Supplementary Appendix

10 February, 2021

Contents

Data and Pre-Processing

Mortality Records

We use public administrative records from the Social Security Administration as provided in the Berkeley Unified Numident Mortality Database (BUNMD) for this analysis. The BUNMD was constructed from the most informative parts of the Social Security Numerical Identification System (Numident). For every person with a Social Security, the Numident number—tracks date of birth, date of death (if applicable), and other background information including birthplace, race, sex, and names of parents. The Social Security Administration transferred a large number of these records to the National Archives and Records Administration (NARA), who publicly released the records in 2019. The BUNMD is a cleaned and harmonized version of the records publicly released by NARA (Goldstein and Breen 2020). Death coverage in the BUNMD is nearly complete for the window of 1988-2005.

Identifying Siblings

To establish two men in the BUNMD as brothers, we identify men who share a common mother and father's first and last name. We have confidence in the accuracy of sibling matches because of the high rate of being born in the same state (90.6%), although we did not use this as a matching criterion.

	Sibling		**Father**		**Mother**	
First (standardized)	Last	Birthplace	First	Last	First	Last
Ernest	Cottman	Ohio	Royal	Cottman	Leona	Jones
Royal	Cottman	Ohio	Royal	Cottman	Leona	Jones
James	Mason	Georgia	Jonas	Mason	Nettie	Jackson
Arthur	Mason	Georgia	Jonas	Mason	Nettie	Jackson
Oscar	Watson	Louisiana	Louis	Watson	Marinda	White
Spencer	Watson	Louisiana	Louis	Watson	Marinda	White

Table 1: Establishing siblingship based on exact matches on parents' names. For exact matches, birthplace was not used as a matching field.

We identify an additional set of siblings by allowing for some flexibility on the exact spelling of parents' first names. This allows for minor misspellings and transcription errors. For these additional siblings, we (conservatively) require birthplace to match. We use the following procedure:

1. Identify potential pairs of brother (restricted to 2) who reported the same place of birth, mother's last name, father's last name, but **different** father's first names and/or mother's first name.

2. If both potential siblings reported the same father's first name but different mother's first names, establish as sibling pair if string distance between the two discrepant mother's first names > 0.7 (based on the Levenshtein distance).

3. If both potential siblings reported the same mother's first name but different father's first names, establish as sibling pair if string distance between the two discrepant father's first names > 0.7 (based on the Levenshtein distance).

4. If neither mother nor father's first name match exactly, establish as a sibling pair if both string distances between the two discrepant fathers' first names and between the two discrepant mothers' first names > 0.8 (based on the Levenshtein distance).

Sibling			**Father**			**Mother**	
First (Standardized)	Last	BPL	First*	Lev Distance	Last	First	Last
Harrison	Jones	Tennessee	Ernest	0.857	Jones	Carrie	Fields
James	Jones	Tennessee	Earnest	0.857	Jones	Carrie	Fields
George	Pittman	Georgia	Deillie	0.714	Pittman	Mattie	Henry
Dwellie	Pittman	Georgia	Dwellie	0.714	Pittman	Mattie	Henry
Arthur	Mitchell	Mississippi	Matthew	0.875	Mitchell	Beatrice	Johnson
Jarvis	Mitchell	Mississippi	Matthews	0.875	Mitchell	Beatrice	Johnson

Table 2: Illustration of establishing sibling pairs allowing for flexibility in father's first name. The same procedure was conducted for mother's first name. A minimum Levenshtein string distance of > 0.7 was used as a threshold to establish a sibling pair.

Name Standardization

We pre-process first names to account for nicknames or minor spelling and transcription errors. We clean first names by (i) removing any non-alpha characters, (ii) selecting the first word for compound string names, and (iii) restricting to first names great than one character in length. After cleaning the names, we standardize first names to account for nicknames (e.g., "Bill" → "William") and misspellings (e.g., "Williaam" → "William"). We construct a master name standardization dictionary by combining name standardization dictionaries from the IPUMS and ABE research teams.

BNI

We use the Black Name Index (BNI), a summary measure of how distinctly black a first name proposed by Fryer and Levitt (2004). We use the standardized names to compute the BNI:

$$BNI_{name_i} = \frac{P(name_i|Black)}{P(name_i|Black) + P(name_i|White)} \tag{1}$$

To compute BNI, we use the BUNMD birth cohorts of 1895-1940 with information on race. This larger sample gives us more precise BNI estimates for first name.

Analyses

Main Model

Table 3 shows our main result: brothers with blacker names die at younger ages. Models are robust across different specifications (e.g., model 2 does not include birth cohort fixed effects; model 5 does not restrict to the high mortality-coverage window of 1988-2005).

For our main analysis, we restrict to Black brothers who were born between 1915 and 1925, died between 1988 and 2005 (the BUNMD window with high death coverage), and had a standardized first name appearing 500 or more times in the BUNMD (N = 30,742).

Dependent Variable:	Death Age				
	Pooled	**Family FE**			
Model:	(1)	(2)	(3)	(4)	(5)
BNI (Standardized)	-0.1483	-0.4922*	-0.6617**	-0.6641**	-0.5547*
	(0.2093)	(0.2963)	(0.2638)	(0.2638)	(0.3363)
Family FE	–	Yes	Yes	Yes	Yes
Birth Year FE	–	–	Yes	Yes	Yes
Birth Cohort FE	–	–	–	Yes	Yes
Mortality Window	1988-2005	1988-2005	1988-2005	1988-2005	1941-2007
Observations	30,742	30,742	30,742	30,742	46,504
R^2	1.63×10^{-5}	0.51585	0.61292	0.61295	0.56430
Within R^2		0.00018	0.00040	0.00040	0.00011

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3: BNI gradient for pooled cohorts of 1915-1925.

Exact vs. Non-Exact Siblings

The estimated BNI gradient for the sample of siblings established by an exact match (model 1) is greater than the estimated BNI gradient for the sample of siblings established by the non-exact matches (model 2). While this may reflect incorrectly-identified sibling pairs biasing results toward population-wide estimates (i.e., our pooled OLS model coefficient), it is likely noise due to the small sample size of non-exact matches. For the combined sample, both the BNI gradient and standard error decreased slightly.

Dependent Variable:	Death Age		
Model:	(1)	(2)	(3)
	Exact	Non-Exact	Combined
BNI (Standard)	-0.7099**	-0.4272	-0.6641**
	(0.2853)	(0.6917)	(0.2638)
Family FE	Yes	Yes	Yes
Birth Order FE	Yes	Yes	Yes
Birth Year FE	Yes	Yes	Yes
Observations	26,140	4,602	30,742
R^2	0.60940	0.63463	0.61295
Within R^2	0.00046	0.00017	0.00040
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 4: BNI gradient for pooled cohorts of 1915-1925.

Sensitivity Analysis: Birth Cohorts

For our main analysis, we used the BUNMD birth cohorts of 1915-1925. The plot below shows the estimated BNI gradients ($\beta_1 \pm 1.96 \times SE(\beta_1)$), where the 10 or 15-year window of birth cohorts was systematically varied. The results are robust across the birth cohort windows corresponding to the BUNMD high death coverage period.

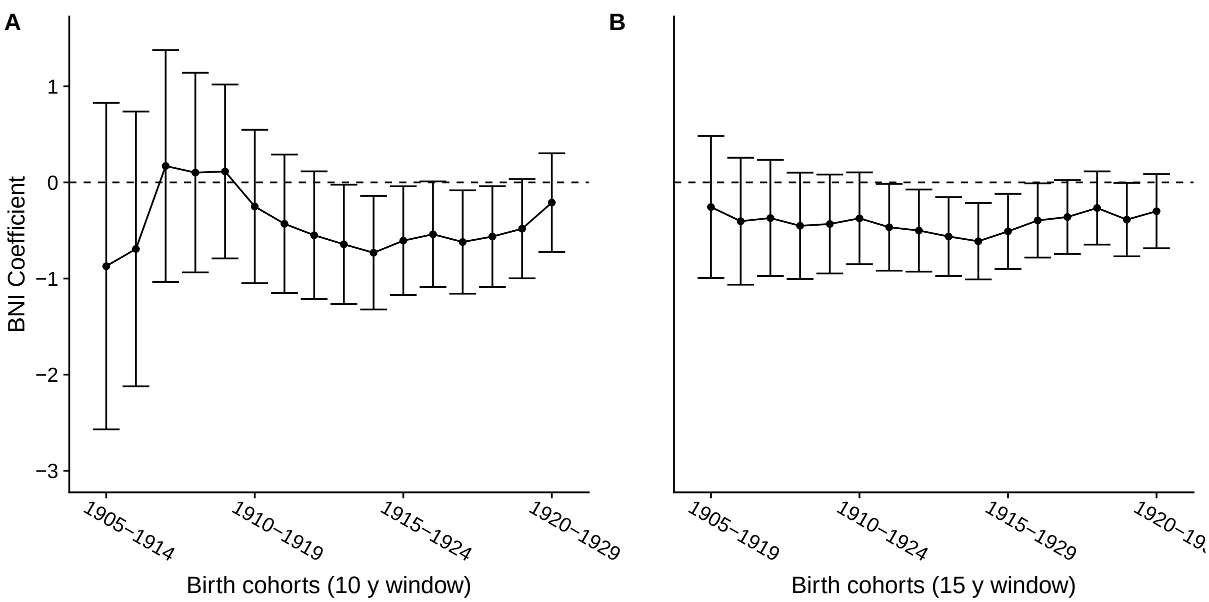

Figure 1: Panel A shows BNI gradients for 10-year windows. Panel B shows BNI gradient for 15-year windows.

Sensitivity Analysis: Minimum Name Threshold

Our analysis is sensitive to our choice of the minimum name frequency for a person to be included in the analysis. Name frequency is measured as the number of records in the original BUNMD used to calculate the BNI for a given name. Figure 2 below shows the estimated BNI gradients ($\beta_1 \pm 1.96 \times SE(\beta_1)$), where the minimum frequency threshold for inclusion was systematically varied (1 to 3,001, increments of 100.) For our final analysis, we used a minimum frequency threshold of 500; that is, we dropped an individual from our analysis if their name was observed 500 times or fewer in the BUNMD (birth cohorts of 1895 - 1940 with non-missing race).

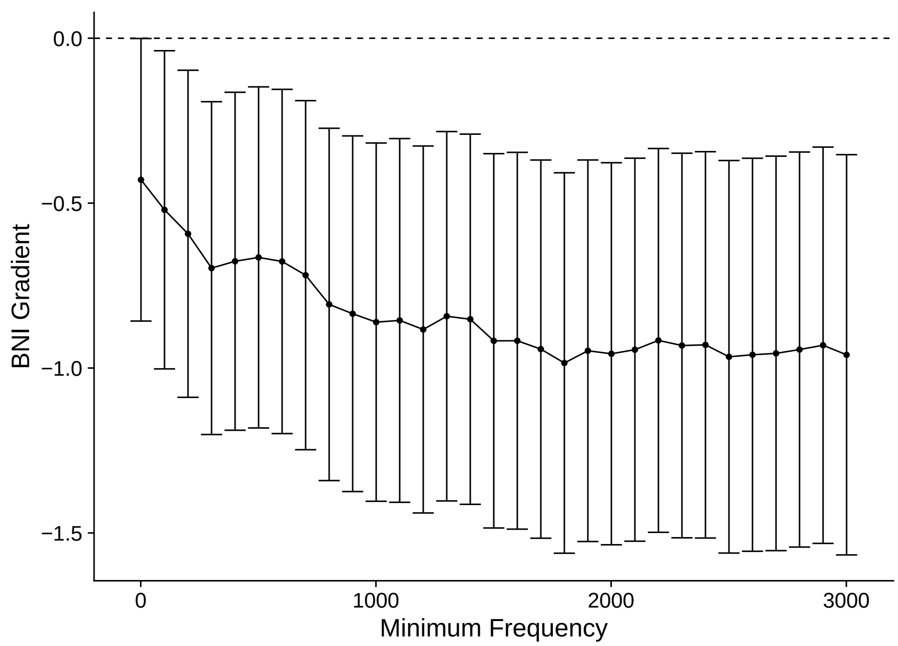

Figure 2: Sensitivity analysis for minimum name threshold.

BNI vs. Percentage Black

For our analysis, we used Black Name Index (BNI) as a measure the "Blackness" of first names. We could have alternatively used the percentage of name-holders as a measure of the "Blackness" of first names. Reassuringly, our results are consistent across both measures.

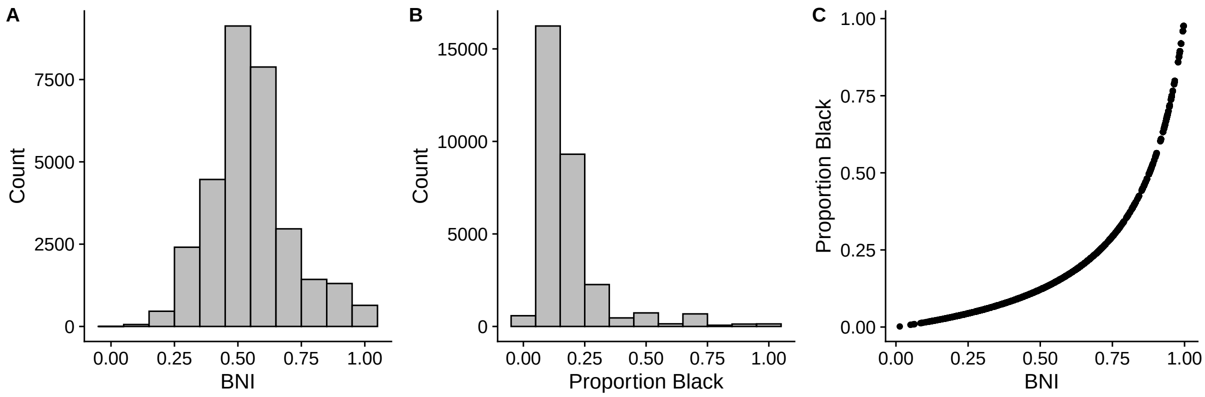

Figure 3: For names in our final sample (N = 30,742), histograms of BNI (A) and percentage Black (B). Panel C shows relationship between BNI and percentage Black.

Dependent Variable:		Death Age	
Model:	(1)	(2)	(3)
Percentage Black	-0.7021**	-0.6809**	-0.6825**
	(0.3150)	(0.2782)	(0.2782)
Family FE	Yes	Yes	Yes
Birth Year FE	—	Yes	Yes
Birth Order FE	—	—	Yes
Observations	30,742	30,742	30,742
R^2	0.51592	0.61291	0.61294
Within R^2	0.00032	0.00037	0.00037
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 5: BNI gradient for pooled cohorts of 1915-1925.

BNI Gradient for Whites

Do we observe a BNI gradient for Whites? A BNI gradient for Whites may suggest that dimensions other than race are driving the BNI effect. Table 6 shows our main model applied to sample of White brothers. There is a trivial BNI gradient for Whites, suggesting the BNI gradient on the longevity of Black Americans is not driven by inferences on social class.

BNI and Neighborhood Affluence

Do people with Blacker names end up living in less affluent ZIP Codes? Table 7 shows BNI is a predictor of the amount of Social Security Benefits dispersed in a ZIP Code; those with Blacker

Dependent Variable:	Death Age	
Model:	(1)	(2)
	Whites	Blacks
BNI (Standardized	-0.0205	-0.6641**
	(0.0727)	(0.2638)
Family FE	Yes	Yes
Birth Year FE	Yes	Yes
Birth Order FE	Yes	Yes
Observations	463,071	30,742
R^2	0.60984	0.61295
Within R^2	3.33×10^{-7}	0.00040
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 6: BNI gradient for pooled cohorts of 1915-1925.

names are less likely to end up in an affluent ZIP Code.

Dependent Variable:	Zip Benefits			
	Pooled	**Sibling FE**		
Model:	(1)	(2)	(3)	(4)
BNI (Standardized)	-0.0119***	-0.0092*	-0.0090*	-0.0091*
	(0.0039)	(0.0049)	(0.0049)	(0.0049)
Family FE	–	Yes	Yes	Yes
Birth Year FE	–	–	Yes	Yes
Birth Order FE	–	–	–	Yes
Observations	28,130	28,130	28,130	28,130
R^2	0.00034	0.65986	0.66032	0.66036
Within R^2		0.00027	0.00025	0.00026
Note:	*p<0.1; **p<0.05; ***p<0.01			

Table 7: BNI gradient for Social Security benefits dispersed in ZIP Code of death. Pooled cohorts of 1915-1925.

References

Bertrand, Marianne, and Sendhil Mullainathan. 2003. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," 40.

Chae, David H., Amani M. Nuru-Jeter, Nancy E. Adler, Gene H. Brody, Jue Lin, Elizabeth H. Blackburn, and Elissa S. Epel. 2014. "Discrimination, Racial Bias, and Telomere Length in African-American Men." *American Journal of Preventive Medicine* 46 (2): 103–11. https://doi.org/10.1016/j.amepre.2013.10.020.

Cook, Lisa D., Trevon D. Logan, and John M. Parman. 2014. "Distinctively Black Names in the American Past." *Explorations in Economic History* 53 (July): 64–82. https://doi.org/10.1016/j.eeh.2014.03.001.

Cook, Lisa, Trevon Logan, and John Parman. 2015. "The Mortality Consequences of Distinctively Black Names," 29.

Gaddis, S. 2017. "How Black Are Lakisha and Jamal? Racial Perceptions from Names Used in Correspondence Audit Studies." *Sociological Science* 4: 469–89. https://doi.org/10.15195/v4.a19.

Garcia, Denia, and Maria Abascal. 2015. "Colored Perceptions: Racially Distinctive Names and Assessments of Skin Color." *American Behavioral Scientist*, October. https://doi.org/10.1177/0002764215613395.

Gerhards, J., and S. Hans. 2009. "From Hasan to Herbert: Name-Giving Patterns of Immigrant Parents Between Acculturation and Ethnic Maintenance1." *American Journal of Sociology*. https://doi.org/10.1086/595944.

Goffman, Erving. 2009. *Stigma: Notes on the Management of Spoiled Identity*. Simon; Schuster.

Goldstein, Joshua, and Casey Breen. 2020. "Berkeley Unified Numident Mortality Database: Public Administrative Records for Individual Level Mortality Research," 33.

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. https://doi.org/10.1177/0003122415621910.

Hacker, J. David. 1999. "Child Naming, Religion, and the Decline of Marital Fertility in Nineteenth-Century America." *The History of the Family* 4 (3): 339–65. https://doi.org/10.1016/S1081-602X(99)00019-6.

Johfre, Sasha. 2020. "What Age Is in a Name?" *Sociological Science* 7: 367–90. https://doi.org/10.15195/v7.a15.

Jr, Roland G Fryer, and Steven D Levitt. 2004. "Understanding the Black-White Test Score Gap in the First Two Years of School." *THE REVIEW OF ECONOMICS AND STATISTICS*, 18.

Lieberson, Stanley. 2000. *A Matter of Taste: How Names, Fashions, and Culture Change*. Yale University Press.

Lindsay, Jo, and Deborah Dempsey. 2017. "First Names and Social Distinction: Middle-Class Naming Practices in Australia." *Journal of Sociology* 53 (3): 577–91. https://doi.org/10.1177/1440783317690925.

Link, Bruce G., and Jo C. Phelan. 2001. "Conceptualizing Stigma." *Annual Review of Sociology* 27 (1): 363–85. https://doi.org/10.1146/annurev.soc.27.1.363.

Obukhova, Elena, Ezra W. Zuckerman, and Jiayin Zhang. 2011. "Why Was There No Fashion in Mao's China? Effects of Politics on Culture in the Case of Chinese Naming Practices." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1878678.

Pager, Devah, and Bruce Western. 2003. "Discrimination in a Low-Wage Labor Market: A Field Experiment," 23.

Stewart, Quincy Thomas, Ryon J. Cobb, and Verna M. Keith. 2020. "The Color of Death: Race, Observed Skin Tone, and All-Cause Mortality in the United States." *Ethnicity & Health* 25 (7): 1018–40. https://doi.org/10.1080/13557858.2018.1469735.