RACHSING and PRED* variables

What they are, why they are needed, and how to use them.

Carolyn Liebler, Ph.D., Associate Professor of Sociology and MPC Faculty Member, UMN
Megan Schouweiler, IPUMS USA

Zoom Logistics

- Webinar is being recorded and will be posted
- Real-time closed captions are being auto-generated
  - Turn off/on by clicking “CC” button in Zoom controls
- Send questions about Zoom directly to host (IPUMS)
- Submit content questions with Q&A tool
- Will post written Q&A document following webinar

Outline

- Brief introduction of IPUMS USA
- Background on race variables in IPUMS USA
- What problem is RACHSING trying to solve?
- Solutions to the problem: Bad ideas, RACESING, RACHSING
- How RACHSING was made
- Using RACHSING or the PRED variables
- How to find RACHSING and the PRED* variables on the IPUMS USA website

Intro to IPUMS USA

- Harmonizes U.S. decennial census and American Community Survey (ACS) microdata across time and space
- Provides the data and accompanying documentation on the IPUMS USA website to access free of charge

Data Available on IPUMS USA

- Datasets:
  - U.S. decennial censuses
    - Household- and county-level (1790-1840)
    - Full count (1850-1940)
    - Samples (1950-2010)
  - American Community Surveys (2000-2020)
  - Supplemental Datasets/Resources

- Topics of Interest:
  - Levels of geography, household composition, family interrelationships, income/poverty, industry/occupation, SES, health insurance, and more!

How to Access the Data

- Create a custom data extract through the website
- Utilize the online analysis tools
  - Online analysis system
  - IPUMS Abacus
- New! IPUMS Microdata Extract API
  - Email ipums+api@umn.edu to become a Beta tester

- Data and documentation freely available to everyone at usa.ipums.org

RACHSING & PRED*

Simplified race/ethnicity variable, in two formats

- Use for consistent denominators for rates
- Show trends over time that begin before 2000
- Five interpretable categories include all people

You can still use the details in RACE and HISPAN to understand the people you're studying.

Background on RACE variables

- 1977 OMB: groups for federal purposes
  - Four single race groups
    - White,
    - Black,
    - American Indian or Alaska Native (AIAN),
    - Asian or Pacific Islander (API)
  - Two “ethnicities”
    - Hispanic and non-Hispanic

Background on RACE variables

- 1997 OMB: groups for federal purposes
  - Five race groups
    - White, Black, AIAN, Asian, Pacific Islander
  - Multiple races now allowed to be reported!
  - Same two “ethnicities”: Hispanic and non-Hispanic
    - but multiple ethnicities are still not acknowledged

What's the problem?

1. States were slow to change from single-race
   - Birth certificates
   - Death certificates
   - Cancer registries

Birth, death, and cancer rates have state data numerator and federal data denominator. ☹

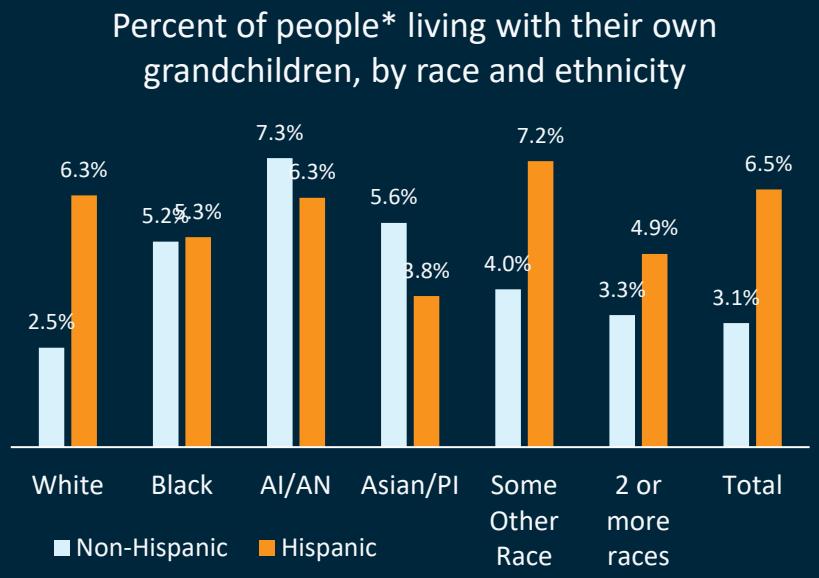
What's the problem?

2. So many categories!
   - 127 possible race/Hispanic categories in the census

Common "solution" is to combine all "Two or more races."

- This is an uninterpretable category ☹
- Half+ of AIAN and Pacific Islanders report two or more races. AIAN & PI single-race are non-random subsets. ☹

What's the problem? Example 1

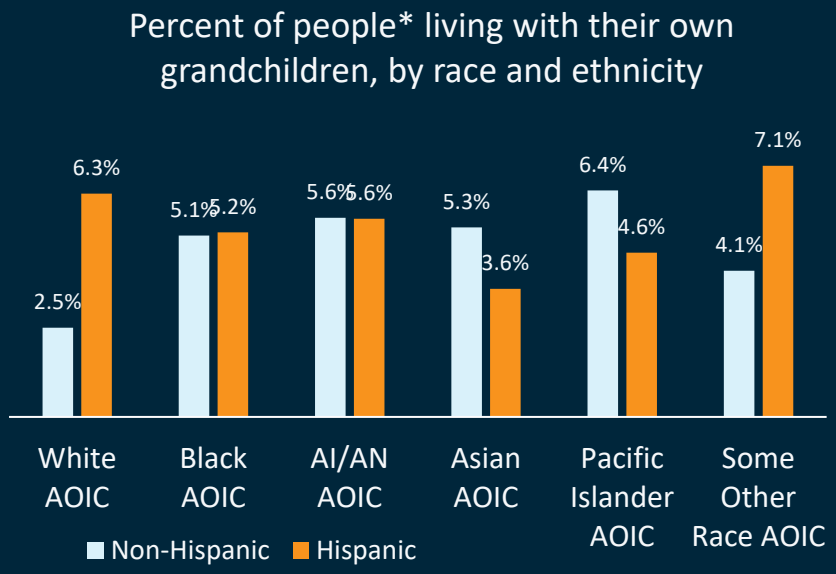


* Includes only people ages 30 and older living in households.
2019 5-year ACS from IPUMS.org.

Some issues:

- 12 categories is a lot
- Who is "2 or more races"?
- Group all Hispanic/Latinx?
- Non-Hispanic SOR is tiny
- Half of AI/AN are in 2+

What's the problem? Example 2



Some issues:

- 10 categories
- Sum of N is more than total population because multiple-race people in multiple categories
- Group all Hispanic/Latinx?
- Non-Hispanic SOR is tiny

AOIC = Alone or In Combination, meaning that single-race and multiple-race people who marked this race group are included.
* Includes only people ages 30 and older living in households.
2019 5-year ACS from IPUMS.org.

First ideas by research community…

1. For all 2-race responses, assign ½ the case to each of the races, etc.

2. For all multi-race responses, assign the whole case to the smallest race group

NCHS to the rescue!

National Center for Health Statistics researchers (Ingram et al. 2003) used the 1997-2000 National Health Interview Survey:

- Individual-level prediction of the single race given by multi-race people who were asked to pick one
- Results can be used for "whole allocation" or "partial allocation" of cases to race groups.

Whole vs partial allocation

- Example prediction equation results for AIAN-White:
  - AIAN = 0.35
  - White = 0.65
- Whole allocation assigns the case to White group.
- Partial allocation assigns 0.35 of the case to the AIAN group and 0.65 to the White group.
  - Similar to a weight or probability

IPUMS USA made RACESING & PROB*

- RACESING & PROB* applied NCHS equations to IPUMS USA data
  - RACESING = whole allocation
  - PROBASIAN, PROBBLK, etc = partial allocation
  - See Liebler & Halpern-Manners (2008) *Demography*
- Continued to 2014

RACESING issues

- RACESING was based on 1997-2000 NHIS data.
- RACESING stopped in 2014 but the need remained
  - Denominators for cancer rates (National Cancer Institute)
- RACESING did not deal with non-Hispanic SOR cases
- RACESING did not match state data (no Hispanic)
  - 5 groups needed: White, Black, AIAN, Asian/PI, Hispanic.

RACHSING categories

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

(NH = Non-Hispanic)

RACHSING is whole allocation

PRED* variables

1. PREDWHT
2. PREDBLK
3. PREDAI
4. PREDAPI
5. PREDHISP

PRED* are partial allocation

Steps to make RACHSING

1. In the FSRDC, use restricted-use data on multi-race people in NHIS
   - Restricted-use NHIS data allow location-based variables
   - 4-year pooled samples, e.g.: 1997-2000, …, 2015-2018

Steps to make RACHSING

2. Compress to 11 types of multiple-races.
   - This involves simplification ☹
     - Hispanic
     - Asian & Pacific Islander
     - “Some other race”
   - This allows separate prediction eqns. ☺

1. AIAN & A/PI
2. AIAN & Black
3. AIAN & White
4. A/PI & Black
5. A/PI & White
6. Black & White
7. AIAN & A/PI & B
8. AIAN & A/PI & W
9. AIAN & B & W
10. A/PI & B & W
11. AIAN & A/PI & B & W

Steps to make RACHSING

3. For each non-Hispanic multiple-race group and each 4 years of data, use individual and county-level characteristics to predict which single race they would have chosen.
   - Get prediction equation results released from FSRDC.
   - These will be released in an MPC working paper soon
   - Available on request

Steps to make RACHSING

4. Apply prediction eqns to IPUMS Census & ACS data
   a. Compress Census & ACS multiple-race answers to 11 groups
   b. Create PUMA-level measures (to substitute for county)
   c. Apply new prediction equations to multiple-race folks
      - Results are PRED* variables
   d. Do "whole allocation" to create RACHSING
   e. Classify everyone else for RACHSING & PRED*(Hispanic; non-Hispanic single-race; non-Hispanic Some Other Race)

RACHSING categories

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

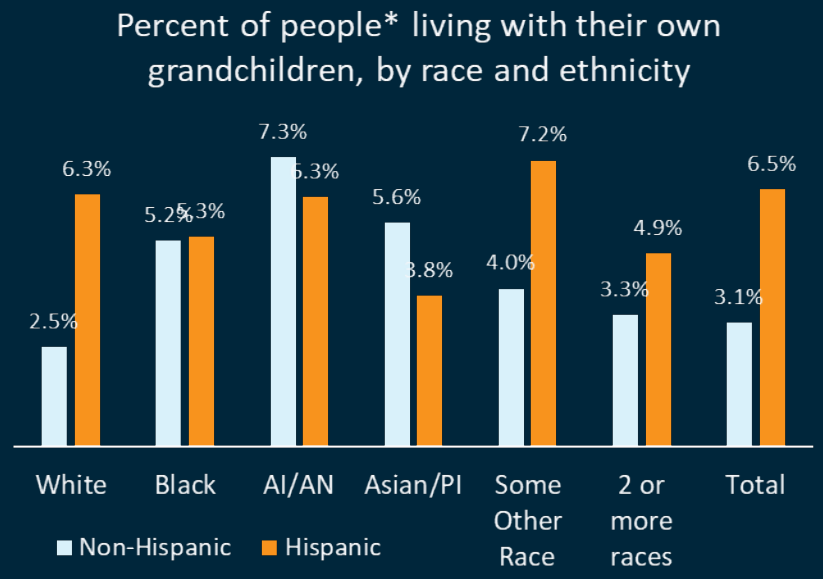
(NH = Non-Hispanic)

PRED* variables

1. PREDWHT
2. PREDBLK
3. PREDAI
4. PREDAPI
5. PREDHISP

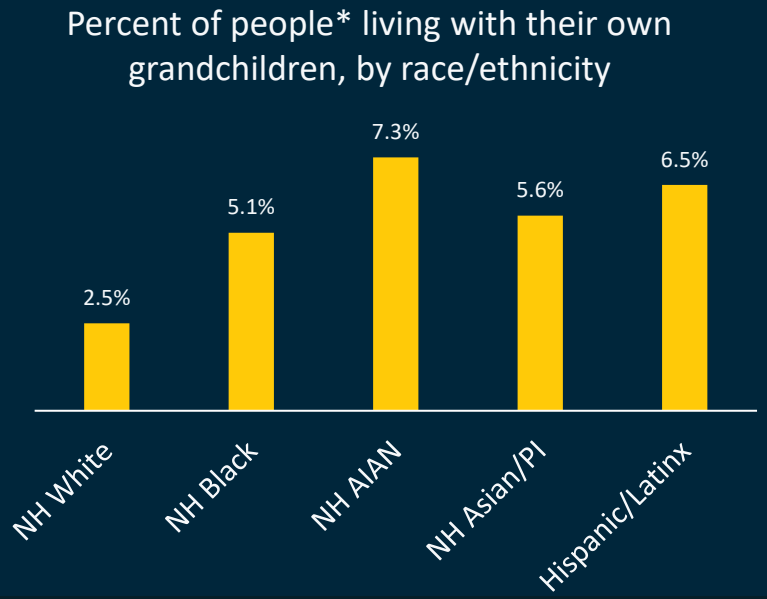
RACE	HISPAN	RACHSING	PREDAI	PREDAPI	PREDBLK	PREDHISP	PREDWHT
Black	No	NH Black	0	0	1	0	0
White & SOR	Cuban	Hispanic	0	0	0	1	0
Chinese & White	No	NH API	0	0.6327	0	0	0.3673
Navajo & Black	No	NH Black	0.4265	0	0.5735	0	0
Some Other Race	No	NH White	0.0136	0.2378	0.036	0	0.7126

RACE x HISPAN



* Includes only people ages 30 and older living in households.
2019 5-year ACS from IPUMS.org.

RACHSING

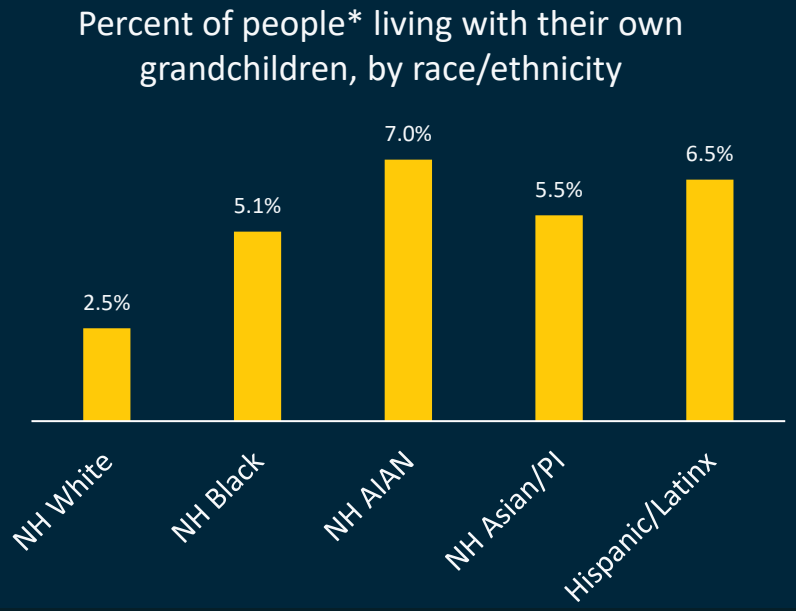


RACHSING, via IPUMS.org: Non-Hispanic multiple-race respondent are included in the category predicted to be their choice if asked to choose one race.
* Includes only people ages 30 and older living in households.
2019 5-year ACS from IPUMS.org.

PRED variables

PREDWHT, PREDBLK, PREDAI, PREDAPI, PREDHISP

- "Partial allocation"
- Each ranges from 0 to 1
- For each person, sum to 1.0
- Multiple-race people are treated as multiple-race



PRED* variables, via IPUMS.org: Non-Hispanic multiple-race respondents are included in each relevant category, as predicted to be their choice if asked to choose one race.

* Includes only people ages 30 and older living in households. 2019 5-year ACS from IPUMS.org.

How to find RACHSING and the PRED* variables on the website

What samples are the RACHSING and PRED* variables available for?

- Currently available:
  - 2000 and 2010 decennial census samples
  - 2000-2019 ACS samples
- Upcoming!
  - Integration of RACHSING and PRED* variables for 1990-prior samples is currently underway

Acknowledgments

- Funding for Carolyn Liebler and IPUMS staff is from the National Cancer Institute

- IPUMS Staff
  - Danika Brockman
  - Stephanie Richards
  - Megan Schouweiler

MPC

Questions?

PREDBLK = 1

- Black or African American
- Black & SOR

0 < PREDBLK < 1

- Black + AIAN + White
- Asian + PI + Black + AIAN
- Asian + PI + SOR + Black
- Etc.

Value of PREDBLK is based on the individual's personal and location characteristics