Social Science History (2022), **46**, 671–691
doi:10.1017/ssh.2022.7

CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

"You're Not from around Here": Regional Naming and Life Outcomes

Alex Beaudin, Elizabeth Kristian, John Robert Warren* and Jonas Helgertz

Department of Sociology, University of Minnesota, Minneapolis, MN, USA
*Email: warre046@umn.edu

(Received 3 April 2020; revised 25 November 2020; accepted 19 January 2021; first published online 25 March 2022)

Abstract

We examine the socioeconomic consequences of discrimination against people of Southern origins during the US Great Migration of the first half of the twentieth century. We ask whether people living in the American North and Midwest in 1940 fared worse with respect to education, occupation, and income if they were perceived to be of Southern origins. We also assess variation in these effects across racial groups and across actual region of origin groups. Using linked data from the 1920 and 1940 US censuses, we compare the life outcomes of about half a million pairs of brothers who differed with respect to the regional origin implied by their first names. For both Whites and Blacks, we find statistically significant associations between outcomes and the regional origin implied by names; regardless of where they were born, men living in the North or Midwest in 1940 did worse if their names implied Southern origins. However, these associations are entirely confounded by family-specific cultural, socioeconomic, and other factors that shaped both family naming practices and life outcomes. This finding—that regional discrimination in the early-twentieth-century United States did not happen based on names—contrasts sharply with findings from research in more recent years that uses names as proxies for people's risk of exposure to various forms of discrimination. Whereas names are a basis for discrimination in modern times, they were not a basis for regional discrimination in an era in which people had more immediate and direct evidence about regional origins.

Keywords: Siblings; education; occupation; wages; discrimination

Introduction

A new era of national identity was emerging in the United States at the beginning of the twentieth century. Between the end of the Civil War and the beginning of the Great Migration, established cultural systems were experiencing both rapid expansion and worsening division. On one hand, the Great Migration increased national labor mobility from the South to the North and Midwest, increasing the cultural exchange between the two regions (Boustan 2008). On the other hand, the Civil War had

© The Author(s), 2022. Published by Cambridge University Press on behalf of the Social Science History Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

aggravated the cultural and ideological differences between regions in a way that lead to a mentality of resentment and isolation in all regions.[1] It comes as no surprise, then, that Southern migrants to the North and Midwest often encountered hostilities along with the new economic opportunities found there (Hume 1977).

What were the social and economic effects of those hostilities for people perceived to be regional "outsiders" in the North and Midwest? To what extent did prejudice against regional outsiders affect people's socioeconomic outcomes? Specifically, we address two questions in our research: First, did people living in the North or Midwest do worse if they were perceived to be of Southern origins (regardless of their actual regional origins)? Second, how did the effects of being perceived to be of Southern origins matter differently for White and Black individuals?

Background

In many ways the economic effects of the Great Migration have been comprehensively studied (Alexander et al. 2017; Boustan 2008; Eichenlaub et al. 2010; Gardner 2016; Smith and Welch 1989). In general, migration was economically beneficial for migrants' economic circumstances through a combination of higher salaries and more prestigious jobs coming within reach. However, this prosperity was often limited along racial lines. For instance, Gardner (2016) found that despite individual increases, the salary of the average Black man in the northern states went down during the Great Migration due to a combination of two factors. First, while Black migrants to the North did make more than they had in the South, they were still kept from many of the higher paying jobs that White migrants dominated, making their income lower overall (Boustan 2008). Second, Blacks already residing in the North experienced a decrease in pay due to the sudden influx into the labor market of Black labor from the South (Eichenlaub et al. 2010). In short, more Black men were competing for a limited number of jobs, leaving many under- or unemployed. Although much of the research literature focuses on Blacks, the Great Migration involved the mass movement of both Whites and Blacks from the South to the North and Midwest. While not as well documented, one can imagine that the White population may have experienced similar, if less extreme, changes in their economic circumstances during this period. With this in mind, resentment and hostility by Northerners toward Southern migrants—or, at least, people perceived to be of Southern origins—is not unexpected.

However, this socioeconomic shift would not have been the only source of tension at the time. While the surrender of the South in 1865 meant the political reunification of the country, distrust and hostility both persisted and grew in the years after the war. The inevitable resentment in the aftermath of a civil war was also aggravated by an ineffective and culturally insensitive reconstruction effort on the part of the North, which delayed the South's economic recovery. The deep suspicion and hostility that this created would persist for generations. One study of

[1]Northern states included Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont, New Jersey, New York, and Pennsylvania. Midwestern states included Illinois, Indiana, Michigan, Ohio, Wisconsin, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota. Southern states included Delaware, the District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia, Alabama, Kentucky, Mississippi, Tennessee, Arkansas, Louisiana, Oklahoma, and Texas.

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

Southern culture found that an "Us vs. Them" mentality was maintained between the North and South well into the twentieth century (McKinney and Bourque 1971). It wasn't until the 1960s, when the South saw a rapid increase in integration and industrialization, that the cultural divide became less pronounced (ibid). With these deeply ingrained hostilities, it would seem natural that any person perceived to be of Southern origins would face social and economic discrimination in the North and Midwest.

Isolating the Effects of Discrimination

Regional migrants—people who move to a region in which they were not born—differ on important life outcomes as compared to nonmigrants living in the same region. For example, Southern-born African Americans moving north in the early twentieth century were very different from African Americans already living in the North with respect to culture, skills, and socioeconomic origins. Differences in observed outcomes, hence, do not represent (just) the effects of region-based discrimination; they also reflect preexisting differences. How can we isolate the effects of region-based discrimination from the effects of cultural, skill, and socioeconomic factors that might confound the association between migrant status and life outcomes? In other words, if people living in the North or Midwest who are perceived to be of Southern origins do worse with respect to education, occupation, and earnings, how can we tell the degree to which this is due to discrimination on the basis of perceived regional origins?

Our strategy is to compare siblings, identical in terms of region of birth and family cultural and socioeconomic factors, who differ with respect to the degree to which their names signal Southern origins. Proper names are an important indicator of group membership, and can convey a great deal about social, economic, cultural, religious, racial, and regional origins (e.g., Arai and Skogman Thoursie 2009; Biavaschi et al. 2017; Busse and Seraydarian 1977; Elchardus and Siongers 2011; Laham et al. 2012; Rubinstein and Brenner 2014). By comparing individuals who are the same with respect to factors that might confound the association between migrant status and life outcomes but who differ with respect to the "Southern-ness" of their names, we seek to isolate the effects of discrimination against perceived regional outsiders. Among brothers who both lived in the North (or Midwest), if the brother with the more Southern-sounding name typically fared worse, then this is evidence of discrimination based on perceived regional origins. This is true regardless of actual regional origins, although we examine differences in effects across actual region of birth.

Why Might People with Southern-Sounding Names Have Faced Discrimination?

Previous studies have demonstrated that socially or culturally nonnormative names—names that identify a person as an "other" or "outsider"—subject the individuals who hold them to negative life outcomes relative to individuals with socially normative names. This literature has considered the effects of holding names that are nonnormative with respect to race, gender, immigrant status, and perceived income in both historical and contemporary contexts (e.g., Bertrand and Mullainathan 2004; Busse and Seraydarian 1977; Carneiro et al. 2020; Cook

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

et al. 2016; Figlio 2005; Fryer Jr. and Levitt 2004; Goldstein and Stecklov 2016). In general, because names often indicate cultural or social group membership, holding a name from a social outgroup can lead a person to experience discrimination. For example, Bertrand and Mullainathan (2004) assessed the impact of racialized names on employment outcomes by submitting nearly identical resumes, differing only on the implied race of the applicant's first name, to various employers. They found that applicants with Black-sounding first names received 50 percent fewer callbacks than those with White-sounding first names. In other words, when all other social and cultural factors were the same, an outgroup name caused individuals to experience worse outcomes compared to their otherwise similar peers.

However, no previous research has used names to identify the effects of being a (perceived) regional outsider in the United States in either historical or contemporary contexts. Our study is thus innovative in that it fills this gap in the current literature and does so in the context of one of the largest migratory movements in American history: the Great Migration. Why might people with Southern-sounding names have faced professional and social discrimination in the North or Midwest?

Given the significant economic, cultural, and political differences between the North and the South in the early twentieth century, it is fair to compare the relationship between the two regions to the relationship between two nations. In this context, we draw on previous research that has considered the effects of holding names associated with international migrants on social and economic outcomes. For example, Biavaschi et al. (2017) examined the consequences for new immigrants to America of changing their names. They found that a more "American" name provided economic benefits for immigrants living in New York City, and could even make up for the economic deficits of being an immigrant (Biavaschi et al. 2017). Similarly, Arai and Skogman (2009) found that immigrants moving to Sweden who changed their surname experienced an increase in economic outcomes in comparison to their nonconforming peers (Arai and Skogman Thoursie 2009; Goldstein and Stecklov 2016).

With all this in mind, it is reasonable to hypothesize that having a Southern-sounding name may have had detrimental effects on life outcomes among people living in the North or Midwest—regardless of their actual origins. People with more Southern-sounding names may thus have had worse life outcomes than their siblings with less Southern-sounding names. Given the unique racial dynamics of Northern and Midwest areas that received large numbers of in-migrants during the Great Migration, it is also reasonable to suppose that the effects of having a Southern-sounding name may have been more pronounced for African Americans than for Whites. That is, in the hostile context faced by African Americans, discrimination against men with Southern-sounding names may have been even more severe.

Why Having a Southern-Sounding Name Might Not Have Mattered

Our hypothesis—that having a Southern-sounding name had detrimental effects on life outcomes among people living in the North or Midwest—is based on extensive research that has found name-based discrimination on the basis of race, nativity, gender, and other attributes. However, the bulk of this research has been conducted

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

in contemporary settings. In modern America, decisions about things like college admissions and hiring often happen without face to face interactions. In such bureaucratic and impersonal contexts, those with the power to make admissions or hiring decisions often make guesses—consciously or unconsciously—about applicants' personal attributes and group memberships (e.g., race, regional origin, gender, nationality). Names can be one important source of information for making such assumptions and guesses. In that space, names—and thus assumptions about personal attributes—can lead to biases and discrimination. However, in the early twentieth century, most such decisions were made after in-person meetings. Job seekers applied in person at the factory or farm, for example, and college applicants more often relied on family and personal connections to gain admission. In these contexts, those with the power to make admissions or hiring or other consequential decisions did not have to make guesses about applicants' regional origin (or other attributes) based on their names. They met applicants in person, heard their accents, saw their manner of dress, knew their families, or observed other cultural cues directly. Consequently, even though extensive prior research conducted in more contemporary contexts implies that people with Southern-sounding names may have fared worse that their siblings with less Southern-sounding names, there is also good reason to doubt that this hypothesis will hold in the context of the early twentieth century.

Research Design

We address our research questions using data that begins with all 11,272,765 US-born males between the ages of 0 and 10 who appeared in the IPUMS version of the complete-count 1920 US Census (Ruggles et al. 2020). From the 1920 Census we obtain information about these boys' first and last names; their demographic characteristics (including age, sex, race, and place of birth); their coresident parents' demographic and socioeconomic characteristics; and their household's geographic location. We then identify those same boys in the IPUMS version of the complete-count 1940 US Census when they were between the ages of 20 and 30. From the 1940 Census we obtain information about these boys' educational attainment, occupation, and wage and salary income.

We begin by restricting our sample to the 11,047,470 boys who were children, grandchildren, stepchildren, or nephews of the head of their household in 1920; this is necessary to ensure that pairs of boys in households are siblings. We further restrict the sample to the 1,188,618 boys (and thus 594,309 pairs of siblings) whose 1920 Census records could be linked to their 1940 Census records and who lived in households in 1920 that contained at least two brothers between the ages of 0 and 10; when there were more than two age-appropriate boys in a household, we picked two of them at random. Finally, we restrict the sample to the 1,022,344 boys (and thus 511,172 pairs of boys) whose names appeared at least 100 times in the 1920 Census and who either: (1) were born in the South and lived in the South in 1940; (2) were born in the South and lived in the North in 1940; (3) were born in the South and lived in the Midwest in 1940; (4) were born in the North and lived in the North in 1940; or (5) were born in the Midwest and lived in the Midwest in 1940.

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

Linking Records

To match records across the 1920 and 1940 censuses we first defined the universe of potential matches. To make the task computationally tractable, we restricted the population of potential matches to records that displayed identical or similar characteristics on features that should be consistent over time (e.g., year and place of birth). For example, when attempting to find the 1940 Census record for Michael Corcoran—born in 1917 in Massachusetts according to the 1920 Census—we limited the population of potential matches in 1940 to males born in Massachusetts between 1914 and 1920. Because age is reported in the 1920 and 1940 censuses rather than year of birth, and because of reporting inaccuracies, we allow for deviations of ±3 years in birth year across data sources.

To identify the correct record from among all possible records in the 1940 Census—for example, the correct "Michael Corcoran" in 1940 from among all the possible Michael Corcorans—we employed a machine learning algorithm that performed probabilistic record linking techniques. Briefly, we trained a computer algorithm to recognize patterns in a dataset of potential matches that are consistent with a true match. We used a modification of Feigenbaum's (2016) probit regression approach, which—like other methods of supervised machine learning—required input from training data. The training data represents a subsample of the population that one wishes to link, but where links have been declared by a trained human to ascertain that confirmed links are as accurate as possible. We used these training data to calibrate the linking algorithm and to evaluate how well it performs in declaring matches and avoiding false positives matches. The goal was to train the algorithm to link as many records as possible while keeping the false positive match rate exceedingly low.

To construct the training data, we randomly selected 1,000 age-appropriate boys from the 1920 Census who we then attempted to link by hand to the relevant universe of possible 1940 matches. Here, the universe of potential matches was limited to cases in which name similarity scores (Jaro 1989, 1995; Winkler 1990, 2006) were at least 0.8. We assessed all potential matches by hand using the wealth of digitized historical information and search features available from Ancestry.com. Historical information about parents, siblings, and the focal individual's places of residence from the time between the 1920 and the 1940 Census allowed us to make confident assessments regarding the validity of potential matches and to minimize the rate of false positive linkages. Using these procedures, and carefully guarding against false positive matches, we were able to manually declare 50.2 percent of the training data sample as uniquely matched across 1920 and 1940.

To calibrate our linking algorithm, we implemented a "train-test-split" procedure using our training data (in which true matches are known). In the first part of the procedure, we split our training data into two equally sized parts. To train the algorithm, we fit a probit regression model on one-half of the sample, and then evaluated its out-of-sample performance on the other half. The model specification we used is similar to the one proposed by Feigenbaum (2016), but we added additional individual- and household-level covariates to reduce the risk of false positives, as also illustrated by Helgertz et al. 2020. Results from the model inform the algorithm as to which, if any, of the universe of possible matches should be considered a valid

link (e.g., which "Michael Corcoran," if any, is the one we are looking for). The algorithm declared a unique link based on threshold values (1) the greatest required similarity between any 1-to-1 match (technically the predicted probability based on the probit regression estimates) and (2) the required relative difference between the best and second-best possible match. When selecting thresholds values, the precision and recall of the resulting data are fundamental metrics to consider. Recall refers to the proportion of true matches that are declared as matches by the algorithm, whereas precision indicates the proportion of the declared matches that are true matches. If the thresholds are set very conservatively, few but accurate matches will be made, translating to a low recall and a high precision. Conversely, less conservative thresholds will result in high recall but a low precision.

Choosing thresholds that favor a high precision or a high recall is not a straightforward task, which is why we used the Matthew's Correlation Coefficient (MCC). Described in Equation 1,[2] the MCC is an especially useful measurement for two-class data where the classes are not well balanced (Chicco 2017); MCC provides a single metric of linking performance (ranging from –1 to +1) and is based on both precision and recall. By looping multiple times over a range of realistic threshold values, we were able to choose values for (1) and (2) that optimized the overall performance of the linking algorithm.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{1}$$

Weighting

Although the rates at which we matched records across data sources are very high by the standards of historical record linkage (see, e.g., Beach et al. 2016; Ferrie 1996), a threat to the external validity of our results is that the sample of boys we were able to match may differ systematically from the sample of boys we were unable to match. To address this potential problem, we constructed and employed poststratification weights that account for each boy's probability of being matched across the 1920 and 1940 censuses.

We take advantage of the fact that we observe several attributes of boys in 1920 regardless of whether we could confidently match them to a record in the 1940 Census. In Table A1 we report descriptive statistics that compare boys who could and could not be matched to their 1940 Census records with respect to race/ethnicity, place of birth, mother's literacy, father's occupation, and homeownership. Boys who could be matched to their records in 1940 are more likely to be White, to come from somewhere other than the South, and to come from more advantaged socioeconomic origins.

To construct the weights, we estimated a probit model predicting whether a boy could be matched as a function of his race, age in 1920, state of residence in 1920, mother's literacy, father's literacy, father's occupation, homeownership, and family

[2] *TP* represents true positive, *TN* represents true negative, *FP* represents false positive, and *FN* represents false negative.

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

size. We then take the inverse of the predicted probability of being linked and normalize the distribution of the resulting values to have a mean of one. Conceptually, these weights count boys who were least likely to be matched more heavily in the analyses without changing the overall sample size.

Name-Based Measures

Our key independent variable describes the percentage of boys ages 0 to 10 with each proper name who were born in a Southern state according to the 1920 Census. That is, for every boy age 0 to 10 in 1920 who had the first name "Thomas" we compute the percentage who were born in the South; we then repeat this for every first name observed among boys in this age range in the 1920 Census. As described in the preceding text, we exclude the relatively small number of boys whose first names were observed fewer than 100 times. Many of these boys had names that were abbreviations or spelling or transcription errors (e.g., "Thomqs" or "TG"). Before constructing our measure, we remove punctuation and middle names/initials (e.g., such that "Thomas A." becomes "Thomas"). We do not account for multiple spellings or nicknames. As a result, "Thomas," "Thos," "Tom," "Thom," and "Tomas" are all treated as distinct names and appear in the data (so long as they each appear at least 100 times). We also drop names that are 0, 1, or 2 characters long (because most are simply sets of initials, like "TG").

Table 1 describes the distribution of the resulting measure. Across our 1,022,344 linked boys, the average first name is held by someone born in the South about 33 percent of the time. The average African American boy's first name was held by someone born in the South about 46 percent of the time. Figure 1 depicts the distribution of the "Southern-ness" of each first name held by the 980,215 White boys (top half of figure) and the 41,081 Black boys (bottom half of figure). It shows that names that are predominantly held by Southerners are more often held by Blacks. Among the names appearing at least 100 times among boys matched between the 1920 and 1940 censuses, those that were most distinctively Southern were Grady (99% Southern born), Odis (95%), and Lonnie (93%). Those that were most distinctively Northern were Carmine (91% Northern born), Regis (87%), and Salvatore (85%).

Although our focus is on the effects of first names, we also constructed parallel measures for each boys' last or surname for the purposes of supplementary analyses described in the following text.

Dependent Variables

Our outcome measures from the 1940 Census include educational attainment (measured in years), wage and salary income (in dollars, which is unavailable for many farmers and self-employed people), and occupational prestige (Siegel 1971). As shown in table 1, there is little missing data on educational attainment or occupational prestige (and no missing data on the name-based measures); those missing on wage and salary income were not generally wage and salary workers. As a result, in each analysis we exclude the small percentage of respondents who are missing on the dependent variable.

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

Table 1. Descriptive statistics, by race

	All Boys			White Boys			Black Boys		
	Mean/%	(sd)	n	Mean/%	(sd)	n	Mean/%	(sd)	N
% with His First Name Born in the South	32.8	(17.9)	1,022,344	31.7	(17.1)	980,215	45.8	(22.6)	41,081
Born in South, Lived in South in 1940	31.0%		317,300	26.8%		262,537	82.9%		34,059
Born in South, Lived in North in 1940	0.5%		5,268	0.3%		3,172	2.9%		1,181
Born in South, Lived in Midwest in 1940	1.2%		12,422	1.0%		9,990	3.6%		1,494
Born in North, Lived in North in 1940	31.7%		323,604	33.8%		331,261	5.8%		2,388
Born in Midwest, Lived in Midwest in 1940	35.6%		363,750	38.1%		373,256	4.8%		1,958
Total	*100.0%*		*1,022,344*	*100.0%*		*980,215*	*100.0%*		*41,081*
Education (Years)	9.68	(2.92)	1,022,344	9.91	(2.77)	980,215	6.89	(3.29)	41,081
Occupation Prestige	30.24	(11.50)	934,143	30.75	(11.49)	896,055	24.02	(9.58)	37,207
Wage and Salary Income	$601.84	($616.10)	962,165	$627.50	($626.31)	923,196	$286.90	($340.35)	38,021

Note: Sample restricted to the 1,022,344 boys who could be linked across the 1920 and 1940 censuses; who were children, stepchildren, grandchildren, or nephews of the household head; who had at least one other age-appropriate brother in the household in 1920; whose first name appeared at least 100 times in 1920 Census; and who lived in one of the five combinations of regions of birth and regions of residence in 1940 considered in our analyses.

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

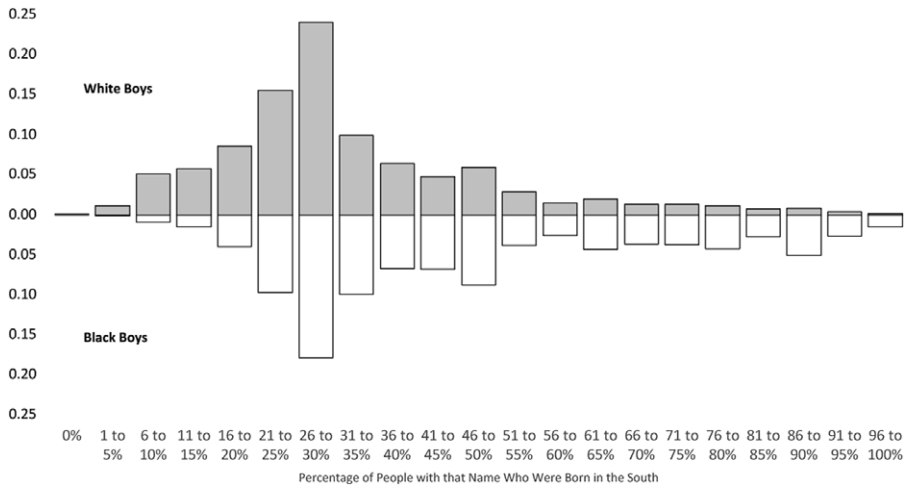

Figure 1. Distribution of Southern-ness of first names, by race, boys linked from 1920 to 1940.
Note: Sample restricted to the 1,022,344 boys who could be linked across the 1920 and 1940 censuses; who were children, stepchildren, grandchildren, or nephews of the household head; who had at least one other age-appropriate brother in the household in 1920; whose first name appeared at least 100 times in the 1920 Census; and who lived in one of the five combinations of regions of birth and regions of residence in 1940 considered in our analyses. All analyses are weighted by the inverse of the probability of selection into the linked sample.

Analytic Methods

We begin by estimating OLS regression models that express the bivariate relationship between our measures of the Southern-ness of names and each outcome. Of course, these associations are confounded by any variable that is both associated with naming patterns and consequential for our several outcomes (e.g., geography, race/ethnicity, cultural preferences, or socioeconomic status). We thus rely most heavily on sibling fixed-effects models that account for all such attributes that are shared by siblings.

To estimate the effects of the Southern-ness of names more robustly on each outcome, we use a standard fixed-effects specification for within-pair estimation:

$$M_{ij} = \alpha + \beta N_{ij} + C_j + \varepsilon_{ij}, \tag{2}$$

In Equation 2, M_{ij} is a continuous outcome measure (i.e., education, occupation, or income) for individual i ($i = 1, 2$) in sibling pair j ($j = 1, 2, \ldots, n$); N_{ij} measures attributes of the name of individual i in sibling pair j; C_j is a measure of unobserved contextual characteristics shared by both members of the pair (e.g., geography, race/ethnicity, cultural preferences, or socioeconomic status); and ϵ_{ij} is a random individual-level error term that is assumed to be uncorrelated with the other explanatory variables in the model.

In such a within-sibling pair model, the unobserved shared components C_j are "controlled away" by modeling differences within pairs of brothers:

$$M_{1j} - M_{2j} = \Delta M_j = \beta \Delta N_j + \Delta \varepsilon_j, \tag{3}$$

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

In Equation 3, the Δs represent differences between variables for the *j*th sibling pair (i.e., $N_{1j} - N_{2j}$). This approach eliminates the confounding effects of unobserved contextual characteristics (C_j)—because siblings have these things in common. In our models, we control for birth order that differs between siblings, may be related to naming practices (e.g., names for firstborn sons may be systematically different than for subsequent sons), and can be associated with our outcomes (e.g., if firstborn sons disproportionately inherit property or businesses).[3]

Both our OLS and within-sibling pair models employ the poststratification weights described in the preceding text to account for selectivity with respect to which boys could be linked across the 1920 and 1940 censuses.

Results

Table 2 describes OLS regressions of each dependent variable on our measure of the Southern-ness of boys' first names. In Models 1, 2, and 3 boys were born in the South and lived in the South, North, and Midwest in 1940, respectively. Regardless of where they lived in 1940 and generally regardless of race, boys born in the South who had more Southern-sounding names fared worse with respect to education, occupational prestige, and income in 1940. For boys who remained in the North (Model 4) or Midwest (Model 5) from birth through 1940, there was a very weak association between name Southern-ness and education or occupational prestige. An exception is that White boys who were born and remained in the North had somewhat more education and higher income the more Southern their name sounded.

The statistically significant—and sometimes sizable in magnitude—associations observed in table 2 suggest that boys with Southern-sounding names often had poorer socioeconomic outcomes in 1940. This may have been due to discrimination against people with Southern-sounding names, but it may also have been due to selection: People with Southern-sounding names may have had other attributes (e.g., actual Southern birth, African American ancestry, fewer skills, family cultural legacies) that were correlated with family naming practices and that mattered for socioeconomic outcomes in 1940. Our sibling fixed-effects models are designed to isolate the effects of discrimination against people with Southern-sounding names—and thus the effects of region-based biases.

Table 3 reports the results of sibling fixed-effects models specified in Equations 2 and 3. Note that each model adjusts for birth order; we suspect that first born boys were more likely to be given their father's (disproportionately regional) name and also to have benefited from greater socioeconomic investments by their parents in terms of educational opportunities, wealth transfer, and inheritances of farms, businesses, or other property. In table 3, the top panel pertains to educational

[3]Although we do not show the results in our tables, there are sizable birth order effects on socioeconomic outcomes. In almost all models, birth order is positively associated with educational attainment but negatively associated with both occupational prestige and wages (such that earlier-born brothers obtain less schooling but hold higher prestige jobs and earn higher wages). This is partly an age effect—firstborns are older, and thus have had more time to launch their careers—but it also likely an inheritance effect. In this era, firstborns were more likely to take over the family farm or business.

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

Table 2. Bivariate OLS regressions of outcomes on Southern-ness of first name, by region of birth and region in 1940

	MODEL 1: Born in SOUTH, Lived in SOUTH in 1940			MODEL 2: Born in SOUTH, Lived in NORTH in 1940			MODEL 3: Born in SOUTH, Lived in MIDWEST in 1940			MODEL 4: Born in NORTH, Lived in NORTH in 1940			MODEL 5: Born in MIDWEST, Lived in MIDWEST in 1940		
	b	se	n	b	se	n	b	se	n	b	se	n	b	se	n
Education (in years)															
All boys	−0.023	(0.000)***	259,792	−0.022	(0.002)***	4,231	−0.017	(0.001)***	10,804	0.004	(0.000)***	325,877	−0.006	(0.000)***	421,640
Black boys	−0.013	(0.001)***	32,562	−0.015	(0.004)***	1,220	−0.011	(0.003)***	1,553	−0.001	(0.003)***	3,013	−0.005	(0.003)	2,733
White boys	−0.020	(0.000)***	226,805	−0.010	(0.003)***	3,009	−0.018	(0.002)***	9,246	0.004	(0.000)***	322,789	−0.006	(0.000)***	418,366
Occupational prestige															
All boys	−0.032	(0.001)***	238,685	−0.081	(0.010)***	3,653	−0.050	(0.006)***	9,907	−0.007	(0.002)***	292,854	−0.011	(0.001)***	389,044
Black boys	0.008	(0.002)**	29,921	−0.008	(0.012)	1,064	−0.022	(0.011)*	1,354	−0.015	(0.012)	2,564	−0.026	(0.011)*	2,304
White boys	−0.028	(0.001)***	208,406	−0.025	(0.016)	2,589	−0.033	(0.007)***	8,548	−0.003	(0.002)*	290,229	−0.008	(0.001)***	386,283
Wage/salary income															
All boys	−2.784	(0.051)***	243,111	−2.279	(0.539)***	4,023	−1.470	(0.313)***	10,306	0.141	(0.091)	307,867	−1.415	(0.069)***	396,858
Black boys	−0.834	(0.073)***	29,958	−0.511	(0.708)	1,140	−0.119	(0.536)	1,486	0.451	(0.584)	2,818	−1.091	(0.521)*	2,619
White boys	−2.772	(0.060)***	212,786	0.663	(0.795)	2,881	−1.003	(0.371)**	8,815	0.259	(0.092)**	304,988	−1.352	(0.070)***	393,726

Note: Sample restricted to the 1,022,344 boys who could be linked across the 1920 and 1940 censuses; who were children, stepchildren, grandchildren, or nephews of the household head; who had at least one other age-appropriate brother in the household in 1920; whose first name appeared at least 100 times in the 1920 Census; and who lived in one of the five combinations of regions of birth and regions of residence in 1940 considered in our analyses. All analyses are weighted by the inverse of the probability of selection into the linked sample.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

Table 3. Sibling fixed effects models of outcomes on Southern-ness of first name, by region of birth and region in 1940

	MODEL 1: Born in SOUTH, Lived in SOUTH in 1940			MODEL 2: Born in SOUTH, Lived in NORTH in 1940			MODEL 3: Born in SOUTH, Lived in MIDWEST in 1940			MODEL 4: Born in NORTH, Lived in NORTH in 1940			MODEL 5: Born in MIDWEST, Lived in MIDWEST in 1940		
	b	se	n	b	se	n	b	se	n	b	se	n	b	se	n
Education (in years)															
All boys	−0.001	(0.000)*	259,792	−0.006	(0.003)	4,231	−0.002	(0.002)	10,804	0.001	(0.000)*	325,877	0.000	(0.000)	421,640
Black boys	−0.001	(0.001)	32,562	−0.009	(0.005)	1,220	−0.005	(0.004)	1,553	−0.001	(0.003)	3,013	−0.001	(0.003)	2,733
White boys	0.000	(0.000)	226,805	−0.002	(0.004)	3,009	−0.002	(0.002)	9,246	0.001	(0.000)*	322,789	0.000	(0.000)	418,366
Occupational prestige															
All boys	−0.001	(0.001)	238,685	−0.009	(0.017)	3,653	−0.003	(0.010)	9,907	−0.005	(0.002)*	292,854	0.000	(0.002)	389,044
Black boys	−0.003	(0.003)	29,921	−0.009	(0.019)	1,064	−0.017	(0.020)	1,354	−0.009	(0.017)	2,564	0.001	(0.015)	2,304
White boys	−0.001	(0.002)	208,406	−0.008	(0.026)	2,589	0.000	(0.011)	8,548	−0.005	(0.002)*	290,229	0.000	(0.002)	386,283
Wage/salary income															
All boys	0.013	(0.068)	243,111	0.526	(0.963)	4,023	0.104	(0.487)	10,306	0.076	(0.121)	307,867	−0.001	(0.085)	396,858
Black boys	0.130	(0.095)	29,958	−1.417	(1.157)	1,140	−0.058	(0.848)	1,486	−0.050	(0.732)	2,818	−0.063	(0.721)	2,619
White boys	−0.011	(0.077)	212,786	2.285	(1.423)	2,881	0.151	(0.569)	8,815	0.079	(0.122)	304,988	−0.004	(0.086)	393,726

Note: Sample restricted to the 1,022,344 boys who could be linked across the 1920 and 1940 censuses; who were children, stepchildren, grandchildren, or nephews of the household head; who had at least one other age-appropriate brother in the household in 1920; whose first name appeared at least 100 times in the 1920 Census; and who lived in one of the five combinations of regions of birth and regions of residence in 1940 considered in our analyses. All analyses are weighted by the inverse of the probability of selection into the linked sample.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

attainment, the middle panel to occupation prestige, and the bottom panel to income. In each panel, we report separate results for all brother pairs, White brother pairs, and Black brother pairs. Recall that our sample is restricted to brother pairs in which both members of the pair lived in the same region in 1940.

As shown in table 3, using this fixed-effects specification we find very few statistically significant relationships between name Southern-ness and any of these outcomes; what is more, coefficient point estimates are typically much smaller in magnitude than the corresponding coefficient estimate in the OLS models in table 2. The exception is that Northern-born boys living in the North in 1940 who have more Southern-sounding names do a bit better with respect to education and a bit worse with respect to occupational prestige. Beyond these exceptions, the overwhelming evidence in table 3 is that associations between the Southern-ness of first names and socioeconomic outcomes—observed in table 2—disappear in models that account for all of the social, economic, cultural, racial, and other factors shared by brothers.

One possibility is that the effects of the regional signaling value of *last* names overwhelms the regional signaling value of first names. That is, people with very Southern-sounding last names may be assumed to be Southerners regardless of their first name. To investigate, we repeated the sibling fixed-effects models described in table 3 separately for brothers with "very Southern" last names (technically, last names for which more than 75 percent of holders were born in the South) and for brothers with "not very Southern" last names (technically, last names for which fewer than 25 percent of holders were born in the South). These results are presented in table 4. Coefficients in each model represent the estimated effect of holding a more Southern-sounding *first* name according to levels of the Southern-ness of *last* names (by region of birth and region of residence in 1940).

With one important exception, the results in table 4 continue to show that there are few significant or substantively large effects of the Southern-ness of boys' first names—regardless of the Southern-ness of their last names. The one exception: Among boys who were born in the South, who lived in the North in 1940, and who had very Southern-sounding *last* names, those with more Southern-sounding *first names* earned substantially lower salaries and wages.

Another possibility is that the Southern-ness of first names typically does not vary much within sibling pairs. In this situation, we might not typically see much of an effect of a single percentage point increase in the Southern-ness of first names. However, we might see larger effects if we focused on brother pairs in which one brother had a very Southern first name and one had a very not-Southern first name.

In the top panel of table 5, we restrict our sample to the subset of brother pairs in which the Southern-ness of their first names differed by at least 25 percentage points (e.g., if one brother's name was held by someone of Southern birth 10 percent of the time, the other brother had to have a name that was held by someone of Southern birth at least 35 percent of the time). In the bottom panel of that table, we restrict our sample to pairs of brothers in which the Southern-ness of their names differed by at least 50 percentage points. Our models in table 5 express the effect of being the brother with the much more Southern-sounding name (not the effect of a single percentage point increase in name Southern-ness as in tables 2 through 4). Here again, we see almost no significant or large effects of holding a

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

Table 4. Sibling fixed effects models of outcomes on Southern-ness of first name, by Southern-ness of last names, region of birth, and region in 1940

	MODEL 1: Born in SOUTH, Lived in SOUTH in 1940			MODEL 2: Born in SOUTH, Lived in NORTH in 1940			MODEL 3: Born in SOUTH, Lived in MIDWEST in 1940			MODEL 4: Born in NORTH, Lived in NORTH in 1940			MODEL 5: Born in MIDWEST, Lived in MIDWEST in 1940		
	b	se	n	b	se	n	b	se	n	b	se	n	b	se	n
Education (in years)															
Not Very Southern Last Name	0.000	(0.001)	26,149	−0.006	(0.007)	836	−0.003	(0.006)	1,571	0.001	(0.000)*	217,304	0.000	(0.000)	261,242
Very Southern Last Name	−0.002	(0.001)**	52,339	−0.013	(0.007)	633	−0.006	(0.004)	1,615	0.005	(0.004)	2,435	0.000	(0.002)	6,776
Occupational prestige															
Not Very Southern Last Name	−0.005	(0.003)	24,038	−0.026	(0.050)	737	0.005	(0.032)	1,463	−0.007	(0.002)**	195,343	−0.004	(0.002)	241,851
Very Southern Last Name	−0.005	(0.003)	48,293	0.013	(0.035)	532	−0.013	(0.023)	1,471	−0.017	(0.024)	2,199	0.017	(0.012)	6,200
Wage/salary income															
Not Very Southern Last Name	−0.264	(0.321)	24,881	−1.509	(2.752)	800	0.586	(1.782)	1,494	0.102	(0.152)	205,334	−0.102	(0.115)	246,552
Very Southern Last Name	0.088	(0.142)	48,936	−4.962	(2.380)*	592	0.207	(1.078)	1,541	0.009	(1.337)	2,301	−0.164	(0.546)	6,316

Note: Sample restricted to the 1,022,344 boys who could be linked across the 1920 and 1940 censuses; who were children, stepchildren, grandchildren, or nephews of the household head; who had at least one other age-appropriate brother in the household in 1920; whose first name appeared at least 100 times in the 1920 Census; and who lived in one of the five combinations of regions of birth and regions of residence in 1940 considered in our analyses. All analyses are weighted by the inverse of the probability of selection into the linked sample.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

Table 5. Sibling fixed effects models of outcomes on having a very Southern first name, by region of birth and region in 1940, within highly discrepant pairs

	MODEL 1: Born in SOUTH, Lived in SOUTH in 1940			MODEL 2: Born in SOUTH, Lived in NORTH in 1940			MODEL 3: Born in SOUTH, Lived in MIDWEST in 1940			MODEL 4: Born in NORTH, Lived in NORTH in 1940			MODEL 5: Born in MIDWEST, Lived in MIDWEST in 1940		
	b	se	n	b	se	n	b	se	n	b	se	n	b	se	n
Brothers' Name "Southern-ness" Differs by at Least 25 Percentage Points															
Education (in years) All boys	0.000	(0.000)	88,067	−0.008	(0.003)*	1,085	−0.002	(0.002)	3,218	0.000	(0.000)	42,225	0.000	(0.000)	67,959
Occupational prestige All boys	−0.002	(0.002)	81,156	0.003	(0.019)	945	0.002	(0.011)	2,966	−0.008	(0.003)**	38,138	0.003	(0.002)	62,802
Wage/salary income All boys	0.034	(0.072)	82,165	0.554	(1.014)	1,031	0.299	(0.549)	3,084	0.020	(0.149)	39,857	0.022	(0.102)	63,544
Brothers' Name "Southern-ness" Differs by at least 50 Percentage Points															
Education (in years) All boys	0.000	(0.000)	22,400	−0.005	(0.005)	232	−0.001	(0.003)	707	0.002	(0.001)*	4,378	0.001	(0.001)	8,411
Occupational prestige All boys	−0.003	(0.003)	20,669	−0.017	(0.042)	188	−0.008	(0.016)	652	−0.011	(0.005)*	3,983	0.007	(0.004)	7,809
Wage/salary income All boys	−0.0023	(0.094)	20,871	2.770	(2.055)	221	−0.701	(0.913)	674	−0.123	(0.287)	4,152	0.137	(0.187)	7,824

Note: Sample restricted to the 1,022,344 boys who could be linked across the 1920 and 1940 censuses; who were children, stepchildren, grandchildren, or nephews of the household head; who had at least one other age-appropriate brother in the household in 1920; whose first name appeared at least 100 times in the 1920 Census; and who lived in one of the five combinations of regions of birth and regions of residence in 1940 considered in our analyses. All analyses are weighted by the inverse of the probability of selection into the linked sample.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

(very) Southern-sounding first name. Within pairs of brothers who differ markedly on the Southern-ness of their first names, the brother with the (much) more Southern-sounding name did no worse (or better).

Discussion

A large body of research uses names as a proxy for people's risk of experiencing discrimination based on social, economic, or demographic characteristics. Previous research—often using experimental designs and nearly all conducted in the context of much more recent cohorts of people—has shown that people discriminate against racial/ethnic minorities, poor people, women, immigrants, and other groups by making inferences about group membership from signals provided by proper names. Given the history of regional hostilities in the late nineteenth and early twentieth century in the United States, we expected to see evidence of discrimination against regional outsiders; our leverage in identifying the effects of discrimination was to take advantage of within-sibling pair comparisons.

On one hand, we hypothesized that we would observe negative consequences for social and economic outcomes of holding a name that implied that a boy was of Southern origins. By studying boys born in the 1910s, our results pertain to cohorts who experienced the Great Migration of Southerners—including many African American Southerners—to the North. In particular, we expected to observe that people living in the North who had Southern-sounding names would fare worse with respect to education, occupation, and wages; we anticipated that these effects might be even more pronounced for African Americans. In fact, we found few such effects; even statistically significant coefficients were small in magnitude. In general, having a Southern-sounding name had little independent impact on boys' life outcomes.

Does this mean that Southerners moving North (and bringing their Southern-sounding names with them) did not, in fact, experience discrimination? More generally, does it mean that American regional outsiders in the decades following the Civil War and in the years of the Great Migration did not face consequential levels of bias? Perhaps, but we suspect it is more likely that such bias did not operate based on people's names. That is, we question whether—in this historical era—names were a primary way in which people inferred people's regional origins.

Prior research using audit study designs—experiments that randomly and artificially manipulate applicants' names to signal group membership—has attempted to isolate the role of interpersonal discrimination in labor market and other outcomes. In a typical design, people with the power to discriminate are asked to evaluate otherwise similar people who differ only with respect to which their name is generally perceived to be typical of women, African Americans, immigrants, or other traditionally marginalized groups. These studies often (but not universally) find pronounced effects of holding a name that is typical of people from marginalized groups on educational, labor market, housing, and other outcomes. That is, in these contexts, names matter.

Although audit studies are strong with respect to their internal validity, they are weak in three other respects. First, they are not typically generalizable to the US population more broadly even in modern times. They usually pertain to a narrow subset of people—for example, applicants for one kind of job or people living in one

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

city or region. Second, they typically pertain to a narrow range of (often hypothetical) labor force or other outcomes; these sorts of studies do not always demonstrate effects on actual realized wages, actual long-term occupational attainment, actual longevity, or other actual long-term outcomes. Third, and perhaps most importantly, in audit and similar designs people with the power to discriminate can only do so on the basis of names and not on the basis of other potentially more powerful signals of group membership—like phenotypical body features (e.g., skin tone), accents (e.g., Southern accents), or relationship to other people whose group memberships are known (e.g., a Southern father).

In our research, we overcome these limitations of audit studies for estimating the effects of discrimination against people of Southern origins. Using nationally representative data, we use a sibling comparison design to estimate the effects of Sothern-ness of boys' first names on their actual realized educational attainment, occupational prestige, and wages and earnings. Our sibling design allows us—like audit studies—to rule out most of the factors that might otherwise confound the association between given names and these sorts of outcomes. Perhaps most importantly, however, our design does not force people at risk of discriminating to infer people's regional origins based on names alone. Potential employers, for example, could guess job applicants' regional origins based on their names—but they could also hear their accents, inquire about their region of birth, know their friends, or otherwise access more direct measures of their Southern-ness.

For real-world outcomes in a large and representative population in the early twentieth century, discrimination against Southerners happened in some other way than inferring regional origins based on names. Northerners in a position to discriminate against Southerners, for example, and to deny them access to education, jobs, higher pay, and other valued resources apparently did so—we reason—based on information about them besides their names.

In the end, our work calls into question the broad applicability and scope conditions of conclusions of research that uses proper names as a marker of people's risk of suffering from discrimination on the basis of regional origins—and also on the basis of race, gender, immigration status, and other attributes. More generally, our work calls into question the conditions and circumstances under which names have independent effects on life outcomes. We conclude that names had no such independent effects in the context of early-twentieth-century America (at least with respect to regional discrimination), and instead that associations between names and life outcomes were confounded by family-level economic and cultural factors—that is, by those factors that are generally observable to people with the power to discriminate.

Acknowledgments. Paper prepared for presentation at the 2019 meetings of the Social Science History Association. The project benefited from support by a grant from the National Institute on Aging (1R21AG054824-01A1). Support has also come from the Minnesota Population Center, which receives core funding (P2C HD041023) from the Eunice Kennedy Shriver National Institute for Child Health and Human Development (NICHD). Noah Greenswig, Anna Harvatine, and Anika Tol were significant contributors to early versions of the research described in this article. We are also very grateful to J. David Hacker, Trevon Logan, Evan Roberts, and two anonymous reviewers for providing helpful comments and suggestions. However, errors and omissions are the responsibility of the authors. Please direct correspondence to John Robert Warren at warre046@umn.edu.

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

References

Alexander, J. Trent, Christine Leibbrand, Catherine Massey, and Stewart Tolnay (2017) "Second-generation outcomes of the Great Migration." Demography **54** (6): 2249–71.

Arai, Mahmood, and Peter Skogman Thoursie (2009) "Renouncing personal names: An empirical examination of surname change and earnings." Journal of Labor Economics **27** (1): 127–47.

Beach, Brian, Joseph Ferrie, Martin Saavedra, and Werner Troesken (2016) "Typhoid fever, water quality, and human capital formation." The Journal of Economic History **76** (1): 41–75.

Bertrand, Marianne, and Sendhil Mullainathan (2004) "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination." American economic review **94** (4): 991–1013.

Biavaschi, Costanza, Corrado Giulietti, and Zahra Siddique (2017) "The economic payoff of name americanization." Journal of Labor Economics **35** (4): 1089–1116.

Boustan, Leah Platt (2008) "Competition in the promised land: Black migration and racial wage convergence in the North, 1940–1970." NBER Working Paper No. 13813, National Bureau of Economic Research, www.nber.org/system/files/working_papers/w13813/w13813.pdf. Accessed March 9, 2022.

Busse, Thomas V., and Louisa Seraydarian (1977) "Desirability of first names, ethnicity and parental education." Psychological Reports **40** (3): 739–42.

Carneiro, Pedro, Sokbae Lee, and Hugo Reis (2020) "Please call me John: Name choice and the assimilation of immigrants in the United States, 1900–1930." Labour Economics **62**: 101778.

Chicco, Davide (2017) "Ten quick tips for machine learning in computational biology." BioData Mining **10** (1): 1–17.

Cook, Lisa D., Trevon D. Logan, and John M. Parman (2016) "The mortality consequences of distinctively black names." Explorations in Economic History **59**: 114–25.

Eichenlaub, Suzanne C., Stewart E. Tolnay, and J. Trent Alexander (2010) "Moving out but not up: Economic outcomes in the Great Migration." American Sociological Review **75** (1): 101–25.

Elchardus, Mark, and Jessy Siongers (2011) "First names as collective identifiers: An empirical analysis of the social meanings of first names." Cultural Sociology **5** (1): 403–22.

Feigenbaum, James J. (2016) "Automated census record linking: A machine learning approach." Working Paper, Harvard University, www.scholar.harvard.edu/files/jfeigenbaum/files/feigenbaum-censuslink.pdf. Accessed March 9, 2022.

Ferrie, Joseph P. (1996) "A new sample of males linked from the public use microdata sample of the 1850 US federal census of population to the 1860 US federal census manuscript schedules." Historical Methods: A Journal of Quantitative and Interdisciplinary History **29** (4): 141–56.

Figlio, David N. (2005) "Names, expectations and the black-white test score gap." NBER Working Paper 11195, National Bureau of Economic Research, www.nber.org/system/files/working_papers/w11195/w11195.pdf. Accessed March 9, 2022.

Fryer Jr., Roland G., and Steven D. Levitt (2004) "The causes and consequences of distinctively black names." The Quarterly Journal of Economics **119** (3): 767–805.

Gardner, John (2016) "Immigration and wages: new evidence from the African American Great Migration." IZA Journal of Migration **5** (1): 1–45.

Goldstein, Joshua R., and Guy Stecklov (2016) "From Patrick to John F.: Ethnic names and occupational success in the last era of mass migration." American Sociological Review **81** (1): 85–106.

Hume, Richard L. (1977) "Carpetbaggers in the Reconstruction South: A group portrait of outside whites in the "black and tan" constitutional conventions." The Journal of American History **64** (2): 313–30.

Jaro, Matthew A. (1989) "Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida." Journal of the American Statistical Association **84** (406): 414–20.

—— (1995) "Probabilistic linkage of large public health data files." Statistics in Medicine **14** (5–7): 491–98.

Laham, Simon M., Peter Koval, and Adam L. Alter (2012) "The name-pronunciation effect: Why people like Mr. Smith more than Mr. Colquhoun." Journal of Experimental Social Psychology **48** (3): 752–56.

McKinney, John C., and Linda Brookover Bourque (1971) "The changing South: National incorporation of a region." American Sociological Review **36** (3): 399–412.

Rubinstein, Yona, and Dror Brenner (2014) "Pride and prejudice: Using ethnic-sounding names and inter-ethnic marriages to identify labour market discrimination." Review of Economic Studies **81** (1): 389–425.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek (2020) "IPUMS USA: Version 10.0." IPUMS, https://doi.org/10.18128/D010.V10.0.

Siegel, Paul Mathew (1971) "Prestige in the American occupational structure." PhD diss., University of Chicago.

Smith, James P., and Finis R. Welch (1989) "Black economic progress after Myrdal." Journal of Economic Literature **27** (2): 519–64.

Winkler, William E. (1990) "String Comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage." US Bureau of the Census.

—— (2006) "Overview of record linkage and current research directions." Research Report Series (Statistics #2006-2). US Bureau of the Census, www.census.gov/srd/papers/pdf/rrs2006-02.pdf (accessed January 27, 2021).

Alex Beaudin completed this research as an undergraduate student in sociology at the University of Minnesota. He has no current academic affiliation and no other publications.

Elizabeth Kristian completed this research as an undergraduate student in sociology at the University of Minnesota. She has no current academic affiliation and no other publications.

John Robert Warren is professor of sociology at the University of Minnesota. His research focuses on life-course determinants of socioeconomic and health outcomes. He is a sociologist, demographer, population health scholar, and education policy researcher with experience and expertise in the collection, production, and dissemination of large-scale data products for research on health, aging, education, and labor force outcomes. He and Dr. Helgertz recently published "The Effects of Education on Mortality: Evidence from Linked US Census and Administrative Mortality Data" in *Demography*.

Dr. Jonas Helgertz is a research scientist at the University of Minnesota's Institute for Social Research and Data Innovation. His research focuses on life-course determinants of demographic, socioeconomic, and health outcomes. He and Dr. Warren recently published "The Effects of Education on Mortality: Evidence from Linked US Census and Administrative Mortality Data" in *Demography*.

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press

Appendix

Table A1. Comparison of boys' 1920 characteristics between those who could and could not be linked to the 1940 Census

	Not Linked	Linked
Race/Ethnicity		
White	85.9%	94.8%
Black	13.7%	5.0%
Others	0.4%	0.2%
Place of Birth		
North	26.2%	27.2%
South	41.2%	28.7%
Midwest	27.6%	37.0%
West	4.1%	6.6%
Non-US	1.0%	0.5%
Mother's Literacy		
Not Literate	18.6%	10.6%
Literate	81.4%	89.4%
Father's Occupation		
Professional, technical	1.5%	1.9%
Farmers	0.2%	0.2%
Managers, officials	4.8%	5.9%
Clerical and kindred	3.6%	4.8%
Craftsmen	11.6%	13.8%
Operatives	10.2%	9.8%
Service workers	1.7%	1.7%
Farm laborers	4.0%	3.0%
Laborers	10.0%	7.9%
Not known	52.4%	50.9%
Homeownership		
Rent	61.4%	53.9%
Own	38.6%	46.1%
N	6,735,013	4,312,380

Note: Sample restricted to all boys who were 0 to 10 in 1920 and who were children, grandchildren, stepchildren, or nephews of the head of their household.

Cite this article: Beaudin, Alex, Elizabeth Kristian, John Robert Warren, and Jonas Helgertz (2022) ""You're Not from around Here": Regional Naming and Life Outcomes," *Social Science History* **46**:671–691. doi:10.1017/ssh.2022.7

https://doi.org/10.1017/ssh.2022.7 Published online by Cambridge University Press