



HHS Public Access

Author manuscript

JAMA Pediatr. Author manuscript; available in PMC 2023 July 01.

Published in final edited form as:

JAMA Pediatr. 2022 July 01; 176(7): 631–632. doi:10.1001/jamapediatrics.2022.0525.

The Complicated Inadequacy of Race and Ethnicity Data

Barbara H. Chaiyachati, MD, PhD,

Division of General Pediatrics, Department of Pediatrics, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania; and Center for Pediatric Clinical Effectiveness, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania.

Michelle-Marie Peña, MD,

Division of Neonatology, Department of Pediatrics, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania.

Diana Montoya-Williams, MD

Division of Neonatology, Department of Pediatrics, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania; and Department of Pediatrics, Perelman School of Medicine, Philadelphia, Pennsylvania.

“Choose from the options below.” This seemingly innocuous prompt embedded in countless daily medical registrations aims to capture race and ethnicity. Yet, as mothers of children with multiple racial and ethnic identities, a prompt that could take just moments sometimes becomes a weighted pause, reminding us of the intersectional reality of our families. How do we represent the Thai and non-Hispanic White; Cuban and Vietnamese; Colombian, non-Hispanic Black, and non-Hispanic White backgrounds, roots, and cultures that our children embody? Which box or boxes do we check off?

Selecting every possibly relevant box might imply that our children are counted as their whole selves, described by each part within. But do we also try to anticipate their eventual self-determined identity that will coalesce with their socially assigned one, as both may contribute to their health status?¹ Or, we can reflex to “other,” recognizing that any collection of small boxes will never equal the sum of their inherited legacies or gathered experiences, a critical acknowledgment as we seek to disentangle the drivers of disparate health outcomes. But, if we designate our children as “other,” how will they be accurately ascribed risk in clinical practice or research?

The weight of the pause and the many different approaches to these questions are not just informed by our own lived experiences and what we understand about the world. They are questions we ponder deeply because, in addition to mothers, we are physician-scientists. We know that our children—with their specific family ancestries—are unlikely to be meaningfully represented in our own result tables. By selecting “other,” we relegate them to exclusion from many data analyses, decreasing not only their contribution to the advancement of health care systems research but also any benefit children like them might

Corresponding Author: Barbara H. Chaiyachati, MD, PhD, Division of General Pediatrics, Bgr 12, Children’s Hospital of Philadelphia, 3500 Civic Center Blvd, Philadelphia, PA 19104 (chaiyachab@chop.edu).

receive because we have obscured them into a box that is uninterpretable. By selecting their specific combinations of races and ethnicities among the options available, they are equally uninterpretable within the typically resultant multiracial category given the infinite combinations contained within.

These are not outlier considerations for just our own children. The 2020 US Census has shown that when given the opportunity, there are many of us who view our family heritage in complex ways that do not fit neatly within a handful of boxes. As of 2020, “some other race” was the second most common racial category selected, an option chosen for 49.9 million people.² The multiracial population in the US increased by 276%, from 9 million in 2010 to 33.8 million in 2020. These large shifts likely represent a combination of real population change along with the improved methodologies to select options that reflect self-perceptions. But it also means that only 66% of infants born in the US in 2020 can be immediately summarized to any single race and ethnicity when you consider both parents’ backgrounds as listed on their birth certificate.³

Many of our own research questions require the careful examination of race and ethnicity and an understanding of how societally constructed variables have been used to group people throughout centuries for political and economic purposes.⁴ We grapple with challenges of feasibility, replicability, and statistical power that arise when we seek to capture the deep granularity that we know exists within racial and ethnic categories. Increasing the number of options to better capture the compositional background of an individual human will inevitably yield smaller and smaller study populations, widening confidence intervals and under-mining associated estimates. Even using the fewest number of categories available from the US Centers for Disease Control and Prevention Wide-ranging Online Data for Epidemiologic Research (CDC WONDER) database of the entire population of infants born in the US in 2020, 448 of 729 potential parent race and ethnicity combinations had 5 or fewer infants, including 385 combined parental race and ethnicity categories with 0 births recorded.³

Historically, attempts to address such issues include various data reduction approaches. Some investigators may exclude or impute the missing data under assumptions that each reason for missing is equivalent and random. For birth data, researchers may choose to ignore paternal race and ethnicity altogether in recognition that missingness could introduce nonrandom bias into analyses.⁵ Other researchers may collapse categories to large, aggregated groups, erasing even continental specificities. This has frequently happened to non-Hispanic Black communities, where African-born and US-born individuals are grouped together, and among Hispanic individuals, where someone born in Spain may be grouped with someone born in the Caribbean or Latin America. However, it is important to consider that while these methodologic choices may be functional and even arbitrary, they may also reflect implicit and explicit biases within investigator teams or within the original data collection process itself. Because the race and ethnicity variables have evolved depending on who is creating the variables, who is making the categorical determination, and how these data are collected, these variables likely reflect sociopolitical contexts in which the work is conducted.⁴

Thoughtful considerations have been provided for the inclusion of race and ethnicity in biomedical results, and medical journals have developed guidance to support standardized reporting of race and ethnicity.^{6–8} However, it is clear that more work is needed.^{4,6,8,9} We offer 3 broadly relevant action items.

1. Increase data transparency, including the how and why and strengths and limitations of decisions to aggregate and/or disaggregate demographic categories. This should include examining sources of missing data and use of proxy measures. For instance, is race actually being used as a proxy for experiences of structural or interpersonal racism? It is possible, in fact probable, that some data assumptions are underpinned by explicit and implicit biases.
2. Advance research and statistical methods to better account for population complexities. For example, researchers may need to account for race, ethnicity, and nativity together given that health outcomes are known to differ when all these characteristics are considered.¹⁰ Researchers may also plan a priori subanalyses of the other category, accounting for the inherent power limitations. Such analyses can help assess whether association estimates found in typical demographic groups are similar to those found within other subgroups to provide exploratory directions. Relatedly, component identities reflected within the multiracial category should be considered and may require partnering with statisticians and methodologists to add to the current cadre of typical approaches.
3. Reflect equity principles within research teams via diversity, equity, and inclusion expertise or lived experience representation.⁹ Beyond making the research process more equitable, these changes may also improve effectiveness and efficiency. Studies developed by diverse teams broaden the consideration of approaches, language, or analyses that may be better aligned with the needs of the population or community being studied, expediting the goal of health improvement underpinning medical research at large.

Ultimately, as proud parents of vibrant, multiracial and multiethnic children, we experience the complicated inadequacy of race and ethnicity, and simultaneously as physician-scientists, we know the importance of data accuracy and the potential harm of small assumptions that undermine the data we build our research on. Thus, we advocate that the scientific community prioritize the work above with urgency to benefit the health of all the children that we serve and love—not only those who we fit nicely into a box.

Conflict of Interest Disclosures:

Dr Chaiyachati receives funding from the National Institutes of Health/National Institute of Mental Health (grant T32 MH019112). Dr Peña receives funding from the National Institutes of Health (grant T32HL098054-11). Dr Montoya-Williams receives funding from the National Institutes of Health (grant K23 HD102526).

Additional Contributions:

We thank Raina Merchant, MD, MSHP, University of Pennsylvania, for her valuable contributions. Dr Merchant was not compensated.

REFERENCES

1. Jones CP, Truman BI, Elam-Evans LD, et al. Using “socially assigned race” to probe white advantages in health status. *Ethn Dis*. 2008;18(4):496–504. [PubMed: 19157256]
2. Jones N, Marks R, Ramirez R, Ríos-Vargas M. 2020 census illuminates racial and ethnic composition of the country. US Census Bureau. Published August 12, 2021. Accessed October 12, 2021. <https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html>
3. US Centers for Disease Control and Prevention. About Natality, 2016–2020 expanded. Accessed November 27, 2021. <https://wonder.cdc.gov/natality-expanded-current.html>
4. Borrell LN, Elhawary JR, Fuentes-Afflick E, et al. Race and genetic ancestry in medicine: a time for reckoning with racism. *N Engl J Med*. 2021;384(5): 474–480. doi:10.1056/NEJMms2029562
5. Tan H, Wen SW, Walker M, Demissie K. Missing paternal demographics: a novel indicator for identifying high risk population of adverse pregnancy outcomes. *BMC Pregnancy Childbirth*. 2004;4(1):21. doi:10.1186/1471-2393-4-21 [PubMed: 15541183]
6. Ioannidis JPA, Powe NR, Yancy C. Recalibrating the use of race in medical research. *JAMA*. 2021;325 (7):623–624. doi:10.1001/jama.2021.0003 [PubMed: 33492329]
7. JAMA Pediatrics. Instructions for authors: reporting demographic information for study participants. Accessed January 12, 2022. <https://jamanetwork.com/journals/jamapediatrics/pages/instructions-for-authors#SecReportingRace/Ethnicity>
8. Kauh TJ. Racial equity will not be achieved without investing in data disaggregation. *Health Affairs Blog*. Published November 29, 2021. Accessed March 11, 2022. <https://www.healthaffairs.org/doi/10.1377/forefront.20211123.426054>
9. Merchant RM, Del Rio C, Boulware LE. Structural racism and scientific journals—a teachable moment. *JAMA*. 2021;326(7):607–608. doi:10.1001/jama.2021.12105 [PubMed: 34402819]
10. Egbe TI, Montoya-Williams D, Wallis K, Passarella M, Lorch SA. Risk of extreme, moderate, and late preterm birth by maternal race, ethnicity, and nativity. *J Pediatr*. 2022;240:24–30.e2. doi:10.1016/j.jpeds.2021.09.035 [PubMed: 34592259]