBER WORKING PAPER SERIES

THE MORTALITY CONSEQUENCES OF DISTINCTIVELY BLACK NAMES

Lisa Cook
Trevon Logan
John Parman

Working Paper 21625
http://www.nber.org/papers/w21625

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2015

We thank Rodney J. Andrews, Alan Barreca, William A. Darity, Robert W. Fogel, Claudia D. Goldin, Darrick Hamilton, Samuel L. Myers, Paul W. Rhode, Richard H. Steckel, seminar audiences at Harvard University, University of Minnesota, University of Wisconsin, The Ohio State University, participants in The Second Wave conference, the NBER Summer Institute, and the Southern Economic Association Meetings for useful comments. Stanley L. Engerman provided particularly detailed comments and advice, for which we are grateful. The usual disclaimer applies. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Mortality Consequences of Distinctively Black Names
Lisa Cook, Trevon Logan, and John Parman
NBER Working Paper No. 21625
October 2015
JEL No. I1,J15,N31,N32

ABSTRACT

Race-specific given names have been linked to a range of negative outcomes in contemporary studies, but little is known about their long term consequences. Building on recent research which documents the existence of a national naming pattern for African American males in the late nineteenth and early twentieth centuries (Cook, Logan and Parman 2014), we analyze long-term consequences of distinctively racialized names. Using over three million death certificates from Alabama, Illinois, Missouri and North Carolina from 1802 to 1970, we find a robust within-race mortality difference for African American men who had distinctively black names. Having an African American name added more than one year of life relative to other African American males. The result is robust to controlling for the age pattern of mortality over time and environmental factors which could drive the mortality relationship. The result is not consistently present for infant and child mortality, however. As much as 10% of the historical between-race mortality gap would have been closed if every black man were given a black name. Suggestive evidence implies that cultural factors not captured by socioeconomic or human capital measures may be related to the mortality differential.

Lisa Cook
Department of Economics
Michigan State University
110 Marshall-Adams Hall
East Lansing, MI 48824
lisacook@msu.edu

Trevon Logan
The Ohio State University
410 Arps Hall
1945 N. High Street
Columbus, OH 43210
and NBER
logan.155@osu.edu

John Parman
Department of Economics
P.O. Box 8795
College of William and Mary
Williamsburg, VA 23187
and NBER
jmparman@wm.edu

"A good name is better than fine perfume, and the day of death better than the day of birth."

- Ecclesiastes 7:1

1 Introduction

Numerous studies have found that those with race-specific first names are negatively affected in terms of birth outcomes, job interview callbacks, and mentoring [Busse and Seraydarian 1977; Bertrand and Mullainathan 2004; Figlio 2005; Ginther et al. 2011; Milkman et al. 2012]. The literature has yet to consider long-term consequences of distinctively racialized names. Racialized names may be related to a host of other factors that play out over the life cycle, and identifying these effects would be important as they may be cumulative. Recently, scholars have uncovered a national racial naming pattern among African Americans that predates the Civil Rights Movement [Cook, Logan and Parman 2014, Goldstein and Stecklov 2014]. We now know that a distinct set of given names were used by African Americans in the late nineteenth and early twentieth centuries. While the finding of an historical racial naming pattern is inherently interesting, the implications of having a black name remain largely unexplored.

In this paper we present the first evidence of long-term consequences of distinctively black names (see Table 1). We concentrate on a straightforward outcome, mortality, using newly-available death certificate data (roughly 3 million records). Mortality is an important dimension of well-being and data are available for many historical settings [Parman 2012]. Key for our analysis, death certificates contain reliable information about race, name, and lifespan.

Our primary objective is to examine whether there is a relationship between having one of the historical black names and *within*-race mortality. We adopt a straightforward empirical strategy, estimating the effect of names on longevity after controlling for the time pattern

of mortality and counties of birth and death. To our knowledge, this is the first study to estimate the effect of racial names on mortality or health outcomes.

We find that the effects of a distinctively African American name on mortality are quite large. Conditional on survival to age 10, African American men with distinctively black names live more than one year longer than other African American men. In elasticity terms, a black name increases lifespan by more than ten percent. The correlation we find between distinctively African American names and lifespan is not sensitive to the functional form used to estimate the relationship. We find mixed evidence that possessing a black name was related to infant or child mortality. The effect was present over the entirety of adulthood, which suggests that the effect was cumulative. We find that as much as ten percent of the historical interracial mortality gap would have been closed if every black man had been given a black name. These results are robust to regional variation, holding over four distinct states– Alabama, Illinois, Missouri, and North Carolina, which guards against the finding being driven by environmental, epidemiological, or contextual factors.

In attempting to uncover evidence on mechanisms behind this mortality differential, we analyze the socioeconomic correlates of given names in census records. The census results provide little evidence that the name effect is due to socioeconomic status or to human-capital differences for those who have African American names. There are, however, demographic differences that are correlated with the names, consistent with historical narrative evidence [Gutman 1976]. While the results do uncover important demographic differences that were previously unknown, they do not conclusively uncover the source of the robust mortality difference. Importantly, men with African American names were more likely to have sons with African American names, and men with fathers who had distinctively black names lived longer than other men even if they did not have a distinctively black name themselves. Overall, the results suggests cultural factors may be at play in both the transmission of distinctively black names and their mortality effects.

2 Empirical Strategy

We estimate the relationship between racial names and mortality in a straightforward way. Since our sample is drawn from death records all deaths are observed. We first estimate a linear regression where lifespan (conditional on survival to age 10) is the dependent variable.

$$Lifespan_i = \alpha + \beta BlackName + X'\gamma + \varepsilon_i$$

To provide an elasticity interpretation we estimate a regression where the natural log of lifespan is the dependent variable.

$$\ln Lifespan_i = \alpha + \beta BlackName + X'\gamma + \varepsilon_i$$

In both regressions *Black Name* is an indicator for the presence of an African American name conditional on being an African American man. This is our coefficient of interest as it estimates the difference in lifespan for African American men with a distinctive name relative to other African American men. We concentrate on intraracial differences in mortality since, over the time period covered, the decline in black mortality was faster than for white mortality. Fully accounting for this difference in a empirical model requires that the results effect for black names be relative to other black men– for that reason we restrict our sample for the regression analysis to black men.[1] The vector X includes controls for year of death and year of death squared to account for general time trends in mortality.[2]

We stress two points for the interpretation of the effect of African American names. First, our measure of names is dichotomous. One either has a distinctive name or not— as such our estimates are at the extensive margin. The coefficient measures the average mortality difference for black men holding one of the names listed in Table 1. Other observational studies have used an index which weights the relative exclusivity of a given name, an intensive

[1] Since relatively few white men have black names by design, the effects of black names for white men are not statistically significant in regressions which include all men.

[2] While one would typically control for year of birth rather than year of death, due to age misreporting year of birth is subject to potentially large measurement error. Year of death, as it is recorded when the death occurs, is measured precisely.

measure. We concentrate on the extensive measure since the names themselves are quite distinct and highly racialized as a class of names by design.[3] We also stress that our estimate of the mortality difference is a mortality differential within race.

We also estimate the survival function using a standard Cox proportional hazard model

$$h_{it}(t) = h_o(t) \exp\left[\beta BlackName + X'\gamma\right]$$

Where i indexes the individual, t time (year) of death, and $h_o(t)$ is the baseline hazard, which is integrated out using the partial likelihood method. In the Cox model, if a coefficient is greater than zero (if the hazard ratio > 1), then the variable is correlated with a shorter life. Similarly, if a coefficient is less than zero (if the hazard ratio < 1), then the variable is correlated with a longer life. The estimates from the hazard model give us the percentage differences in the waiting time to mortality.

Before turning to the results, it is useful to consider the possible biases of the estimate of β that could be due to selection. It could certainly be the case that those with African American names are more likely to be aged to the extent that these names represent a nineteenth century naming pattern. For this reason we explicitly control for year of death in the specifications. With these controls included the estimated name effect would have to be attributable to the name itself and not the time in which the name was assigned.[4]

[3]Given the methodology in Cook, Logan and Parman [2014], each name identified would have a high index value if a names index were used.

[4]Another issue of selection with a hazard estimate is truncation. In our case, those born in North Carolina, for example, but dying elsewhere are not included in the data. In a basic sense, these estimates are permanently missing, but if their distribution of deaths is different our estimates will not be applicable. We note that this bias will be present only if the death distribution for those truncated is different. As we noted earlier, there is no evidence that migrants have a different death distribution than non-migrants [Sanders and Muszynska 2009]. In addition, recent methods of proportional hazard estimation have made it possible to correct for the potential of truncation to impact the results [Huber-Carol and Vonta 2004, Tsai 2009, Copas and Farewell 2001, Tsai, Jewell and Wang 1987, Vardi 1982]. Also, since we focus on within-race differences by the presence of a distinctively African American name our hazard estimates would be biased only if those with distinctively African American names were selectively missing, which we view as unlikely.

3 Results

3.1 Death Record Summary Statistics

We use one measure of length of life — total lifespan measured by year of death minus year of birth. In our primary specification we restrict attention to adult mortality (lifespan conditional on survival to age ten) due to the large declines in infant and child mortality early in the twentieth century. The total sample of death records for each state is large, with nearly 100,000 males for Missouri, 300,000 for Alabama, 1.3 million for North Carolina, and over 1.5 million for Illinois. The states offer a broad range of racial compositions, with Illinois and Missouri having relatively low percentages of African Americans, 5.5 and 8.2 percent respectively, and North Carolina being over 30 percent black and Alabama being nearly 50 percent black. Table 2 summarizes the sample sizes for each state and the degree of racial distinction for the black names.[5] In all four states, the distinctively African American names identified in Cook, Logan and Parman [2014] are far more frequently held among African Americans than among whites.[6] The shares of black individuals with an African American name are 1.7, 1.4, 1.7 and 1.3 percent for Alabama, Illinois, Missouri and North Carolina, respectively. The shares of white individuals with an African American name are 0.6, 0.7, 0.6 and 0.4 percent for Alabama, Illinois, Missouri and North Carolina, respectively.[7]

Average lifespan, conditional on survival to age 10, varies for each state, from being under 40 in Alabama (38.73), over 40 in Missouri (43.8), under 50 in Illinois (48.90), to nearly 60 in North Carolina (58.79). For the deaths we observe, which, on average, occurred between 1925 and 1945, the average person was born in the late nineteenth century. The summary statistics reveal some differences by race in each state. Whites could expect to live more than four years longer than African Americans, on average, in Alabama and Missouri and

[5]Table A1 in the appendix gives summary statistics for length of life by state.

[6]Importantly, the analysis here includes a state, Missouri, which was not used in Cook, Logan, and Parman [2014]. As such, the distinctive name pattern documented there holds in an independent data source with a different racial composition.

[7]The disproportionality of the names is similar to that seen in modern analysis of black names [Bertrand and Mullainathan 2004; Fryer and Levitt 2004].

more than a decade longer than African Americans in Illinois and North Carolina.[8] These estimates agree broadly to other summary measures of the population of each state for the early twentieth century.[9]

3.2 Distinctively Black Names and Mortality

Table 3 shows the estimates of the regression models described above. We analyze each state separately due to different time periods covered and to give easily interpreted results of the within-state black mortality differential due to black names. We also control for county of death to act as a control for later-life geographic factors. The results show that having an African American name (column I) increases the lifespan by more than three years in Alabama (3.48), two years in Illinois (2.48) more than seven years in Missouri (7.52) and nearly four years in North Carolina (3.93). All of the estimates are statistically significant at all conventional levels.

Column II uses the semi-logarithmic specification, and the general pattern seen in Column I is present. The elasticity, the percentage increase in the length of life due to the presence of a distinctively African American name, ranges from a low of five percent percent in Illinois to nearly 17.9 percent in Missouri.[10] Column III shows the estimates of the hazard model, which shows that the hazard of mortality was substantially lower for those with a distinctive African American name. The estimates range from a fifteen-percent (in Illinois) to a thirty-three-percent (in Missouri) decrease in the hazard of mortality due to an African

[8]Full summary statistics are provided in the appendix.

[9]See *Historical Statistics of the United States, Millennial Edition.* One question for the generalizability of our analysis is the degree to which one can draw inferences from any state to the rest of the nation. To do this we compared the deaths in the 1880 Federal Death Census to those for the nation as a whole (not reported). What we found was that the differences for death (age at death and differences by sex and race) were similar for each state by region. It is true, however, that Southern states were different from the rest of the nation. Importantly, we found no difference in white mortality (age at death) for the Carolinas or Alabama when compared to the rest of the nation. There are differences for blacks when compared to the rest of the nation, but blacks in Alabama and North Carolina are no different in their average age at death from blacks in the South more generally. Given the results from the 1880 Federal Death Census comparison, we are confident that our analysis can be extended, with some caveats, to the general pattern for the Southern United States.

[10]Note that we follow Halvorsen and Palmquest [1981] in the interpretation of a dichotomous indicator in semilogarithmic models.

American name. The correlations suggest that, if every African American man was assigned a distinctively African American name, the racial mortality gap at the time would have been cut by more than ten percent.

Table 3 shows that there was a substantial increase in the length of life correlated with having a black name. This difference in mortality is striking and quite large and adds a new dimension to the existing analysis of racial differences in mortality in the American past. Previous studies of racial differences in mortality have found significant racial differences, even by cause [Costa 2005, Costa et al. 2007], but these studies have not linked to the socioeconomic or cultural factors nor have they looked within race to uncover those racial differences. For example, Logan [2009] found differences by migratory status, but how this was linked to other factors remains unclear. We analyzed the migration issue with the North Carolina data (in keeping with choosing a state where there would have been significant out-migration during this time period). For all men born in North Carolina, the black name effect on mortality in a linear regression is 0.762 years (s.e. 0.286, t=2.66). For all men who were born outside of North Carolina who died in North Carolina, the black name effect is 0.897 (s.e. 0.369, t=2.43).

It could be the case that the hazard varies by age, such that the mortality effects we estimate above are concentrated in advanced ages. If this were the case, the average effect we estimate may be due to an error in age reporting or some other factor that would be related to ages reported in death records. We address this issue directly in Figure 1, where we show Kaplan-Meier estimates of the survival function for black men with and without a black name. The results by state show that the survival function by age is not age-variant. The increased survival rate for black men with black names is seen over the entire age structure.[11]

[11]We stress that it is unlikely that the results of Table 3 are driven by age-misreporting for those with distinctive names. As noted earlier, our coefficient of interest is for those who were named on death records and who were African American males. If racial names are correlated with extended family networks, those with distinctive names would be *more* likely to have correctly reported ages at death as they would likely have more local representatives who could report accurate years of birth. Since the stylized fact is that African Americans are assumed to be older than they truly are at death, any effect of age misreporting would go in the opposite direction. This implies that our mortality estimates for names could be a lower bound estimate.

3.3 Infant and Child Mortality

Given the results in Table 3, it is natural to ask if the effects extend to infant and child mortality.[12] Table 3 also shows the estimates of the regression models where the dependent variable is the length of life conditional on dying before age 11. In all specifications, the effect of a distinctive name is relatively small and is not statistically significant in all states. In the regression estimates (column IV), the effect of a distinctive name is anywhere from a negative impact of three quarters of a year of life (in Missouri) to an increase of one year (in Illinois). The elasticity estimates (column V) range from less than 2 percent (in Alabama) to more than 25 percent (in Illinois). In contrast to the estimates of later mortality, there is considerable heterogeneity in the estimates.[13]

One additional question, related to infant and child mortality, would be to investigate whether men with black names were less likely to die before age 10. To investigate this issue we estimated a linear probability model where the dependent variable was an indicator for survival to age 10. For Alabama the coefficient on black name was 0.0501 (s.e. 0.00824, t=6.07), for Illinois the coefficient on black name was 0.0839 (s.e. 0.00817, t=10.27), for Missouri the coefficient was 0.0506885 (s.e. 0.3967, t=1.28), and for North Carolina the coefficient was -0.0033763 (s.e. 0.003947, t=-0.86). Overall, the results imply that men with black names were less likely to die before age 10 in Alabama and Illinois, but that there was no effect in Missouri and North Carolina. One issue with this interpretation is that many children are unnamed if they died at particularly young ages, and the proper counterfactual would need to account for the names that would have been assigned to children who died

[12] As we described earlier, if the result is due to cumulative effects we would suspect that distinctive names would confer few advantages at early life, where mortality is more likely due to exogenous factors such as disease environment. This would be especially true during this time period. On the other hand, there could be effects after infancy yielding effects for early life mortality. While there are known racial differences in child mortality in the past [Costa 2004], there is nothing to suspect that the effects of names would be large at young ages. There is little evidence of intraracial differences in infant mortality in the past. Gutman [1976] describes the usual practice of naming children after their deceased siblings. If this trend continued during the time period we study, then we would be more likely to observe deaths to distinctive names if children born in earlier cohorts were subject to a high-infant mortality environment.

[13] Figure 1 also shows these effects. Part of this could be due to differences in infant mortality coverage in death records, the preponderance of unnamed children in death records, or other measurement issues.

before they were named. In general, however, the name effect is not present at early ages in a systematic way and does not exhibit the robustness of the later-life mortality relationship. There is some evidence that black names are related to survival before age 10.

4 Considering Possible Mechanisms

4.1 Name Selection

It could be the case that the name effect is a figment of selection. If those with distinctive names who experienced positive outcomes chose to retain their names while others discarded them, the result could be endogenous. This would be an interesting fact as the current literature discusses the ways that African Americans attempt to avoid the racial stigma of black names. If African Americans in the past chose to use those names due to positive outcomes that would, in and of itself, be worthy of note. The historical record, however, does not provide any evidence of African Americans adopting different first names after the Reconstruction era. In fact, the very lack of any literature documenting this practice suggests that it was rare.[14] To the extent that the races separated after the Reconstruction era [Woodward 1955], fewer interracial interactions would have given African American less incentive to change names, and it is unclear why those with the most distinctive African American names would retain them.

Overall, unlike the literature on European immigrants, whose name changes during the late nineteenth and early twentieth centuries are well documented, there is no evidence that African Americans did the same. While the lack of a literature on this subject does not mean it did not occur, the lack of a discussion stands in stark contrast to the literature on name changes after the Civil War and the literature on racial passing. Similarly, the literature on name changes consistently shows that name changes were made by those seeking to avoid the

[14]While Litwack [1979] describes the power of names and the ability to choose names after emancipation, the later history of African American life in the South provides no discussion of this issue [Litwack 1998, Hahn 2003, Ritterhouse 2006, Hale 1998].

stigma of an ethnic name. The fact that our result points to a positive effect of a racialized name casts further doubt on the notion that a significant fraction of the result would be due to selection itself. If those who ceased to use their racialized given names in their lives did so in hopes of better life outcomes, those hopes appear to have been misplaced.

4.2 Socioeconomic Effects

We consider the implications of wealth and socioeconomic effects more generally in Table 4. As in Cook, Logan and Parman [2014], we do not find differences in occupation due to having a black name. There are differences in education, but those differences show that those with black names were *less* likely to be literate and *less* likely to be enrolled in school. The direction of these effects would run counter to the mortality differences we find. To the extent that these names reflected longstanding family structures we would expect there to be a positive effect of socioeconomic status due to the names. We would predict that those with black names would be more likely to come from intact families, for example, since the names we analyze are passed down from father to son. We find no differences in family size, school attendance, literacy, single parent households, or occupation (adult or child) that are related to black names in Table 4. Another factor could be migration, a primary means of investing in human capital in the past. We also analyzed the census data to see if migration was related to black names. In a linear probability model, the effect of having a black name on migration (pooled 1900-1920, with year fixed effects) was -.000283 (s.e. 0.00113, t=-0.25) for all men over the age of 15. In 1910 the effect was -0.0012559 (s.e. 0.00245, t=-0.51) and in 1920 the effect was -0.002306 (s.e. 0.00483, t=-0.53). This evidence suggests that having a black name was not related to interstate migration.

This lack of differences in socioeconomic outcomes runs counter to the findings in Goldstein and Stecklov [2014], who find that black names were associated with lower socioeconomic status via occupation. One reason why our results may differ from theirs is that they use a names index which rates each individual name by its distinctiveness. Unlike the

method used in Cook, Logan, and Parman [2014], a names index is not conditioned on frequency/commonness. Empirically, it is difficult to distinguish unique name effects and racial name effects. Both would have the same index value.[15] Extending the previous analysis, we also consider the demographic differences that could be attributed to names in Table 4. Here, we do find some important distinctions. First, we find that black men with black names have fewer children, on average. Also, black men with black names are less likely to be (currently) married and more likely to never have been married. We find that there are no distinctions by black name on being divorced, however. The differences in demographic outcomes that we find here by name are a new finding that leads to a range of questions. The widower results appear to be consistent with the mortality estimates, and yet the marriage and never-married results suggests that men with black names had different family structure. Given the lack of an occupational difference, it is difficult to believe that a standard marriage-market explanations would apply. On the other hand, the results for literacy and school attendance suggest that men with black names may have been less valued as spouses. This, too, stands in contrast to the divorce results– we would expect lower-quality marital matches to be more likely to divorce.

4.3 Family/Cultural Effects

We explored the potential for a cultural transmission by analyzing the first names of fathers and sons in the death certificate data where both were available. The death records for Alabama have digitized father and son names. Those with a distinctively black name were no more or less likely to have a missing father in the death records. For black men with black names, 55.11% had fathers names in the records, and 56.83% of black men without

[15] Another problem of using such an index to uncover within-race differences is that, by construction, the majority of blacks will have high index value names relative to whites. The slope of the relationship over all names may obscure the difference within race which is our primary interest. As such, even when estimating a relationship (such as income) that varies by race, the inclusion of the name index and race may yield a biased estimate. We explicitly estimate the effect of a black name *within* race to avoid such confounders. An additional difference is that our measure is at the extensive margin, since we use a relatively small set of names that have been verified in non-census data sources and which we establish were common among African American men at the time.

black names had fathers listed in the records. We then looked to see if fathers who had black names were more likely to have sons who also had black names.[16] We found that for black men who had black names, 13.95% had a father with a black name, while only 1.58% of black men without black names had a father with a black name. We also see this in the reverse. For black fathers with black names in the Alabama data, 12.67% had a son with a black name, while only 1.42% of other black fathers had black sons in the data with black names. We view this as further evidence of family-cultural effects and evidence of intergenerational transmission of black names.

We also modified the regressions for Alabama to include a black father name to see if a father having a black names was related to a longer lifespan. We found that a men with fathers who had black names (whether they were black named themselves) did live longer than others. The elasticity estimates suggest that having a black father resulted in an 8.95% longer lifespan, even when controlling for whether the deceased himself had black name himself. In effect, it does appear as if the names are related to intergenerational transmission, are perhaps cumulative, and reflect an omitted cultural factor which is related to within-race mortality differences. This suggestive evidence is difficult to reconcile with the results in Table 4, because we find few socioeconomic effects there. If there is a cultural mechanism at play, it does not appear to operate through socioeconomic status. Although there appears to be a family-cultural effect, it is difficult to see how it produces the mortality effects we find.

5 Conclusion

We find that that having a distinctively African American name was strongly correlated with mortality. Our estimates imply that those with distinctively African American names

[16] We did so because the naming pattern suggested by Gutman suggests that male family members could be named for fathers and other male relatives. As such, the mechanism behind this result would be the same (family/cultural effects), and restricting the analysis to father-son name matching results in a much smaller sample size in the Alabama records.

lived nearly a year longer than other African Americans. This difference is striking in that it is a within-race and within-gender effect. Unlike the negative outcomes associated with black names today, we find a large and positive effect from having a black name in the past. The literature gives us few strong clues about potential mechanisms at this point, and we therefore argue that this robust correlation should be a springboard for future work into socioeconomic differences in mortality among African Americans. As this exploration has shown, concentrating on interracial differences can obscure important and neglected intraracial differences in outcomes.

At present, the existing analysis suggest that the result is likely not due to either selection on names nor to a wealth effect. Uncovering the mechanism underlying the mortality result will require a wealth of additional empirical evidence as well. If it is cultural factors that explain this result, this implies that empirical evidence must come from both quantitative and qualitative sources. This would not only include more quantifiable data but also detailed narrative analysis of the names of prominent individuals, analysis of church registers, lists of African Americans in prestigious occupations, and the like. That this hitherto unknown fact appears to have such a large effect on mortality suggests that there are likely several pieces of the African American demographic experience which remain hidden from contemporary scholarship and which require serious and sustained investigation. The discovery of the specific causes of this relationship will go hand in hand with the development of the nascent literature on the political and social histories of African Americans in the late nineteenth and early twentieth centuries which could uncover further robust within-race differences in outcomes. This period has been relatively neglected in quantitative historical and demographic scholarship, and findings such as the mortality relationship presented here should stimulate further research into this period of American history.

References

[1] Bertrand, M. and S. Mullainathan (2004). "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination" American Economic Review 94: 991-1013.

[2] Busse, T.V. and L. Seraydarian (1977). "Desirability of First Names, Ethnicity and Parental Education." Psychological Reports 40: 739-742.

[3] Cook, L.D., T.D. Logan and J.M. Parman (2014) "Distinctively Black Names in the American Past." Explorations in Economic History 53: 64-82.

[4] Costa, D. L. (2004) "Race and Pregnancy Outcomes in the Twentieth Century." Journal of Economic History 64: 1056-1086.

[5] Costa, D.L. (2005) "Race and Older Age Mortality: Evidence from Union Army Veterans." NBER Working Paper No. 10902.

[6] Costa, D.L., L. Helmchen and S. Wilson (2007) "Race, Infection, and Arteriosclerosis in the Past." Proceedings of the National Academy of Sciences 104: 13219-13224.

[7] Figlio, D.N. (2005). "Names, Expectations and the Black-White Test Score Gap." NBER Working Paper No. 11195.

[8] Fryer, R.G. and S.D. Levitt. (2004). "The Causes and Consequences of Distinctively Black Names" Quarterly Journal of Economics 119: 767-805.

[9] Ginther, D.K., W.T. Schaffer, J. Schnell, B. Masimore, F. Liu, L.L. Haak, and R. Kington (2011). "Race, Ethnicity, and NIH Awards." Science 333: 1015-1019.

[10] Goldstein, J. and G. Stecklov (2014). "Contours and Consequences of Black First Names in the Historical United States." Working Paper, University of California, Berkeley.

[11] Gutman, H.G. (1976). The Black Family in Slavery and Freedom, 1750-1925. New York: Vintage.

[12] Hahn, S. (2003). A Nation Under Our Feet: Black Political Struggles in the Rural South from Slavery to the Great Migration. Cambridge: Harvard University Press.

[13] Hale, G.E. (1998). Making Whiteness: The Culture of Segregation in the South, 1890-1940. New York: Vintage.

[14] Halvorsen, R. and R. Palmquist (1980). "The Interpretation of Dummy Variables in Semilogarithmic Equations" American Economic Review 70: 474-475.

[15] Litwack, L.F. (1979) Been in the Storm so Long: The Aftermath of Slavery. New York: Knopf.

[16] Litwack, L.F. (1998) Trouble in Mind: Black Southerners in the Age of Jim Crow. New York: Knopf.

[17] Logan, T. D. (2009) "Health, Human Capital, and African American Migration Before 1910." Explorations in Economic History 46: 169-185.

[18] Milkman, K.L., M. Akinola and D. Chugh (2012) "Temporal Distance and Discrimination: An Audit Study in Academia." Psychological Science 23: 710-717.

[19] Parman, John (2012). "Gender and Intergenerational Mobility: Using Health Outcomes to Compare Intergenerational Mobility Across Gender and Over Time." Working Paper 122, College of William and Mary Department of Economics Working Paper Series.

[20] Ritterhouse, J.L. (2006). Growing Up Jim Crow: How Black and White Southern Children Learned Race. Chapel Hill, NC: UNC Press.

[21] Sanders, S. and M. Muszynska (2009) "The Great Migration and Mortality of African Americans." Abstract for the Population Association 2010 Annual Meeting.

[22] U.S. Bureau of the Census (1902). Mode of Statement of Cause of Death and Duration of Illness upon Certificates of Death. Washington, DC: Government Printing Office.

[23] Woodward, C.V. (1955). The Strange Career of Jim Crow. New York: Oxford.

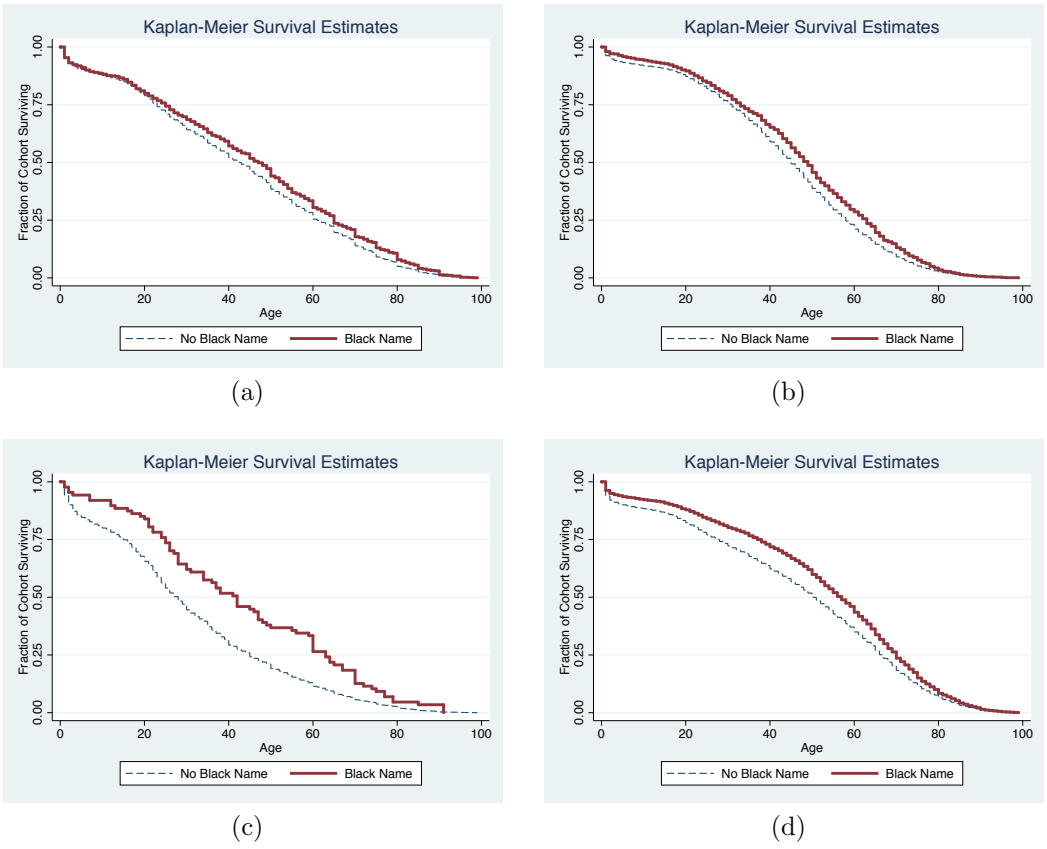

Figure 1: Survival Function Estimates for African American Men with and without African American Names in: (a) Alabama; (b) Illinois; (c) Missouri; and, (d) North Carolina.

Table 1: The Historical African American Names

Historical African American First Names
Abe
Abraham
Alonzo
Ambrose
Booker
Elijah
Freeman
Isaac
Isaiah
Israel
King
Master
Moses
Pearlie
Percy
Perlie
Purlie
Presley
Presly
Prince
Titus

Table 2: Summary of Sample Sizes for Death Certificates by State

	Alabama	Illinois	Missouri	North Carolina
Year range for death certificates	1908-1959	1916-1947	1802-1910	1910-1970
Number of observations	309,121	1,533,135	86,696	1,256,111
Percentage who are African American	44.5%	5.5%	8.2%	32.7%
Percentage of African American individuals with an African American name	1.7%	1.4%	1.7%	1.3%
Percentage of white individuals with an African American name	0.6%	0.7%	0.6%	0.4%

Note: Data include all males in the death certificate records with race and first name reported. African American names are those given in Table 1.

Table 3: The Correlation of African American Names with Mortality

Dependent variable:	Lifespan	Log Lifespan	Lifespan	Lifespan	Log Lifespan
Estimation method:	OLS	OLS	Hazard	OLS	OLS
Includes individuals surviving to age 10:	X	X	X		
Includes individuals dying by age 10:				X	X
	I	II	III	IV	V
			Panel A: Alabama 1908-1959		
African American Name	3.476***	0.073***	-0.165***	0.363***	0.011
	[0.554]	[0.013]	[0.024]	[0.121]	[0.056]
			(0.848)		
Constant	1,113***	1,174***		1,111***	1,172***
	[1.127]	[1.127]		[6.335]	[6.192]
Observations	90,581	90,581	90,581	31,667	12,635
R-squared	0.021	0.025	---	0.012	0.011
			Panel B: Illinois 1916-1947		
African American Name	2.478***	0.047***	-0.159***	1.110***	0.264**
	[0.560]	[0.013]	[0.030]	[0.285]	[0.107]
			(0.853)		
Observations	65,248	65,248	65,248	14,214	5,702
R-squared	0.051	0.052	---	0.003	0.006
			Panel C: Missouri 1802-1910		
African American Name	7.522***	0.179***	-0.328***	-0.750**	0.211
	[2.594]	[0.063]	[0.104]	[0.354]	[0.335]
			(0.720)		
Constant	1,113***	1,174***		1,111***	1,172***
	[1.127]	[1.127]		[6.335]	[6.192]
Observations	3,332	3,332	3,332	1,930	764
R-squared	0.007	0.014	---	0.017	0.011
			Panel D: North Carolina 1910-1970		
African American Name	3.928***	0.085***	-0.174***	0.509***	0.101**
	[0.290]	[0.006]	[0.014]	[0.090]	[0.041]
			(0.840)		
Observations	290,853	290,853	290,853	100,491	38,934
R-squared	0.055	0.062	---	0.002	0.004

Robust standard errors in brackets *** p<0.01, ** p<0.05, * p<0.1

Note: Regression samples are restricted to black males. All models include controls for year of death and year of death squared. For hazard models, exponentiated coefficient estimates are in parentheses.

Table 4: Socioeconomic Correlates of African American Names

	(I)	(II)	(III)	(IV)	(V)
Outcome	Occupational Score	Literate	Attends School	Parent Not in Household	Number of Siblings
Black Name	-0.1813	-0.0224*	-0.0222*	0.00213	-0.145**
	-0.147	(0.0124)	(0.0121)	(0.0114)	(0.0718)
Observations	1,931,557	1,931,557	812,181	1,160,244	1,160,244
R-squared	0.228	0.156	0.286	0.062	0.060

	(VI)	(VII)	(VIII)	(IX)	(X)
Outcome	Number of Children	Married	Divorced	Widowed	Never Married
Black Name	-0.141**	-0.0455***	-0.000596	0.00831*	0.0378***
	(0.0558)	(0.00934)	(0.00131)	(0.00443)	(0.00762)
Observations	1,931,557	1,931,557	1,931,557	1,931,557	1,931,557
R-squared	0.239	0.333	0.003	0.109	0.435

Note: Data is males in 1900, 1910, and 1920 IPUMS sample. All standard erros are clustered at the state level. All regressions include state fixed effects, age, age squared, age cubed, and census year. * p<0.1; ** p<0.05; *** p<0.01
Sample Restrictions -- Occupational Score: Only those above age 15. Literacy: Only those above age 15. Attends School: Only those aged 5-18. Missing Parent: Only those under age 18. Number of Siblings: Only those under age 18. Number of Children: Only those above age 15. Married, Divorced, Widowed and Never Married: Only those above age 15. Literacy, Attends School, Missing Parent, Married, Divorced, Widowed and Never Married are estimated with linear probability models. All regressions are estimaed by OLS.

A Appendix

A.1 Data from Death Records

We use the death records from four states: Alabama, Illinois, Missouri, and North Carolina. These are the only four states that have sizable numbers of death records available for the time period of interest with both name and race digitized. Each state had different death registration histories, different racial compositions, and cover different regions of the country. The basic information about death, cause of death, age at death, occupation, and parental information is available for all years for all four states. However, the time spans and underlying sources of the records vary across the states.

The Alabama records are drawn from the Alabama Deaths and Burials Index created by the Genealogical Society of Utah for the years 1881 to 1959. For the early years, the index is drawn from multiple sources including church, civil and family records of Alabama deaths and burials. Beginning with 1908, state law required that all deaths within the state be registered with death certificates being filed with the Alabama Center for Health Statistics. A compliance rate of 90 percent was achieved by 1925 at which point the state was admitted to the federal death registration area, an indication that the state had achieved a high standard of performance in registration standards.[17] The index for 1908 through 1974 is based on these death certificates.

The Illinois names are drawn from all of the available records in the Illinois deaths and stillbirths index for 1916 to 1947. This index includes information transcribed from one-page pre-printed death certificate forms. The 1916 start date for the records is the result of a 1915 statute that required the State Board of Health (succeeded by the Illinois Department of Public Health) and county clerks to record deaths and stillbirths. To aid in achieving high compliance rates, the statute created a system of financial incentives for registrars. Statewide compliance with this statue was at 95 percent by 1919.

[17] A brief history of the death registration area including the years in which states were admitted to the area is available in the Census Bureau's *Physicians' Handbook* [1939].

The death records for the state of Missouri are taken from the Missouri Birth and Death Records Database maintained by the Missouri Secretary of State's office. The database contains information from individual death certificates transcribed from microfilm stored at the Missouri State Archives. The database consists of over 185,000 death records. The death records extend back to the early 1850s but widespread coverage does not begin until the early 1880s when Missouri passed legislation requiring the Board of Health to supervise the registration of births and deaths.[18] The coverage of the database extends up to 1909. In this year, state legislation introduced mandatory statewide collection of death certificates with the records maintained by the Missouri Bureau of Vital Records. By 1911, collection was sufficiently uniform for Missouri to be admitted to the federal death registration area. While these post-1909 records have been partially transcribed, the transcribed information does not contain the year of birth or age at death data required for this study.

The North Carolina data are constructed from the universe of death certificates for individuals who died between the years of 1910 and 1975.[19] The upper end of this range is determined by the availability of publicly available digitized death certificates. The lower end of the range is chosen such that most individuals will have fully recorded death certificates, and as such we start our period after the standardization of causes of death. Before this time death registrations and the policies related to death registrations were not uniform.[20]

Summary statistics for the death records in each state are provided in Table A1.

A.2 Advantages and Disadvantages of Death Records

The advantages of death certificates for analyzing the relationship between names and mortality are numerous. First, death certificates are person-specific records while census enumer-

[18]Despite this supervision, there were still many problems with non-compliance. This can be seen in the data with several counties either not appearing in the database at all or having far fewer deaths than expected given the county populations and historical mortality rates. It is also confirmed by Missouri's legislative history.

[19]This process is more fully described in Logan and Parman [2014].

[20]Even once the death registration was standardized, compliance still varied. North Carolina would not achieve the standard of registration performance needed to be admitted to the federal death registration area until 1916.

ation is household-based. Second, for each set of death records that we use death certification was required early in the twentieth century, so those individuals born in the late nineteenth and early twentieth centuries corresponding to the period in which the names were identified in the census records (1890-1920) are also highly likely to appear in the death records.

There are disadvantages to death certificates data as well. While we can capture intrastate migration (the dominant migratory pattern early in the century), we cannot capture the effects of selective migration. While there certainly was selective migration– migrants have been shown more likely to be urban and more educated in a variety of studies, it does not appear that migration itself was related to longer life– there is no statistical difference in the mortality of black migrants versus non-migrants during the Great Migration for cohorts born 1905-1925, either overall or for age specific mortality [Black et al. 2015]. Recent work has also documented that blacks migrating out of the South during the Great Migration had *worse* socioeconomic outcomes than those who stayed within the South, counter to the conventional wisdom that migration was beneficial to black migrants [Eichenlaub et al 2010]. Furthermore, we concentrate on *within* race differences in mortality. Unless one could argue that a distinctly black name was strongly related to the probability of migration (which itself could be investigated in subsequent work) our mortality results would not be influenced by migration itself. This also helps to avoid the thorny issue of age-misreporting in death registration data.[21] Key for us is that fact that race of the deceased is known and observable

[21] An additional concern about the quality of death certificate data is the fact that ages at death are known to be biased for the African American population. Birth and death registrations early in the century are incomplete and official counts of the African American population and number of deaths in that population are known to be biased [US Department of Health, Education, and Welfare 1956, Eblen 1974, Coale and Rives 1973, Elo 2001, Elo and Preston 1994, Preston at al. 1998, Rosenberg et al. 1999, Zelnik 1969]. Researchers have also documented significant measurement error in black ages among the aged, making inference about racial differences in older age mortality, precisely where mortality is concentrated, difficult [Elo et al. 1996, Hill et al. 1997]. Unfortunately, demographic research cannot escape the racial stain of the past: while whites are found to have extremely low rates of age misreporting and generally excellent population coverage throughout the twentieth century [Rosenwaike and Logue 1983, Hill et al. 2000], our historical demographic data on the African American population is lacking [Ewbank 1987]. For example, Elo [2001] notes that there exist no official lifetables for the black population from 1935 to 1970. The "mortality crossover," where at older ages the mortality of blacks has been shown to be *lower* than whites, has been challenged as a figment of age misreporting among the African American population [Coale and Kisker 1986, Preston et al. 1996, Rosenwaike and Hill 1996]. Others, however, argue that the finding is robust and extends to specific causes of death for the late twentieth century [Lynch et al. 2003, Eberstein and Nam

at the time of death and therefore the name of the deceased is not a signal of race itself.

Another concern would be that the death records used here also formed part of the independent verification of the names in Cook, Logan, and Parman [2014]. That is, the death records from Alabama, Illinois, and North Carolina were used to determine the distinctiveness of black names. To guard against the possibility of an spurious correlation, we include a fourth state, Missouri, that was not used to confirm name distinctiveness. As such, the Missouri death records serve as an additional check of the black naming pattern and a check for the mortality effects.

2008].

References

[1] Black, D., S. Sanders, E. Taylor, and L. Taylor (2015). "The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South" American Economic Review 105: 477-503.

[2] Coale, A. and N.W. Rives (1973). "A Statistical Reconstruction of the Black Population of the United States, 1880-1970: Estimates of True Numbers by Age and Sex, Birth Rates, and Total Fertility." Population Index 39: 3-36.

[3] Eberstein, I. and C. Nam (2008) "Causes of Death and Mortality Crossovers by Race." Biodemography & Social Biology 54: 214-228.

[4] Eblen, J.E. (1974) "New Estimates of Vital Rates of United States Black Population During the Nineteenth-Century." Demography 11:301–319

[5] Eichenlaub, S.C., S.E. Tolnay and J.T. Alexander (2010) "Moving Out but Not Up: Economic Outcomes in the Great Migration." American Sociological Review 75: 101-125.

[6] Elo, I.T. (2001). "New African American Life Tables from 1935-1940 to 1985-1990." Demography 38: 97-114.

[7] Elo, I.T. and S.H. Preston (1994). "Estimating African American Mortality from Inaccurate Data." Demography 31: 427-258.

[8] Elo, I.T., S.H. Preston, I. Rosenwaike, M. Hill and T. Cheney (1996). "Consistency of Age Reporting on Death Certificates and Social Security Records Among Elderly African Americans." Social Science Research 25: 292-307.

[9] Ewbank, D.C. (1987) "History of Black Mortality and Health Before 1940." Milbank Quarterly 65(S1):100–28

[10] Hill, M.E., S.H. Preston, and I. Rosenwaike (2000). "Age Reporting Among White Americans Ages 85+: Results of a Record Linkage Study." Demography 37: 175-186.

[11] Hill, M.E., S.H. Preston, I.T. Elo, and I. Rosenwaike (1997). "Age-Linked Institutions and Age Reporting Among Older African Americans." Social Forces 75: 1007-1030.

[12] Logan, T.D. and J. Parman (2014). "The Dynamics of African American Health: An Historical Perspective." Review of Black Political Economy 41: 299-318.

[13] Lynch, S.M., J.S. Brown, and K.G. Harmsen (2003). "Black-White Differences in Mortality Compression and Deceleration and the Mortality Crossover Reconsidered." Research in Aging 25: 456-483.

[14] Preston, S.H., M. Hill, and G.L. Drevenstedt (1998). "Childhood Conditions that Predict Survival to Advanced Ages Among African-Americans." Social Science and Medicine 47: 231-246.

[15] Preston, S.H., I.T. Elo, A. Foster, and H. Fu (1998). "Reconstructing the Size of the African American Population by Age and Sex, 1930-1990." Demography 35: 1-21.

[16] Preston, S.H., I.T. Elo, I. Rosenwaike, and M. Hill (1996). "African American Mortality at Older Ages: Results from a Matching Study." Demography 35: 1-21.

[17] Rosenberg, H.M., J.D. Maurer, P.D. Sorlie, N.J. Johnson, M.F. MacDorman, D.L. Hoyert, J.F. Spitler, and C. Scott (1999). Quality of Death Rates by Race and Hispanic Origin: A Summary of Current Research, 1999. Washington, D.C.: US Government Printing Office.

[18] Rosenwaike, I. and M.E. Hill (1996). "Accuracy of Age Reporting Among Elderly African Americans: Evidence of a Birth Registration Effect." Research on Aging 18: 310-324.

[19] Rosenwaike, I. and B. Logue (1983). "Accuracy of Death Certificate Ages for the Extreme Aged." Demography 16: 279-288.

[20] U.S. Bureau of the Census (1939). Physicians' Handbook on Birth and Death Registration. Washington, DC: Government Printing Office.

[21] U.S. Department of Health Education and Welfare, National Office of Vital Statistics (1956). "Death Rates by Age, Race, and Sex, United States, 1900–1953" Washington, DC: Government Printing Office.

[22] Zelnik, M. (1969). "Age Patterns of Mortality of American Negroes 1900-02 to 1959-61." Journal of the American Statistical Association 64: 433-451.

Table A1: Summary Statistics from Death Records

Alabama 1908-1959

	African Americans		African American Names	
	Mean	s.d.	Mean	s.d.
Lifespan	36.01	27.56	41.15	28.22
Year of Birth	1888.75	29.27	1883.05	30.12
Year of Death	1924.94	13.32	1924.33	28.22

Illinois 1916-1947

	African Americans		African American Names	
	Mean	s.d.	Mean	s.d.
Lifespan	39.40	23.74	46.11	22.21
Year of Birth	1893.41	23.66	1886.93	22.15
Year of Death	1932.53	8.81	1932.71	8.92

Missouri 1802-1910

	African Americans		African American Names	
	Mean	s.d.	Mean	s.d.
Lifespan	22.63	23.51	33.54	27.77
Year of Birth	1859.89	26.92	1853.20	29.87
Year of Death	1885.24	16.66	1889.90	11.43

North Carolina 1910-1970

	African Americans		African American Names	
	Mean	s.d.	Mean	s.d.
Lifespan	40.51	46.68	44.26	40.45
Year of Birth	1903.27	46.92	1901.53	40.55
Year of Death	1943.34	18.74	1945.52	18.45