ORDSA

Office for Research Development and Scholarly Activity

ORDSA@ouhsc.edu

Guidelines for Using Race as a Variable in Your Research

ORDSA Guidelines on Quantitative Analysis of Race as a Variable

Along all stages of the research process, investigators should consider how race—*a social construct*—plays a role in the research question being explored. The unassailable connection of race to power, resources, and privilege (or lack thereof) is a fundamental cause of health inequities,[1] virtually irrespective of the health outcome studied. As scholars of community medicine and public health, ORDSA endorses the following guidelines and recommended considerations for all quantitative research that includes race as a variable. This document begins with a checklist summarizing the information in the detailed guidelines on the following pages.

Summary Guidelines
Using Race as a Variable in Your Research

Planning Your Study
☐ Carefully consider what role race plays in your research questions.
☐ Consult with people who are part of the population you are studying.

Performing Your Study
☐ Allow participants to self-identify race when possible.
☐ Alphabetize races in demographic questions.
☐ Use dual terms to describe race when applicable, i.e., Black/African American and Hispanic/Latinx.
☐ Consider listing Hispanic/Latinx as a race in your study rather than an ethnicity (while making sure people can choose more than one race on your survey). These results can be recoded into ethnicity variables during analysis if the funder or target journal requires it.

Reporting Your Study
☐ It is preferable not to combine race categories in analyses. If this is necessary for statistical power reasons, acknowledge this as a limitation in your manuscript.
☐ Do not blindly adjust for race as a confounder.
☐ Only use race as a covariate when race does not interact with any of the other variables or the dependent variable.
☐ It is appropriate to include race as a covariate when the research question relates to race.
☐ Contextualize race in a broader societal context when describing your results.
☐ Do not assume that racial differences in your results relate to genetics.
☐ Involve members of the affected community in results interpretation.

The Office for Research Development and Scholarly Activity, ORDSA@ouhsc.edu **ORDSA**

Detailed Guidelines for Using Race as a Variable in Research

Developing a Research Question

At the beginning of the research process, the researcher must consider how their research could have practice and policy implications. For example, a study that aims to identify causes of racial disparities in diabetes management will undoubtedly conclude that access to medical care plays a significant role if the primary measures are limited to healthcare. In contrast, a study that contextualizes daily life experiences that are central to diabetes management and race, such as psychosocial stressors and healthy food access, in addition to healthcare access, will likely conclude that psychosocial stressors and food access contribute more to diabetes disparities than access to healthcare contributes to these disparities. The research question(s) and study design should consider context and incorporate perspectives of the affected population. Researchers with an outsider perspective should consult with members of the affected population to determine whether the group's lived experiences are adequately reflected in the research question(s). This is an important step in shifting of power to minority groups who often do not play a role in formulation of research questions and resulting policy recommendations.

Best Practices for Methods and Reporting Results

Measuring Race

Since race is a social construct, race should be self-identified by the participant, whenever possible, with the option to report more than one race, which is consistent with NIH policy. Additionally, instructions should explicitly state to choose all options that are applicable. An additional variable may include self-reported socially-assigned race[2] to further elucidate the role of social discrimination and white privilege on opportunity and health.

In surveys, we recommend to alphabetize races to avoid perceptions of order bias. Use the term "Black or African American" rather than only "African American" to avoid offending people who prefer to be referred to as Black. Capitalize Black; it is a proper noun describing an identity. Do not hyphenate African American, as this can imply Black people are not fully American.[3] There are arguments for and against capitalizing white as a race option on surveys. Some people prefer to capitalize all races in their document for consistency.[4] Arguments against capitalizing white are related to a perception among some that white people are not a distinct cultural group.[5] Choosing either of these options is appropriate.

Hispanic/Latinx people often view Hispanic/Latinx as their racial identity rather ethnicity.[6] Including Hispanic/Latinx as a race rather than an ethnicity in your research shows you recognize that Hispanic/Latinx people considered "white" do not experience white privilege[6]. However, NIH views being Hispanic as an ethnicity and not a race. A tactic that values both of these perspectives is to use Hispanic as a race in your survey but report it as an ethnicity in work that requires use of NIH's current race/ethnicity categories. While it is preferable to report Hispanic as a race and not an ethnicity, this may not be possible for secondary datasets

that already have Hispanic listed as an ethnicity rather than a race. In this case, the author may choose to briefly report this as a limitation in their manuscript. One place to report this limitation would be in a footnote under the demographics table.

Some Native American people may prefer the term Native American to American Indian[4]. The researcher should use the preferred term from the group of people they are referring to when possible[4]. Some guidelines suggest avoiding using the word tribe and instead using people or nation[4]. It is preferred refer to the specific nation or people if possible (e.g., Cherokee, Chickasaw, Muscogee)[4].

Statistical Analysis
<u>Race as a Predictor Variable</u>

A common analysis uses race as a predictor for a specific health behavior or health outcome. For example, Race X is more likely to have Outcome Y than Race Z (often whites). This type of analysis is important for documenting health disparities, but it will often fall short of being able to explain a health inequity, which is the underlying cause for the disparity.[7] Researchers collecting race information should statistically explore social determinants of health, such as disparities in income, education, healthy food access, and discrimination[7]. These may explain how these racial disparities are fundamentally linked to these determinants of health.[7] This statistical approach allows the discussion to be framed toward opportunities for social action.

<u>Race as a CoVariate and stratified analysis</u>

It is acceptable to use race as a covariate, or confounding variable, only when there is no interaction between race and other independent variables or the dependent variable of interest[9].Statistical interaction or effect modification between race and other variables should be assessed prior to using race as a covariate[9]. If statistical interaction is present, then stratified analyses should be reported[9]. If race is a key component of the analyses or the researcher anticipates possible differences in the outcome by race, then the researcher needs to work with a statistician to estimate an adequate study sample size to accommodate appropriately powered stratified analysis. If an interaction is not present, race can be used as a covariate, but the interpretation of the results should be carefully considered[9]. For example, if race confounds results between an outcome and a risk factor, investigators need to consider what social determinants may be operating as upstream or distal determinants of health[9].

Additionally, if other social determinants of health are available in the data and are related to race, it may be more appropriate to use these variables as the primary covariates[9]. Statistical analysis must also be carefully designed to not over-adjust in the multiple variable models[9]. For example, if income confounds the same relationship as race and when both variables are included in the model the relationship needs to be evaluated to determine if over-adjustment is occurring[9]. Consultation with a statistician is advised early in the study design stage to appropriately plan for data collection of all variables to evaluate relationships of the carefully crafted study question(s).

Combining Racial Categories in Analysis

Whenever possible, do not to combine non-white races into a single group for comparisons against white people. People of Color represent diverse lived experiences and deserve distinct representation.[5] Sometimes there will not be enough statistical power to analyze data across all races separately due to low sample size (either overall or in a particular racial group). In these cases, the researcher should carefully whether race is an important variable to include in analyses. If the researcher determines race is important to include, or it is required by a funder or journal, the researcher should work with a statistician to determine which racial variables to collapse and how they should be combined, especially with regard to multi-racial people. It is important to acknowledge in the study's limitations section that combining groups for statistical power reasons may be problematic. Be careful to avoid implications in your discussion that racial minorities are a monolith without distinct cultures and experiences.[5]

Intervening Mechanisms

Although identification of intervening mechanisms linking race to health outcomes is necessary for most research due to funding or other stakeholder expectations, we discourage the use of analyses that emphasize intervening mechanisms in racial disparities without recognition of race as a fundamental cause itself due to a culture of systematic oppression. The intervening mechanisms are used to document the racial inequity, which must remain central to the discussion. Emphasis on intervening mechanisms distracts from the root causes of health disparities, which must stay central to policies that advance health equity.

Reflection Questions for the Researcher:
Have my analyses implicitly or explicitly assumed that white is the normative, standard, or default position?[7]

Highly Discouraged

Do not blindly adjustment for race as a confounder. Avoid the "subtraction method", which is controlling for standard socioeconomic and health behavior variables and assuming unexplained factors relate to race[2]. In particular, keep in mind that race does not represent underlying genetic differences[2,10] . Race is a multi-faceted *social* construct that inherently promotes advantage or disadvantage across the life course. Adjustments for income and/or education will not adequately capture this (dis)advantage of race[2].

Stratifying Results by Race

- If you have a large sample where race is a key component of the analyses, stratifying results by race would be acceptable. Examples would include studies specifically about race, such as attitudes towards adopting Black children, or the effect of racial bias on prescribing pain medication[2].
- When studies expect different outcomes for different racial groups, they may need to plan for a larger sample size in order to have enough people representing different groups from a statistical power perspective[7].

Discussing Findings in the Context of Race

Discussing Results, including Implications for Practice and Policy

- Note that a community medicine perspective requires researchers to recognize and respond to those factors in the community that shape individual and population health. Individual racial disparities should be contextualized to the broader societal and/or built environment. Whenever possible, discuss racial disparities in the context of power and privilege differentials and/or structural determinants of health.[1]
- Name variables that were not included that likely play an important role in shaping racial inequities in health, such as racial discrimination, minority stress, or citizenship status.[1]
- Identify limitations of analyses that required collapsing of race variables to preserve power, if applicable.[5]
- Involve a member or members of the affected community during interpretation of the results for their opinions on whether the results accurately and appropriately describe their community.[6]

Additional Guidance

We strongly recommend review of the following guidance on promoting racial/ethnic equity in research:

1. Andrews K., Parekh, J., Peckoo, S. (2019). How to Embed a Racial and Ethnic Equity Perspective in Research: Practical Guidance for the Research Process. Available at: https://www.childtrends.org/wp-content/uploads/2019/09/RacialEthnicEquityPerspective_ChildTrends_October2019.pdf
2. NIMHD Research Framework Adaptation. Available at: https://www.nimhd.nih.gov/about/overview/research-framework/adaptation-framework.html

This living document will continually evolve as new best practices are developed for racial inclusivity in research.

Acknowledgments

Many people and groups assisted in the development of this document from the Office for Research Development at OU-Tulsa. A social psychologist and a public health professor were the lead writers on this document. We drew from governmental agency guidelines including but not limited to the National Institute of Health, the National Science Foundation, the American Psychological Association, and the American Medical Association. Academic literature as cited also contributed greatly to the development of this document. An Inclusivity in Research committee at OU-Tulsa involving members from multiple colleges and multiple perspectives gave extensive feedback on this document and its revisions. The President's Office

for Diversity, Equity, and Inclusion also served as a resource for expert opinions on parts of this document as it evolved. Different members of the OU-Tulsa Office for Research Development and Scholarly Activity including epidemiologists also provided feedback on this document, and particularly its statistical content. In addition, professors who are national experts on implicit bias across multiple institutions advised on sections of this document. Most importantly, we greatly thank all of the people of color who provided their emic perspective in the development of these guidelines.

References

1. Link, B.G. & Phelan, J. (1995)Social conditions as fundamental causes of disease. *J Health Soc Behav,* Spec No:80-94.
2. Jones, C.P., Truman, B.I., Elam-Evans L.D., et al. (2008). Using "socially assigned race" to probe white advantages in health status. *Ethn Dis.* 8(4):496-504.
3. Berkley Political Review. The politics of the hyphen. https://bpr.berkeley.edu/2013/11/25/the-politics-of-the-hyphen/. Published 2013. Accessed.
4. American Psychological Association. "Race and Ethnic Identity." https://apastyle.apa.org/style-grammar-guidelines/bias-free-language/racial-ethnic-minorities/
5. Crenshaw, K. (1991). Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Color. Stanford Law Review, 43(6), 1241-1299.
6. Parker K, Menasce Horowitz J, Morin R, Hugo Lopez M. (2015).The many dimensions of Hispanic racial identity. In: Multiracial America: Pew Research Center.
7. Andrews, K., Parekh, J., Peckoo, S. (2019). *How to Embed a Racial and Ethnic Equity Perspective in Research: Practical Guidance for the Research Process.* Available at: https://www.childtrends.org/wp-content/uploads/2019/09/RacialEthnicEquityPerspective_ChildTrends_October2019.pdf
8. National Research Council (US) Panel on Race, Ethnicity, and Health in Later Life; Anderson, N.B., Bulatao,R.A., Cohen, B. editors. Critical Perspectives on Racial and Ethnic Differences in Health in Late Life. Washington (DC): National Academies Press (US); 2004. 8, Genetic Factors in Ethnic Disparities in Health.Available from: https://www.ncbi.nlm.nih.gov/books/NBK25532/.
9. VanderWeele & Robinson (2014). "On Causal Interpretation of Race in Regression for Confounding and Mediating Variables". *Epidemiology, 25*(4), 473-484.
10. American Association of Physical Anthropologists (2019). Executive Summary: AAPA Statement on Race and Racism. https://physanth.org/documents/199/AAPA_Race_statement_March_2019.pdf.