

Research on Race Bridging for 2020

Ben Bolender
Assistant Division Chief
Population Estimates and Projections

Federal Committee on Statistical Methodology
March 8, 2018

Overview

1. Race bridging and reverse bridging
2. Bridging and imputation
3. Why would we use bridging after 2020?
4. What methods could we use?
5. How would we test these new methods?

Something Old and Something New

Current Procedures

Current Procedures

1. Race bridging

Why do we bridge races?

What is race bridging?

How do we bridge race data?

2. How do we bridge backwards?

3. How is bridging different from imputation?

Why do we bridge races?

Compatibility

The National Center for Health Statistics (NCHS) and the Census Bureau make extensive use of each other's data

5 Number of race alone or in combination categories used in Census estimates

4 Number of race categories that some states still collect in vital records

For us to effectively share data with each other, we needed a way to convert back and forth

Why do we bridge races?

1977 OMB race standards

- American Indian or Alaska Native
- Asian or Pacific Islander
- Black
- White

1997 OMB race standards

- Separated Native Hawaiian or Pacific Islander from Asian
- Allowed multiple race categories
- Greatly increased diversity that people could report

31

The total number of current population estimates race groups, because estimates do not account for “Some Other Race”

What is race bridging?

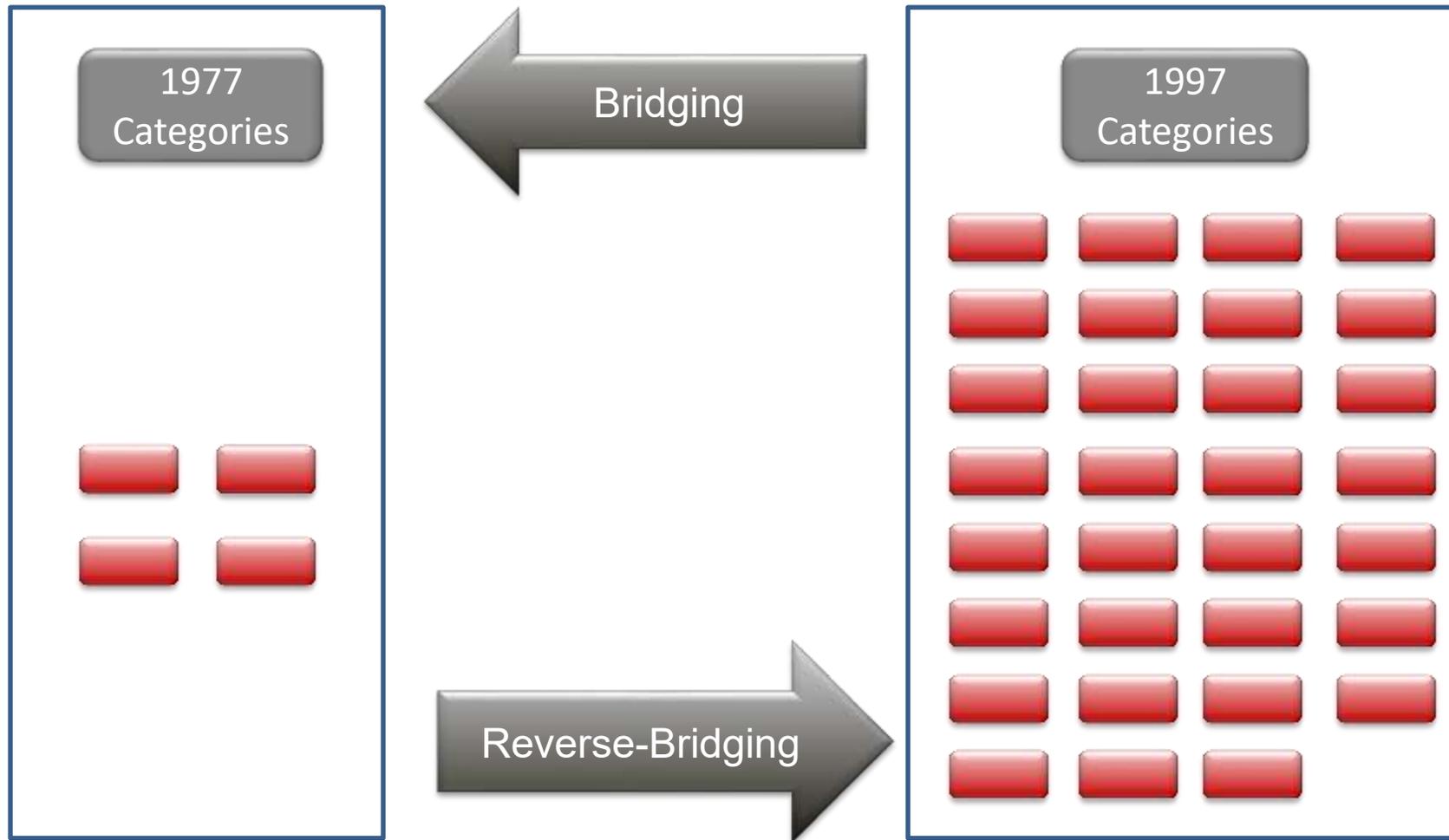
Race bridging is a way to convert one set of categories into another using aggregate data and proportions

The proportions

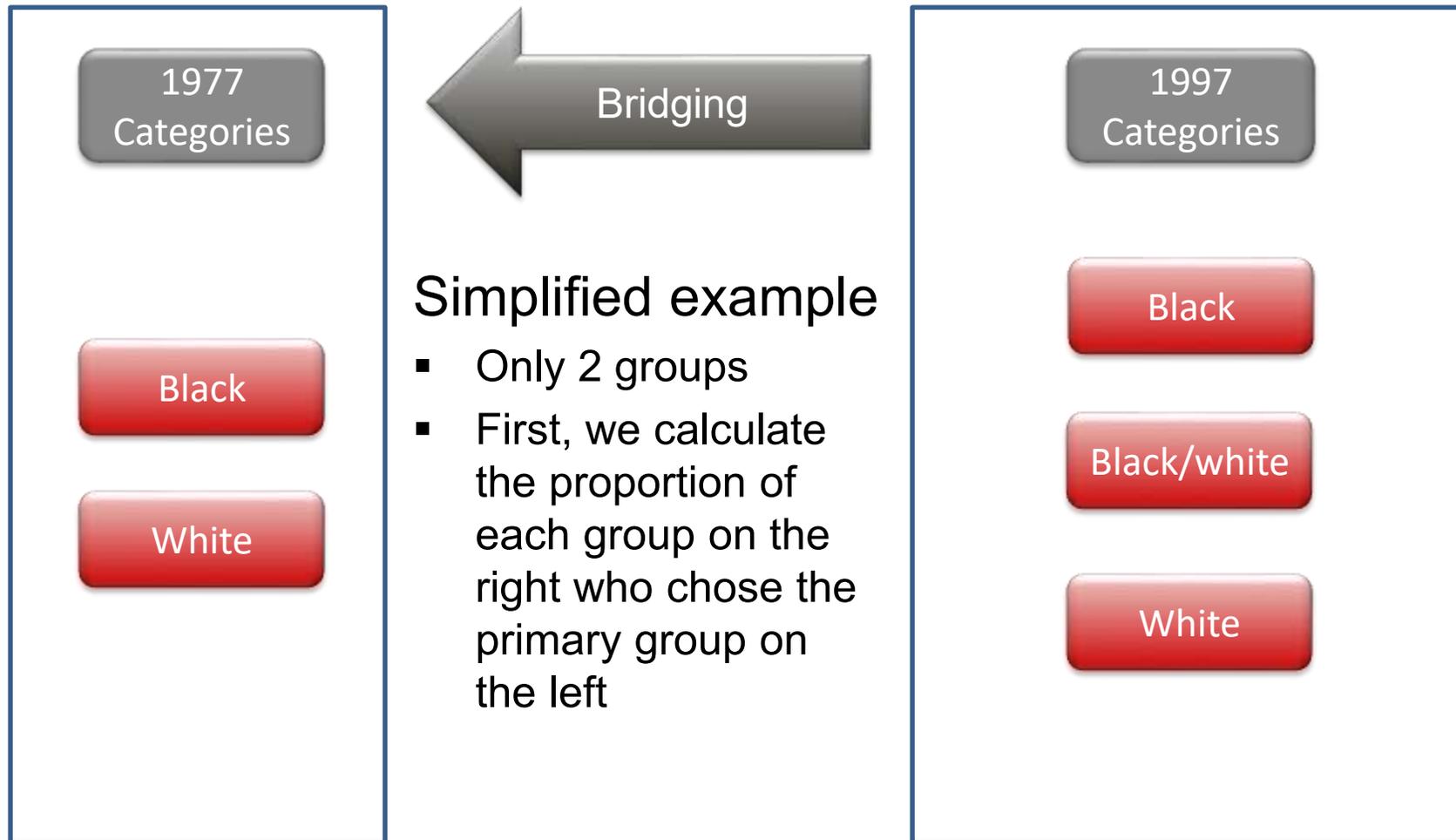
Come from work NCHS did with the National Health Interview Survey (NHIS) data from 1997-2000

- The survey asked respondents their races in the new 1997 categories, then asked them to choose a “primary race” from the 1977 list
- This work allowed for the creation of “bridging factors”

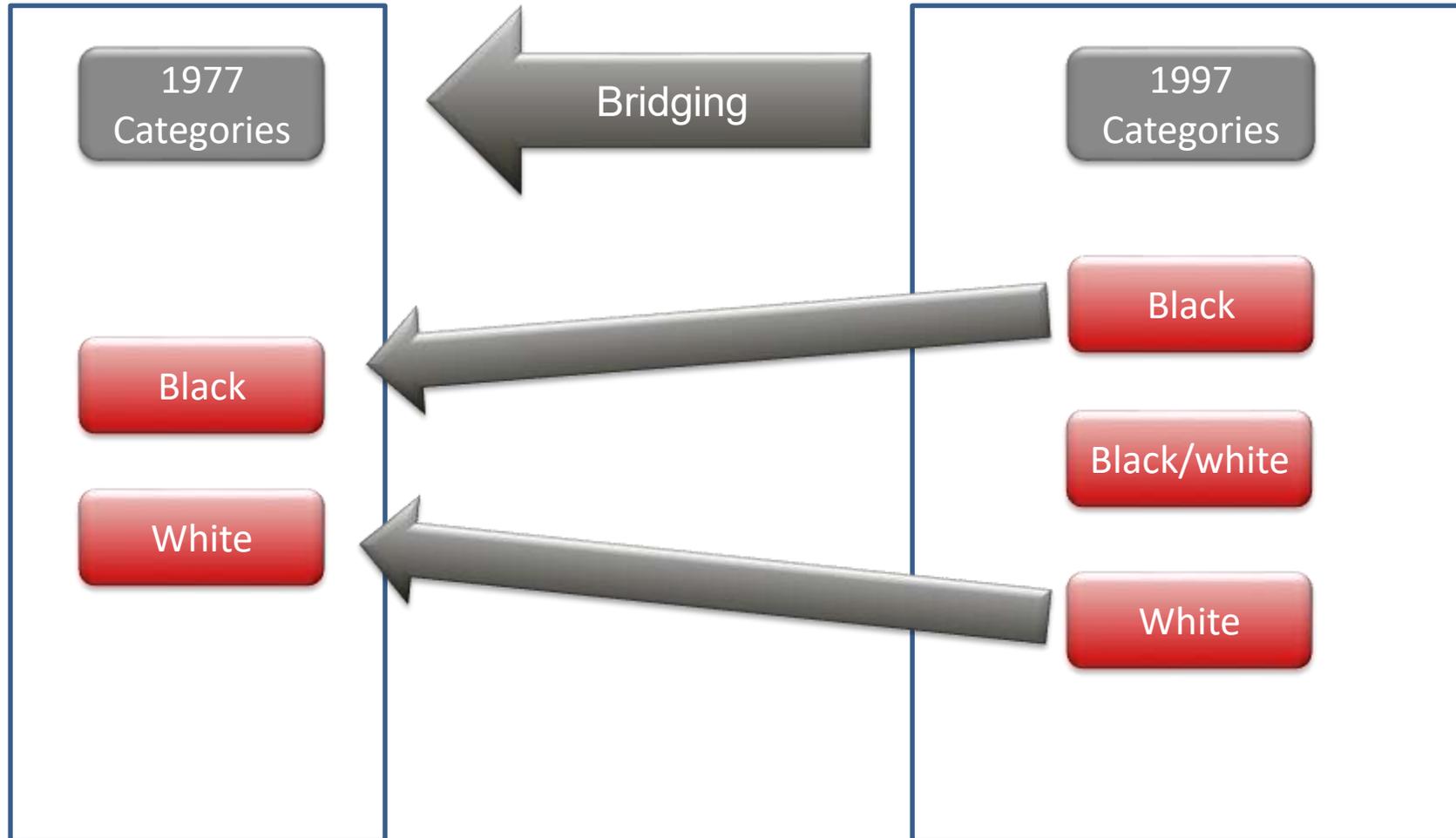
How do we currently bridge races?



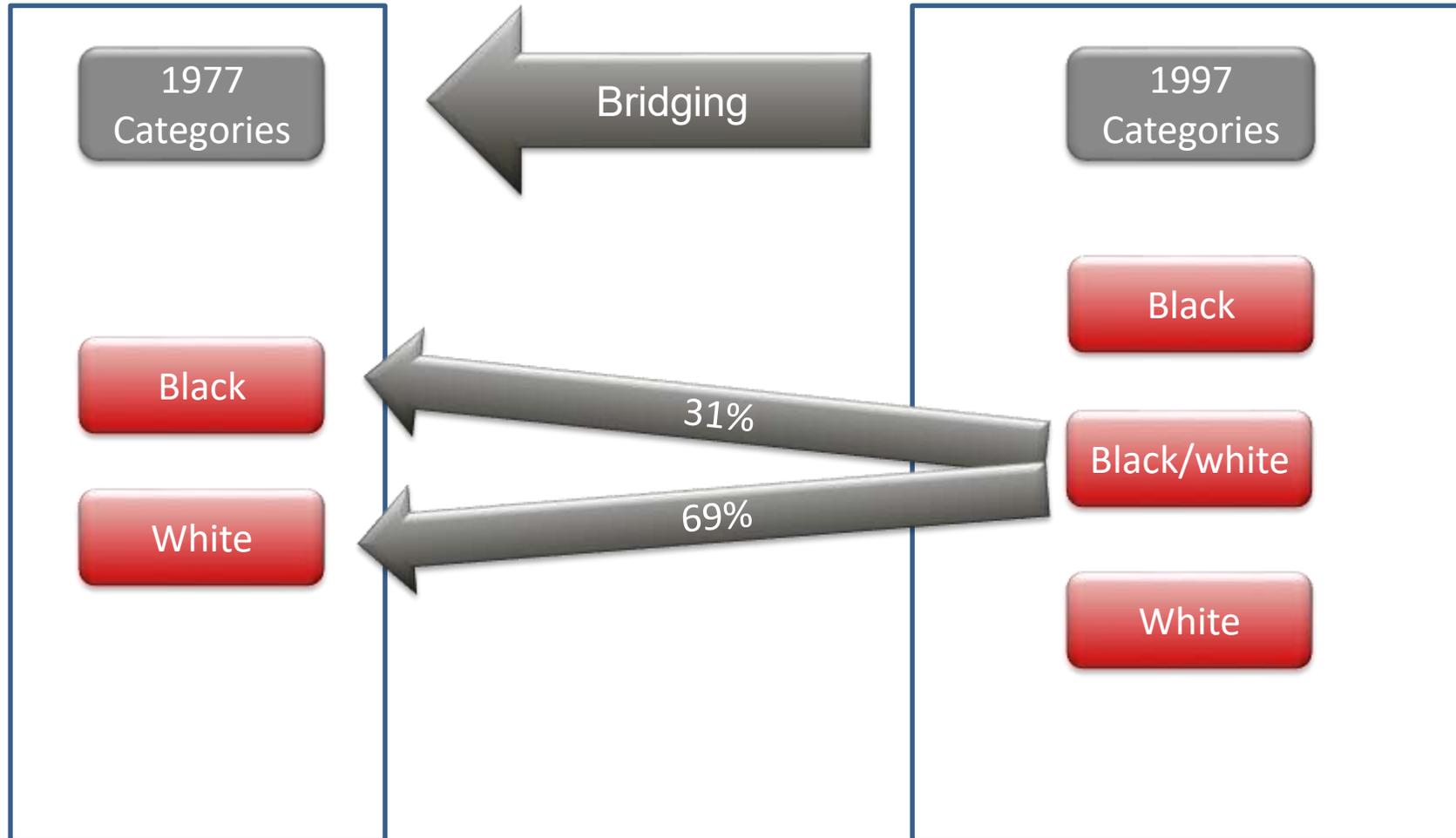
How do we currently bridge races?



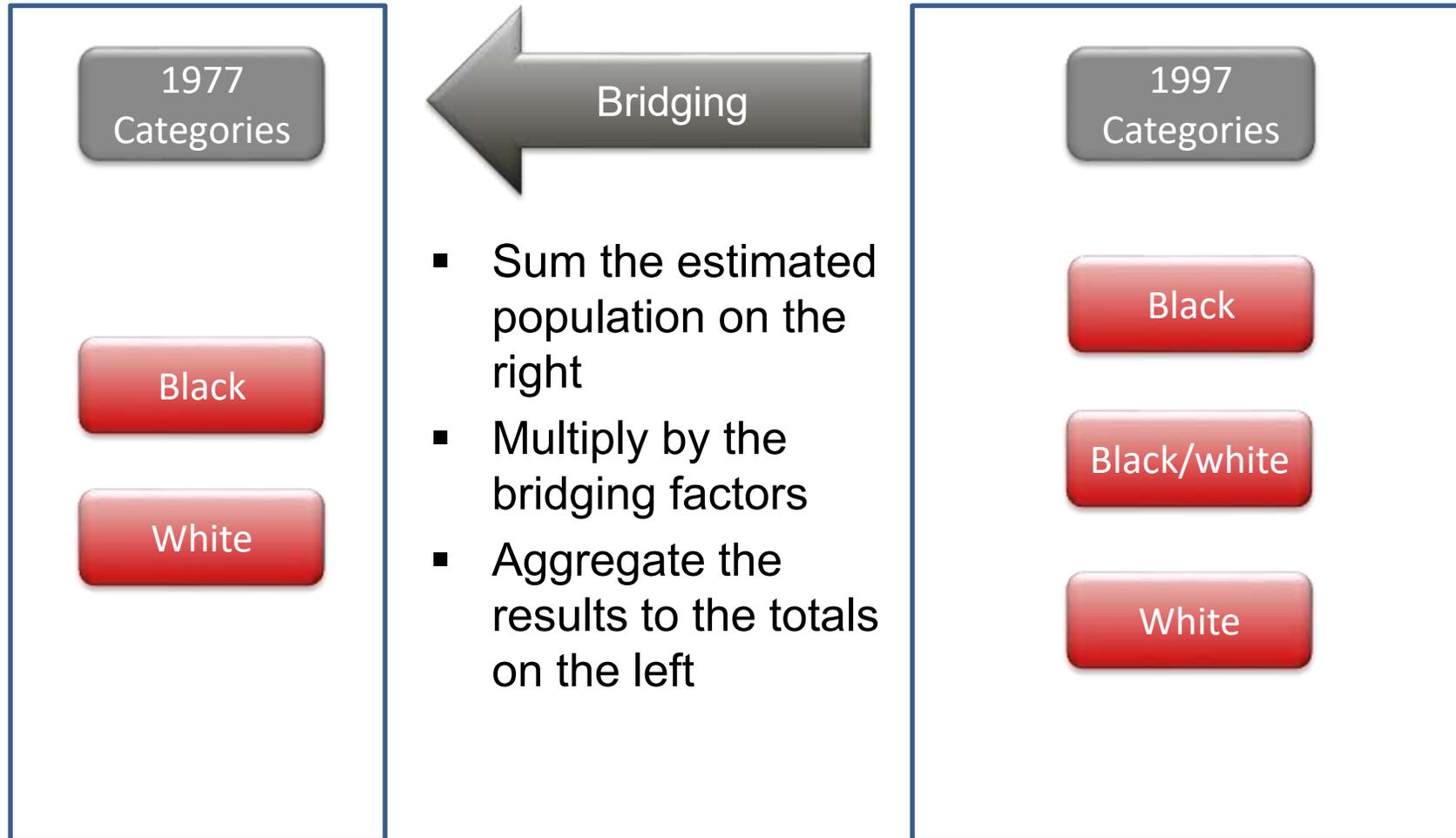
How do we currently bridge races?



How do we currently bridge races?



How do we currently bridge races?



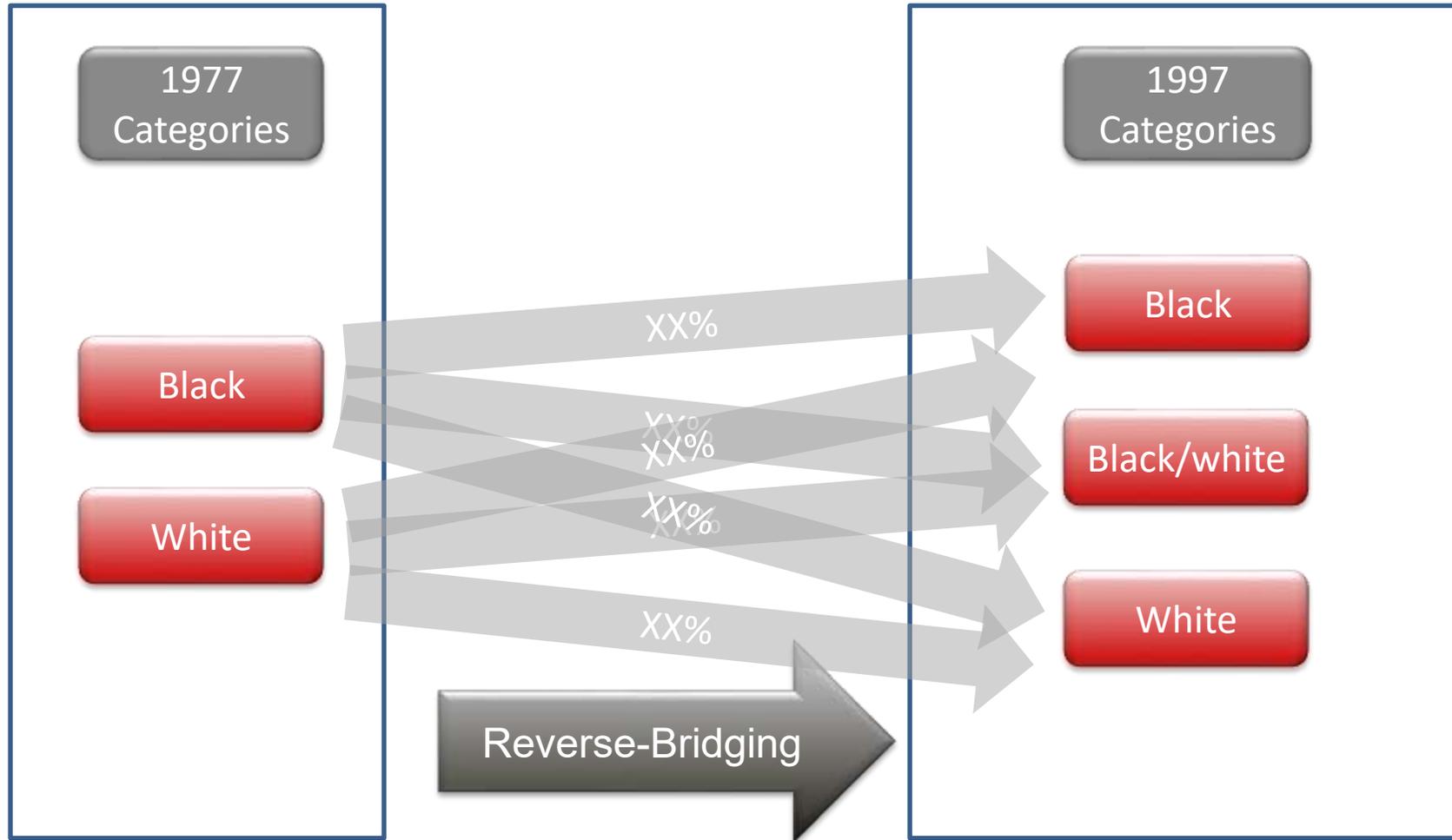
How do we bridge backwards?

The Census Bureau needs 31 groups for its estimates, so it built on NCHS work to develop “reverse-bridging”

Bridging from 4 back to 31 groups

- We start by bridging the most current decennial Census
- This gives us a population count in both race systems
- The ratios between those two populations are used to convert birth and death data from 4 races to 31

How do we bridge backwards?



What are bridging and imputation?

Bridging and imputation are two major ways that we have long used to convert from one classification to another

Bridging uses proportions and aggregate data



Imputation assigns a value to micro records



What are bridging and imputation?

Bridging

- Converts one characteristic distribution to another
- Applies proportions to an aggregate population

Example: Converting 31 race PEP data to 4 race NCHS controls

Imputation

- Generates new or different characteristics for responses
- Operates on individual records
- Relies on criteria or “hot-decking”

Example: Modifying individual “Some Other Race” responses in the decennial Census into OMB standards

What are bridging and imputation?

Census imputes race for “Some Other Race”

Imputation

- Primarily done on the decennial Census (base)
- Process drops SOR from multiple race responses or assigns a race from a record with similar characteristics
- Most likely sources are people within the household or neighborhood

97 % of the SOR alone population is Hispanic

40 % of Hispanics have their race imputed through this process

Something Borrowed, What to Do?

Proposed Improvements to Race Bridging in 2020

Proposed Improvements

1. Background
2. How could we bridge from Some Other Race alone?
3. Putting it all together
4. How do we know if it's good?

Who might need a conversion and why?

The short answer is “almost everybody”

- Many agencies and researchers use the OMB standard race groups (as a maximum)
- We develop population estimates only for the OMB standard race groups

National Center for Health Statistics

Bureau of Labor Statistics

National Cancer Institute

Department of Justice

Department of Education

Numerous Census Operated Surveys

How would we bridge from SOR alone?

Option A: 2020 Census

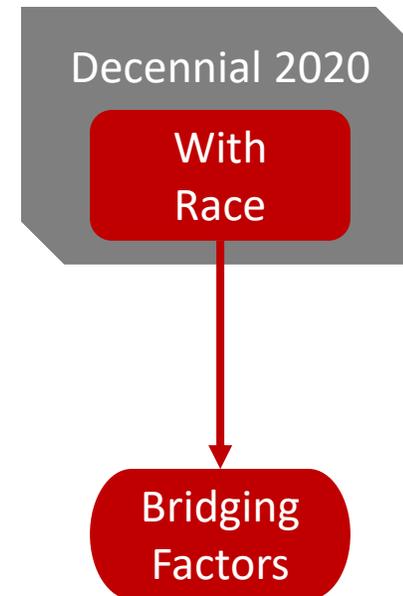
Develop bridging factors using the 2020 Census responses that have a non-imputed race

Pro

- + Largest sample
- + Simplest method
- + Allows best geographic resolution

Con

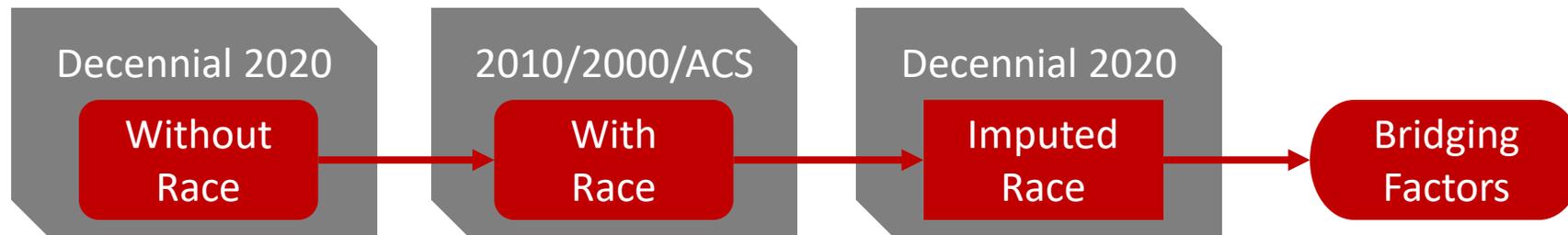
- Updates would require ACS data
- Most disconnect from micro data



How would we bridge from SOR alone?

Option B: Linking Records

Link 2020 micro records to previous responses to decennial census and American Community Survey (ACS)



Pro

- + Allows us to link micro records
- + Similar to original methodology
- + Linkage work already planned

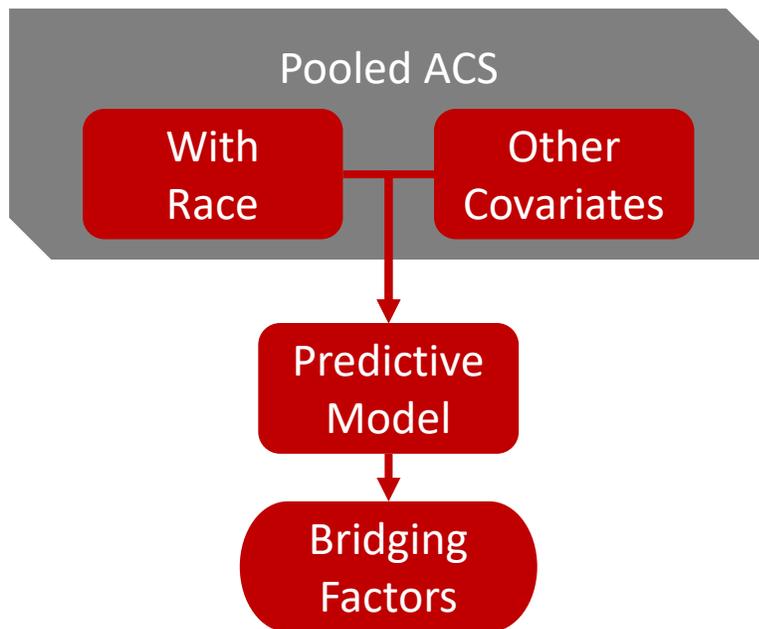
Con

- Impossible to update after 2020
- Relies on smaller sample
- Assumes race identification does not change over time

How would we bridge from SOR alone?

Option C: ACS Model

Model bridging factors using pooled ACS data on the covariates of the population who chose each race



Pro

- + Allows for increased specification
- + Can be updated regularly

Con

- Smallest sample
- Sampling variability year to year

How would we bridge from SOR alone?

Option D: Demographic Characteristics File (DCF)

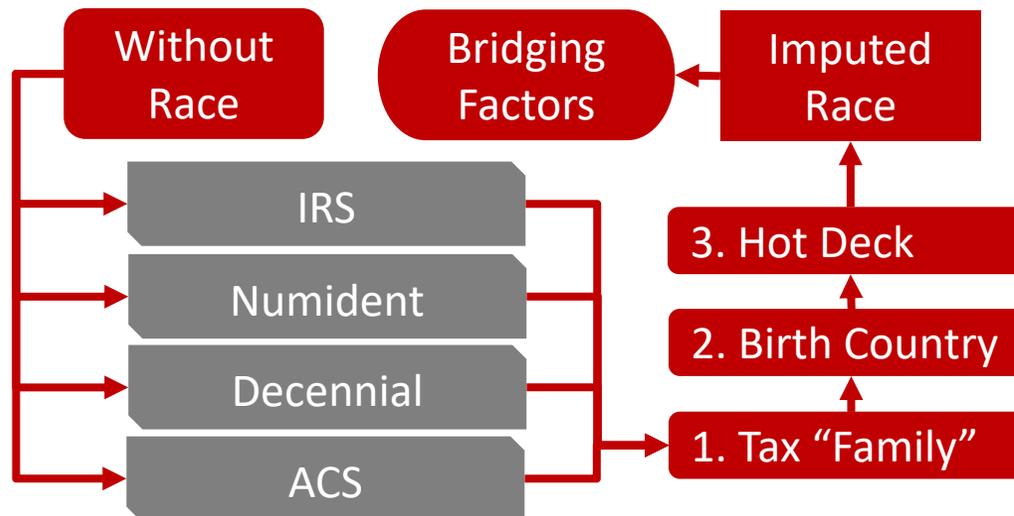
Link multiple data files and impute like migration records

Pro

- + Currently in production for another estimates product
- + Can be updated regularly
- + Allows imputation of micro data if required

Con

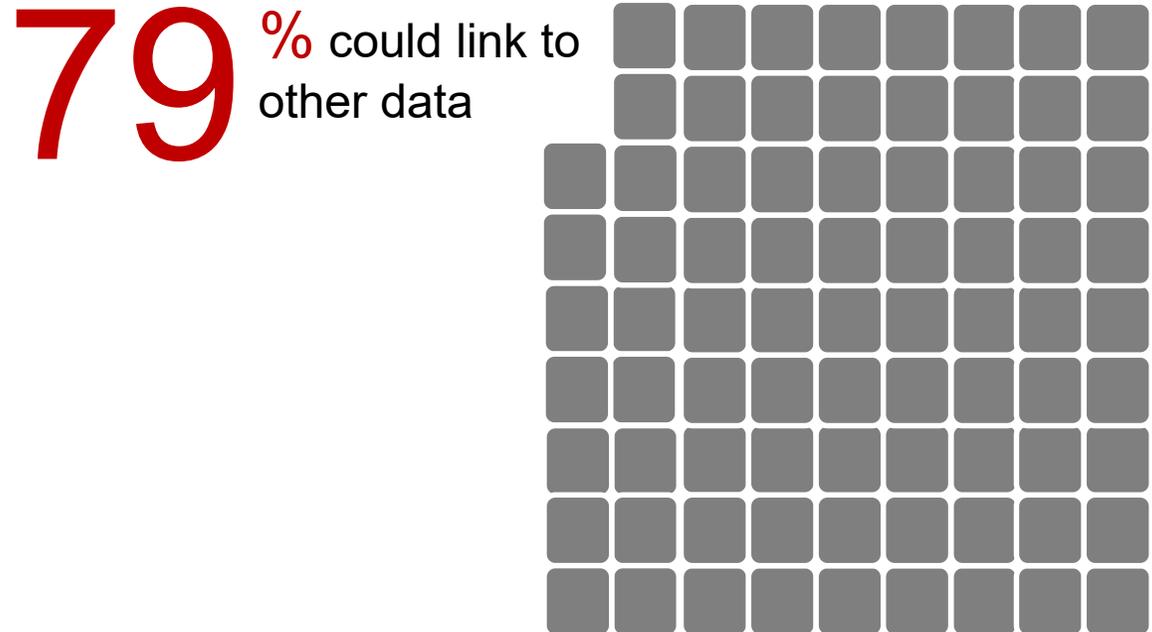
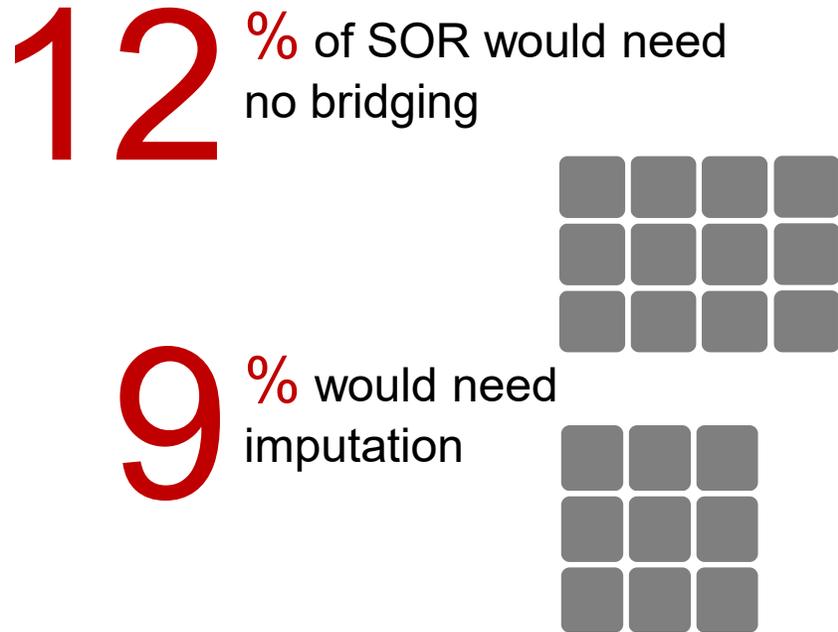
- Most technically complicated
- Highest data requirements
- New data linkages



Putting it all together

Step 1: 2020-Based Bridging Factors

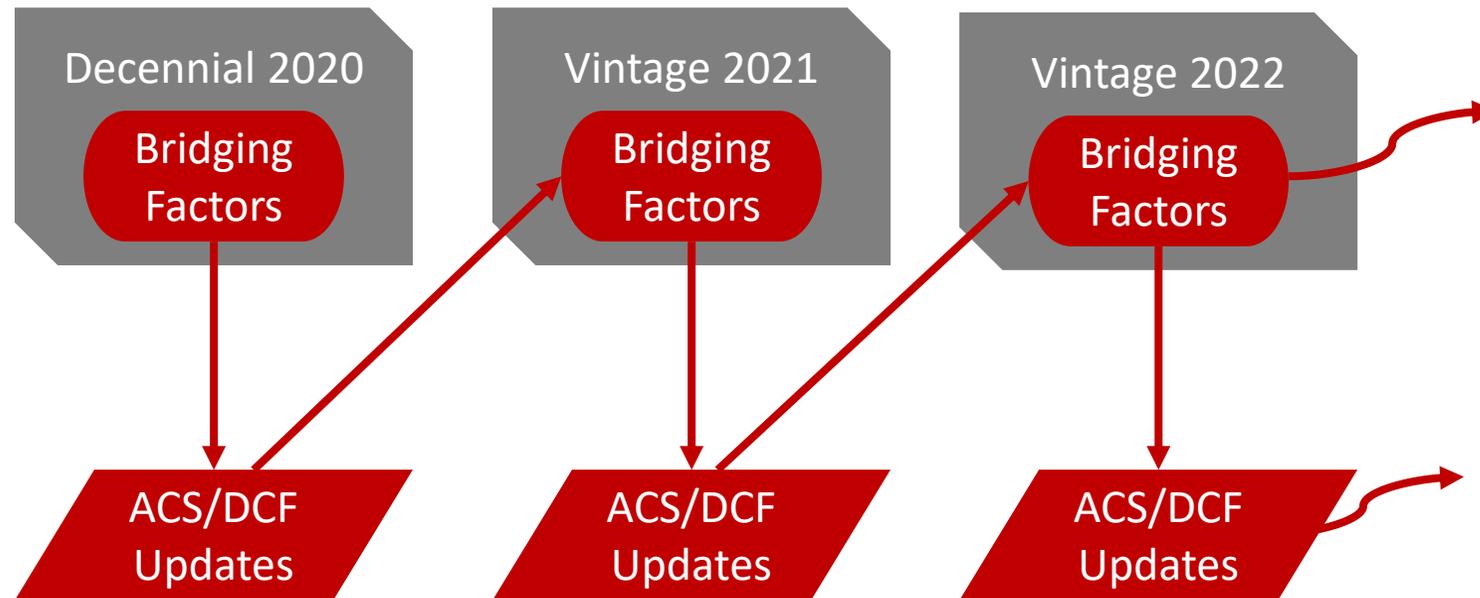
Take race responses as they are, link what we can to ACS or decennial data, impute the rest based on DCF method



Putting it all together

Step 2: Continual Updating with ACS/DCF

Research ways to update these bridging factors over time using new input from ACS and DCF data linkages

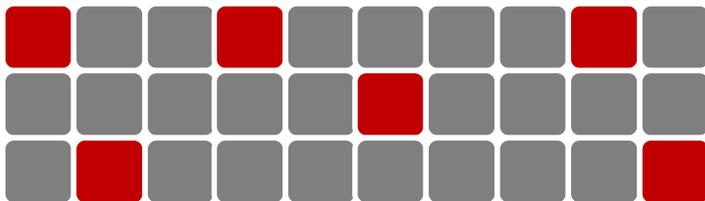


How do we know if it's good?

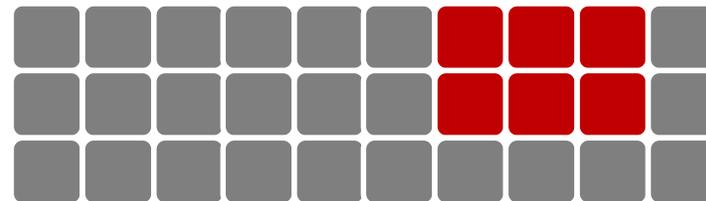
Reproduce distributions

Next we “blank out” races and see how well we could reproduce the reported race distribution by characteristics such as geography, age, or sex

We may block out responses randomly



We may block out particular groupings



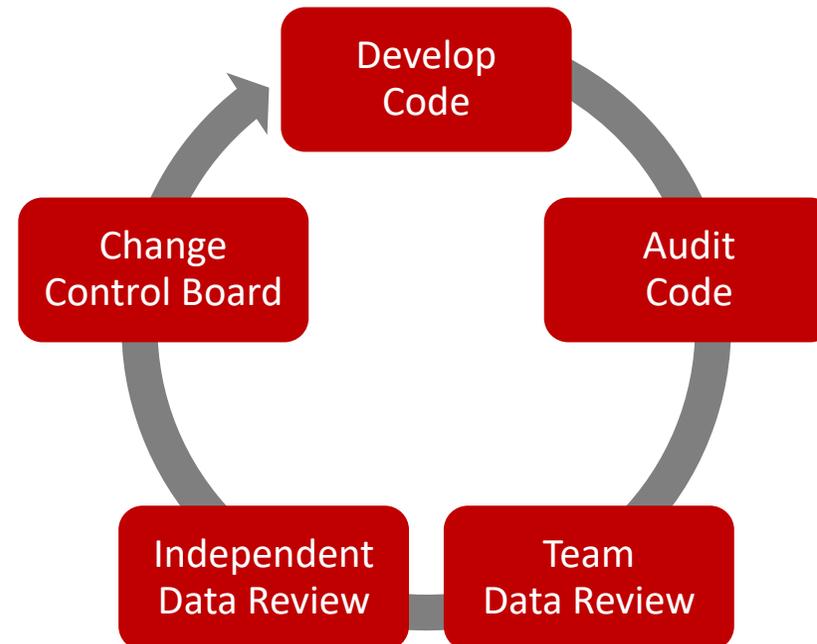
How do we know if it's good?

Iterative review and continuous improvement

Quality is central to the Census Estimates program

We plan to test

- Simple proportions
- Each option individually
- Combinations
- Sequencing
- Bridging vs imputation
- Reverse bridging



Testing allows us to refine our method for 2020

Research on Race Bridging for 2020

Questions and Discussion

Ben Bolender

Assistant Division Chief: Population Estimates and Projections

benjamin.c.bolender@census.gov

301-763-9733