



# HHS Public Access

Author manuscript

*Gastroenterology*. Author manuscript; available in PMC 2024 April 01.

Published in final edited form as:

*Gastroenterology*. 2023 March ; 164(3): 320–324. doi:10.1053/j.gastro.2023.01.008.

## Disaggregating racial and ethnic data: a step toward diversity, equity, and inclusion

Peter S. Liang<sup>1,2,3</sup>, Simona Kwon<sup>2</sup>, Ilseung Cho<sup>1</sup>, Chau Trinh-Shevrin<sup>1,2</sup>, Stella Yi<sup>2</sup>

<sup>1</sup>Department of Medicine, NYU Grossman School of Medicine

<sup>2</sup>Department of Population Health, NYU Grossman School of Medicine

<sup>3</sup>Department of Medicine, VA New York Harbor Health Care System

### Introduction

Data on race and ethnicity are crucial for identifying and addressing health disparities. Incomplete or inaccurate data constrain the ability of researchers, administrators, and policymakers to provide targeted assistance to groups who could otherwise be overlooked. Unfortunately, a large discrepancy exists between the many racial and ethnic categories individuals use to self-identify, and available categories in typical databases. In the healthcare setting, categories for race and ethnicity generally follow the Office of Management and Budget (OMB) minimum standards for reporting federal data, which have not been updated since 1997.<sup>1</sup> The OMB minimum standards for classification of race and ethnicity (Table 1) use 5 categories for race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White) and 2 categories for ethnicity (Hispanic or Latino and Not Hispanic or Latino). These broad categories combine subgroups of individuals with distinct ancestries and cultures as well as socioeconomic realities. Many individuals—especially those who identify with minoritized subgroups—do not have the option to choose the specific ethnic group they prefer. Similarly, those who are of Middle Eastern or North African descent have not been provided options but have historically chosen White as their race. In this context, individuals become anonymous members of a larger group, and potential health disparities within these broader racial and ethnic categories are concealed in the aggregate data. In this article, our objectives are to show the importance of disaggregating racial and ethnic data in gastroenterology and hepatology as well as to suggest solutions for addressing this problem at the individual, institutional, and societal levels (Table 2).

### Disaggregating Racial and Ethnic Data Reveals Health Disparities

In the following section, we present 2 of the many examples in gastroenterology and hepatology that illustrate the importance of disaggregating race and ethnicity.

**Colorectal cancer screening**—Asian and Hispanic/Latino individuals, on aggregate, have among the lowest colorectal cancer screening uptake in the United States. In the 2015 National Health Interview Survey (NHIS), 62.4% of the overall population were up-to-date with screening, compared with 52.1% of Asian and 47.4% of Hispanic/Latino persons.<sup>2</sup> Because there are 49 Asian nations and 19 nations and territories in the Americas and

Caribbean with predominantly Hispanic/Latino populations, it is unsurprising that screening uptake in Asian and Hispanic/Latino subgroups varies substantially. These nations and territories have distinct geopolitical histories and patterns of emigration, which have resulted in vastly different socioeconomic circumstances for immigrants who arrive in the United States. Disaggregated data collected by NHIS indicate that screening uptake in the Hispanic/Latino population ranged from 63.2% for Puerto Rican to 36.0% for Mexican individuals.<sup>2</sup> Similarly, data from the 2014 New York City Community Health Survey showed that although 61.7% of Asian and Pacific Islander persons were up-to-date with colonoscopy, this ranged from 70.4 % for Chinese to 45.1% for Asian Indian individuals.<sup>3</sup> These granular details highlight disparities in communities that would benefit from targeted screening campaigns, and that otherwise would have been hidden in the aggregate data.

**Screening for Hepatitis B virus**—Hepatitis B virus (HBV) is a major risk factor for cirrhosis and hepatocellular carcinoma, and the risk of HBV is higher for many individuals who are born outside of the United States. For this reason, the US Preventive Services Task Force recommends screening persons for HBV if they are either born in a country where the prevalence of HBV exceeds 2%, or are US-born persons who are unvaccinated and whose parents were born in countries where HBV prevalence exceeds 8%.<sup>4</sup> The prevalence of HBV surpasses 8% in some Asian countries (Kyrgyzstan, Laos, Mongolia, Vietnam, and Yemen), and falls short of 2% in others (eg, India, Malaysia, and Japan). Similarly, the prevalence exceeds 2% in a handful of Latin American countries (Belize, Colombia, Ecuador, Peru, and Suriname). Without detailed disaggregated data on race, ethnicity, and country of origin for recent immigrants, it would be difficult to follow these recommendations in clinical practice and impossible to measure adherence or impact across a health care system. In this case, the likely consequence of relying on broad racial and ethnic categories would be both overscreening low-risk individuals and underscreening high-risk individuals.

### **Additional Reasons for Collecting Disaggregated Racial and Ethnic Data**

In addition to identifying hidden health disparities, there are other reasons to collect disaggregated racial and ethnic data. First, nuanced demographic information provides clarity in patient-centered care. Understanding the prevalence of disease in a specific group or community is intrinsically valuable, regardless of how it compares to other groups. Second, giving individuals a chance to select a racial or ethnic category that they self-identify with promotes a culture of inclusivity and counters the sense of invisibility that many communities feel. It may serve as a momentary but meaningful acknowledgement of a small community in the larger fabric of society. Third, disaggregated data help to challenge stereotypes, such as the heterogenous Asian American population being uniformly considered the model minority, having minimal health disparities, or universally benefiting from the healthy immigrant effect.<sup>5</sup> These stereotypes perpetuate a false sense of well-being that deprioritizes research and funding for these communities.

### **Potential Solutions**

**Individual level**—Clinicians should recognize that screening and management recommendations for certain conditions may differ depending on a patient's race, ethnicity, and country of origin. More nuanced data also allow clinicians to be aware of potential

cultural sensitivities of their patients. Therefore, creating a workflow that incorporates the collection of disaggregated racial and ethnic data is a clinically valuable endeavor. If a provider uses a patient intake form or questionnaire and has control over its content, then adding additional racial and ethnic categories based on local demographics would be an efficient way to capture this data. Alternatively, updating clinical note templates to add a question about patient self-identified race and ethnicity in the Social History section will accomplish a similar goal.

Researchers should collect and report disaggregated racial and ethnic primary data whenever possible. When using secondary data, researchers should consider using data sources such as the NHIS or American Community Survey, which provide subcategories of the American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, and Hispanic/Latino classifications. In instances where aggregated data must be used—whether because of the method of collection or because of small sample sizes for reporting—use of aggregated data should be acknowledged as a study limitation. Groups within the aggregated data category should also be specified when possible.

**Institutional level**—We strongly encourage systems-level change in the electronic health record. Collecting disaggregated data is consistent with patient-centered care and upholds the principles of diversity, equity, and inclusion. Change at the level of an institution or healthcare system requires engagement of and support from administrators, and the impact will far exceed individual efforts. The prospective and retrospective data disaggregation initiatives at NYU Langone Health serve as illustrative examples. A collaboration between health equity researchers and the institutional leadership led to the introduction of multiple Asian and Hispanic/Latino subgroup options in the electronic health record for new patients. Simply providing the new categories, however, did not lead to a rapid shift in how patients were classified. From this experience, we learned that raising awareness about the update among clinical staff and patients was a crucial step in the implementation process, and training and support are necessary for a successful rollout. Additional engagement with data managers responsible for reporting to federal agencies, clinicians, and community members have been critical towards building the new data capture system.

In addition to prospective efforts, we are also exploring several methods to disaggregate existing data in the electronic health record. We have used a name list algorithm successfully to identify a large number of Arab Americans who often previously were misclassified as White, Other, or Unknown. The algorithm also doubled the size of Asian Americans in the dataset and permitted specific identification of ethnic groups (eg, Chinese, Asian Indian). We also are using a statistical method called Bayesian Improved Surname Geocoding (BISG) to predict the probability of an individual belonging to broad racial and ethnic groups based on surname and residential address.<sup>6</sup> BISG is the best algorithm available for classifying individuals with missing race and ethnicity, and it has been shown to have high predictive accuracy for Asian, Black, Hispanic/Latino, and White individuals.<sup>7,8</sup> Patient self-reported data remains the gold standard. However, evidence-based methods such as name lists and BISG can help direct quality improvement activities including translation services and educational support toward specific communities and geographic areas.

At the institutional level, it is important to recognize that racial and ethnic data categories vary by region. For instance, individuals of Dominican descent account for 23% of the Hispanic/Latino community in the New York metropolitan area but make up less than 1% of the Hispanic/Latino population in the Los Angeles and Houston metropolitan areas.<sup>9</sup> Institutions should either adopt a comprehensive list of racial and ethnic subgroups<sup>10</sup> or build one using local population demographics and preferably with feedback from the community. Regardless of how many categories are available for selection, it is also important that they can rollup or combine into the 5 race and 2 ethnicity groups in the OMB minimum standard to maintain comparability with other data sources.

### Societal level

With support from the Robert Wood Johnson Foundation, the research institute PolicyLink produced a comprehensive report on methods for collecting and analyzing data and government policies that enable data disaggregation.<sup>11</sup>

At the federal level, modifying the OMB minimum standards remains the shortest path to widespread change. Although the US Census Bureau's internal research in 2015 concluded that "it is optimal to use a dedicated" Middle Eastern or North African category,<sup>12</sup> and a 2017 proposal from the Federal Interagency Working Group for Research on Race and Ethnicity concurred, the OMB ultimately rejected these recommendations and the 2020 Census did not incorporate the proposed change. However, the Biden administration OMB recently signaled a willingness to reassess the minimum standards by reconvening an Interagency Working Group and beginning a series of listening sessions with the public.<sup>13</sup> It is worth noting that the Department of Health and Human Services, which oversees the NHIS, has adopted an expanded data standard for race and ethnicity (Table 1).<sup>14</sup> The Centers for Medicare & Medicaid Services, which operates within the Department of Health and Human Services, also has announced that new enrollment forms for Medicare Advantage and Medicare Prescription Drug (Part D) plans are required to include disaggregated racial and ethnic categories in January 2023.<sup>15</sup> The majority of federal agencies, however, continue to follow the OMB minimum standard.

Although federal reform has been hindered, states have led efforts for data disaggregation. In 2000, the Massachusetts Department of Public Health developed a new data collection form that offered additional racial and ethnic groups choices, including Cape Verdean, Haitian, and Puerto Rican.<sup>11</sup> In 2016, the Accounting for Health and Education in API Demographics Act became law in California and required the State Department of Public Health to collect more detailed data on Asian American, Native Hawaiian, and Pacific Islander groups, including Bangladeshi, Indonesian, Taiwanese, Fijian, and Tongan.<sup>16</sup> In 2021, New York passed a law that required every state agency that collected racial and ethnic data to include options for at least 20 Asian and 6 Native Hawaiian and Pacific Islander groups (Table 1).<sup>17</sup>

### Concerns about Disaggregation

Critics of disaggregation argue the following: (1) these distinctions are a manifestation of identity politics that fragment US society, (2) disaggregation weakens the political

strength of a larger umbrella group, and (3) smaller populations may be more susceptible to privacy concerns.<sup>11,18</sup> In addition, smaller groups may make data analysis and interpretation more challenging. However, each legislative success described earlier required years of persistent advocacy from a coalition of community-based organizations while contending with opposing viewpoints. For groups that historically have been subjected to surveillance and monitoring, working with community organizations to explain the rationale for disaggregation and underscore a commitment to privacy is crucial. A number of strategies for suppressing data that do not meet statistical reliability, data quality, or confidentiality criteria have been developed.<sup>19</sup> Smaller groups that have an insufficient sample size also can be combined in an aggregate category for analysis.

## Conclusion

Collecting and reporting disaggregated racial and ethnic data is an important, practical step toward building a more diverse, equitable, and inclusive society and to improve health care quality. Going beyond the broad demographic categories of the OMB minimum standard allows us to unmask and address health disparities in our patients and bring more resources and attention to smaller communities. From modifying how we record patient or participant data to advocating for policy changes at the state and national levels, there are numerous ways to contribute to and advance health equity across our health systems.

## References

1. Office of Management and Budget. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. 1997. Available at: [https://obamawhitehouse.archives.gov/omb/fedreg\\_1997standards](https://obamawhitehouse.archives.gov/omb/fedreg_1997standards) [Accessed October 20, 2022].
2. White A, Thompson TD, White MC, et al. Cancer Screening Test Use — United States, 2015. *MMWR Morb Mortal Wkly Rep* 2017;66:201–206. [PubMed: 28253225]
3. Rastogi N, Xia Y, Inadomi JM, et al. Disparities in colorectal cancer screening in New York City: An analysis of the 2014 NYC Community Health Survey. *Cancer Med* 2019;8:2572–2579. [PubMed: 30843666]
4. US Preventive Services Task Force, Krist AH, Davidson KW, et al. Screening for Hepatitis B Virus Infection in Adolescents and Adults: US Preventive Services Task Force Recommendation Statement. *JAMA* 2020;324:2415. [PubMed: 33320230]
5. Yi SS, Kwon SC, Suss R, et al. The Mutually Reinforcing Cycle Of Poor Data Quality And Racialized Stereotypes That Shapes Asian American Health: Study examines poor data quality and racialized stereotypes that shape Asian American health. *Health Aff (Millwood)* 2022;41:296–303. [PubMed: 35130076]
6. Elliott MN, Morrison PA, Fremont A, et al. Using the Census Bureau’s surname list to improve estimates of race/ethnicity and associated disparities. *Health Serv Outcomes Res Methodol* 2009;9:69–83.
7. Yee K, Hoopes M, Giebultowicz S, et al. Implications of missingness in self-reported data for estimating racial and ethnic disparities in Medicaid quality measures. *Health Serv Res* 2022;57:1370–1378. [PubMed: 35802064]
8. Sartin EB, Metzger KB, Pfeiffer MR, et al. Facilitating research on racial and ethnic disparities and inequities in transportation: Application and evaluation of the Bayesian Improved Surname Geocoding (BISG) algorithm. *Traffic Inj Prev* 2021;22:S32–S37. [PubMed: 34402327]
9. Noe-Bustamante L. Key facts about U.S. Hispanics and their diverse heritage. 2019. Available at: <https://www.pewresearch.org/fact-tank/2019/09/16/key-facts-about-u-s-hispanics/> [Accessed November 1, 2022].

10. Agency for Healthcare Research and Quality. Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement. 2018. Available at: <https://www.ahrq.gov/research/findings/final-reports/iomracereport/reldataaptabe1.html> [Accessed November 2, 2022].
11. Rubin V, Ngo D, Ross A, et al. Counting a Diverse Nation: Disaggregating Data on Race and Ethnicity to Advance a Culture of Health. Available at: [https://www.policylink.org/sites/default/files/Counting\\_a\\_Diverse\\_Nation\\_08\\_15\\_18.pdf](https://www.policylink.org/sites/default/files/Counting_a_Diverse_Nation_08_15_18.pdf) [Accessed October 18, 2022].
12. US Census Bureau. 2015 National Content Test Race and Ethnicity Analysis Report. 2017. Available at: <https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2015nct-race-ethnicity-analysis.pdf> [Accessed November 2, 2022].
13. Orvis K OMB Launches New Public Listening Sessions on Federal Race and Ethnicity Standards Revision. 2022. Available at: <https://www.whitehouse.gov/omb/briefing-room/2022/08/30/omb-launches-new-public-listening-sessions-on-federal-race-and-ethnicity-standards-revision/> [Accessed November 2, 2022].
14. US Department of Health and Human Services. Explanation of Data Standards for Race, Ethnicity, Sex, Primary Language, and Disability. 2021. Available at: <https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=3&lvlid=54#:~:text=The%20OMB%20minimum%20categories%20for,and%20Not%20Hispanic%20or%20Latino.> [Accessed October 20, 2022].
15. Centers for Medicare & Medicaid Services. Model Individual Enrollment Request Form to Enroll in a Medicare Advantage Plan (MA) or a Medicare Prescription Drug Plan (Part D), and Advance Announcement of January 2023 Software Release - Addition of Race and Ethnicity Data Fields on Enrollment Transactions. 2022. Available at: <https://www.cms.gov/files/document/hpms-announcement-memo-race-and-ethnicity.pdf> [Accessed November 29, 2022].
16. State of California. AB-1726. 2016. Available at: [http://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill\\_id=201520160AB1726](http://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201520160AB1726) ( [Accessed November 2, 2022].
17. The New York State Senate. Senate Bill S6639A. 2021. Available at: <https://www.nysenate.gov/legislation/bills/2021/S6639> [Accessed November 2, 2022].
18. Kader F, Doan LN, Lee M, et al. Disaggregating Race/Ethnicity Data Categories: Criticisms, Dangers, And Opposing Viewpoints. 2022. Available at: <http://www.healthaffairs.org/doi/10.1377/forefront.20220323.555023/full/> [Accessed October 12, 2022].
19. Klein RJ, Proctor SE, Boudreault MA, et al. Healthy People 2010 criteria for data suppression. Healthy People 2010 Stat Notes Cent Dis Control Prev Cent Health Stat 2002:1–12.

**Table 1.****OMB Racial and Ethnic Categories and Alternative Disaggregated Categories**

<b>OMB</b>	<b>HHS</b>	<b>New York State<sup>a</sup></b>
American Indian or Alaska Native		
Asian	7 subgroups: Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, and Other Asian	20 subgroups: 7 HHS subgroups, Laotian, Cambodian, Bangladeshi, Hmong, Indonesian, Malaysian, Pakistani, Sri Lankan, Taiwanese, Nepalese, Burmese, Tibetan, and Thai
Black or African American		
Hispanic or Latino	4 subgroups: Mexican/Mexican American/Chicano/a, Puerto Rican, Cuban, and Another Hispanic/Latino/Spanish origin	
Native Hawaiian or Other Pacific Islander	4 subgroups: Native Hawaiian, Guamanian or Chamorro, Samoan, and Other Pacific Islander	6 subgroups: 4 HHS subgroups, Fijian, and Tongan
White		

HHS, Department of Health and Human Services; OMB, Office of Management and Budget;

<sup>a</sup>These categories are required by NY S.6639-A/A.6896-A, which specifically addressed the Asian, Native Hawaiian, and Pacific Islander populations.

**Table 2.**

## Potential Solutions for Disaggregating Racial and Ethnic Data

Level of Action	Potential Solutions	Impact
Individual	Clinician: use disaggregated categories for intake forms or in clinical notes Researcher: Collect disaggregated primary data, use existing disaggregated datasets, acknowledge limitation of aggregated data, and specify component groups	+ <sup>a</sup>
Institutional	Disaggregate categories in electronic health record (using a comprehensive list or based on local demographics), use evidence-based algorithms to reclassify existing data	++ <sup>a</sup>
Societal	Advocate for policy change on a local, state, or federal level by working with community-based or national organizations or by participating in Office of Management and Budget listening session with the public	+++ <sup>a</sup>

<sup>a</sup>These are qualitative markers of impact, with + representing the least impact and +++ representing the most impact.