



# The Critical Role of Racial/Ethnic Data Disaggregation for Health Equity

Tina J. Kauh<sup>1</sup> · Jen'nan Ghazal Read<sup>2</sup>  · A. J. Scheitler<sup>3</sup>

Received: 10 December 2020 / Accepted: 16 December 2020 / Published online: 8 January 2021  
© The Author(s), under exclusive licence to Springer Nature B.V. part of Springer Nature 2021

## Abstract

Population-level health outcomes and measures of well-being are often described relative to broad racial/ethnic categories such as White or Caucasian; Black or African American; Latino or Hispanic; Asian American; Native Hawaiian and Pacific Islander; or American Indian and Alaska Native. However, the aggregation of data into these groups masks critical within-group differences and disparities, limiting the health and social services fields' abilities to target their resources where most needed. While researchers and policymakers have recognized the importance of disaggregating racial/ethnic data—and many organizations have advocated for it over the years—progress has been slow and disparate. The ongoing lack of racial/ethnic data disaggregation perpetuates existing inequities in access to much-needed resources that can ensure health and well-being. In its efforts to help build a Culture of Health and promote health equity, the Robert Wood Johnson Foundation has supported activities aimed to advance the meaningful disaggregation of racial/ethnic data—at the collection, analysis, and reporting phases. This special issue presents further evidence for the importance of disaggregation, the technical and policy challenges to creating change in practice, and the implications of improving the use of race and ethnicity data to identify and address gaps in health.

**Keywords** Health equity · Culture of health · Data disaggregation · Race/ethnicity

---

✉ Tina J. Kauh  
tkauh@rwjf.org

<sup>1</sup> Research-Evaluation-Learning Unit, Robert Wood Johnson Foundation, 50 College Road East, Princeton, NJ 08543, USA

<sup>2</sup> Department of Sociology, Global Health Institute, Duke University, Durham, NC, USA

<sup>3</sup> Center for Health Policy Research, University of California Los Angeles, Los Angeles, CA, USA

## Context of the Special Issue

Health outcomes or measures of well-being related to social determinants of health are often discussed in reference to one of five broad U.S. racial/ethnic categories: White or Caucasian; Black or African American; Latino or Hispanic; Asian American; Native Hawaiian and Pacific Islander; and American Indian or Alaska Native.<sup>1</sup> These categories are defined by the Office of Management and Budget and guide the collection and presentation of all federally collected data (OMB n.d.). Despite these broad categories, however, there is growing recognition that significant variation exists within each of these racial/ethnic groups.

Asian Americans, for example, account for 17 million people and nearly half of all refugees who arrived in the U.S. between 2000 and 2010. Use of aggregated data in health, income, and other social determinants of health perpetuate the model minority myth—namely, that all Asian Americans are healthy, affluent, and well-educated—and can obscure the very real challenges that many people within Asian American communities face. In fact, a 2017 study released by the National Bureau of Economic Research examining race and income inequality showed that Asian Americans are one of the most economically divided racial groups in the U.S. (Akee et al. 2017). Asian Americans demonstrate a bimodal distribution, being overrepresented relative to the total U.S. population in both the top and bottom 10% of the income spectrum. Conversely, the aggregation of income data among Asian Americans misleadingly suggests that they, as a singular population, are actually thriving (Kochhar and Cilluffo 2018). The diversity among Asian Americans is not unique:

- There are 562 federally recognized Indian nations in the U.S. In addition to members of these tribes differing ethnically, culturally, and linguistically, they vary in terms of whether they live on or off reservations, which influences their access to health services and other major resources.
- Forty-two million people in the U.S. self-identify as Black or African American. While most of them have lived in the United States for generations, more than three million are immigrants, mostly from different parts of Africa or the Caribbean.
- The Hispanic/Latino population makes up about 16% of the U.S. population, and about three-quarters self-identify as Mexican, Puerto Rican, or Cuban, which represent strikingly different cultures and histories.
- Individuals self-identifying as White represent more than two-thirds of the U.S. population. Though Western Europeans were once the predominant ancestry, persons of Middle Eastern, North African, and Eastern European descent make up an increasing share of this culturally diverse population.

Factors including nativity, immigrant status, language, socioeconomic status, and experiences with structural and interpersonal racism all significantly influence

---

<sup>1</sup> Although the OMB guidelines separates Asian Americans from Native Hawaiian and Pacific Islanders, many data sources often aggregate them into a single group.

health outcomes, and that nuance is lost once data are aggregated to those five or six broad categories of race and ethnicity. And that lack of systematic disaggregation—whether it be at the collection, analysis, or reporting phase—limits the health and social services fields' ability to target their resources where they are most needed, to all communities experiencing significant disparities.

The groundwork has been laid for advancing the field of racial/ethnic data disaggregation for years. In 2009, for instance, the Institute of Medicine (now the National Academy of Medicine) examined similar methodological issues and recommendations for the healthcare field (AHRQ 2018).

While researchers recognize the importance of disaggregating data and many organizations have advocated for it, progress has been slow and disparate in terms of seeing actual change in data reported to decision makers and to the public. The collection of more meaningfully disaggregated racial/ethnic data is uncommon, but the analysis and reporting of those data are even more rare. A reality exists that there are numerous methodological challenges to doing this well. For example, disaggregating data in research studies, even after more nuanced data have been collected, can result in small sample sizes, which risks generalizability and creates data privacy concerns. While efforts to advocate for policy change in data collection are critical, more also needs to be done to ensure that data are analyzed and reported in ways that do justice to the racial/ethnic diversity in the nation.

Recently, the Robert Wood Johnson Foundation (RWJF) began commissioning a series of activities to advance disaggregation practices of health data across the life cycle of data. These activities began in 2016 with reports summarizing the state of the field for data disaggregation across those broad racial/ethnic categories and convenings of experts in research and advocacy to build cross-sector networks (PolicyLink 2020). In addition, RWJF has supported advocacy efforts for policy change, research to better communicate the importance of policy change, and technical assistance and training to help the field overcome the oft-cited methodological barriers to meaningful disaggregation. All of these activities are rooted in the Foundation's efforts to achieve its vision for a Culture of Health.

A society that thrives through a Culture of Health is one that consists of policies and systems that prioritize health and put individuals' well-being at the center (RWJF 2020). It requires collaboration from a wide range of partners across diverse sectors to contribute towards a national culture change. As Braveman et al. (2017) note, "health equity surrounds and underpins [this] vision of a society in which everyone has an equal opportunity to live the healthiest life possible." Much of decision makers' efforts to promote health equity focuses on the healthcare system or the social determinants of health. However, a critical source of inequity that receives far less attention is the lack of meaningfully disaggregated data for health and healthcare outcomes as well as those key social determinants of health.

"Data inequity," or the continued lack of visibility and representation in data for some populations, takes on a particular urgency during health and public health crises, such as the recent COVID-19 pandemic. Media outlets and community advocates have demanded data to show which communities were being most negatively impacted by the deadly virus. In many cases, states are required to report disaggregated data along the five broad racial/ethnic categories, but many failed to do

so initially. Once states began to comply, those data were summarized by publicly available resources (The Atlantic 2020). However, the ways in which these statistics were reported and subsequently discussed in the media failed to acknowledge the disproportionate impact of the virus on certain populations (e.g., Native Hawaiians and Pacific Islanders).

COVID-19 has also put a dramatic spotlight on how systemic racism exacerbates the existing health disparities that exist for vulnerable populations. Data inequity should effectively be considered a form of systemic racism wherein data collection, analysis, and reporting policies, practices, and norms continue to disproportionately exclude certain communities from access to opportunity and resources. The needs of racial/ethnic subpopulations are often ignored because the data simply do not exist to inform policy or resource allocation decisions.<sup>2</sup>

One thing is clear: when data are disaggregated by detailed racial/ethnic subgroups, a more representative picture of the health and well-being of the nation emerges. Changing the ways in which we collect, analyze, and report data can advance health equity and reduce health disparities, but it requires collaboration across sectors to push for better data collection policies and, of course, adequate resources to change systems and to responsibly implement new practices. Otherwise, some of the most disadvantaged in our communities will remain invisible to decision makers, leaving their critical needs unmet.

## Contents of the Special Issue

The five articles in this special issue include a mixture of conceptual pieces, systematic reviews, and empirical investigations around the central theme of racial and ethnic data disaggregation. Each paper focuses on one of the five major racial/ethnic categories that are defined by the Office of Management and Budget's standards and used in all federally collected data efforts. Because challenges and opportunities for data disaggregation vary across the five groups, each article takes a slightly different analytical and methodological approach. Together, the articles present further evidence for the importance of disaggregation, the technical and policy challenges to creating change in practice, opportunities to overcome these challenges, and the implications of improving the use of race and ethnicity data to identify and address gaps in healthcare.

The first paper by Read, Lynch, and West focuses on data disaggregation within the non-Hispanic White population. The issue starts here because Whites are almost always, if not exclusively, treated in the aggregate and used as the reference group when measuring U.S. racial/ethnic health disparities—with little attention paid to disparities *among* White subgroups. As the paper demonstrates, this approach is problematic because the ethnic origins of Whites have changed drastically over the

---

<sup>2</sup> Data inequity does not only come in the form of race and ethnicity. The needs of populations based on other demographic characteristics (e.g., gender, sexual orientation, and disability) often similarly lack equitable representation in data to inform policy and resource allocation decisions.

past few decades, with individuals of Middle Eastern, North African, and Eastern European descent making up an increasing proportion of the population as those of Western European descent decline. The authors use nationally representative data from the American Community Survey (2008–2016) to disaggregate the non-Hispanic White population by ancestry and other major racial/ethnic groups (non-Hispanic Black, non-Hispanic Asian, and Hispanic) by common subgroupings to examine and compare heterogeneity in disability. A key distinction they make is between Whites of Western European descent and those of Eastern European, Middle Eastern/North African (MENA region), North American, and other origins (i.e., major U.S. Census ancestry categories for Whites).

The results revealed that health disparities within the White population are almost as large as disparities within other large racial/ethnic groups. In fact, when Whites were disaggregated by ancestry, mean health appeared to be more varied *among* Whites than between Whites and members of other racial/ethnic groups in many cases. Compositional changes in the ancestry of Whites, particularly declines in Whites of Western European ancestry and increases in Whites of Eastern European and Middle Eastern ancestry, contributes to this diversity. Importantly, these newer groups often fare worse than Western Europeans, suggesting the need for greater attention to diversity among Whites.

The second paper by Larimore and colleagues proposes that fully understanding health disparities experienced by Blacks is limited by the tendency to view the population as a single, homogenous group when compared with other racial/ethnic populations. They offer an argument for considering disaggregation among Blacks by examining the lived experiences of different groups combined under this category. Using national birth records data from 2013 to 2016, they examine differences in low birth weight, preterm birth, gestational hypertension, and gestational diabetes within the Black population by nativity (i.e., U.S.- or foreign-born), region of origin (e.g., Sub-Saharan Africa, the Caribbean), and current U.S. division of residence (e.g., Middle Atlantic, South Atlantic).

The study highlights important differences in the experiences of Black women of birthing age in several critical measures of health for both the mother and the infant. Specifically, although foreign-born Blacks experience a lower risk of low birth weight, preterm birth, and gestational hypertension, they have a higher risk of gestational diabetes compared to U.S.-born Blacks. The difference in outcomes among subgroups by nativity, region of origin, and region of U.S. residence offers compelling evidence that data disaggregation within the Black population provides greater insight into health and well-being than when they are treated as a monolithic group.

The third article by Alcántara and colleagues examines the extent to which research on Latina/o health disaggregates population health estimates by key social determinants, therefore providing *within*-Latina/o group comparisons. The authors conducted a systematic review of the contemporary scientific literature and searched biomedical electronic databases (Ovid MEDLINE, EMBASE, PsycINFO, JSTOR, Sociological Abstracts) for observational U.S. studies published between January 2006 and June 2016. They identified 573 full-text articles on Latina/o health, 175 of which further disaggregated the data along the following five categories of social determinants: socio-demographics, socioeconomic status, migration factors, place-based factors,

and individual characteristics. Three-fourths of the articles focused on adults, with the remaining one-fourth focused on children.

The number of mean articles published per year was 15.9, with some slight variation over the 10-year period. They argue that the seemingly low percentage may stem from limitations in research design and data collection, as well as the lack of clear guidelines or a standardized set of survey items that reflect disaggregation categories most relevant to the Latina/o community. They conclude by suggesting the need for programmatic initiatives to promote and standardize Latina/o health data disaggregation across the life course and across the research process from design, data collection, analysis, to reporting and publication.

The fourth paper by Shimkhada, Scheitler, and Ponce focuses on the challenges of disaggregating public health data for the fastest growing population group: Asian Americans, Native Hawaiians, and Pacific Islanders (AANHPI). Through a literature review and interviews with survey leaders, the authors found that the individual groups are often too small to produce reliable sample sizes for analysis in major federal and state data sets. However, when the subpopulations are parsed into an Asian American category separate from Native Hawaiians and Pacific Islanders, some significant inequities are revealed. While challenges to collection exist, the authors also found that when the availability of disaggregated data was improved, often through policy initiatives, the result was an increase in studies that were able to contribute insightful evidence on these communities. The paper speaks to both the importance of disaggregating these subgroups, and also the role of policy in driving disaggregation efforts for racial and ethnic data.

The final paper by Becker and colleagues discusses the challenges researchers face in analyzing population health data among American Indian and Alaska Native (AIAN) populations. They argue that AIAN health data capacity is a function of the granularity in collection, release, and reporting in population-based surveys. They identify strategies to improve data capacity for AIAN in federal health surveys by exploring current approaches to collecting and coding across nine population-based health surveys (seven federal surveys, the California Health Interview Survey, and the American Community Survey). The results find that commonly used racial/ethnic tabulations lead to an undercount and/or a limited understanding of the health challenges faced by AIAN populations because they rely on single race responses to AIAN and exclude respondents who also identify as Hispanic/Latino or with another race. Yet, AIANs are the second most likely racial group to identify as two or more races, second only to Native Hawaiians and Pacific Islanders. The authors discuss the implications of the limited accessibility of AIAN information in population data sets and provide recommendations for expanding the classification categories to include multiracial and multiethnic responses. Doing so would improve availability of AIAN health information and our understanding of disparities within the AIAN population.

## Dedication

We dedicate this special issue to the late Dr. James S. Jackson, a pioneer in increasing the understanding of health in Black communities. His early career helped increase representation of Blacks in sociological studies, significantly improving the understanding of the unique—and important—intersections between social determinants of health and health outcomes. His advocacy to view the Black community as many subgroups with diverse lived experiences both advanced research practice and brought about policy change to improve lives. We appreciate his vision, mentorship, and dedication to giving a voice to marginalized populations and dedicate this issue as a testament to his enduring example to which population health researchers aspire.

**Funding** Funding was supported by Robert Wood Johnson Foundation.

## References

- AHRQ. (2018). Race, ethnicity, and language data: Standardization for health care quality improvement, Rockville, MD: Agency for Healthcare Research and Quality. Retrieved November 4, 2020 from [www.ahrq.gov/research/findings/final-reports/iomracereport/index.html](http://www.ahrq.gov/research/findings/final-reports/iomracereport/index.html).
- Akee, R., Jones, M. R., & Porter, S. R. (2017). Race matters: Income shares, income inequality, and income mobility for all U.S. races. *National Bureau of Economic Research*. <https://doi.org/10.3386/w23733>.
- Braveman, P., Arkin, E., Orleans, T., Proctor, D., & Plough, A. (2017). "What Is Health Equity?": Robert Wood Johnson Foundation. Retrieved November 4, 2020 from [www.rwjf.org/en/library/research/2017/05/what-is-health-equity-.html](http://www.rwjf.org/en/library/research/2017/05/what-is-health-equity-.html).
- Kochhar, R., & Cilluffo, A. (2018). Income inequality in the U.S. is rising most rapidly among Asians, Washington, D.C.: Pew Research Center. Retrieved November 4, 2020, from [www.pewsocialtrends.org/2018/07/12/income-inequality-in-the-u-s-is-rising-most-rapidly-among-asians/](http://www.pewsocialtrends.org/2018/07/12/income-inequality-in-the-u-s-is-rising-most-rapidly-among-asians/).
- OMB. n.d., Office of Management and Budget (Omb) Standards, Bethesda, MD: National Institutes of Health. Retrieved November 4, 2020 from <https://orwh.od.nih.gov/toolkit/other-relevant-federal-policies/OMB-standards>.
- PolicyLink. (2020). Making the case for data disaggregation to advance a culture of health: PolicyLink. Retrieved November 4, 2020 from [www.policylink.org/our-work/community/health-equity/data-disaggregation](http://www.policylink.org/our-work/community/health-equity/data-disaggregation).
- RWJF. (2020). About a Culture of Health: Robert Wood Johnson Foundation. Retrieved November 4, 2020 from [www.rwjf.org/en/cultureofhealth/about.html](http://www.rwjf.org/en/cultureofhealth/about.html).
- The Atlantic. (2020). The Covid Tracking Project: The Atlantic Monthly Group. Retrieved November 4, 2020 from <https://covidtracking.com/>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.