UNIVERSITY OF MINNESOTA

Building New Bridges

Developing and Disseminating a Simplified Race/Ethnicity Measure for Working with Complex or Contradictory Race Data

Carolyn A. Liebler
University of Minnesota

May 2022

Working Paper No. 2022-02
Cross listed as IPUMS Working Paper No. 2022-02

DOI: https://doi.org/10.18128/MPC2022-02

The National Cancer Institute provided financial support (Contract No. 75N91019P00786). Mandi Yu of the National Cancer Institute offered leadership and guidance throughout the project. I am also grateful to Megan Schouweiler of IPUMS USA for her efforts and suggestions.

Summary

Overview: For 25 years, the US federal standards for collecting data on race have required that people be allowed to report multiple races. The current federal race/ethnicity data collection standards have proven challenging for analysts who find the large number of categories impractical. Also, some data sources have been slower to collect high-quality and analyzable multiple-race data. To enable studies that use data sources with different race/ethnic collection strategies, multiple-race responses must be "bridged" into a small number of simplified categories. The simplified categories are meaningful and inclusive, making them useful for other analytic purposes as well. This report describes the development of updated coefficients for bridging and provides them to interested readers. It also describes three ways that resulting data are being disseminated: as microdata with whole allocation, as microdata with fractional allocation, and as aggregate data based on fractional allocation. The race/ethnic categories in the simplified, bridged race/ethnicity variables are: Hispanic/Latino, non-Hispanic/Latino (NH) American Indian or Alaska Native, NH Asian or Pacific Islander, NH Black, and NH White. To illustrate the utility of the simplified race/ethnicity variables, the report ends by showing the estimated total number of US persons who would have been reported as one of the five simplified race/ethnicity groups in 2000 to 2019.

Development of updated bridging coefficients: From 1997 to 2018, the National Health Interview Survey (NHIS) asked multiple-race respondents to report their "best" single race. This research used the restricted-use 1997-2018 NHIS to estimate a series of logistic and multinomial logistic regression models predicting which "best" single race a non-Hispanic multiple-race person would choose, based on their age, sex, urbanization level, and county race composition. The resulting "bridging coefficients" are included in an appendix and can be applied to a researcher's microdata to create bridged race/ethnicity variables.

Dissemination in microdata through IPUMS USA: To be useful to researchers using microdata, IPUMS USA (https://usa.ipums.org/usa/) collaborated to disseminate the results of applying the bridging coefficients to non-Hispanic multiple-race respondents in the public-use decennial census (2000 and 2010) microdata and American Community Survey (2000 to 2019) microdata, with modifications for the less detailed available geographic information in public-use data. Results are given in two types of variables in IPUMS USA. First, the results of applying the bridging equations to the individual's case are given in the PRED variables: PREDHISP, PREDAI, PREDAPI, PREDBLK, and PREDWHT. All people in these data resources have values in the PRED variables that sum to 1; this makes the PRED variables especially useful for multivariate analyses. Second, all people are assigned to a single category in the five-category variable RACHSING based on their PRED variable values.

Dissemination in aggregate data through SPARC: To support researchers requiring bridged race/ethnicity variables in aggregate data, population estimates and standard errors based on these updated bridging equations are being disseminated on the SPARC website hosted by the National Cancer Institute (https://surveillance.cancer.gov/sparc/). SPARC provides these estimates and standard errors for each state, sex, foreign-born status, and age group. Bridged population estimates are particularly useful for analysts working with state-collected data such as cancer records, birth records, and death records because these are often provided with only simplistic race/ethnicity categories.

Table of Contents

1. Introduction

1.1. Overview of the project

For 25 years, the US federal standards for collecting data on race have required that people be allowed to report multiple races. Some data products in the United States (e.g., the census) allow for multiple-race reporting, while others (e.g., cancer registries) have been slow to collect high-quality and analyzable data with multiple-race options. The 1997 standards have also proven challenging for analysts who find the large number of race/ethnic categories impractical.

When data collection methods differ, multiple-race responses must be "bridged" into simplified categories. This allows the researcher to create consistent denominators when calculating rates such as race-specific cancer incidence rates. Simplified categories are meaningful and inclusive, making them very useful for other analytic purposes as well.

Not all bridging strategies are similar. *Deterministic* bridging strategies assign all individuals from a particular multiple-race group to a particular simplified category; for example, assigning all Asian-Black individuals to the Black category. *Probabilistic* bridging strategies use characteristics of the individual multiple-race respondent to predict the individual's likely single-race choice; for example, a bridging equation might predict that a particular Asian-Black person would have a 0.63 probability of choosing Asian and 0.37 probability of choosing Black, given their demographic and contextual characteristics. To apply this method to research, the analyst would either assign this person to the Asian group (i.e., use "whole allocation") or use the probabilities in their analysis (i.e., use "fractional allocation").

This report describes the development of a series of probabilistic bridging equations which predict the likely single-race response of non-Hispanic multiple-race respondents to the National Health Interview Survey (NHIS) between 1997 and 2018. Equations were estimated for each of 19 four-year periods from 1997 to 2018. Detailed results are provided so that interested readers can apply them to their own data.

After bridging, the entire US population can be grouped into five simplified race/ethnic categories:

1) Hispanic/Spanish/Latino (simplified to "Hispanic" in this report)
2) Non-Hispanic American Indian or Alaska Native (AIAN)
3) Non-Hispanic Asian or Pacific Islander (API)
4) Non-Hispanic Black or African American (B or Black)
5) Non-Hispanic White (W or White)

Each non-Hispanic category includes people who reported the group directly, as well as those who are predicted to have reported the group IF only these categories were offered.

All people who reported Hispanic origins are grouped together, regardless of race response. The practical reason for this aspect of the procedure is that the NCHS assessed Hispanic origin separately from race and therefore the bridging procedure cannot be applied. A respondent who reported Black in the race question and Hispanic in the ethnicity question, for example, was not asked which of these "best"

describes them (as was done for people who reported multiple races); thus, there is no measure to serve as the dependent variable in a bridging prediction equation. People of Hispanic origin often do not report a race,[1] and often change their race responses (but remain consistent in their Hispanic origin response).[2]

The updated and expanded bridging equation coefficients were then applied to nationally-representative data in the US Census and American Community Survey (ACS) to create analysis-ready variables, estimates, and standard errors. The bridging equations use information about respondents' county of residence, but county-level geography is not available in public-use US Census and ACS microdata; characteristics of the person's Public Use Microdata Area (PUMA) were used instead. These simplified race/ethnic variables are being disseminated by IPUMS USA[3] as RACHSING, PREDHISP, PREDAI, PREDAPI, PREDBLK, and PREDWHT. The National Cancer Institute intends to use the results for their research and reporting on cancer rates. To this end, population estimates and standard errors were generated using the IPUMS USA fractional allocation variables (PRED*), and standard errors were calculated using replicate weights.[4] The National Cancer Institute's SPARC tool provides population estimates and standard errors by sex, five-year age group (with top-coding at age 85), US/foreign birthplace,[5] state, and year for each simplified race/ethnicity category.

The report concludes by showing the estimated total number of US persons who would have been reported as one of the five simplified race/ethnicity groups in 2000 to 2019.

1.2. Relationship to prior work

In 1977, the US federal government's Office of Management and Budget (OMB) created standards for collecting race and ethnicity data to assist with enforcing civil rights legislation. The 1977 standard named

---

[1] Ríos M, Romero F, and Ramírez R. 2014. *Race Reporting among Hispanics: 2010.* Population Division Working Paper #102. US Census Bureau. https://www.census.gov/content/dam/Census/library/working-papers/2014/demo/shedding-light-on-race-reporting-among-hispanics/POP-twps0102.pdf

[2] Almost all people in Census 2000 who reported Hispanic origins also reported Hispanic origins in 2010, though a consistent race response across the decade was relatively rare in this group. Liebler CA, Porter SR, Fernandez LE, Noon JM, and Ennis SE. 2017."America's Churning Races: Race and Ethnic Response Changes between Census 2000 and the 2010 Census." *Demography*. 54(1):259-284.

[3] Ruggles S, Flood S, Foster S, Goeken R, Pacas J, Schouweiler M, and Sobek M. IPUMS USA: Version 11.0 [dataset]. Minneapolis, MN: IPUMS, 2021. https://doi.org/10.18128/D010.V11.0

[4] See https://usa.ipums.org/usa/repwt.shtml.

[5] SPARC defines US-born and foreign-born following the Census Bureau's definition. A foreign-born person is anyone who was not a US citizen at birth. US-born refers to anyone born in the United States, Puerto Rico, a US Island Area, or abroad of a US citizen parent or parents (see: https://www.census.gov/topics/population/foreign-born/about.html).

four race categories (White, Black, American Indian, and Asian or Pacific Islander) and required a single-race response. Hispanic (including Latino and Spanish) and non-Hispanic were defined as ethnicities.[6] In 1997, OMB revised the federal standard for race categories[7] to require that multiple-race responses be allowed and to split the Asian/Pacific Islander category into two categories: Asian and Native Hawaiian and other Pacific Islander. The ethnicity standard was not revised. The revised standards created the need to re-categorize post-1997 multiple-race reports into the pre-1997 single-race categories for effective cross-time comparisons and for research that uses multiple datasets which adhere to differing standards.

When the revised standards were issued, analysts searched for strategies to deal with mismatches between categories from data collected under the different standards. The choice of a bridging strategy has a powerful impact on the population size and characteristics of groups with many multiple-race members (e.g., Pacific Islander and American Indian/Alaska Native). *Deterministic* bridging strategies assign all people from the same multiple-race group to the same simplified race group, regardless of their characteristics. Many *deterministic* bridging strategies were suggested, such as assigning all multiple-race individuals to the smallest of their named groups for analysis (e.g., all Black-Asian people are assigned to the Asian category). As a rule, deterministic bridging strategies are not empirically based.

In this context, a team of researchers (Ingram and colleagues)[8] working for the National Center for Health Statistics (NCHS) developed a high-quality *probabilistic* bridging method to create empirically-based simplified race/ethnic categories for analysis. The team leveraged the National Health Interview Survey's 1997-2000 pooled data[9] in which all multiple-race respondents were asked to nominate a single "best" race. They estimated regression models predicting multiple-race respondents' "best" race using characteristics of the individual and their county context. The resulting coefficients were applied to nationally-representative microdata to "bridge" the large variety of multiple-race responses into a few simplified categories. The bridging equations and resulting population estimates have been widely used by the Census Bureau, NCHS, and other analysts for almost two decades.

---

[6] See: OMB. 1977. Race and Ethnic Standards for Federal Statistics and Administrative Reporting. *Statistical Policy Directive 15*.

[7] See: OMB. 1997. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. *Federal Register* 62FR58781-58790.

[8] Ingram DD, Parker JD, Schenker N, Weed JA, Hamilton B, Arias E, and Madans JH. 2003. United States Census 2000 Population with Bridged Race Categories. National Center for Health Statistics. *Vital Health Statistics* 2(135).

[9] Data were pooled to increase case counts because bridging equations are calculated using specific multiple-race responses (e.g., Asian-Black or AIAN-White), most of which are rare.

A practical limitation of the work by Ingram and colleagues is the reliance on county-level context characteristics, which are not available in public microdata. Liebler and Halpern-Manners[10] applied the Ingram et al. regression coefficients to public-use data and coordinated its release through IPUMS USA as six public-use microdata variables.[11] Although Ingram and colleagues' work was very high quality, the bridging coefficients are based on data from 1997-2000. In recognition of the aging coefficients, IPUMS USA discontinued the six bridged race variables in 2014.

Simplified race/ethnicity variables serve important functions. They enable work with data collected under varying schemes, and they allow researchers to include multiple-race respondents in analysis in meaningful ways (as opposed to using a vague "two or more races" category). This research describes a renewed effort to make updated variables public. The research described here builds on prior work by generating a full series of bridging equation coefficients (including a replication of prior work), applying them to nationally-representative data, and disseminating results in multiple formats. The National Cancer Institute commissioned the present work to update the bridging coefficients and generate population estimates by simplified race/ethnicity because these are used as denominators for their cancer rate calculations.

2. Estimate updated race bridging coefficients using NHIS data

The first steps for creating updated bridging coefficients involved work with the restricted-use National Health Interview Survey. The restricted-use version of the data was required because the regression equations use variables about the individual's county, as well as the individual themselves. Research in the Minnesota Federal Statistical Research Data Center (MnRDC) was completed in the summer of 2020. Procedures described in this section, completed within the MnRDC, were:

1) Reduce the number of race/ethnicity categories
2) Code or import individual-level and county-level characteristics (independent variables)
3) Estimate regression coefficients for four-year periods, 1997-2018, and
4) Compare coefficient results to prior results

[10] Liebler CA and Halpern-Manners A. 2008. A Practical Approach to Using Multiple-Race Response Data: A Bridging Method for Public-Use Microdata. *Demography* 45: 143–155.

[11] These variables are available in Census and ACS data from 2000 to 2014 in IPUMS-USA. They are: RACESING, PROBAI, PROBAPI, PROBBLK, PROBOTH, and PROBWHT. Note that these are different groups than the five developed here. In this report (like in Ingram et al., 2003), single-race non-Hispanic Some Other Race respondents are allocated to one of the five simplified categories rather than having their own category (PROBOTH). See usa.ipums.org/usa.

2.1. Reduce the number of race/ethnicity categories

The current research follows the strategies used by Ingram and colleagues (2003) and the Census Bureau's Modified Race Data Summary File[12] to combine some response categories before estimating bridging equations.

There are currently five federally-defined races, and respondents are invited to mark one or more. There are 31 unique combinations of these five race categories. In the 1977 definition of race groups, Asian, Native Hawaiian, and other Pacific Islander groups were categorized together in the "Asian and/or Pacific Islander" category. Therefore, any NHIS who reported any combination of these groups was coded into a simplified "Asian/Pacific Islander" (API) category. Reducing the categories to only four simplified race group categories (AIAN, API, B, and W) is also useful because the number of multiple-race respondents to the NHIS in any specific multiple-race group is relatively small.

There are 11 possible combinations of the four remaining race categories (AIAN, API, B, W).

1) AIAN-API
2) AIAN-B
3) AIAN-W
4) API-B
5) API-W
6) B-W
7) AIAN-API-B
8) AIAN-API-W
9) AIAN-B-W
10) API-B-W
11) AIAN-API-B-W

Using procedures described below, these 11 multiple-race categories were bridged to the four non-Hispanic simplified race/ethnicity categories.

2.2. Code or import individual-level and county-level characteristics

Because this research replicates and expands upon prior work, the same variables were used to produce the estimates.[13] The county-level independent variables were derived from public data and imported into the MnRDC for use with the restricted-use NHIS data.

[12] https://www.census.gov/programs-surveys/popest/technical-documentation/research/modified-race-data.html

[13] Ingram and colleagues did not include the names of the restricted-use variables in their work and this project was unable to exactly match their published case count. Regression results are similar but not identical, probably because there are slightly different cases in the samples and the samples are small.

Individual-level independent variables:

- Age (in years) per 10 years
- Hispanic origin (yes/no) and
- Sex (male/female)

County-level independent variables

- NCHS Urbanization Level (discussed below)
- County racial composition (discussed below) and
- Census region (West, Northeast, Midwest, and South)

Variables were calculated for each individual year from 1997 to 2018 because the NHIS data span this period. Data were then pooled into 19 four-year periods for the regression models after the year-specific context variables were assigned to each case.

2.2.1. Urbanization level

The urbanization level of all counties in the US is defined by the National Center for Health Statistics Urban-Rural Classification Scheme (NCHS UR Codes).[14] These codes are revised occasionally and there are currently three versions: 1990-based, 2006, and 2013.

For the NHIS data from years 1997 to 2000, the 1990-based NCHS UR codes were used. These codes and years match those used by Ingram and colleagues, increasing the ability of the current analysis to replicate those analyses. The analyses using NHIS data from 2001 to 2007 used the 2006 Urban-Rural codes.[15] And the analyses using NHIS data from 2008 to 2018 used the 2013 NCHS UR codes. These three sets of codes are distributed by the National Center for Health Statistics on their public use data files website.[16]

The current analysis combined the UR codes to create the four codes used by Ingram and colleagues. Definitions are those given by NCHS on their website. Codes used in the present analysis were:

- *Large central metropolitan counties*: in a metropolitan statistical area (MSA) of 1 million population that: 1) contain the entire population of the largest principal city of the MSA, or 2) are

[14] The codes are defined in Ingram DD, Franco SJ. 2012. NCHS Urban-Rural Classification Scheme for Counties. National Center for Health Statistics. *Vital Health Stat* 2(154). Also see: Eberhardt MS, Ingram DD, Makuc DM, et al. 2001. *Urban and Rural Health Chartbook*. Health, United States. Hyattsville, Maryland: National Center for Health Statistics.

[15] These are, in turn, based on the 2003 USDA urban-rural classification codes.

[16] See https://www.cdc.gov/nchs/data_access/urban_rural.htm

completely contained within the largest principal city of the MSA, or 3) contain at least 250,000 residents of any principal city in the MSA.

- *Large fringe metro counties*: in an MSA of 1 million or more population that do not qualify as large central.
- *Medium and small metro counties*: medium metro counties: in MSA of 250,000-999,999 population. Small metro counties: in MSAs of less than 250,000 population.
- *Nonmetropolitan counties*: micropolitan counties in micropolitan statistical area as well as noncore counties not in micropolitan statistical areas.

2.2.2. County racial composition

The bridging regression equations also used county-level contextual variables describing specific aspects of the racial composition of the county. The Census Bureau Population Estimates Program publishes county-level estimates for the resident population as of July 1 for a given year.[17] Based on these county-level estimates, the categories used in the regression were:

- Percent of the county population reporting single-race American Indian or Alaska Native. Where it improved model fit (AIAN-B and AIAN-W), logged percent was substituted.
- Percent of the county population reporting single-race Asian, single-race Pacific Islander, or both Asian and Pacific Islander.
- Percent of the county population reporting single-race Black. This number was instead squared where this transformation improved model fit (AIAN-B).
- Percent of the county population reporting two or more census races (i.e., two or more of: AIAN, Asian, Black, NHPI, White, or "Some Other Race"). This variable is set to 0 in the 1997-1999 data because multiple-race reporting was not yet allowed in census data.

2.3. Estimate regression coefficients for four-year periods, 1997-2018

Each of the respondents in the NHIS who reported multiple races was asked a follow up question[18] about which single race they would have reported. In this report, this is referred to as the "best" single race. Not all multiple-race respondents gave a "best" single race; Ingram and colleagues (Table 5) found that 19.2% of respondents in the 1997-2000 NHIS did not give a single race response in their pooled sample from 1997-2000 and similar levels of non-response were found in the current work. Because the regression

[17] See https://www.census.gov/data/datasets/time-series/demo/popest/intercensal-1990-2000-state-and-county-characteristics.html for 1997-1999 intercensal estimates. The 2000-2010 intercensal estimates are from: https://www.census.gov/data/datasets/time-series/demo/popest/intercensal-2000-2010-counties.html. And 2010-2018 intercensal estimates are from: https://www.census.gov/programs-surveys/popest/technical-documentation/file-layouts.html.

[18] The question is: ''Which of these groups would you say best represents your race?''

equations predict the person's response to this item, the bridging equations described in this report include only multiple-race respondents who provided a "best" single race.

Using cases where a multiple-race person did report a best race, logistic and multinomial logistic regression models were run separately for six multiple-race groups with sufficient case counts (at least 100 multiple-race people in this group reported a "best" race in the four-year pooled data). Groups with sufficient case counts to support separate models were:

- AIAN-B
- AIAN-W
- API-B
- API-W
- B-W
- AIAN-B-W

These are the same six groups as were estimated separately by Ingram and colleagues.

Bridging coefficients for the remaining five multiple-race groups were generated with a seventh "all groups" model – a composite multinomial logistic regression model based on the "best" race given by *all* multiple-race respondents who reported a "best" race. Ingram and colleagues present a thoughtful discussion about using a composite model; see that publication for more information.

2.4. Compare coefficient results to each other and Ingram et al.

The Appendix to this report shows the regression coefficients from the five logistic and two multinomial logistic models for each of 19 four-year periods between 1997 and 2018.

These results are summarized in Figure 1, which shows a line for each independent variable. The line connects 20 points, the first of which is Ingram et al.'s coefficient for that variable. The second is the current analysis's coefficient for that variable in 1997-2000, the third is the current analysis's coefficient for 1998-2001, and so on.

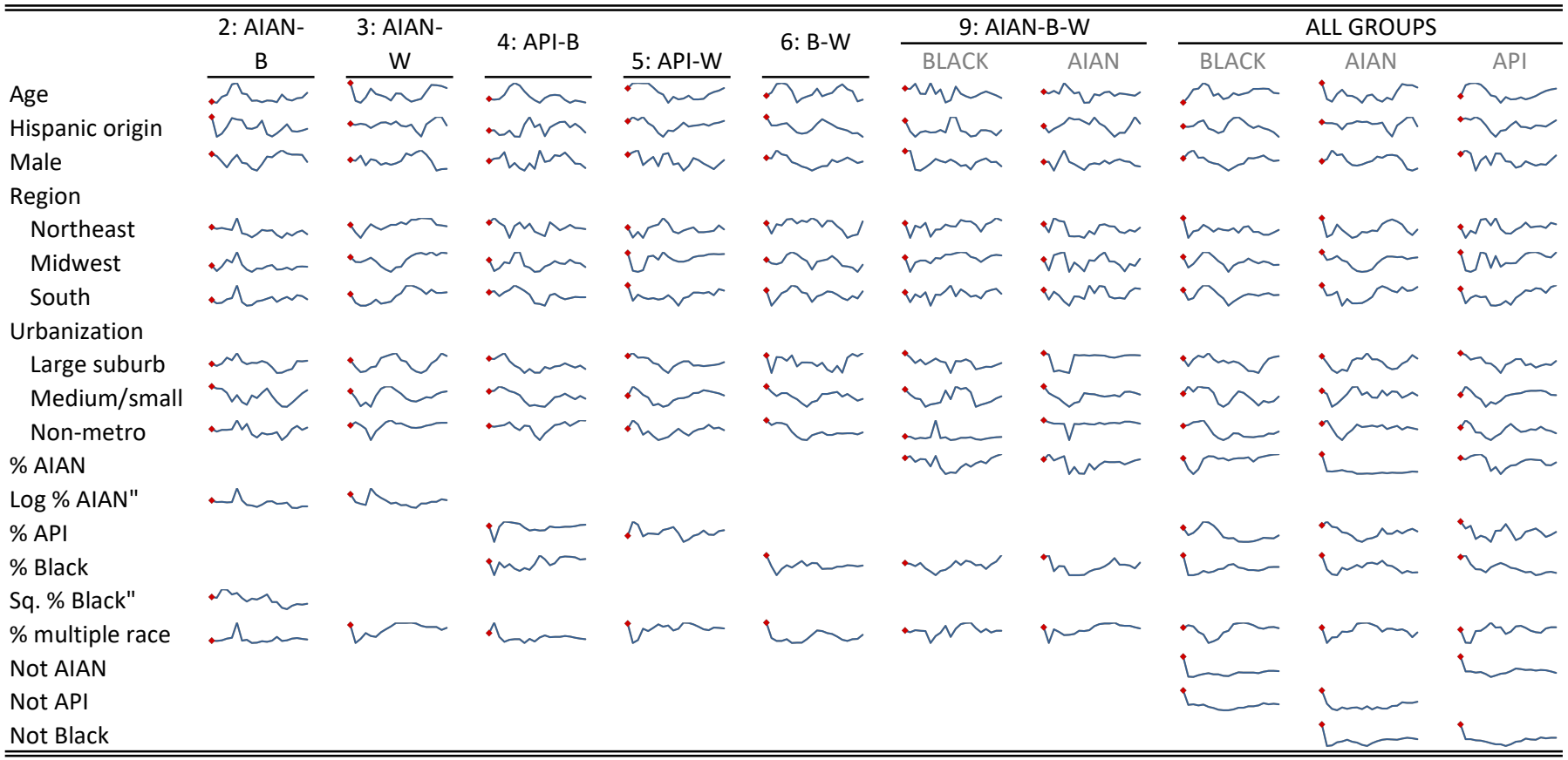
Two conclusions are apparent in Figure 1. First, the present results are in the same general range as Ingram et al.'s results. Second, the coefficients are not stable over time. This instability is likely caused by multiple factors, including the very small sample sizes for the regressions (usually 100 to 300) and the reality that the regression equations contain only a few limited measures that might influence a multiple-race person's reported "best" race. Analysts using these results should keep in mind that these are imperfect estimates of a complex reality.

2.5. Impact of new coefficients on population estimates

How similar are population estimates generated by the revised bridging coefficients to estimates based on bridging coefficients developed by Ingram and colleagues? Table 1 shows the impact of the revised bridging coefficients (using fractional allocation, explained below) on the estimated populations of the

United States with simplified race/ethnicity categories applied to the 2006 ACS data (from IPUMS USA). In columns labeled “Liebler,” the population totals are based on the procedures described in this report and using the updated bridging coefficients from the 2003-2006 NHIS data. In columns labeled “Ingram et al.,” all aspects of the calculations are identical except the bridging coefficients are from Ingram and colleagues’ 2003 work. Comparing the two sets of columns reveals that the estimated population sizes by sex, age group, foreign/US-born status, and simplified race/ethnicity are very similar. Most subpopulation estimates categories are within 2%, with larger difference among foreign-born subpopulations and especially impacting the AIAN category.

Figure 1: Bridging coefficients from 19 models compared to Ingram et al.'s bridging coefficients



◆ = Ingram et al.'s coefficient using 1997-2000 pooled NHIS data

= Regression coefficients for each of 19 similar models with four years of pooled NHIS data, 1997-2018

Table 1: Impact of revised bridging coefficients on the estimated populations of the United States with simplified race/ethnicity categories, 2006

Year = 2006	Non-Hispanic Populations								Hispanic Population	
	AIAN		Asian/PI		Black		White			
	Liebler	Ingram et al.	Liebler	Ingram et al.	Liebler	Ingram et al.	Liebler	Ingram et al.	Liebler	Ingram et al.
US-Born Males										
Ages 0-4	88,138	87,620	439,501	430,826	1,554,192	1,566,667	5,850,193	5,846,911	2,311,571	2,311,571
Ages 5-9	84,964	83,308	378,792	374,653	1,513,661	1,522,846	5,825,014	5,821,624	1,887,035	1,887,035
Ages 10-14	107,744	106,346	344,648	341,003	1,585,899	1,594,044	6,227,394	6,224,292	1,668,613	1,668,613
Ages 15-19	118,042	116,866	311,159	309,040	1,629,279	1,634,465	6,685,685	6,683,793	1,377,057	1,377,057
Ages 20-24	94,143	92,395	245,862	243,962	1,344,116	1,347,689	6,396,117	6,396,192	1,082,780	1,082,780
Ages 25-29	77,837	76,668	175,187	174,682	1,155,655	1,159,021	5,911,023	5,909,330	932,641	932,641
Ages 30-34	73,878	72,485	126,539	126,476	1,023,876	1,025,392	5,518,680	5,518,620	752,488	752,488
Ages 35-39	77,481	76,190	97,855	97,607	1,085,625	1,086,769	6,424,195	6,424,590	640,088	640,088
Ages 40-44	81,477	80,291	81,338	81,834	1,158,946	1,159,921	7,197,617	7,197,334	588,811	588,811
Ages 45-49	85,667	84,568	78,209	78,697	1,117,699	1,118,217	7,661,723	7,661,816	501,564	501,564
Ages 50-54	70,456	69,227	66,598	66,572	939,821	940,533	7,078,325	7,078,867	391,250	391,250
Ages 55-59	57,316	56,411	55,836	56,111	761,309	761,489	6,458,617	6,459,067	317,260	317,260
Ages 60-64	38,724	38,204	34,025	34,319	508,468	508,641	4,888,785	4,888,839	219,754	219,754
Ages 65-69	30,865	30,588	28,068	28,242	381,590	382,053	3,619,403	3,619,043	153,362	153,362
Ages 70-74	20,397	20,104	22,588	22,973	284,238	284,108	2,933,475	2,933,512	126,365	126,365
Ages 75-79	12,419	12,248	23,231	23,451	214,565	214,593	2,499,091	2,499,014	98,668	98,668
Ages 80-84	7,364	7,314	19,146	19,317	131,908	131,858	1,755,790	1,755,719	59,184	59,184
Ages 85-89	2,733	2,732	9,499	9,568	59,175	59,051	861,725	861,780	24,659	24,659
Ages 90-94	1,255	1,283	4,526	4,557	28,047	28,015	387,772	387,745	8,687	8,687
Ages 95 +	-	-	-	-	-	-	3,898	3,898	-	-
Average difference		805		1,196		2,399		797		-

Note: The Liebler column uses fractional assignment and the updated bridging coefficients from the 2003-2006 NHIS data. The Ingram et al. column uses fractional assignment and the bridging coefficients calculated by Ingram et al. (2003) using the 1997-2000 NHIS data.

Table 1, continued.

Year = 2006	Non-Hispanic Populations								Hispanic Population	
	AIAN		Asian/PI		Black		White			
	Liebler	Ingram et al.	Liebler	Ingram et al.	Liebler	Ingram et al.	Liebler	Ingram et al.	Liebler	Ingram et al.
US-Born Females										
Ages 0-4	85,642	84,711	408,681	402,786	1,479,219	1,482,260	5,523,827	5,527,613	2,240,343	2,240,343
Ages 5-9	88,730	88,176	354,681	351,516	1,445,514	1,446,544	5,515,468	5,518,156	1,824,526	1,824,526
Ages 10-14	96,571	95,144	327,824	325,038	1,551,523	1,550,330	5,904,723	5,910,130	1,616,118	1,616,118
Ages 15-19	112,393	110,825	293,222	292,355	1,586,507	1,585,786	6,337,010	6,340,166	1,365,409	1,365,409
Ages 20-24	99,592	98,605	226,005	225,263	1,342,095	1,341,077	6,051,858	6,054,604	1,036,432	1,036,432
Ages 25-29	86,322	85,523	163,157	162,905	1,251,403	1,250,134	5,829,282	5,831,602	907,002	907,002
Ages 30-34	73,801	72,708	120,334	120,535	1,117,676	1,117,401	5,429,079	5,430,246	728,068	728,068
Ages 35-39	88,643	87,473	92,538	92,519	1,284,010	1,284,278	6,320,144	6,321,065	658,944	658,944
Ages 40-44	87,260	85,992	79,267	79,387	1,350,502	1,350,416	7,278,353	7,279,586	590,972	590,972
Ages 45-49	92,013	90,232	76,949	77,082	1,302,699	1,302,875	7,731,906	7,733,377	521,894	521,894
Ages 50-54	78,708	77,526	62,882	62,808	1,130,334	1,130,424	7,270,039	7,271,205	428,482	428,482
Ages 55-59	68,807	67,459	48,620	48,611	932,687	932,142	6,666,709	6,668,610	342,407	342,407
Ages 60-64	45,367	44,634	38,444	38,456	647,770	647,382	5,163,199	5,164,307	243,365	243,365
Ages 65-69	31,913	31,653	29,565	29,602	531,006	531,067	3,983,973	3,984,136	176,709	176,709
Ages 70-74	24,169	23,876	24,265	24,460	412,465	412,324	3,501,448	3,501,687	148,241	148,241
Ages 75-79	18,647	18,301	28,064	27,902	326,166	326,179	3,272,922	3,273,418	134,269	134,269
Ages 80-84	10,860	10,761	24,946	25,170	255,836	255,822	2,781,212	2,781,101	95,875	95,875
Ages 85-89	8,432	8,408	14,275	14,543	146,109	146,026	1,680,719	1,680,558	44,641	44,641
Ages 90-94	4,842	4,768	8,795	8,791	110,366	110,339	1,077,357	1,077,462	24,800	24,800
Ages 95 +	-	-	-	-	-	-	10,710	10,710	-	-
Average difference		797		758		522		1,517		-

Note: The Liebler column uses fractional assignment and the updated bridging coefficients from the 2003-2006 NHIS data. The Ingram et al. column uses fractional assignment and the bridging coefficients calculated by Ingram et al. (2003) using the 1997-2000 NHIS data.

Table 1, continued.

Year = 2006	Non-Hispanic Populations								Hispanic Population	
	AIAN		Asian/PI		Black		White			
	Liebler	Ingram et al.	Liebler	Ingram et al.	Liebler	Ingram et al.	Liebler	Ingram et al.	Liebler	Ingram et al.
Foreign-Born Males										
Ages 0-4	543	539	42,922	42,815	13,733	13,777	49,971	50,038	88,971	88,971
Ages 5-9	1,093	1,051	84,893	84,556	33,328	33,518	103,240	103,429	210,156	210,156
Ages 10-14	977	972	108,985	108,823	60,257	60,282	136,588	136,728	318,937	318,937
Ages 15-19	1,398	1,353	184,319	184,713	86,746	86,984	214,606	214,021	533,995	533,995
Ages 20-24	792	764	267,251	267,476	125,033	125,162	253,936	253,609	987,873	987,873
Ages 25-29	2,081	2,067	372,666	373,004	137,653	137,177	262,016	262,168	1,331,976	1,331,976
Ages 30-34	2,230	2,217	537,180	537,536	151,708	151,640	355,848	355,574	1,366,115	1,366,115
Ages 35-39	2,130	2,120	544,015	544,227	152,476	152,506	413,240	413,009	1,237,194	1,237,194
Ages 40-44	2,038	1,966	489,039	488,950	169,787	169,734	438,134	438,350	1,058,587	1,058,587
Ages 45-49	2,038	2,008	414,923	415,363	159,810	159,786	450,734	450,347	791,176	791,176
Ages 50-54	2,340	2,208	362,473	362,409	129,694	129,615	393,392	393,667	580,597	580,597
Ages 55-59	1,432	1,384	289,646	289,840	84,610	84,491	343,874	343,847	407,784	407,784
Ages 60-64	846	815	213,805	213,927	55,349	55,294	247,905	247,869	276,448	276,448
Ages 65-69	475	438	158,342	158,620	32,881	32,841	221,478	221,277	189,866	189,866
Ages 70-74	631	644	114,748	114,933	24,683	24,618	181,457	181,323	138,787	138,787
Ages 75-79	264	257	74,118	74,173	14,737	14,719	153,862	153,832	94,755	94,755
Ages 80-84	161	169	45,020	45,113	8,466	8,466	120,689	120,588	65,934	65,934
Ages 85-89	31	26	19,186	19,200	4,255	4,242	56,857	56,859	24,150	24,150
Ages 90-94	4	4	8,888	8,922	1,442	1,442	29,827	29,793	13,716	13,716
Ages 95 +	-	-	-	-	-	-	23	23	181	181
Average difference		27		185		83		170		-

Note: The Liebler column uses fractional assignment and the updated bridging coefficients from the 2003-2006 NHIS data. The Ingram et al. column uses fractional assignment and the bridging coefficients calculated by Ingram et al. (2003) using the 1997-2000 NHIS data.

Table 1, continued.

Year = 2006	Non-Hispanic Populations								Hispanic Population	
	AIAN		Asian/PI		Black		White			
	Liebler	Ingram et al.	Liebler	Ingram et al.	Liebler	Ingram et al.	Liebler	Ingram et al.	Liebler	Ingram et al.
Foreign-Born Females										
Ages 0-4	524	510	59,502	59,242	16,395	16,747	50,364	50,286	74,650	74,650
Ages 5-9	1,216	1,194	104,742	104,547	30,898	30,942	92,651	92,823	195,090	195,090
Ages 10-14	1,114	1,087	112,882	112,779	53,316	53,525	132,167	132,089	303,149	303,149
Ages 15-19	1,359	1,302	171,368	171,674	98,121	97,959	187,639	187,552	430,434	430,434
Ages 20-24	704	704	274,134	274,156	126,161	126,381	240,364	240,122	682,623	682,623
Ages 25-29	1,866	1,714	419,637	419,738	153,696	153,287	270,883	271,341	913,499	913,499
Ages 30-34	2,391	2,357	573,871	574,347	166,701	166,303	348,582	348,537	1,023,291	1,023,291
Ages 35-39	3,625	3,608	567,322	567,846	176,855	176,235	411,346	411,458	1,037,769	1,037,769
Ages 40-44	2,039	2,003	530,983	530,694	176,936	176,512	445,040	445,788	900,085	900,085
Ages 45-49	1,935	1,858	480,272	480,968	155,729	155,109	442,346	442,346	723,039	723,039
Ages 50-54	1,202	1,137	426,884	427,665	122,150	121,628	389,960	389,765	552,239	552,239
Ages 55-59	1,723	1,597	361,397	362,120	88,388	87,791	373,532	373,532	419,648	419,648
Ages 60-64	611	571	249,705	249,840	71,146	70,963	305,023	305,110	315,970	315,970
Ages 65-69	864	877	194,980	195,273	49,580	49,482	287,550	287,343	241,170	241,170
Ages 70-74	841	858	145,657	145,763	38,636	38,484	269,859	269,887	187,339	187,339
Ages 75-79	162	159	112,374	112,603	27,213	27,057	232,417	232,347	136,187	136,187
Ages 80-84	60	30	68,659	68,694	16,960	16,966	198,728	198,716	93,401	93,401
Ages 85-89	38	33	29,583	29,803	7,721	7,483	105,240	105,263	48,985	48,985
Ages 90-94	56	78	21,427	21,413	5,979	5,989	81,966	81,948	32,754	32,754
Ages 95 +	-	-	-	-	-	-	179	179	121	121
Average difference		38		276		271		133		-

Note: The Liebler column uses fractional assignment and the updated bridging coefficients from the 2003-2006 NHIS data. The Ingram et al. column uses fractional assignment and the bridging coefficients calculated by Ingram et al. (2003) using the 1997-2000 NHIS data.

3. How to apply bridging coefficients to microdata

To put the estimated equations to use, they are applied to each multiple-race person's case. Based on the person's individual and contextual characteristics, the equation gives a number between 0 and 1, representing the predicted result if that specific person were asked for their "best" race. Once the predicted value for each race response is generated for each multiple-race person, the analyst can choose to use the predicted values as they are (this is "fractional allocation") or assign the entire case to the race category with the highest predicted value (this is "whole allocation").

Whole allocation is formatted in a way that is familiar to most researchers.[19] It is a categorical variable, with each person in one category. However, the fractional allocation approach is preferable when possible because of the enhanced level of detail and reduced loss of race/ethnicity information provided by the respondent. Fractional allocation retains three types of important information: (1) the fact that the person reported multiple races, (2) which races the person reported, and (3) the predicted values generated by the bridging calculation. Fractional allocation accounts for the reality that individual multiple-race people have only a fractional probability of choosing a particular single race if required to choose. Fractional allocation can be seen as a more ethical choice because it better represents respondents' intentions (i.e., the intention to be recorded as having multiple races).

To use the fractional allocation information, each race/ethnicity category must have a separate variable which ranges from 0.0 (for a person who did not report this race/ethnicity) to 1.0 (for a person who only reported this race/ethnicity) and the sum of these variables is 1.0 for all people.[20] Multiple-race respondents are assigned values on these variables using the bridging equations; the results depend on their personal and locational characteristics.

Table 2 shows hypothetical bridging results for five cases, as an example, and shows the IPUMS USA variable names for reference. The first row illustrates results for people who mark a single race; the fractional allocation is 1.0 for one of the race/ethnicity variables, and the whole allocation matches the original response. The middle three rows show results for multiple-race individuals. Based on their personal and locational characteristics, the person who reported both Chinese and White is predicted (based on their personal and local characteristics) to choose their Asian/Pacific Islander race if required to choose one race but has some likelihood of reporting White in this scenario. Because the predicted value is higher for the API answer, the person is assigned to API in the whole allocation variable (this is RACHSING in IPUMS USA). The fourth row illustrates a case where predicted values are equal for two groups; in IPUMS USA, the case is assigned to the less populated RACHSING category. The fifth row highlights the case of a person reporting Hispanic origins.

[19] This strategy was used by IPUMS USA in the variable RACHSING.

[20] In IPUMS USA, these variables are: PREDAI, PREDAPI, PREDBLK, PREDHISP, and PREDWHT.

An analyst using whole allocation with this five-person dataset would calculate an estimate of two Asian or Pacific Islander people, while an analyst using fractional allocation would calculate an estimate of 0.91651 Asian or Pacific Islander people. In analysis, researchers using fractional allocation can treat the predicted values in a manner similar to weights.

Table 2: Fractional allocation and whole allocation in microdata

Race and Hispanic origin responses	FRACTIONAL ALLOCATION					WHOLE ALLOCATION
IPUMS VARIABLE:	PREDAI	PREDAPI	PREDBLK	PREDHISP	PREDWHT	RACHSING
Black	0.0	0.0	1.00000	0.0	0.0	Black
Chinese & White	0.0	0.54637	0.0	0.0	0.45363	API
Navajo & Black	0.42371	0.0	0.57629	0.0	0.0	Black
Thai & Black & White	0.0	0.37014	0.37014	0.0	0.25972	API
Mexican, White	0.0	0.0	0.0	1.00000	0.0	Hisp

Note: IPUMS variables can be found at https://usa.ipums.org/usa/

4. Dissemination: Microdata variables on IPUMS USA

This section of the report explains the procedures used to apply the new bridging coefficients to the 2000 and 2010 public-use decennial census microdata and the 2000 to 2019 American Community Survey microdata for distribution via IPUMS USA. The result is that every person in these data resources has values for the five fractional allocation variables (PREDAI, PREDAPI, PREDBLK, PREDHISP, and PREDWHT) that sum to 1.0, and is coded into one of five categories for the whole allocation variable (RACHSING).[21] This section describes the procedures used to generate these useful microdata variables.

4.1. Reduce the number of ACS multiple race categories

The decennial census and ACS race and Hispanic ethnicity questions have many response options.[22] Respondents are invited to report multiple races and give detailed information about American Indian tribe, Asian or Pacific Islander group, or type of Hispanic origin. In the IPUMS USA data, these details can be found in the RACE and HISPAN variables; researchers are encouraged to use the other race/ethnicity information about individuals to supplement their use of the simplified race/ethnicity measure. In these Census Bureau data products, a supplemental race category called "Some Other Race" (SOR) is included as well. There are 63 unique combinations of the six race categories (AIAN, Asian, Black, Pacific Islander, White, and SOR; not counting race non-response), and there are 126 unique categories when Hispanic/non-Hispanic are also considered. With simplification and bridging, these are reduced to the five simplified race/ethnicity categories. In the categorical variable RACHSING, the five categories are:

1 White
2 Black/African American
3 American Indian/Alaska Native
4 Asian/Pacific Islander
5 Hispanic/Latino

There are several steps to simplify the 126 combinations of federally-defined race/ethnicity into these five categories, besides applying the bridging equations. *First*, people who reported any Hispanic origin,

[21] These IPUMS-USA (https://usa.ipums.org/usa/) variables can be found here:
- RACHSING: https://usa.ipums.org/usa-action/variables/RACHSING#description_section
- PREDAI: https://usa.ipums.org/usa-action/variables/PREDAI#description_section
- PREDAPI: https://usa.ipums.org/usa-action/variables/PREDAPI#description_section
- PREDBLK: https://usa.ipums.org/usa-action/variables/PREDBLK#description_section
- PREDHISP: https://usa.ipums.org/usa-action/variables/PREDHISP#description_section
- PREDWHT: https://usa.ipums.org/usa-action/variables/PREDWHT#description_section

[22] See https://www.census.gov/programs-surveys/acs/methodology/questionnaire-archive.html for ACS questionnaires.

regardless of race response, were coded into the simplified Hispanic category. *Second*, SOR responses were disregarded *if* the person also reported a federally defined race. Thus, for example, someone who reported both White and SOR was recoded as White. Note that most people who report single-race SOR also report Hispanic origins and were therefore coded as Hispanic. Single-race non-Hispanic SOR respondents – a small group – were allocated to one of the four non-Hispanic simplified race groups after multiple-race responses were bridged; see below for details. *Third*, non-Hispanic respondents who reported one or more American Indian or Alaska Native groups, but no other race group, were coded as AIAN. The same procedure was applied to people who reported multiple Asian groups or multiple Pacific Islander groups. *Fourth*, non-Hispanic respondents who reported any combination of Asian and/or Pacific Islander groups (but not AIAN, Black, or White) were recoded into API group.

These simplification procedures resulted in a decrease in the number of respondents designated as multiple-race for the bridging procedure because people who reported SOR and a federally-defined race (or races) were recoded to be only the other race(s), and people who reported both an Asian group and a Pacific Islander group were recoded to be in the Asian/Pacific Islander group.

After these simplifying recodes were complete, there were only 11 (non-Hispanic) multiple-race categories.[23] These are listed above in Section 2.1. Next, independent variables were coded so that the bridging coefficients could be applied to these cases.

4.2. Definitions and calculations of local characteristics

The Census Bureau safeguards privacy, in part, by providing only limited geographic information about respondents in decennial and ACS data. This presents a challenge when applying the bridging coefficients because they were designed to use county-level measures of urbanicity and racial composition.

The most detailed level of geography available in the decennial and ACS microdata from this period is a Public-Use Microdata Area, or PUMA, which is a geographically contiguous area within a state containing 100,000-199,999 population.[24] PUMA boundaries are created by the Census Bureau and are contiguous with county boundaries in many cases. Where possible, IPUMS USA has identified counties in the census and ACS; about 60% of ACS respondents that live in identifiable counties, which are generally in urban and suburban areas.[25] The county-level data described in Section 2.2.2 was used when the county was known.

---

[23] Census Bureau-provided dichotomous race variables (RACASIAN, RACAMIND, RACBLK, RACOTHER, RACPACIS, RACWHT, and RACNUM in IPUMS USA) were used.

[24] For more about PUMAs, see https://usa.ipums.org/usa-action/variables/PUMA#description_section

[25] See https://usa.ipums.org/usa-action/variables/COUNTYFIP#description_section

To substitute for county-level urbanization information when it is not identifiable, the IPUMS USA variable METRO was used and recoded as follows:

- *Large central metropolitan PUMA*: In a central or principal city in a metropolitan area (METRO=2)
- *Large fringe metro PUMA*: In a metropolitan area, not in a central or principal city (METRO=3)
- *Medium and small metro PUMA*: In a metropolitan area, but central or principal city status is mixed or indeterminable (METRO=4)
- *Nonmetropolitan PUMA*: Not in a metropolitan area or metropolitan status is indeterminable or mixed (METRO=0 or METRO=1)

To calculate the racial composition of the PUMA for use with the regression coefficients, the dataset being bridged was used to calculate the contemporary percent of the total PUMA population reporting:

- Single-race AIAN
- Single-race Black
- Single-race API (Any combination of Asian and Pacific Islander, but no other races)
- Two or more races

4.3. Apply coefficients from a 4-year period to a dataset

Bridging equations were calculated using pooled NHIS data from sequential 4-year periods, yet the decennial census is in a single year and ACS data are released in 1-year, 3-year, and 5-year datasets. The mismatch was resolved by matching the last year of the NHIS period to the last year of the microdata series. For example, the bridging coefficients from the pooled 2015-2018 NHIS analysis were applied to the 2018 1-year ACS data, the 2018 3-year ACS data (which is from 2016-2018), and the 2018 5-year ACS data (which is from 2014-2018). Because the NHIS series stops in 2018, the 2015-2018 bridging coefficients were applied to later data as well.

4.4. Race allocation for single-race non-Hispanic Some Other Race

The five simplified race/ethnicity categories, and the 11 simplified multiple-race categories, leave out one type of ACS response: Non-Hispanic Some Other Race (NH SOR). There were about 79,000 NHSOR people total in the 2005-2018 American Community Surveys.

The NH SOR respondents were allocated values on the PRED variables and RACHSING. This was accomplished in four steps. First, bridging coefficients were applied to all other cases in the dataset. Second, for each unique combination of 5-year age group, sex, US/foreign birthplace, state, and year, the mean value of each PRED variable was calculated. Third, the four non-Hispanic PRED variables were rescaled to sum to 1.0 because the person indicated that they are not Hispanic. Fourth, each NH SOR respondent was assigned a Hispanic probability of 0 and the rescaled mean probabilities of the four simplified race groups coinciding with their age group, sex, US/foreign birthplace, state, and year. From there, RACHSING codes were assigned. Most US-born NH SOR people were allocated to the White

RACHSING category, while foreign-born NH SOR people were more often allocated to the API category. Very few were allocated to Black or AIAN.

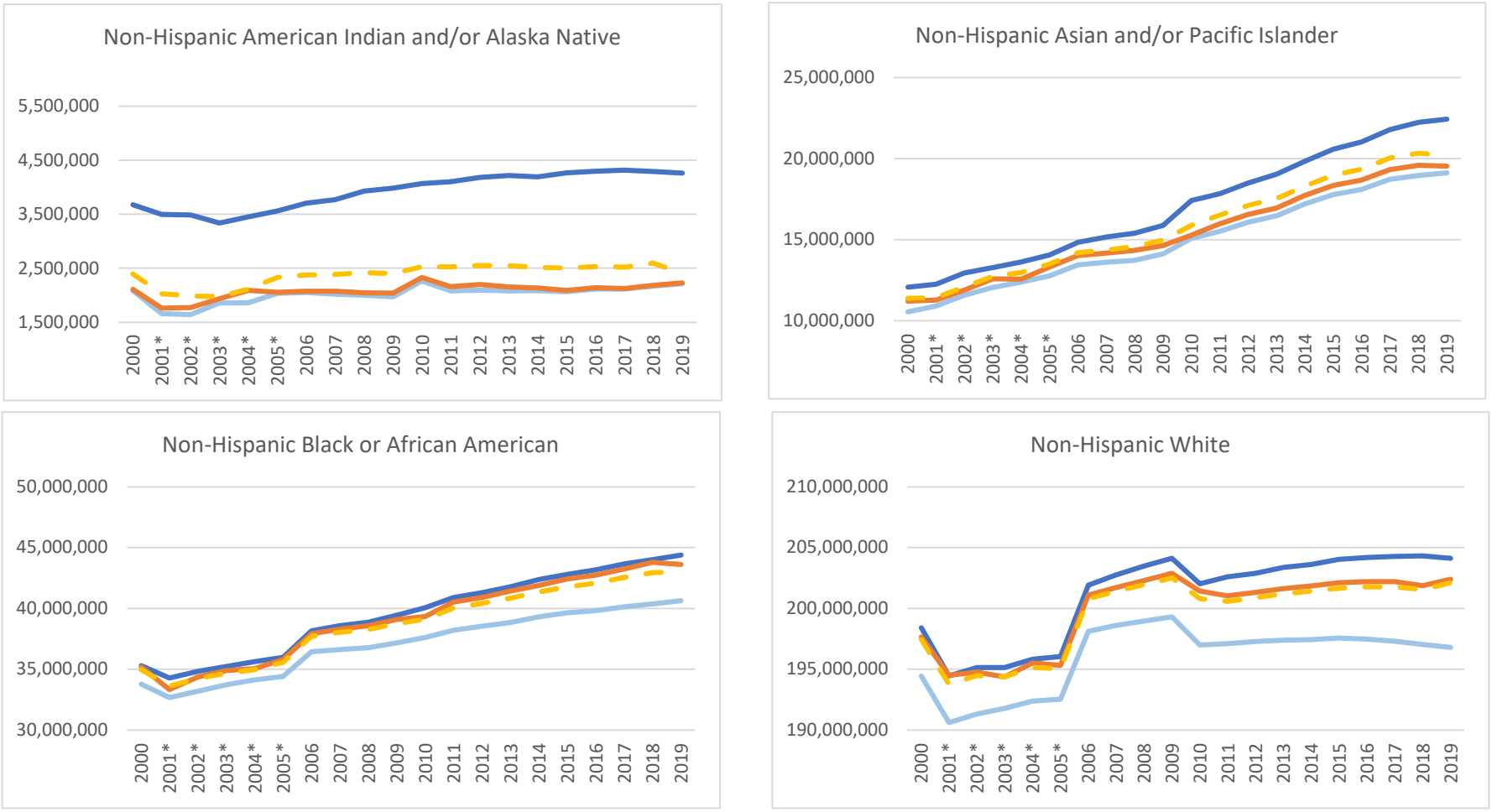
4.5. RACHSING and PRED variables in comparison

Using the simplified race/ethnicity measures provided in IPUMS USA (RACHSING and PRED variables) gives an inclusive alternative to the all-too-common approach of studying only single-race respondents and using an uninterpretable "two or more races" category. In this scenario, multiple-race people are essentially excluded from the benefits of research because they are not disaggregated. Because single-race respondents are a non-random subset of the people who reported a race group, this approach can also lead to biased results.

The simplified race/ethnicity measures area also a more practical solution to complex data than using the "one or more races" approach in which all people who marked a race category (whether they marked other race categories or not) are included in a population. The "one or more race" approach places multiple-race people in multiple categories such that a two-race person is seen in twice as many categories as a single-race person. The categories add to more than 100% of the population, which is inconvenient for many calculations.

Figure 2 shows the estimated total population using each of these strategies. The darkest and highest line represents the number of people in this non-Hispanic race group if the "one or more races" approach is used. The lightest and lowest line shows the population estimate of only single-race people (the "two or more races" population is excluded). In between these two lines are the estimates given by RACESING (whole allocation, shown in solid orange) and the PRED variables (fractional allocations, shown in a dashed yellow line). Fractional allocation gives consistently higher estimates of the total AIAN and API populations and consistently lower estimates of the total Black and White populations. This reflects the pattern within NHIS respondents (and thus the bridging equations) of more often choosing Black or White as their "best" single race than choosing AIAN or API.

Figure 2: Estimates of race/ethnic group total populations using four strategies, 2000 to 2019



(1) All people in this non-Hispanic race group, including both single- and multiple-race responses

(2) Non-Hispanic single race responses only (a response of both Asian and Pacific Islander is treated as a single race)

(3) RACHSING (whole allocation of non-Hispanic multiple-race and Some Other Race (SOR) responses)

(4) PREDAI, PREDAPI, PREDBLK, or PREDWHT (partial allocation of non-Hispanic multiple-race and SOR responses)

Data: Weighted Census 2000 1% PUMS, Census 2010 1% PUMS, and 2001-2009 and 2011-2019 one-year ACS data, via IPUMS-USA.

* Note that 2001-2005 data are from ACS years that were not nationally-representative.

5. Dissemination: Population estimates and standard errors on National Cancer Institute's SPARC

A second type of dissemination of this work is in the form of population estimates and standard errors generated using the new bridging equations. Using the 1-year ACS microdata, the National Cancer Institute disseminates results on a platform titled Survey-based Population-adjusted Rate Calculator or SPARC; see https://surveillance.cancer.gov/sparc/.

For dissemination in SPARC, population estimates for each of the five simplified race/ethnicity groups were calculated separately by 5-year age group (with top-coding), sex, US/foreign birthplace, state, and year. Replicate weight standard errors were also calculated and are included in SPARC. A foreign-born person is anyone who was not a US citizen at birth. US-born refers to anyone born in the United States, Puerto Rico, a US Island Area, or abroad of a US citizen parent or parents. These are Census Bureau definitions[26] and are applied in SPARC as well as elsewhere in this research when the distinction is made.

The estimates in SPARC use the fractional allocation strategy for calculating subpopulation estimates. Unlike whole allocation, fractional allocation does not introduce rounding error by disregarding the smaller predicted values. Also, the fractional predicted values represent individuals who chose to report multiple races – they did not choose to report within very small set of categories in the simplified race/ethnicity list. The fractional predicted values retain information about the answer the respondent *did* choose to report. Therefore, fractional allocation is more realistic and more respectful to the respondent.

The standard errors provided in SPARC were calculated using the ACS replicate weights provided by IPUMS-USA. This website explains why replicate weights are preferred and how they work: https://usa.ipums.org/usa/repwt.shtml.

[26] See https://www.census.gov/topics/population/foreign-born/about.html.

6. Conclusion

Race is a complex feature of the United States, impacting most aspects of life. Researchers aiming to understand incidence and prevalence rates often need to use data from multiple sources. For example, states gather data about cause of death while the federal government gathers population data (via the Census Bureau). Variation in race measurement strategies must be managed before, for example, cause-specific mortality rates can be calculated.

The research described in this report involved developing new bridging equations to enable the creation and of updated, simplified race/ethnicity measures in two forms (fractional and whole allocation). The five simplified race/ethnicity categories are: Hispanic, and non-Hispanic American Indian/Alaska Native, Asian/Pacific Islander, Black, and White. These simplified measures are very practical for working with complex race/ethnicity data and contradictory data collection systems.

The simplified race/ethnicity categories are a helpful alternative for analysts working with complex race data such as that given in the decennial census or ACS data. These data give immense detail on the race/ethnic responses of individuals, with not only the many federally-defined general race and ethnicity categories but also subcategories such as AIAN tribe and Asian, PI, and Hispanic country of origin. The detail is retained in other variables and is worth exploration, but many analyses have a practical need for a categorization scheme that includes all people and has a small number of meaningful categories.

The simplified race/ethnicity categories made possible with bridging are also required for working with contradictory data sources within the same equation. For example, cancer rates are calculated using state registry data in the numerator and federally-collected population data in the denominator. State cancer registries have been slow to transition to collecting high-quality and analyzable multiple-race data. Rates can be calculated, however, using a simplified race/ethnicity categorization scheme in both the numerator and denominator.

The work described in this report involved two major steps. First, multivariate models were developed using the “best” single race indicated by multiple-race respondents to the (restricted-use) National Health Interview Survey. Models included demographic and contextual characteristics of the multiple-race respondents. This process was repeated for 19 four-year periods from 1997 to 2018, generating “bridging” coefficients to be applied to other data. Results were similar to those found by Ingram and colleagues. Second, the bridging coefficients were applied to nationally-representative microdata. The total population in each simplified race/ethnicity group was calculated and disseminated in microdata format at IPUMS USA (using the RACHSING and PRED variables) and in the form of population estimates and standard errors in the National Cancer Institute’s SPARC tool.

Appendix

Model coefficients from logistic regression models and multinomial logistic regression models predicting "best" race chosen by multiracial people in the National Health Interview Survey, 1997-2018

19 four-year pooled samples with seven models each

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **1997-2000**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	-0.01940	-0.10857	0.03770	0.15437 *	0.07236
Hispanic origin (Not Hisp. = ref.)	-1.62043 *	0.85140 *	0.07690	0.45795 #	-0.45630
Male (Female = reference)	-0.10551	-0.00210	0.48474	0.05424	0.09937
Region (West = reference)					
Northeast	-0.30721	-0.26644	0.76117	-0.62111	-0.45503
Midwest	-1.19803 *	-0.21743	-3.11630 #	-0.77434 *	0.03367
South	-0.62206	-0.35048	0.25457	-0.53052 #	-0.63809 *
Urbanization level (Large urban = reference)					
Large suburban	-0.32461	-0.37661	1.36400 #	0.73258 *	-0.39237
Medium/small metro	0.69566	-0.27109	1.88186 *	0.67829 *	0.03998
Non-metro	-0.36877	0.25020	-0.25920	0.98910 #	0.12887
County pop: % AIAN					
Log of "County pop: % AIAN"	-0.12808	0.24033 *			
County pop: % API			-0.44541	0.04241 #	
County pop: % Black			-0.06475		
Square of "County pop: % Black"	-0.00001				0.00026
County pop: % multiple race	-0.02601	-0.25746 *	1.23671	-0.05814	-0.00334
Not AIAN					
Not API					
Not Black					
Constant	2.09787 *	0.18769	-0.29339	-0.99709 *	0.73717 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **1997-2000**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.35479 *	0.19359	0.04722	-0.08173 *	0.15270 *
Hispanic origin (Not Hisp. = ref.)	-1.11872	-0.71367	-0.60330 *	0.80118 *	0.28650
Male (Female = reference)	0.58427	0.00332	0.15536	0.05009	0.07831
Region (West = reference)					
Northeast	-3.38580 *	-2.67760	-0.64297 *	-0.47218	-0.68292 #
Midwest	-4.01993 *	-2.62648	-0.41293	-0.14124	-0.71914 *
South	-1.87597	-2.64553	-0.56225 *	-0.35391	-0.64151 *
Urbanization level (Large urban = reference)					
Large suburban	1.31286	1.09637	-0.20343	-0.14554	0.35808
Medium/small metro	0.06208	0.54898	0.19873	-0.10646	0.47557 *
Non-metro	1.57269	1.78258	0.06135	0.47762	0.94280 *
County pop: % AIAN	0.49931	1.79384 *	-0.11635	0.08113 *	0.04202
Log of "County pop: % AIAN"					
County pop: % API			0.01692	0.06679 *	0.02326
County pop: % Black	0.04517	0.12610 *	-0.00299	0.00514	0.04090 *
Square of "County pop: % Black"					
County pop: % multiple race	-0.09742	-2.73301 *	0.04186	-0.22856 *	-0.03095
Not AIAN			-2.29033 *	---	-0.36419
Not API			-1.29620 *	-0.31434	---
Not Black			---	-2.64858 *	-1.81267 *
Constant	2.04139	1.15723	4.41443 *	1.89212 *	1.29927 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **1998-2001**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	0.08735	-0.12051	0.05779	0.15870 *	0.11877 *
Hispanic origin (Not Hisp. = ref.)	-1.37111 *	0.89628 *	-0.79615	0.27944	-0.45029
Male (Female = reference)	-0.40955	0.06897	0.51218	0.08301	0.35477 *
Region (West = reference)					
Northeast	-0.15007	-0.61096	-0.01361	-0.26286	-0.03432
Midwest	-0.15792	-0.24272	-2.26691	-0.82593 *	0.00007
South	-0.54644	-0.47027	-0.79743	-0.26230	-0.38029
Urbanization level (Large urban = reference)					
Large suburban	0.04392	-0.66164 #	2.00112 *	0.37224	-0.03609
Medium/small metro	0.67331	-0.62012 *	2.92659 *	0.45646 #	-0.14544
Non-metro	-0.14065	0.03625	0.12297	0.51546	0.32385
County pop: % AIAN					
Log of "County pop: % AIAN"	-0.10691	0.18220 *			
County pop: % API			-0.02377	0.03467 *	
County pop: % Black			0.01518		
Square of "County pop: % Black"	0.00042				-0.00015
County pop: % multiple race	0.03811	-0.19753 *	-0.02446	-0.03459	-0.06642
Not AIAN					
Not API					
Not Black					
Constant	1.50591 #	0.15947	-0.64767	-0.96362 *	0.41742

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **1998-2001**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.45335 *	0.30127	0.10715 *	-0.09581 *	0.17822 *
Hispanic origin (Not Hisp. = ref.)	-1.66225	-0.11663	-0.57045 *	0.80829 *	0.18688
Male (Female = reference)	-0.89229	-1.15143	0.22679	0.19350	0.02405
Region (West = reference)					
Northeast	-0.56773	2.07396	-0.29754	-0.32367	-0.31914
Midwest	-1.40007	1.05569	-0.20025	-0.07402	-0.77726 *
South	-0.36049	-0.41605	-0.45519 #	-0.32674	-0.39103
Urbanization level (Large urban = reference)					
Large suburban	0.03183	-20.84581 *	0.08618	-0.48058	0.08040
Medium/small metro	-0.46962	-0.18681	0.06020	-0.52783 *	0.25261
Non-metro	-0.99928	-0.22569	0.25974	-0.03210	0.49547
County pop: % AIAN	0.13499	0.35970	-0.17159 #	0.07094 *	0.11194 *
Log of "County pop: % AIAN"					
County pop: % API			0.02769	0.03500 *	0.02828 #
County pop: % Black	0.02865	-0.01178	-0.00309	-0.00665	0.01707
Square of "County pop: % Black"					
County pop: % multiple race	0.04599	-0.33217	0.02532	-0.14756 *	-0.03293
Not AIAN			-2.33823 *	---	-0.35884
Not API			-1.12498 *	-1.52303 *	---
Not Black			---	-2.59357 *	-1.86648 *
Constant	1.31960	-0.46180	3.95637 *	3.36176 *	1.34936 #

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **1999-2002**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	0.12010	-0.08815	0.16004	0.15590 *	0.12526 *
Hispanic origin (Not Hisp. = ref.)	-0.80292	0.85920 *	-0.67020	0.47865 #	-0.37315
Male (Female = reference)	-0.71327 #	-0.07336	0.84421	-0.13652	0.23040
Region (West = reference)					
Northeast	-0.35447	-0.12679	-0.53269	-0.64144	-0.09051
Midwest	1.34757 *	-0.11537	-0.60273	-0.75050 *	0.06213
South	0.13802	-0.46539	0.01917	-0.39745	-0.11052
Urbanization level (Large urban = reference)					
Large suburban	1.02885 #	-0.62691 #	2.52607 *	0.40538	-0.04055
Medium/small metro	0.28226	-0.46544 #	2.52428 *	0.13911	-0.01763
Non-metro	-0.12413	-0.24961	0.52010	-0.64862	0.31328
County pop: % AIAN					
Log of "County pop: % AIAN"	-0.15657	0.14500 *			
County pop: % API			0.12262 #	0.00623	
County pop: % Black			-0.01708		
Square of "County pop: % Black"	0.00041				0.00014
County pop: % multiple race	0.18847	-0.11305 *	-0.43859 *	0.05669	-0.06765
Not AIAN					
Not API					
Not Black					
Constant	0.78674	-0.16288	-1.36219	-0.65960 *	0.28230

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **1999-2002**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.26722 #	0.18867	0.11601 *	-0.05574	0.18012 *
Hispanic origin (Not Hisp. = ref.)	-1.19442 #	0.31454	-0.43675 *	0.76645 *	0.37042
Male (Female = reference)	-0.95220	0.29731	0.02623	0.11405	-0.25434 #
Region (West = reference)					
Northeast	-1.03625	1.11360	-0.23762	0.23931	-0.45685
Midwest	-0.84762	1.44965	0.15054	0.13447	-0.68293 *
South	-0.91014	-1.54019	-0.01896	-0.21927	-0.29360
Urbanization level (Large urban = reference)					
Large suburban	0.80828	-19.40678 *	0.23601	-0.63138 *	0.16874
Medium/small metro	-0.60290	-0.86558	0.19960	-0.41909 #	-0.07071
Non-metro	0.09220	-0.34218	0.29444	-0.50055	-0.57691
County pop: % AIAN	0.32437	0.89609 *	-0.11790	0.08658 *	0.14523 *
Log of "County pop: % AIAN"					
County pop: % API			0.05176 *	0.02975 *	0.00210
County pop: % Black	0.05238 *	0.03292	0.00088	-0.00842	0.00299
Square of "County pop: % Black"					
County pop: % multiple race	-0.01183	-0.81130	-0.06831	-0.13980 *	0.05541
Not AIAN			-1.93501 *	---	-0.62749
Not API			-1.42404 *	-1.89170 *	---
Not Black			---	-2.05202 *	-2.10987 *
Constant	1.13111	-1.09395	3.54693 *	3.10154 *	2.10403 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2000-2003**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	0.30859 *	-0.03507	0.37300 *	0.15516 *	0.11776 #
Hispanic origin (Not Hisp. = ref.)	-0.09566	0.70135 *	0.00000	0.08454	-0.32863
Male (Female = reference)	-0.32744	0.01898	-0.03203	-0.07484	0.19596
Region (West = reference)					
Northeast	-0.51257	0.25845	-2.48653 *	-0.02198	0.17504
Midwest	0.68335	0.04770	-0.89552	-0.15343	0.35931
South	0.38515	-0.39024 #	0.50959	-0.45104	-0.01818
Urbanization level (Large urban = reference)					
Large suburban	0.60287	-0.40220	0.96101	0.33317	-0.11947
Medium/small metro	-0.54191	-0.64749 *	1.78474 #	-0.02058	-0.01671
Non-metro	-0.03717	-1.11127 *	1.80012	-0.26611	0.13583
County pop: % AIAN					
Log of "County pop: % AIAN"	-0.11480	0.71858 *			
County pop: % API			0.10799 #	0.01357	
County pop: % Black			0.00442		
Square of "County pop: % Black"	0.00016				0.00032 *
County pop: % multiple race	0.25255	-0.15480 *	-0.27582 #	0.03724	-0.03117
Not AIAN					
Not API					
Not Black					
Constant	0.28303	-0.28997	-1.33529	-0.75021 *	0.06291

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2000-2003**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.25829 #	0.61578 *	0.16108 *	-0.03413	0.16002 *
Hispanic origin (Not Hisp. = ref.)	-1.07337 #	0.83418	-0.38269 #	0.56229 *	0.21441
Male (Female = reference)	-0.71900	1.57109 #	0.02717	0.13094	-0.08847
Region (West = reference)					
Northeast	0.09577	0.78846	0.04608	-0.07431	0.25033
Midwest	-0.84452	1.79284	0.48713 #	-0.17780	0.01436
South	0.04019	-3.27884 *	0.26203	-0.71363 *	-0.37200
Urbanization level (Large urban = reference)					
Large suburban	-0.29509	-20.97397 *	0.16705	-0.46442 #	0.21546
Medium/small metro	-1.66982 *	-2.00102 #	0.16925	-0.26308	-0.23460
Non-metro	-0.46244	-0.52983	0.38727	-0.75319 *	-0.61501
County pop: % AIAN	0.30145	1.34844 *	-0.00522	0.11543 *	0.15967 *
Log of "County pop: % AIAN"					
County pop: % API			0.08675 *	0.00070	0.00814
County pop: % Black	0.01228	0.02923	0.00999	0.00918	0.00190
Square of "County pop: % Black"					
County pop: % multiple race	-0.01908	-1.33785 *	-0.11151 *	-0.05693	0.05750 #
Not AIAN			-1.70948 *	---	-0.48929
Not API			-1.16696 *	-1.22177 *	---
Not Black			---	-1.88448 *	-2.16245 *
Constant	1.85669 *	-2.38633	2.48936 *	2.18106 *	1.77641 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2001-2004**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	0.33396 *	-0.06790	0.46386 *	0.10765 *	0.08159
Hispanic origin (Not Hisp. = ref.)	-0.21356	0.82735 *	-0.99974	-0.12160	-0.58494 *
Male (Female = reference)	-0.03882	0.00215	0.23294	-0.01562	-0.01346
Region (West = reference)					
Northeast	1.88316 #	0.05747	-0.52203	0.10838	0.19858
Midwest	3.06222 *	-0.18568	2.03735	-0.11910	0.47622
South	2.86774 *	-0.22027	1.46137	-0.42150	0.22043
Urbanization level (Large urban = reference)					
Large suburban	1.75683 *	-0.34837	0.09548	-0.04693	0.00440
Medium/small metro	0.07753	-0.18720	1.44966 #	-0.52742 *	-0.17128
Non-metro	1.34592	-0.43746	-0.45886	-0.81802	-0.52300
County pop: % AIAN					
Log of "County pop: % AIAN"	1.00061 *	0.50072 *			
County pop: % API			0.07859	0.01223	
County pop: % Black			-0.02391		
Square of "County pop: % Black"	0.00023				0.00021
County pop: % multiple race	1.60940 *	-0.17085 *	-0.23213	0.06309 *	-0.10983
Not AIAN					
Not API					
Not Black					
Constant	-3.59733 *	-0.45107	-0.99364	-0.56847 *	0.40124

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2001-2004**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.46670 *	0.40354 *	0.13980 *	-0.06460 #	0.10439 *
Hispanic origin (Not Hisp. = ref.)	-1.21642 #	1.74767	-0.65242 *	0.78133 *	-0.15874
Male (Female = reference)	-0.42107	0.01830	-0.07525	0.02038	-0.01943
Region (West = reference)					
Northeast	-3.15006 #	-4.07339 *	-0.10135	-0.29652	0.28205
Midwest	-1.61953	-2.93542 #	0.46566 #	-0.30061	-0.05135
South	-2.58879 #	-4.39213 *	0.23800	-0.47544 #	-0.34290
Urbanization level (Large urban = reference)					
Large suburban	0.36652	-22.51602 *	0.34138	-0.53331 *	-0.10900
Medium/small metro	-1.36345	-3.01366 *	0.16263	-0.09931	-0.65314 *
Non-metro	0.87806	-21.38469 *	-0.07179	-0.36224	-1.23483 *
County pop: % AIAN	-0.15295	-2.83062 #	0.02243	0.07791 *	0.07728
Log of "County pop: % AIAN"					
County pop: % API			0.08256 *	-0.00559	0.00486
County pop: % Black	-0.00999	-0.04557	0.00515	0.00432	-0.00572
Square of "County pop: % Black"					
County pop: % multiple race	-0.88080 #	-1.28928 #	-0.25520 *	-0.06127	0.08279 *
Not AIAN			-1.51835 *	---	-0.91576 #
Not API			-1.87373 *	-1.53020 *	---
Not Black			---	-1.59775 *	-2.43424 *
Constant	4.49413 *	7.17041 *	3.47172 *	2.34669 *	2.82103 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2002-2005**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	0.14590	-0.07569	0.39745	0.04993	0.01772
Hispanic origin (Not Hisp. = ref.)	-0.25348	0.94881 *	-1.07272	-0.55751 #	-0.82223 *
Male (Female = reference)	-0.46965	-0.06471	-0.06220	-0.15499	-0.06856
Region (West = reference)					
Northeast	-1.22073	-0.07837	0.01951	0.20100	-0.00605
Midwest	0.17454	-0.52185 *	2.29521	-0.29184	0.38963
South	-0.42893	-0.38767 #	0.92681	-0.47644	0.16004
Urbanization level (Large urban = reference)					
Large suburban	0.36062	0.12785	-0.39423	-0.15869	-0.24386
Medium/small metro	-0.60093	0.10238	0.79620	-0.42099	-0.29249
Non-metro	-0.08635	-0.10285	-1.76059	-1.26697 *	-0.98647 *
County pop: % AIAN					
Log of "County pop: % AIAN"	0.00609	0.39109 *			
County pop: % API			0.05847	0.01080	
County pop: % Black			-0.03760		
Square of "County pop: % Black"	-0.00003				0.00037 *
County pop: % multiple race	0.00534	-0.10271 *	-0.22995 #	0.09286 *	-0.10183
Not AIAN					
Not API					
Not Black					
Constant	2.00715 *	-0.60393 #	0.24053	-0.49457 #	0.77340 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.

Complex survey sampling was taken into account.

Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White

* Coefficient differs from zero, $p < 0.05$

Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2002-2005**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.25936 *	0.19724	0.05693	-0.06661 *	0.02889
Hispanic origin (Not Hisp. = ref.)	-1.18317	1.47403	-0.71596 *	0.80276 *	-0.60394 #
Male (Female = reference)	-0.07177	-0.27958	-0.12570	-0.03998	-0.16546
Region (West = reference)					
Northeast	-1.47819	-4.13188 #	-0.30740	-0.51353 #	-0.05021
Midwest	-0.13848	-0.69369	0.23380	-0.73502 *	-0.62361
South	-0.44128	-1.77961	0.01444	-0.63351 *	-0.73615
Urbanization level (Large urban = reference)					
Large suburban	1.07016	0.64313	0.08364	-0.03795	-0.14453
Medium/small metro	-1.21719	-2.03684 #	0.02595	0.11590	-0.38922
Non-metro	21.16691 *	-0.42501	-0.55401 #	-0.12227	-1.51356 *
County pop: % AIAN	0.50887	-1.15405	0.01755	0.06531 #	-0.22686
Log of "County pop: % AIAN"					
County pop: % API			0.06596 *	-0.02422	-0.00407
County pop: % Black	-0.03669	-0.04674	0.00835	0.00980	0.00813
Square of "County pop: % Black"					
County pop: % multiple race	-0.37984	-1.13001 #	-0.22621 *	-0.04786	0.11660 *
Not AIAN			-1.76066 *	---	-1.55325 *
Not API			-2.21238 *	-1.29797 *	---
Not Black			---	-1.95444 *	-2.77299 *
Constant	2.52058	4.69185 #	4.55782 *	2.57678 *	4.20775 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2003-2006**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	0.13383	-0.08585	0.24658	0.02312	0.04007
Hispanic origin (Not Hisp. = ref.)	-0.93569	0.95876 *	1.10302	-0.79901 *	-0.88596 *
Male (Female = reference)	-0.50919	-0.03552	-0.19272	0.02025	-0.15144
Region (West = reference)					
Northeast	-1.46575 #	0.06726	-0.96001	0.60316	0.12746
Midwest	-0.91595	-0.71436 *	-1.67303	0.03807	0.23104
South	-1.09123	-0.35590 #	0.31800	-0.43472	-0.06868
Urbanization level (Large urban = reference)					
Large suburban	-0.01058	0.22159	-1.32264 #	-0.54137 #	-0.15806
Medium/small metro	-0.83097	0.20211	-0.71247	-0.61183 *	-0.46740 *
Non-metro	0.79834	0.26054	0.22292	-1.10957 *	-1.25697 *
County pop: % AIAN					
Log of "County pop: % AIAN"	-0.36390	0.30409 *			
County pop: % API			-0.04513	0.02162	
County pop: % Black			-0.01950		
Square of "County pop: % Black"	-0.00020				0.00052 *
County pop: % multiple race	0.08543	-0.06147	-0.55293	0.07350 *	-0.10309 #
Not AIAN					
Not API					
Not Black					
Constant	2.56298 *	-0.81822 *	3.17610 *	-0.56009 *	0.96791 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.

Complex survey sampling was taken into account.

Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White

* Coefficient differs from zero, $p < 0.05$

Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2003-2006**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.34501 *	0.23867	0.07261 #	-0.07817 *	0.01519
Hispanic origin (Not Hisp. = ref.)	-0.96562	1.49188	-0.81022 *	0.88744 *	-0.89545 *
Male (Female = reference)	-0.22456	-0.70380	-0.14596	-0.04912	0.02567
Region (West = reference)					
Northeast	-1.42548	-3.99519 *	-0.14728	-0.44167	0.33022
Midwest	-0.28776	-1.55906	-0.02667	-0.96955 *	-0.14389
South	-0.48417	-1.85404	-0.25214	-0.63633 *	-0.74683 *
Urbanization level (Large urban = reference)					
Large suburban	0.46049	-0.12474	-0.04542	0.06680	-0.36247
Medium/small metro	-1.04134	-1.73786	-0.27394	0.14892	-0.49980
Non-metro	-0.11806	-1.06406	-0.79640 *	0.20706	-1.05597 *
County pop: % AIAN	-0.24628	-2.69194 #	-0.00459	0.00712	-0.13675
Log of "County pop: % AIAN"					
County pop: % API			0.03983 *	-0.05087 *	0.01873
County pop: % Black	-0.00577	-0.04659	0.01568 *	0.01613	0.01436
Square of "County pop: % Black"					
County pop: % multiple race	-0.15571	-0.56975	-0.18654 *	0.05757	0.07997 *
Not AIAN			-1.77492 *	---	-1.27031 *
Not API			-2.69892 *	-1.21798 #	---
Not Black			---	-2.07446 *	-3.07415 *
Constant	2.08818 #	5.01904 #	5.36292 *	2.51673 *	3.82268 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2004-2007**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	0.01619	-0.10744	0.11992	-0.07140	0.05127
Hispanic origin (Not Hisp. = ref.)	-0.96979	0.86983 *	2.40755 *	-0.40577	-0.83621 *
Male (Female = reference)	-0.80774 *	-0.06116	0.22117	0.04178	-0.20118
Region (West = reference)					
Northeast	-2.02593 *	0.21728	-0.15241	0.28045	0.05817
Midwest	-1.40054	-0.87122 *	-2.13813	-0.17116	0.11882
South	-0.98742	-0.30433	-0.45575	-0.28714	-0.12033
Urbanization level (Large urban = reference)					
Large suburban	0.21578	0.31372	-0.56014	-0.81248 *	-0.15572
Medium/small metro	0.04803	0.20093	-1.85077 *	-0.51288 #	-0.63061 *
Non-metro	-0.79758	0.40184		-0.92292 #	-1.30867 *
County pop: % AIAN					
Log of "County pop: % AIAN"	-0.40416	0.19177 *			
County pop: % API			-0.12870 *	0.02429	
County pop: % Black			-0.03087		
Square of "County pop: % Black"	-0.00006				0.00029
County pop: % multiple race	-0.20559	-0.01644	-0.18721	0.08169 *	-0.02786
Not AIAN					
Not API					
Not Black					
Constant	3.56303 *	-0.90919 *	3.88208 *	-0.50864 #	1.02331 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2004-2007**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.09841	-0.17454	0.02961	-0.10114 *	-0.05881
Hispanic origin (Not Hisp. = ref.)	-1.23448	1.06515	-0.74639 *	0.84893 *	-0.51931 #
Male (Female = reference)	-0.32693	-1.06349	-0.23894 *	-0.03118	0.02143
Region (West = reference)					
Northeast	-0.25772	-4.61065 *	-0.21944	-0.18009	-0.23161
Midwest	0.80277	-2.79803	-0.22915	-1.02381 *	-0.56862
South	0.32010	-2.46372	-0.47990 #	-0.61794 *	-0.70101 #
Urbanization level (Large urban = reference)					
Large suburban	0.21442	0.28587	0.03688	0.04694	-0.69156 *
Medium/small metro	0.60992	0.52268	-0.35628	-0.13315	-0.31502
Non-metro	0.88309	0.21090	-0.93646 *	0.05513	-0.69552
County pop: % AIAN	-0.63626 #	-2.79444 *	0.00127	0.01011	-0.29887
Log of "County pop: % AIAN"					
County pop: % API			0.00528	-0.05952 *	0.01912
County pop: % Black	0.00283	-0.04418	0.01710 *	0.01004	0.01989 *
Square of "County pop: % Black"					
County pop: % multiple race	0.36037	-0.63497	-0.07737	0.07314	0.08167 *
Not AIAN			-2.04924 *	---	-0.87944 *
Not API			-2.86955 *	-1.62132 *	---
Not Black			---	-2.46789 *	-3.01469 *
Constant	0.71185	6.77001 *	6.00634 *	3.49088 *	3.62296 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, p < 0.05
# Coefficient differs from zero, p < 0.10
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2005-2008**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	0.02183	-0.06345	0.02923	-0.00257	0.06006
Hispanic origin (Not Hisp. = ref.)	-0.84037	0.99690 *	0.63832	-0.43064	-0.66429 *
Male (Female = reference)	-0.89486 *	-0.04288	-0.08814	-0.12827	-0.27808 #
Region (West = reference)					
Northeast	-1.35510 #	0.21529	-1.47292	-0.24345	0.23606
Midwest	-0.66450	-0.57230 #	-3.35155 *	-0.30485	-0.21498
South	-0.25230	-0.08608	-2.28123 #	-0.45956 #	-0.47694
Urbanization level (Large urban = reference)					
Large suburban	0.14590	0.39831	-1.28570	-0.32963	-0.16352
Medium/small metro	-0.20383	0.04416	-1.65835 *	-0.19908	-0.41663 #
Non-metro	-1.28048	0.14543	-4.10044 *	-0.43006	-0.84756 *
County pop: % AIAN					
Log of "County pop: % AIAN"	-0.52774 *	0.21944 #			
County pop: % API			-0.10545 *	0.02863 *	
County pop: % Black			0.00474		
Square of "County pop: % Black"	-0.00011				0.00043
County pop: % multiple race	-0.17226 *	0.03421	-0.41867	0.04220	0.05810
Not AIAN					
Not API					
Not Black					
Constant	2.81981 *	-1.22636 *	5.11619 *	-0.50229 #	0.84296 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.

Complex survey sampling was taken into account.

Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White

* Coefficient differs from zero, $p < 0.05$

Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2005-2008**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.14391	-0.16485	0.04077	-0.06447 *	0.00835
Hispanic origin (Not Hisp. = ref.)	0.38713	0.67276	-0.45826	0.86251 *	-0.47008 #
Male (Female = reference)	-0.12302	-0.33832	-0.32223 *	0.00330	-0.10040
Region (West = reference)					
Northeast	-0.66638	-2.27639	-0.25556	-0.21559	-0.45458
Midwest	0.84510	-0.14895	-0.65042 *	-0.98910 *	-0.42499
South	1.40638	1.37471	-0.88389 *	-0.53108 *	-0.53279
Urbanization level (Large urban = reference)					
Large suburban	-2.19480	-0.30334	-0.07013	0.31007	-0.21197
Medium/small metro	-0.61312	0.30588	-0.26898	-0.04836	-0.05074
Non-metro	-2.43819 #	-0.99151	-0.78522 *	0.09731	0.03577
County pop: % AIAN	-0.52164 *	-0.47393	-0.02277	0.01642	-0.17263
Log of "County pop: % AIAN"					
County pop: % API			-0.00020	-0.04112 #	0.03221 *
County pop: % Black	0.02638	-0.01439	0.02863 *	0.02003 *	0.01482
Square of "County pop: % Black"					
County pop: % multiple race	-0.43823	-0.48017	-0.05068	0.07585	0.03318
Not AIAN			-2.16596 *	---	-0.74398 *
Not API			-2.90642 *	-0.96876	---
Not Black			---	-2.66517 *	-3.44127 *
Constant	2.93979 *	2.63278	6.20677 *	2.61931 *	3.61436 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2006-2009**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	-0.00690	-0.06562	-0.03749	-0.03044	0.09450 #
Hispanic origin (Not Hisp. = ref.)	-0.30246	0.80984 *	1.13348	-0.19510	-0.35401
Male (Female = reference)	-0.54906 #	-0.00581	0.95591	-0.11406	-0.20769
Region (West = reference)					
Northeast	-0.52008	0.38968	-1.85052	-0.35740	0.12622
Midwest	-0.70844	-0.53645 #	-3.21631 *	-0.37944	0.09554
South	-0.04570	0.04087	-2.41495 #	-0.73705 *	-0.24601
Urbanization level (Large urban = reference)					
Large suburban	0.48213	0.00100	-1.29132	-0.07972	-0.23576
Medium/small metro	0.27084	-0.09581	-1.96632 #	-0.08502	-0.35760
Non-metro	-0.61870	0.15766	-6.35357 *	0.01468	-0.69133 *
County pop: % AIAN					
Log of "County pop: % AIAN"	-0.34039	0.14006			
County pop: % API			-0.13096 *	0.01399	
County pop: % Black			0.05981		
Square of "County pop: % Black"	0.00001				0.00041
County pop: % multiple race	-0.06111	0.03213	-0.07427	0.05668 #	0.15362
Not AIAN					
Not API					
Not Black					
Constant	2.28712 *	-1.20710 *	3.74565 *	-0.45005 #	0.27578

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2006-2009**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.39692 *	0.16827	0.07816	-0.05914 #	-0.01689
Hispanic origin (Not Hisp. = ref.)	0.35795	1.78251	-0.30566	0.77682 *	-0.15941
Male (Female = reference)	-0.30865	-0.16947	-0.24337 *	0.02338	-0.11355
Region (West = reference)					
Northeast	-0.74832	-3.33363 #	-0.13958	-0.10093	-0.46719
Midwest	1.22640	-0.33080	-0.28501	-0.83071 *	-0.43245
South	-0.73790	-0.87822	-0.65101 *	-0.40783	-0.70407 *
Urbanization level (Large urban = reference)					
Large suburban	0.50495	0.12949	0.00027	0.00804	-0.07548
Medium/small metro	1.19029	0.26436	-0.08252	-0.21413	0.02353
Non-metro	-1.63578	-1.16549	-0.47924	0.14000	0.48931
County pop: % AIAN	-0.56692	-1.75064	-0.00652	0.01278	-0.09946
Log of "County pop: % AIAN"					
County pop: % API			0.00185	-0.03671	0.01731
County pop: % Black	0.06576 *	0.00190	0.03171 *	0.02128 *	0.00652
Square of "County pop: % Black"					
County pop: % multiple race	0.15548	0.20240	0.06197	0.04851	0.05068
Not AIAN			-1.66936 *	---	-0.32729
Not API			-2.68008 *	-1.72073 *	---
Not Black			---	-2.16346 *	-3.03979 *
Constant	-0.98489	1.17871	4.68836 *	2.85368 *	2.87842 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2007-2010**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	0.01905	-0.09533	0.08426	-0.02085	0.05226
Hispanic origin (Not Hisp. = ref.)	-1.41727	0.76959 *	-0.95356	0.11182	-0.11828
Male (Female = reference)	-0.13644	0.10448	0.29794	-0.22023 #	-0.14355
Region (West = reference)					
Northeast	-0.70715	0.26440	-2.22445	-0.18519	-0.15512
Midwest	-0.15704	-0.07585	-1.84793	-0.37913	0.13674
South	0.27913	0.33107	-2.79329 *	-0.41638 #	-0.16913
Urbanization level (Large urban = reference)					
Large suburban	0.18643	-0.08125	-0.16808	0.10235	-0.15928
Medium/small metro	0.70298	-0.30280	-2.07782 *	-0.06243	-0.04923
Non-metro	-0.83013	0.00109	-3.43135 *	-0.42528	-0.54935 #
County pop: % AIAN					
Log of "County pop: % AIAN"	-0.08308	0.14866			
County pop: % API			-0.11690 *	-0.00741	
County pop: % Black			0.03441		
Square of "County pop: % Black"	0.00016				0.00015
County pop: % multiple race	-0.10910	0.03231	0.10439	0.10635 *	0.13803
Not AIAN					
Not API					
Not Black					
Constant	1.63654 #	-1.18203 *	3.45439 *	-0.58607 *	0.29471

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2007-2010**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.26290 *	0.11336	0.06295	-0.08545 *	-0.01538
Hispanic origin (Not Hisp. = ref.)	-1.06074	0.77524	-0.27330	0.78566 *	0.08730
Male (Female = reference)	-0.59040	-0.30129	-0.16463	0.09234	-0.20766 #
Region (West = reference)					
Northeast	0.75474	-0.27502	-0.36867	-0.13622	-0.26043
Midwest	1.54254 #	0.88953	-0.12894	-0.40275	-0.42481
South	0.60941	1.23706	-0.47167 #	-0.11817	-0.39221
Urbanization level (Large urban = reference)					
Large suburban	0.43467	-0.20540	-0.05712	0.01111	0.11843
Medium/small metro	0.83231	-0.16641	0.06442	-0.22589	0.06520
Non-metro	0.29739	1.05806	-0.49810	0.07689	-0.07461
County pop: % AIAN	-0.20871	-0.67428	0.00081	0.02669 *	-0.08795
Log of "County pop: % AIAN"					
County pop: % API			-0.01703	-0.02337	-0.00055
County pop: % Black	0.03757	0.01069	0.02103 *	0.01836 *	0.00432
Square of "County pop: % Black"					
County pop: % multiple race	0.48326	0.59202	0.07958 *	0.01982	0.08888 *
Not AIAN			-1.30287 *	---	0.25958
Not API			-2.10494 *	-1.13226 *	---
Not Black			---	-1.83196 *	-2.44549 *
Constant	-2.12193	-2.08936	3.84231 *	1.84255 *	1.59128 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2008-2011**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	0.01517	-0.11434	0.14870	0.00962	0.06289
Hispanic origin (Not Hisp. = ref.)	-1.59822 #	0.86173 *	0.55509	-0.14407	-0.26551
Male (Female = reference)	-0.17242	0.07591	0.36442	-0.15227	-0.13084
Region (West = reference)					
Northeast	-0.55929	0.54311	0.04230	-0.07024	0.11226
Midwest	0.13794	0.16130	-1.26838 #	-0.31942	0.18525
South	0.59324	0.35311	-0.61256	-0.35433	-0.03293
Urbanization level (Large urban = reference)					
Large suburban	-0.54710	-0.45670	-0.49583	0.09359	-0.31769
Medium/small metro	-0.01371	-0.39962	-0.95309	0.02492	-0.05082
Non-metro	-0.72915	-0.16783	-1.41581	-0.75073 #	-0.82646 *
County pop: % AIAN					
Log of "County pop: % AIAN"	-0.04767	0.06579			
County pop: % API			-0.01367	-0.00173	
County pop: % Black			-0.00270		
Square of "County pop: % Black"	-0.00024				0.00016
County pop: % multiple race	0.00939	0.03681 #	-0.03227	0.10328 *	0.09923
Not AIAN					
Not API					
Not Black					
Constant	1.88675 #	-1.07611 *	1.89003 *	-0.75548 *	0.21172

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2008-2011**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.21491 #	0.23920	0.07271 *	-0.09490 *	0.00543
Hispanic origin (Not Hisp. = ref.)	-1.19739	0.30745	-0.40212 *	0.83075 *	-0.26477
Male (Female = reference)	-0.54447 #	0.16786	-0.13411	0.07561	-0.18119 #
Region (West = reference)					
Northeast	0.44283	-0.02071	-0.08051	0.25724	-0.12755
Midwest	1.39670 #	1.74076	-0.08195	-0.08309	-0.21940
South	-0.01584	0.96148	-0.30911	-0.03449	-0.24077
Urbanization level (Large urban = reference)					
Large suburban	0.90355	-0.45248	-0.21715	-0.10489	0.11766
Medium/small metro	0.98829	-0.33334	0.02249	0.01238	0.11174
Non-metro	0.36473	-0.76737	-0.70104 *	0.17168	-0.48405
County pop: % AIAN	0.00647	0.04639	0.00011	0.03873 *	-0.01645
Log of "County pop: % AIAN"					
County pop: % API			-0.01816	0.02331	0.00405
County pop: % Black	0.06875 *	0.02996	0.01833 *	0.02605 *	-0.00061
Square of "County pop: % Black"					
County pop: % multiple race	0.55436	0.67483 #	0.08770 *	-0.05719	0.09402 *
Not AIAN			-1.37599 *	---	0.11594
Not API			-2.14881 *	-1.46012 *	---
Not Black			---	-1.60521 *	-2.29342 *
Constant	-2.63888 *	-3.65060 #	3.80668 *	1.68438 *	1.37150 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2009-2012**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	-0.01121	-0.10480	0.16357	-0.03223	0.07134
Hispanic origin (Not Hisp. = ref.)	-1.33156 #	0.54924 *	1.05545	-0.08172	-0.40842 *
Male (Female = reference)	0.03592	0.12257	0.67419 *	-0.03551	0.05627
Region (West = reference)					
Northeast	-1.54750 *	0.55338	-0.55326	0.01057	0.09396
Midwest	-0.90940	0.30701	-0.31391	-0.18245	0.31651
South	-0.34354	0.28609	-0.29740	-0.18461	-0.06533
Urbanization level (Large urban = reference)					
Large suburban	-1.45251 *	-0.63935 *	-0.04285	-0.24170	-0.37383
Medium/small metro	-0.55886	-0.43437 #	0.30675	0.25702	-0.00350
Non-metro	-0.46065	-0.19060	0.56832	-0.29599	-0.83697 *
County pop: % AIAN					
Log of "County pop: % AIAN"	-0.26337	0.04635			
County pop: % API			-0.04300	0.00784	
County pop: % Black			0.01046		
Square of "County pop: % Black"	-0.00023				0.00016
County pop: % multiple race	0.34558	0.02293	-0.01538	0.09258 *	0.07936
Not AIAN					
Not API					
Not Black					
Constant	2.37639 *	-1.02549 *	1.00670 #	-0.91896 *	0.14483

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2009-2012**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.18715	0.26764 *	0.08344 *	-0.07456 *	-0.01649
Hispanic origin (Not Hisp. = ref.)	-1.63895 *	-0.56263	-0.53841 *	0.51538 *	-0.27259
Male (Female = reference)	-0.87290 *	-0.22754	0.01845	0.11679	-0.10129
Region (West = reference)					
Northeast	0.42740	-0.70283	-0.02671	0.42225	0.10158
Midwest	0.18042	-0.43037	0.11174	0.10979	-0.05716
South	-0.84415	-1.84067	-0.22384	-0.08725	-0.05597
Urbanization level (Large urban = reference)					
Large suburban	0.36065	-2.10937 *	-0.40317 *	-0.34814	-0.24340
Medium/small metro	0.38540	-0.20528	-0.07535	-0.15491	0.19489
Non-metro	1.09057	0.58059	-0.72330 *	-0.07194	-0.44053
County pop: % AIAN	-0.16109	-0.02880	0.01007	0.02827 *	0.02821
Log of "County pop: % AIAN"					
County pop: % API			-0.01812	-0.00815	0.01694
County pop: % Black	0.06020 *	0.05662 #	0.01327 *	0.01537 #	-0.00542
Square of "County pop: % Black"					
County pop: % multiple race	0.58472 *	0.69366 *	0.07132 *	-0.01453	0.07905 *
Not AIAN			-1.35723 *	---	-0.07774
Not API			-1.69621 *	-1.46766 *	---
Not Black			---	-1.53706 *	-1.71372 *
Constant	-1.43872	-2.47917	3.34326 *	1.80718 *	0.83802

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2010-2013**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	0.12527	-0.09571	0.14716	-0.03125	0.10077 *
Hispanic origin (Not Hisp. = ref.)	-0.99058	0.32137	1.52789	-0.05520	-0.56496 *
Male (Female = reference)	0.21085	0.15253	0.63411 #	-0.06537	-0.01671
Region (West = reference)					
Northeast	-2.12806 *	0.65737 #	-1.21194	-0.27754	-0.12257
Midwest	-0.82630	0.36395	-0.96851	-0.13302	0.16076
South	0.22024	0.17294	-1.05931	-0.18099	-0.17805
Urbanization level (Large urban = reference)					
Large suburban	-1.39081 *	-0.71524 *	0.04440	-0.22547	-0.22908
Medium/small metro	-1.00282 #	-0.34623 #	0.48595	0.27334	0.03494
Non-metro	-1.68023 *	-0.14233	0.91454	-0.04510	-0.65411 *
County pop: % AIAN					
Log of "County pop: % AIAN"	-0.24591	0.19876 *			
County pop: % API			-0.03561	0.01083	
County pop: % Black			0.04681 #		
Square of "County pop: % Black"	-0.00052				0.00025 *
County pop: % multiple race	-0.01872	-0.00967	-0.02090	0.07246 *	0.02466
Not AIAN					
Not API					
Not Black					
Constant	2.81620 *	-0.95191 *	0.59752	-0.89067 *	0.31550

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2010-2013**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.21736 *	0.09869	0.11268 *	-0.08494 *	0.00073
Hispanic origin (Not Hisp. = ref.)	-1.55594 *	-1.45360 #	-0.63633 *	0.29092	-0.37736
Male (Female = reference)	-0.33984	0.18151	0.01026	0.12248	-0.14256
Region (West = reference)					
Northeast	-0.33303	-0.79967	-0.32568	0.50983 #	-0.14574
Midwest	-0.19806	-0.48881	-0.04162	0.09157	0.00319
South	-0.57217	-2.32654 *	-0.26986	-0.19442	-0.08722
Urbanization level (Large urban = reference)					
Large suburban	-1.38215 *	-2.91622 *	-0.41654 *	-0.45298 #	-0.17033
Medium/small metro	-1.66839 *	-0.00906	-0.19240	-0.01997	0.23028
Non-metro	-1.44337	0.98470	-0.75563 *	0.04298	-0.17787
County pop: % AIAN	0.05334	-0.05041	-0.03509	0.04328 *	0.04478
Log of "County pop: % AIAN"					
County pop: % API			-0.01483	-0.00852	0.01978
County pop: % Black	0.03532	0.10239 *	0.01250 *	0.01310	-0.00425
Square of "County pop: % Black"					
County pop: % multiple race	0.18909	0.83588 #	0.04349	-0.01371	0.05765 *
Not AIAN			-1.36654 *	---	-0.02844
Not API			-1.70417 *	-0.91162 *	---
Not Black			---	-1.58286 *	-1.92274 *
Constant	1.29036	-3.35519	3.60802 *	1.34579 *	0.95488

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2011-2014**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	0.04274	-0.05300	0.03616	-0.02440	0.11821 *
Hispanic origin (Not Hisp. = ref.)	-0.61229	0.87611 *	1.56824	0.02265	-0.64976 *
Male (Female = reference)	0.03439	0.08882	0.83217 *	-0.11498	0.04368
Region (West = reference)					
Northeast	-1.15209 #	0.64707 #	-0.84393	-0.28671	-0.30333
Midwest	-0.44255	0.25573	-1.09867	-0.12312	-0.07161
South	0.63171	0.00359	-1.44338 *	-0.20282	-0.32533
Urbanization level (Large urban = reference)					
Large suburban	-1.01952 *	-0.57534 *	0.53344	-0.28513	-0.38313 #
Medium/small metro	-1.02546 *	-0.25157	1.15818 *	0.44712 *	-0.16563
Non-metro	-1.02101	-0.01508	1.99108	0.04080	-0.69265 *
County pop: % AIAN					
Log of "County pop: % AIAN"	-0.18334	0.19264 *			
County pop: % API			-0.02357	0.02005 #	
County pop: % Black			0.05219 *		
Square of "County pop: % Black"	-0.00060 #				0.00025 *
County pop: % multiple race	0.03283	-0.02754	0.06487	0.05157 *	-0.04269
Not AIAN					
Not API					
Not Black					
Constant	2.57181 *	-1.07960 *	-0.16481	-0.94360 *	0.65464 #

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.

Complex survey sampling was taken into account.

Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White

* Coefficient differs from zero, $p < 0.05$

Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2011-2014**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.26982 *	0.19529 #	0.11356 *	-0.04033	0.02227
Hispanic origin (Not Hisp. = ref.)	-0.98362 *	-0.76407	-0.56826 *	0.71711 *	-0.17316
Male (Female = reference)	-0.15313	0.37096	0.03777	0.10643	-0.17349 #
Region (West = reference)					
Northeast	-1.91414 *	-2.59266 *	-0.32441	0.41326	-0.10334
Midwest	-1.23438	-1.31279	-0.16240	0.05844	0.01138
South	-1.16646	-2.15095 #	-0.32746	-0.24465	0.04205
Urbanization level (Large urban = reference)					
Large suburban	-1.12956 #	-1.39096	-0.48718 *	-0.41449 #	-0.37243 #
Medium/small metro	-1.39409 *	-0.40810	-0.32062 *	-0.03577	0.23552
Non-metro	-2.07018 *	0.33011	-0.69074 *	0.11925	-0.24392
County pop: % AIAN	0.16664	-0.01138	0.00571	0.03331 *	0.02961
Log of "County pop: % AIAN"					
County pop: % API			-0.00748	0.01086	0.02328 *
County pop: % Black	0.03973 #	0.04570	0.01074 #	0.00614	-0.01328
Square of "County pop: % Black"					
County pop: % multiple race	-0.10765	0.36310	-0.00129	-0.03903	0.04530 *
Not AIAN			-1.49470 *	---	-0.18430
Not API			-1.47904 *	-0.80019 *	---
Not Black			---	-1.61281 *	-1.99152 *
Constant	2.09005 #	-1.05989	3.60898 *	1.19734 *	1.27377 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2012-2015**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	0.02033	-0.01229	-0.03629	0.02779	0.09070 *
Hispanic origin (Not Hisp. = ref.)	-1.16466 #	1.03449 *	0.54437	-0.08166	-0.71893 *
Male (Female = reference)	-0.02449	-0.00616	0.53498	-0.15523	0.13053
Region (West = reference)					
Northeast	-1.68480 *	0.59677 #	-0.36523	-0.25235	-0.63834 *
Midwest	-0.94036	0.42542 #	0.09172	-0.06557	-0.09004
South	0.03180	0.17675	-1.19216 #	-0.15731	-0.40355 #
Urbanization level (Large urban = reference)					
Large suburban	-0.78924 #	-0.17883	0.10536	-0.24215	-0.04390
Medium/small metro	-0.64879	-0.27293	0.43155	0.38541 *	-0.05014
Non-metro	-0.01104	0.05025	0.16821	0.46185	-0.65583 *
County pop: % AIAN					
Log of "County pop: % AIAN"	-0.59687 *	0.24836 *			
County pop: % API			-0.01977	0.01011	
County pop: % Black			0.05096 *		
Square of "County pop: % Black"	-0.00043				0.00032 *
County pop: % multiple race	0.17340	-0.02792	0.01628	0.04641 *	-0.06916
Not AIAN					
Not API					
Not Black					
Constant	2.24078 *	-1.33144	0.31537	-0.94013 *	0.72730 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.

Complex survey sampling was taken into account.

Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White

* Coefficient differs from zero, $p < 0.05$

Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2012-2015**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.30872 *	0.17673	0.11474 *	-0.00626	0.05848 #
Hispanic origin (Not Hisp. = ref.)	-1.00205 *	-0.77430	-0.64740 *	0.85675 *	-0.19228
Male (Female = reference)	-0.03599	0.10383	0.08842	0.03049	-0.14861
Region (West = reference)					
Northeast	-0.25370	-3.03518 *	-0.48497 *	0.23021	-0.08502
Midwest	-0.24889	-2.82678 *	-0.04571	0.08113	-0.04887
South	0.16666	-2.41034	-0.23694	-0.05982	0.00963
Urbanization level (Large urban = reference)					
Large suburban	-0.74895	0.01522	-0.13910	-0.18220	-0.25692
Medium/small metro	-1.18693 *	0.83138	-0.19440	-0.23080	0.24253
Non-metro	-0.77484	2.21791	-0.67094 *	-0.10216	0.12802
County pop: % AIAN	0.03975	-0.55803 #	0.01902	0.03401 *	0.02363
Log of "County pop: % AIAN"					
County pop: % API			0.00211	-0.00672	0.01589
County pop: % Black	0.00758	0.04401	0.01507 *	-0.00371	-0.01091
Square of "County pop: % Black"					
County pop: % multiple race	0.11376	0.27598	-0.01554	-0.02298	0.04138 *
Not AIAN			-1.19220 *	---	-0.07281
Not API			-0.77863 #	0.01834	---
Not Black			---	-1.42847 *	-1.46966 *
Constant	0.60627	-0.73124	2.40185 *	0.29921	0.59931

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2013-2016**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	0.06066	-0.01348	0.01675	0.05466	0.07968 *
Hispanic origin (Not Hisp. = ref.)	-1.19298 *	1.19960 *	1.20065	0.05994	-0.87515 *
Male (Female = reference)	-0.01982	-0.14881	0.19828	-0.19756 #	0.04355
Region (West = reference)					
Northeast	-0.97972	0.21376	-0.84279	-0.19902	-0.54392 *
Midwest	-0.24297	0.19922	-0.74754	-0.05199	-0.13845
South	0.97033 #	0.04899	-0.94824	-0.29405	-0.23808
Urbanization level (Large urban = reference)					
Large suburban	0.45350	0.01924	-0.28331	-0.19073	0.03471
Medium/small metro	-0.19723	-0.13139	-0.18399	0.30366 #	-0.20879
Non-metro	0.46243	0.18837	1.37374	0.03873	-0.66415 *
County pop: % AIAN					
Log of "County pop: % AIAN"	-0.60912 *	0.24936 *			
County pop: % API			-0.00377	0.00983	
County pop: % Black			0.04819 *		
Square of "County pop: % Black"	-0.00034				0.00030 *
County pop: % multiple race	0.23809	-0.02785	-0.10140	0.06536 *	-0.03300
Not AIAN					
Not API					
Not Black					
Constant	0.73560	-1.31067 *	0.85603	-1.03418 *	0.66889 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2013-2016**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.26904 *	0.13620	0.11350 *	-0.01111	0.08547
Hispanic origin (Not Hisp. = ref.)	-1.09579 *	0.34808	-0.69732 *	1.02129 *	-0.24125
Male (Female = reference)	-0.31641	0.02744	-0.00633	-0.10282	-0.16107
Region (West = reference)					
Northeast	0.24010	-1.90047 #	-0.47580 *	-0.14453	-0.18102
Midwest	0.49802	-1.07624	-0.10992	-0.07736	-0.16122
South	0.48730	-0.31742	-0.17215	-0.18946	-0.17500
Urbanization level (Large urban = reference)					
Large suburban	-0.82156	0.42711	0.10566	-0.07219	-0.18967
Medium/small metro	-0.93102 #	0.93874	-0.23413	-0.14813	0.18092
Non-metro	0.23675	2.41760 *	-0.50336 #	0.05912	-0.24829
County pop: % AIAN	0.36963 #	-0.15162	0.03146	0.04982 *	-0.08675 #
Log of "County pop: % AIAN"					
County pop: % API			0.00074	0.00998	0.00615
County pop: % Black	0.04195 *	0.03678	0.02155 *	0.00186	-0.00525
Square of "County pop: % Black"					
County pop: % multiple race	-0.13070	0.28455	0.00908	-0.05791	0.07488 *
Not AIAN			-1.00037 *	---	-0.13094
Not API			-1.09541 *	-0.16719	---
Not Black			---	-1.26077 *	-1.67309 *
Constant	-0.21061	-2.68983	2.31924 *	0.31477	0.91657

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, p < 0.05
# Coefficient differs from zero, p < 0.10
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2014-2017**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	0.07384	-0.01953	-0.00359	0.07058 #	0.02482
Hispanic origin (Not Hisp. = ref.)	-1.11659 *	1.20131 *	0.54562	0.11462	-0.85413 *
Male (Female = reference)	-0.04330	-0.12651	0.18523	-0.12775	-0.04144
Region (West = reference)					
Northeast	-0.65190	0.18386	-0.89444	0.13578	-0.48437 #
Midwest	-0.20506	0.39073 #	-0.89733	-0.05626	-0.30764
South	0.76910	0.05039	-1.03128	-0.03170	-0.33252
Urbanization level (Large urban = reference)					
Large suburban	0.44112	0.37493	0.11829	-0.13642	-0.01130
Medium/small metro	0.23449	-0.04074	0.54694	0.24071	-0.35312
Non-metro	-0.10872	0.20286	2.72979 *	0.14919	-0.77635 *
County pop: % AIAN					
Log of "County pop: % AIAN"	-0.47633 *	0.34143 *			
County pop: % API			0.02542	0.01864 #	
County pop: % Black			0.03266 #		
Square of "County pop: % Black"	-0.00037				0.00026 *
County pop: % multiple race	0.15915	-0.06669	-0.17475 #	0.06294 *	-0.02286
Not AIAN					
Not API					
Not Black					
Constant	0.83828	-1.32224 *	1.04172	-1.30837 *	0.97621 *

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.

Complex survey sampling was taken into account.

Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White

* Coefficient differs from zero, $p < 0.05$

Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2014-2017**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.23197 *	0.10720	0.08031 *	-0.01186	0.10007 *
Hispanic origin (Not Hisp. = ref.)	-1.52448 *	1.89458 *	-0.77615 *	1.04621 *	-0.05422
Male (Female = reference)	-0.29210	-0.48334	-0.06338	-0.12725	-0.08123
Region (West = reference)					
Northeast	1.22574	-2.51566 *	-0.36686	-0.35085	0.09008
Midwest	0.74194	-1.51515	-0.25605	-0.01094	-0.20228
South	0.74429	0.53205	-0.29982	-0.17565	0.08745
Urbanization level (Large urban = reference)					
Large suburban	-0.38495	0.27007	0.17406	0.22117	-0.22917
Medium/small metro	-0.48210	0.63954	-0.20373	-0.04944	0.00294
Non-metro	1.05955	1.10620	-0.52005 *	-0.01378	-0.38022
County pop: % AIAN	0.51018 *	0.35799	0.04010	0.07757 *	-0.09517 #
Log of "County pop: % AIAN"					
County pop: % API			0.00042	0.02080	0.01000
County pop: % Black	0.06204 *	0.01236	0.02217 *	-0.00248	-0.00377
Square of "County pop: % Black"					
County pop: % multiple race	0.00795	-0.00010	0.01288	-0.14111 *	0.08320 *
Not AIAN			-0.98623 *	---	-0.35944
Not API			-0.93487 *	-0.15696	---
Not Black			---	-1.36581 *	-1.49414 *
Constant	-1.61065	-1.89779	2.29413 *	0.52004	0.76195

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$
--- Coefficient constrained to zero.

Model coefficients from logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2015-2018**

Multiple-race category:	2: AIAN-B	3: AIAN-W	4: API-B	5: API-W	6: B-W
predicted race:	BLACK	AIAN	BLACK	API	BLACK
Covariates:					
Age (in yrs.) per 10 yrs.	0.15869 *	-0.03034	-0.03131	0.09842 *	0.03405
Hispanic origin (Not Hisp. = ref.)	-0.98046 #	0.82237 *	-0.30467	0.22550	-1.05309 *
Male (Female = reference)	-0.41707	-0.12088	-0.04742	-0.06579	0.00602
Region (West = reference)					
Northeast	-1.40559 *	0.13859	-1.04963	-0.11426	0.08443
Midwest	-0.22309	0.35116 #	-1.97425 *	-0.04656	-0.02579
South	0.15088	0.06911	-1.04845	-0.09180	-0.03475
Urbanization level (Large urban = reference)					
Large suburban	0.54055	0.24528	-0.52948	0.03439	0.05658
Medium/small metro	0.53859	-0.00940	-0.06014	0.10520	-0.27083
Non-metro	0.21595	0.20301	2.56108	-0.13413	-0.59071 *
County pop: % AIAN					
Log of "County pop: % AIAN"	-0.47102 *	0.31072 *			
County pop: % API			0.03515	0.02110 *	
County pop: % Black			0.03585 #		
Square of "County pop: % Black"	-0.00033				0.00030 *
County pop: % multiple race	0.10412	-0.03898	-0.22078 #	0.05867 *	0.06349
Not AIAN					
Not API					
Not Black					
Constant	1.03453	-1.25388 *	1.98294 *	-1.34380 *	0.41842

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.

Complex survey sampling was taken into account.

Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White

* Coefficient differs from zero, $p < 0.05$

Coefficient differs from zero, $p < 0.10$

Model coefficients from multinomial logistic regression models predicting best race among those who reported one, NHIS

Four years pooled NHIS: **2015-2018**

Multiple-race category:	9: AIAN-B-W		ALL GROUPS		
predicted race:	BLACK	AIAN	BLACK	AIAN	API
Covariates:					
Age (in yrs.) per 10 yrs.	0.18655 *	0.24698 *	0.07622 *	-0.02262	0.11248 *
Hispanic origin (Not Hisp. = ref.)	-0.99487	0.81497	-0.93093 *	0.68721 *	0.15597
Male (Female = reference)	-0.61207 #	-0.60293	-0.04434	-0.09680	-0.00940
Region (West = reference)					
Northeast	0.67811	-1.14693	-0.22168	-0.05473	0.01907
Midwest	0.59824	0.14072	-0.24393	0.04148	-0.00591
South	-0.25698	0.42663	-0.28420	-0.05825	0.16770
Urbanization level (Large urban = reference)					
Large suburban	0.08142	0.02518	0.21613	0.05045	-0.00068
Medium/small metro	-0.19950	0.16103	-0.07686	-0.15214	-0.00925
Non-metro	1.35702	-0.22324	-0.36636	-0.09250	-0.65802 #
County pop: % AIAN	0.62546 *	0.60318 *	0.04347 #	0.06608 *	-0.03022
Log of "County pop: % AIAN"					
County pop: % API			0.01524	0.00615	0.01593
County pop: % Black	0.10675 *	0.06266 #	0.02242 *	-0.00168	-0.00677
Square of "County pop: % Black"					
County pop: % multiple race	-0.00564	-0.20574	-0.00499	-0.07241	0.08145 *
Not AIAN			-1.03543 *	---	-0.71083 *
Not API			-1.09311 *	0.11923	---
Not Black			---	-1.52658 *	-1.51089 *
Constant	-1.56029	-2.06973	2.44035 *	0.40628	0.88838

Samples include people who (a) reported the specific listed multiple races, and (b) reported a single best race response. All ages.
Complex survey sampling was taken into account.
Abbreviations: AIAN = American Indian / Alaska Native; API = Asian / Pacific Islander; B = Black /African American; W = White
* Coefficient differs from zero, $p < 0.05$
# Coefficient differs from zero, $p < 0.10$
--- Coefficient constrained to zero.