

Accuracy in Measuring Racial Disparities during the COVID-19 Pandemic Needs Improvement

Aditi G.M. Patel,^{a,b,*} Chu J. Hsiao,^{c,d} Henrietta O. Fasanya,^{c,e} Michelle Stoffel,^{f,g} and Martha E. Lyon^h

Calculations of morbidity, mortality, and laboratory testing in the wake of the Coronavirus Disease 2019 (COVID-19) pandemic were marked by the stark absence of demographic information. Many early studies of documented cases did not include race or socioeconomic status (SES). In earlier COVID-19-related reports from the Centers for Disease Control (CDC), not only were there significant data missing regarding race and SES, the available data were likely affected by misclassification bias, which is the most common type of bias and occurs when an exposure or disease is categorized incorrectly (1, 2). In a study from the CDC of 1,482 laboratory-confirmed COVID-19-associated hospitalizations in March 2020, only 39.1% of those hospitalizations had race/ethnicity data (1). According to *The Atlantic's* COVID Racial Data Tracker, there is substantial variability among the states in reporting racial and demographic data of COVID-19 cases, ranging from self-reported race/ethnicity in 6% of cases (Texas) to self-reported race in 99% of cases and self-reported ethnicity in 92% of cases (Washington, DC) (3). Additionally,

this dashboard indicates one of the states, New York, is not reporting the race or ethnicity data of COVID-19 cases at all when this publication went to press (3). This dearth of race and SES data led to multiple calls for the accurate collection of demographic data, as well as cautionary notes on the use of race/ethnicity data.

Racial data in laboratory testing matters because COVID-19 is magnifying preexisting racial health disparities in the United States (US), and testing is the entry point to access healthcare and receive necessary local policy interventions. In April 2020, New Orleans health officials recognized that the drive-through testing strategy for COVID-19, a main source of testing states across the US, was not effective because hot spots for the virus were located in predominantly low SES, Black, and Latino neighborhoods, where many of these residents rely heavily on public transportation and lack cars (4). In response, health officials sent mobile vans to these neighborhoods to increase testing because “data is the only way we can see the virus,” says Thomas LaVeist, dean of

^aDepartment of Medicine, College of Medicine, University of Florida, Gainesville, FL, USA; ^bDepartment of Health Outcomes & Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL, USA; ^cMD-PhD Program, College of Medicine, University of Florida, Gainesville, FL, USA; ^dDepartment of Anthropology, College of Liberal Arts & Sciences, University of Florida, Gainesville, FL, USA; ^eDepartment of Pathology, College of Medicine, University of Florida, Gainesville, FL, USA; ^fDepartment of Laboratory Medicine, University of Washington, Seattle, WA, USA; ^gDepartment of Biomedical Informatics, University of Washington, Seattle, WA, USA; ^hDepartment of Laboratory Medicine, Royal University Hospital, Saskatchewan Health Authority, Saskatoon, SK, Canada.

*Address correspondence to this author at: 2199 Mowry Road (Building 2020 - Modular 1), Gainesville, FL 32611. E-mail aditi.gm.patel@ufl.edu WC: 1642.

Received July 1, 2020; accepted October 26, 2020.

DOI: 10.1093/jalm/jfaa212

© American Association for Clinical Chemistry 2020. All rights reserved.

For permissions, please email: journals.permissions@oup.com.

Tulane University's School of Public Health and Tropical Medicine and co-chair of Louisiana's COVID-19 Health Equity Task Force (4). In major urban areas, rates of positive COVID-19 infections are higher in racial minority groups, groups that simultaneously have greater preexisting disease comorbidities and less access to care (5). The drive to better characterize and quantify these disparities underlies the call for better collection of racial data related to COVID-19 healthcare (6). Better awareness of at-risk populations helps our healthcare systems to allocate resources better and to prepare for surges. In laboratory testing, contextualizing a test result to an individual, process, system, and environment enables identification of areas for improvement; thus, race data can directly impact the accuracy of measurements. Apart from COVID-19 laboratory testing validity, racial data also have a crucial impact on research and awareness of the magnitude of racial disparities in healthcare delivery. As a group of multisite collaborators across North America working to investigate race and COVID-19 testing, we faced several challenges in acquiring optimal COVID-19 data related to race.

The first challenge is the accuracy of race data within a single healthcare system. It is not uncommon for clinicians to interface with the electronic health record (EHR) before seeing a patient—and in some cases, clinicians ever see only the EHR. Any research that uses EHR data must carefully consider the validity of race data. Patients' racial identities are often assigned at intake where race is occasionally presumed from name or is sometimes missing because race is often not a "required" data element. This type of misclassification bias between a patient's self-reported and EHR race disproportionately affects minorities, including Black and Hispanic Americans, which could lead to underreporting of poorer health outcomes associated with these populations (7). Additionally, a significant and growing population in the United States identify with more than one racial or ethnic identity: 30% of self-

reported whites, 37% of Hispanics, and 41% of African Americans identified with more than one racial group (8). In a study by Lee et al., most of the "errors" in race/ethnicity data are caused by missing or "Unknown" data values (9). This suggests that for EMRs which offer multiselection options for the collection of race and ethnicity data to improve accuracy, this race/ethnicity data is still significantly less likely to be available compared to sex and insurance status (9).

Attempts to use other data elements such as language and country of origin to validate EHR race has limitations and can introduce a specific type of misclassification bias, called differential misclassification, where errors in classifying race/ethnicity data are related to errors in classifying outcome that can inflate or deflate the real effect depending on who is misclassified. For instance, if race/ethnicity data among COVID-19 cases has lower accuracy (i.e., use of proxy variables such as language and country of origin) compared to controls or uninfected persons, then differential misclassification would bias the results toward the null, masking the actual impact. Note, the different degree of accuracy between two compared groups (cases versus controls) is the key feature in differential misclassification (2).

In addition to Black and Hispanic minorities, indigenous minorities in the US, American Indian and Alaskan Natives (AI/AN), are also disproportionately impacted by COVID-19 and are also likely to be affected by inaccuracies in race/ethnicity data collection, especially because the Indian Health system is a separate entity from most healthcare systems in the US. Among 23 states with at least 70% complete race/ethnicity information, which was defined as "adequate" race/ethnicity data, the cumulative incidence of laboratory-confirmed COVID-19 among AI/AN persons was 3.5 times that among non-Hispanic white persons (10).

Testing strategies that select away from a representative sample suggest that race data may not be missing at random, but due to poor selection

(selection bias) and data inaccuracies (information bias), which is reflected even at the point of access to COVID-19 testing. Tackling the challenge of accurate documentation of race and ethnicity in EHRs needs to occur through more training of vested stakeholders, including clinicians involved in direct (hospitalists, surgeons) and indirect (pathologists, radiologists) patient care, EHR vendors, researchers devising evidence-based interventions, and administrative staff entering patient information into the EHR at intake. Current training already encourages critique of the validity of laboratory test values in developing a patient's "clinical picture" and projected health outcomes. However, equally essential is the need to encourage correction of the EHR to accurately reflect patients' racial identities, which can serve as proxies for the stresses of experiencing discrimination, microaggressions, and systematic disadvantages, structural components of life that also significantly impact patient health outcomes.

A second challenge is the ability to compare data across multiple institutions and countries. Differing policies and laws introduce selection bias. This is error introduced by the protocol for selecting subjects or from factors influencing study or testing participation, which limits both the merging of demographic data elements and the generalizability of the results. At an international level, we faced challenges collaborating with Canadian colleagues. Canada passed the *Canadian Human Rights Act* in 1977 with the specific intent of ensuring equal opportunity to individuals who belong to populations that were historically victims of discriminatory practices based on myriad characteristics (including race, ethnicity, religion, age, sex, sexual orientation, marital status, disability, or conviction for an offence that is pardoned or completed probation). The creation and implementation of this law was well-intentioned, but ultimately limited the collection of race or ethnicity data in their healthcare system (11). During a time-sensitive public health crisis such as COVID-19, such policy creates obstacles to

collaborative efforts to evaluate the effect of race on morbidity and mortality of infected persons. The simple task of describing any healthcare disparities across different populations becomes extremely challenging, masking the problems created by systemic racism, and further disenfranchising underrepresented minorities, such as the indigenous population in the province of Saskatchewan. Ultimately, we were unable to accomplish a collaboration between US- and Canada-based collaborators because we could not answer the primary research question of racial disparities in COVID-19 testing without race data. Therefore, we recommend increasing the strength and number of international multidisciplinary collaboration between governmental agencies, academic centers, and community practices, to create and align cohesive strategies of improving demographic data collection. Institutional policies and legislation should mandate collection of demographic data, including race and ethnicity across healthcare systems.

A pandemic presents a real-time challenge to gathering accurate demographic data, where rapid response and action are necessary. However, there is a trade-off between the accuracy of demographic data and the expediency of acquiring it. In the United States, the first COVID-19 cases were reported in January 2020, yet there was no systematic collection of demographic data until it became clear that racial disparities in testing were mirrored racial disparities in morbidity and mortality of those infected. A federal mandate for all US laboratories to collect demographic data was issued June 4, 2020, to take effect on August 1, 2020 (12). While the mandate may introduce additional data collection issues for laboratories, it is a major step in the right direction. A comprehensive definition of measurement, with both objective and subjective terms, aids in more rigorous study parameters and will lead to greater health equity. When health care organizations commit to *systematically collect* race/ethnicity and language—as well as *correct* its inaccuracies—it enhances our

care processes and improves patient health outcomes, not only at the individual level, but also for all communities and populations (13).

While we were able to anticipate and address some challenges in the design of our study investigating racial and socioeconomic disparities in COVID-19 testing, the issues related to lack of race

data at national and international levels were impossible to mitigate in the short-term. However, we believe discussing these challenges and suggesting practical recommendations will foster more effective interventions directed at improving access to COVID-19 testing.

Nonstandard Abbreviations: COVID-19, coronavirus disease-2019; EHR, electronic health record.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 4 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.

M. Stoffel, administrative support, provision of study material or patients.

Authors' Disclosures or Potential Conflicts of Interest: No authors declared any potential conflicts of interest.

REFERENCES

1. Garg S, Kim L, Whitaker M, O'Halloran A, Cummings C, Holstein R, et al. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed Coronavirus Disease 2019 – COVID-NET, 14 States, March 1-30, 2020. *MMWR Morb Mortal Wkly Rep* 2020; 69:458–64.
2. Aschengrau A, Seage GR. Chapter 10: Bias. In: Aschengrau A and Seage, GR, editors. *Essentials of Epidemiology in Public Health*. 3rd Ed. Jones & Bartlett Learning; 2014:283–89 pp.
3. The Atlantic COVID Tracking Project and Center for Antiracist Research. The COVID Racial Data Tracker Dashboard. <https://covidtracking.com/race/dashboard> (Accessed October 12, 2020).
4. Godoy M, Wood D. What do coronavirus racial disparities look like state by state? 2020. NPR Health. <https://health.wusf.usf.edu/post/what-do-coronavirus-racial-disparities-look-state-state#stream/0> (Accessed October 12, 2020).
5. Webb Hooper M, Napoles AM, Perez-Stable EJ. COVID-19 and racial/ethnic disparities. *JAMA* 2020;323:2466-67
6. Khalatbari S, Cumming RG, Delpierre C, Kelly-Irving M. Importance of collecting data on socioeconomic determinants from the early stage of the COVID-19 outbreak onwards. *J Epidemiol Community Health* 2020; 74: 620–23.
7. Fremont A, Lurie N. The role of racial and ethnic data collection in eliminating disparities in health care. In: Ver Ploeg M. and Perrin E, editors. *Eliminating Health Disparities: Measurement and Data Needs*. National Research Council Report. Washington DC: National Academy Press, p. 202–29.
8. Laurencin CT, McClinton A. The COVID-19 pandemic: a call to action to identify and address racial and ethnic disparities. *J Racial and Ethnic Health Disparities* 2020;7: 398–402.
9. Lee SJ, Grobe JE, Tiro JA. Assessing race and ethnicity data quality across cancer registries and EMRs in two hospitals. *J Am Med Inform Assoc*. 2016;23:627–34.
10. Hatcher SM, Agnew-Brune C, Anderson M, Zambrano LD, Rose CE, Jim MA, et al. COVID-19 among American Indian and Alaska Native persons – 23 States, January 31-July 3, 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:1166-69.
11. Grant T, Balkissoon D. How Canada's racial data gaps can be hazardous to your health. <https://www.theglobeandmail.com/canada/article-how-canadas-racial-data-gaps-can-be-hazardous-to-your-health-and/> (Accessed June 16, 2020).
12. U.S. Department of Health and Human Services. HHS Announces New Laboratory Data Reporting Guidance for COVID-19 Testing. <https://www.hhs.gov/about/news/2020/06/04/hhs-announces-new-laboratory-data-reporting-guidance-for-covid-19-testing.html> (Accessed October 12, 2020).
13. Hasnain-Wynia R, Baker DW. Obtaining data on patient race, ethnicity, and primary language in health care organizations: current challenges and proposed solutions. *Health Serv Res* 2006;41:1501–18.