

Advancing Equity in Race and Ethnicity Data in Population Surveys: Findings from Expert Interviews



By **Alyssa Ghirardelli**,¹ **Petry Ubri**,¹ and **Praveen Karunatileka**¹

¹ NORC at the University of Chicago, Health Care Programs

Abstract

This Equity Brief describes findings from a qualitative study using interviews with subject matter experts in survey research and measurement and collection of demographic data on race and ethnicity. The purpose of the interviews was to identify and share best practices on equitable methods to collect and analyze demographic data on race and ethnicity.

Keywords: race and ethnicity data collection, equitable research methods

Introduction

Race and ethnicity definitions and data collection have changed over time due to social and political changes in the United States.¹ Recent efforts highlight the need for changes to current Office of Management and Budget (OMB) racial and ethnic categories. Established in January 2021, the [Equitable Data Working Group](#), co-chaired by OMB and the Office of Science and Technology Policy (OTSP), released [A Vision for Equitable Data: Recommendations from the Equitable Data Working Group](#), which recommended revisions to the OMB standards while preserving the ability for comparable data in [Statistical Policy Directive 15](#). In January 2023, OMB released [Initial Proposals for Updating OMB's Race and Ethnicity Statistical Standards](#) for public comment. The Federal Committee on Statistical Methodology (FCSM) published [The Equitable Data Toolkit](#) to support equity analyses on historically underserved populations, which includes a section on [Race and Ethnicity Data Tools](#).

NORC at the University of Chicago (NORC) prioritizes accurately capturing the diversity of the U.S. population in population surveys. NORC conducted this research on equitable practices in survey research in the fall of 2022 with funding from NORC's Diversity Equity and Inclusion

Research Innovation Fund. We interviewed eight subject matter experts (SMEs, see Appendix) from September to October to understand best practices and approaches in the collection and disaggregation of race and ethnicity data, including with research design, sampling, and analysis. We also asked SMEs about the need for equitable methods, discussed the race and ethnicity categories, and asked about recommendations for researchers to improve race and ethnicity data. NORC identified SMEs based on a literature review, a search for data disaggregation experts, and suggestions made by internal experts at NORC. We sought to ensure representation from SMEs who work with diverse populations in survey research to allow for discussion of challenges and best practices for each OMB racial-ethnic category. We used Dedoose, a qualitative data software, to code interview data and synthesize findings using thematic analysis.

The purpose of this Equity Brief is to identify best practices to integrate future revisions of the OMB standards into current practice while considering additional opportunities in the equitable measurement and collection of demographic data on race and ethnicity. Findings facilitate rigor for innovations in conducting inclusive and equitable survey research.

Corresponding Author: Alyssa Ghirardelli, Senior Research Scientist, NORC at the University of Chicago, ghirardelli-alyssa@norc.org

Best Practices

SMEs identified several best practices for advancing equity in survey research. We describe these findings below.

RESEARCH DESIGN AND PLANNING

Use mixed methods approaches to hear from voices that are often underrepresented in research.

Some SMEs described the need to supplement survey research with qualitative data (e.g., interviews, focus groups) to capture and share the experiences of populations with small sample sizes. SMEs noted that qualitative research methods are an additional tool to understand a broader set of experiences. In conjunction with quantitative data, qualitative data can help tell a story and provide more in-depth information about key issues.

Engage communities to build trust for collection of race and ethnicity data in surveys, reach populations underrepresented in research, and promote more equitable survey research methods.

SMEs highlighted the importance of working with trusted community partners and members to:

- Understand populations and stay abreast of proper terminology for how people describe their identities
- Use focus groups to pilot survey questionnaires and ways to ask about race, ethnicity, and identity
- Explain the purpose of data collection and the need for race and ethnicity data collection, specifically to build trust and understanding; this includes explaining how researchers will use the data and protect the privacy of research participants
- Reduce racial and ethnic bias in how researchers approach surveys and ask survey questions

“It is the responsibility of researchers and scholars to build capacity. More often than not, there aren’t Indigenous people involved in decision-making, especially in leadership positions. Working towards equity equals having representation and investment in training Indigenous people and all folks that are underrepresented.”

– Desi Small-Rodriguez, University of California, Los Angeles

RACE AND ETHNICITY CATEGORIES IN INSTRUMENT DEVELOPMENT

Consider multiple ways to ask about identity.

Race is a social construct, and multiple factors influence identity. Self-reporting is the most common collection method of race and ethnicity data in large population survey. However, how people identify may differ from how other people perceive them. For example, measures of “street race, or how you believe other ‘Americans’ perceive your race at the level of the street” and “socially assigned race or...ascribed race, which refers to how you believe others usually classify your race in the United States” can inform how identity affects people’s well-being and access to opportunity.² When developing survey questions, researchers should consider multiple ways of asking about identity and investigate how people identify both as individuals and as part of a community and/or culture.

Be aware of how different individuals and groups identify and how those identifications may change over time.

SMEs noted that race and ethnicity categories in which people are unable to identify themselves creates the potential for biased data because they often exclude the experiences and perspectives of racially and ethnically minoritized groups. How researchers capture race and ethnicity should change over time as identity evolves for racial and ethnic groups.

Understand the limitations to comparability that result from updating racial and ethnic categories.

While SMEs acknowledged that updating racial and ethnic categories is often at odds with being able to compare data to prior years, they stated a need to balance time, resources, and the tradeoffs of limited comparability of data with ensuring that people are represented.

Ask identity questions that allow for data disaggregation beyond race.

Equitable collection of data facilitates the ability for researchers to disaggregate data and understand experiences of populations beyond broader racial-ethnic groups. However, SMEs noted that planning for data disaggregation in the design of survey research increases the use of equitable practices in research. SMEs also noted that data disaggregation generates insights on different segments of the population, describes experiences of people with different backgrounds, and identifies discrepancies and disparities within races.

SME Recommendations

Overall

- Allow individuals to check multiple categories when identifying their race and ethnicity*
- Use a “top six” most prevalent check-box option in main categories with additional write-in option³

Black

- Distinguish the experiences of U.S.-born Black Americans from African and Caribbean immigrants, a fast-growing population in the United States

American Indian/Alaskan Native (AI/AN)

- Include write-in options that allow individuals to report all the ways in which they identify[±]
- Based on sample availability and protection of personally identifying information, include skip patterns to choose tribal identity
- Report disaggregated data by tribe as available, based on sample size
- Work with tribal institutional review boards (IRBs⁴) and support data sovereignty⁵

Asian

- Report disaggregated data by country of origin

Hispanic/Latino

- Include Hispanic/Latino as part of the race question as opposed to asking it separately^{±‡}
- Ask if individuals consider themselves Afro-Latino
- Ask individuals how they identify (e.g., Hispanic, Latino/a/x/e/@) and use that language throughout, when using a web survey
- Report disaggregated data by country of origin

Multi-Racial

- Allow individuals to “check all that apply” so that they can select more than one racial category*
- Add a question directly asking individuals whether they consider themselves to be multi-racial
- Consider wording of the question (e.g., multi-racial vs. mestizo/a vs. mulatto/a) because not all individuals who identify with multiple races consider themselves to be multi-racial or mixed race

Native Hawaiian or Other Pacific Islander

- Consider creative outreach and sampling techniques for this historically marginalized group that is often combined with the Asian category

White

- Disaggregate Middle East and North African (MENA) populations from the white category[±]
- Include questions about nativity to understand the experiences of the immigrant white population

*Part of existing OMB guidance

± Part of 2023 proposed OMB guidance.

‡ There are mixed perspectives on whether combining Hispanic/Latino ethnicity with race will mask racial differences among this population.^{6,7}

Consider the role of other demographic factors that intersect with race and ethnicity.

Education, language spoken, country of origin, nativity, immigration status, income, and gender are other important factors to consider when teasing out inequities. SMEs suggested that, in large surveys, researchers should ask about and disaggregate data by these other factors to provide more context about the structural factors that affect populations.

SAMPLING

Oversample populations that are typically underrepresented in research.

SMEs acknowledged that one of the main challenges with data disaggregation is lack of sufficient sample sizes for some populations; small numbers may require data suppression to protect individual privacy. For smaller surveys, SMEs noted that increased granularity or disaggregated categories in a data collection instrument may take up too much valuable real-estate and still not yield enough sample for comparisons. SMEs recommended oversampling populations that are often underrepresented in research and using weighting strategies (with caution) to ensure that these groups are represented in the data. They also noted that IRBs should consider privacy and confidentiality issues when examining inclusion of typically underrepresented populations.

Pool data across years.

When surveys are repeatedly fielded across multiple years, consider pooling data to increase sample size and power to facilitate data disaggregation by race and ethnicity.

Use purposive sampling but understand its limitations.

SMEs recommended using purposive sampling methods to reach underrepresented populations, such as identifying populations through connections with local community-based organizations. While purposive sampling can limit the generalizability of findings, researchers should consider oversampling under-represented populations and using weighting or statistical calibration to adjust the overall sample.

DATA CLEANING AND ANALYSIS

Allow sufficient characters for write-in options in online surveys and granular coding categories to improve data cleaning practices.

Limiting the number of characters for write-in options in web surveys makes it difficult for people to describe who they are and how they identify. SMEs advocated for using more granular coding categories for back coding based on write-in and open-ended responses. For example, SMEs stated that it is often difficult to disaggregate data from individuals who identify as both AI/AN and Hispanic/Latino because Hispanic/Latino is often automatically top-coded, and their race may be excluded. For tribal nations, it is important to ensure that unique, federally and non-federally recognized tribal names can be coded and tabulated.

Imputation is a solution to missing race and ethnicity data at a population level, but researchers should recognize its limitations and consider consultation with data equity experts when establishing protocols.

Some researchers impute race and ethnicity for missing data. However, SMEs noted that imputing an individual's race can perpetuate racial stereotypes and assumptions about specific geographies and heritage. Imputations based on last names have limitations because some names are generic and provide little insights into race and ethnicity; marriage also affects name change. Furthermore, imputing missing data on race and ethnicity overrides an individual's choice to decline to answer the question.

Consider alternative models of analysis.

SMEs recommended using separate stratified, analysis models, even if there is a loss in the ability for comparison. One SME stated that there is bias in comparing racial and ethnic groups to a white reference group. Using white as a reference group creates an inherently racially biased model that sets white as the standard. Separate, stratified models consist of running separate models for racial and/or ethnic groups, which allows each group to get its own intercept and slopes.⁸ Survey researchers can compare differences across models.

“When you have a dummy variable for race and you mark white people as zero on that dummy variable...you’ve got a white intercept, which means that...the model...becomes racially biased.”

– Mosi Ifatunji, University of Wisconsin

REPORTING AND DISSEMINATION

Acknowledge the diversity of specific racial or ethnic populations and the overlap in identities when reporting findings.

Researchers must acknowledge that race is not always sharply defined in their data reporting. They should consider first describing the makeup of different populations beyond their race or ethnicity. In addition, they should acknowledge that there is often overlap in the race and ethnicity categories that people use to describe their identities—even if that does not create mutually exclusive categories and makes comparisons more difficult.

Represent findings from populations with small numbers in reporting.

SMEs suggested sharing descriptive findings of groups with small sample sizes, when possible, even if the findings from inferential analyses do not reach the point of statistical significance. When using this approach, one SME recommended reporting that data are purely descriptive and to interpret, with caution, and noted that this approach helps honor the contributions of groups with small sample sizes to the research.

Give data back to communities.

SMEs emphasized the importance of sharing data with communities in ways and formats that communities prefer (e.g., culturally responsive, linguistically appropriate). One SME noted that demystifying research is about proving its utility and ensuring that populations served by the research get value out of the data. For example, when working with Indigenous populations, consider the importance of data sovereignty. It is important to talk up front to Indigenous populations and their IRBs or governing structures about what happens to the data after collection and how communities will receive results of the surveys.

Summary

Researchers must think about the purpose and intent of their data collection efforts to capture current and accurate population trends for drawing causal inferences. Current OMB categories insufficiently capture the identities of U.S. populations.⁹ There is a need for more granular race and ethnicity categories that more accurately reflect the diversity of the country's population. Researchers should include additional variables and demographic factors that shed light on the complexity of identity beyond standard race and ethnicity categories to improve equitable practices.¹

Appendix: Participating Subject Matter Experts

Expert Name and Organization	Description	Population(s) of Focus
Ignatius Bau, JD Independent Consultant	Advancing health equity through data disaggregation	Asian American
Tiffany Burkhardt, PhD University of Chicago	Creator of the Racial Bias in Data Assessment Tool	N/A
Rashida Dorsey, PhD Federal Housing Finance Agency	Utility of federal data assets to advance equity	N/A
Mark Hugo Lopez, PhD Pew Research Center	Issues of racial and ethnic identity, Latino politics and culture	Hispanic/Latino; Asian American
Mosi Adesina Ifatunji, PhD University of Wisconsin	Racial and ethnic theory and associated methodologies	Black/African American
Nicholas A Jones, MA U.S. Census Bureau	Race reporting patterns and the demographic characteristics of children in inter-racial families	Multi-racial
Jen'nan G. Read, PhD Duke University	Secondary data disaggregation; diversity of white populations/immigrants	White
Desi Small-Rodriguez, PhD University of California, Los Angeles (UCLA)	Social demography of Indigenous communities	American Indian and Alaska Native

References

- ¹ Brown A. (2020). The Changing Categories the U.S. Census Has Used to Measure Race. Pew Research Center, <https://www.pewresearch.org/fact-tank/2020/02/25/the-changing-categories-the-u-s-has-used-to-measure-race/>.
- ² Gonzalez D, Lopez N, Karpman M, et al. (2022). Observing Race and Ethnicity through a New Lens. Urban Institute, <https://www.urban.org/research/publication/observing-race-and-ethnicity-through-new-lens>.
- ³ Mathews K, Phelan J, Jones NA, et. al. (2017). 2015 National Content Test Race and Ethnicity Analysis Report: A New Design for the 21st Century. United States Census Bureau, <https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2015nct-race-ethnicity-analysis.pdf>.
- ⁴ Around Him D, Andalcio Aguilar T, Frederick A, et al. (2019). Tribal IRBs: A Framework for Understanding Research Oversight in American Indian and Alaska Native Communities, *American Indian and Alaskan Native Mental Health Research*, 26(2), 71-95.
- ⁵ Garcia J. (2018). Support of US Indigenous Data Sovereignty and Inclusion of Tribes in the Development of Tribal Data, National Congress of American Indians, <https://www.ncai.org/resources/resolutions/support-of-us-indigenous-data-sovereignty-and-inclusion-of-tribes-in-the-development-of-tribal-data>.
- ⁶ Franco M. (2023). U.S. Government Considers Changing How It Asks About Latinos' Race, Axios, <https://www.axios.com/2023/01/31/census-latino-hispanic-race-ethnicity>.
- ⁷ The Leadership Conference Education Fund. (2023). Fact Sheet: Why do we Need a Combined Race and Ethnicity Question? <https://civilrights.org/wp-content/uploads/2023/04/FAQCombinedQuestion-1.pdf>.
- ⁸ Jackson JS, Neighbors HW, Nesse RM, et al. (2006). Methodological Innovations in the National Survey of American Life. *International Journal of Methods in Psychiatric Research*, 13(4), 289-298. <https://onlinelibrary.wiley.com/doi/abs/10.1002/mpr.182>.
- ⁹ Compton E, Bentley M, Ennis S, & Rastogi S. (2013). 2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment. (2013). U.S. Census Bureau, <https://www2.census.gov/programs-surveys/decennial/2010/program-management/5-review/cpex/2010-cpex-211.pdf>.

ABOUT NORC

NORC at the University of Chicago conducts research and analysis that decision-makers trust. As a nonpartisan research organization and a pioneer in measuring and understanding the world, we have studied almost every aspect of the human experience and every major news event for more than eight decades. Today, we partner with government, corporate, and nonprofit clients around the world to provide the objectivity and expertise necessary to inform the critical decisions facing society.