

Race and Ethnicity: What Graduate Medical Education Programs Should Know About the Physician Data Initiative's Data Collection and Reporting Standard

Pilar Ortega , MD, MGM

Renee Bauschke, MA, PMP

William McDade , MD, PhD

Background

The race and ethnicity of the physician workforce has been shown to impact health care outcomes for racially and ethnically minoritized populations, including enhanced access to care,¹ improved therapeutic relationships,² reduced health care expenditures,² and increased scholarly activity and clinical trial participation.³ Physician race and ethnicity have implications beyond the individual patient-physician relationship; a recent study shows a mortality benefit for Black persons in communities with greater representation of Black primary care physicians.⁴

Workforce race and ethnicity data can also be used to ensure nondiscrimination in educational and career opportunities. Nondiscrimination by race in employment is mandated by Title VII of the 1964 Civil Rights Act⁵ and enforced by the Equal Employment Opportunity Commission.⁶ Institutions may use race and ethnicity data to track whether the opportunities and benefits they offer are being utilized equitably by employees, such as resident and fellow physicians and faculty members.⁷ Data are also important in identifying patterns such as whether discrimination is more commonly being reported among certain groups.^{8,9}

Collecting data about physician race and ethnicity is necessary for evaluating the impact of these variables on medical education and health care outcomes. Yet, collecting, tracking, and reporting these data are not without challenges. As individuals progress throughout the continuum of medical education—from applying to medical school, through residency and fellowship, and throughout their careers as physicians—such data

has been historically collected and reported in varied ways. These inconsistencies make it difficult to compare the demographics of aspiring, training, and practicing physicians and to explore trends over time and across organizations. Moreover, race and ethnicity data are complex, with definitions and categorizations expected to evolve and be potentially impacted by the sociopolitical climate.¹⁰ Developing a shared standard on race and ethnicity data is a strategy to enable organizations to conduct more accurate comparative workforce research, track progress, and evaluate impact.¹¹⁻¹³

To this end, the Association of American Medical Colleges (AAMC), the Accreditation Council for Graduate Medical Education (ACGME), and the American Medical Association (AMA) created a collaborative working group known as the Physician Data Initiative (PDI) with the aim of developing a shared standard for collecting and reporting race and ethnicity data across the 3 organizations. In this article, we provide insights into the PDI's development of the race and ethnicity standard, with a focus on the ACGME's perspective. We aim to enhance the graduate medical education (GME) community's understanding of the rationale for the new race and ethnicity standard, facilitate other organizations' alignment of their own data collection processes with the new standard, and encourage future collaborative research.

Developing the Race and Ethnicity Standard

Representatives from each organization with expertise in race and ethnicity categories, policy, and data convened monthly starting in 2021. First, the group compared each organization's existing approach to collecting and reporting race and ethnicity data and evaluated multiple sources. Data sources included each organization's previously collected responses on race and ethnicity, their data collection policies and procedures, as well as guidelines from the US Office of Management and Budget (OMB) Statistical Policy

DOI: <http://dx.doi.org/10.4300/JGME-D-25-00376.1>

Editor's Note: The ACGME News and Views section of JGME includes reports, initiatives, and perspectives from the ACGME and its review committees. This article was not reviewed through the formal JGME peer review process. The decision to publish this article was made by the ACGME.

Directive No. 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity.¹⁴ Next, the PDI proposed and iteratively revised a draft standard until consensus was reached.

Race and Ethnicity in Data Collection

The AAMC, ACGME, and AMA differed in some aspects as to how they each had previously approached data collection on race and ethnicity. For example, in academic year 2013-2014, the AAMC began using a combined race and ethnicity question allowing respondents to select one or all that apply.¹⁵ This differed from ACGME's question, which permitted a single selection.

Organizations also differed in data collection practices, such as at what point in a physician's career data are gathered and who is the respondent. For example, given that the ACGME oversees the accreditation of US residency and fellowship programs, the ACGME's primary source of data collection is at the residency program level; hence, program directors and coordinators input substantial portions of data.¹⁶ By contrast, the AAMC collects self-reported data from applicants through instruments such as the Electronic Residency Application Service.¹⁷ The AMA's Physician Professional Data draws from multiple sources.¹⁸ Self-reported race and ethnicity data are considered to be most reliable,¹⁹ but may not be feasible to gather by all organizations. The PDI's discussions centered around arriving at a consensus regarding race and ethnicity categories as well as increasing the transparency of how data are gathered to improve comparability across organizations.

Race and Ethnicity in Data Reporting

Issues that affect data reporting—how organizations make their data available—include how groups are aggregated and how individuals are categorized if they select more than one race or ethnicity. Historically, the ACGME has allowed for only a single race and ethnicity category selection as a second question following a separate Hispanic ethnicity question; however, in 2019, the ACGME began augmenting resident physician demographic data provided by the program in the resident roster with corresponding race and ethnicity data provided by the AAMC through a data sharing agreement. The augmentation resulted in a reduction of unknown identity observed in the resident roster of 20 to 25% to only 4% after being paired with AAMC data. The AAMC collection instrument is self-reported and allows respondents to select more than one category in a single race and ethnicity question. Since the AAMC reports multiple races as an identity category, the augmentation arrangement led to the ACGME's initiation of a Multiple Race

reporting category for any individual selecting more than one race or ethnicity. It was noted, however, that this approach hid many Hispanic individuals due to the substantial number of Hispanic people who selected more than one race or ethnicity and were therefore reported under Multiple Race. To address this, starting with the 2021-2022 academic year, the ACGME changed its data reporting to align with the Integrated Postsecondary Education Data System (IPEDS).²⁰ Using IPEDS guidelines,²¹ individuals selecting Hispanic and any other race or ethnicity are reported as Hispanic. Individuals selecting any other 2 or more race or ethnicity categories are reported as Multiple Race or Ethnicity. By comparison, the AAMC reports Hispanic or Latino similarly to racial categories, which consists of individuals who have selected only Hispanic or Latino and no other race or ethnicity. Summing all Hispanic or Latino categories the AAMC's *alone and in combination* reporting of Hispanic people enables reconciliation with the number of Hispanic people reported by the ACGME and AMA.²²

PDI panelists also discussed the need to balance between the desire to disaggregate race and ethnicity categories and the importance of making categories meaningful.²³ The benefit of greater disaggregation is that it allows individuals to self-identify in the way that most closely describes them. This permits subgroup analysis because the overarching categories often mask large intergroup differences. The downside of this approach is that it may result in some race and ethnicity groups being too small to report without risking identification of individuals. Additionally, the lack of numerical power can limit the quantitative research methodologies that can be applied.

Alignment With US OMB Guidelines

Updates from the OMB informed the PDI's decision-making regarding some aspects of the race and ethnicity standard. For example, the OMB's recent Statistical Policy Directive-15 (SPD-15) update combined race and ethnicity in a single question,¹⁴ and the PDI also followed this combined approach. The SPD-15 modified the Native Hawaiian or Other Pacific Islander category to remove "Other" from in front of Pacific Islander¹⁴; the same wording was used by the PDI. Also, the category of Middle Eastern North African (MENA) was newly added by both the OMB and the PDI.

One area in which the PDI differed from the SPD-15 was in the category "Some other race or ethnicity." Unlike the OMB, the PDI chose to maintain this write-in category to give individuals the option to self-describe their race in their own words. This category is meant to be used to inform future changes to the standard if patterns emerge regarding new race

TABLE

Physician Data Initiative Race and Ethnicity Data Collection Standard—Long Version

How do you self-identify? Please check as many categories and subcategories that may apply:	
Category	Subcategories
American Indian or Alaska Native	Tribal affiliation: [write-in]
Asian	Bangladeshi
	Cambodian
	Chinese
	Filipino
	Indian
	Indonesian
	Japanese
	Korean
	Laotian
	Pakistani
	Taiwanese
	Vietnamese
	Some other Asian: [write-in]
Black or African American	African
	African American
	Afro-Caribbean
	Ethiopian
	Haitian
	Jamaican
	Nigerian
	Somali
	Some other Black or African American: [write-in]
Hispanic or Latino	Argentinean
	Colombian
	Cuban
	Dominican
	Mexican or Mexican American
	Peruvian
	Puerto Rican
	Salvadoran
	Some other Hispanic or Latino: [write-in]
Middle Eastern or North African	Arab
	Egyptian
	Iranian
	Israeli
	Lebanese
	Moroccan
	Palestinian
	Syrian
	Some other Middle Eastern or North African: [write-in]

TABLE

Physician Data Initiative Race and Ethnicity Data Collection Standard—Long Version (continued)

How do you self-identify? Please check as many categories and subcategories that may apply:	
Category	Subcategories
Native Hawaiian or Pacific Islander	Chamorro
	Fijian
	Marshallese
	Native Hawaiian
	Samoan
	Tongan
	Some other Native Hawaiian or Pacific Islander: [write-in]
White	English
	French
	German
	Irish
	Italian
	Polish
	Some other White: [write-in]
Some other race or ethnicity: [write-in]	

Note: This table was prepared using Physician Data Initiative Standards developed by the Accreditation Council for Graduate Medical Education, American Medical Association, and Association of American Medical Colleges.

or ethnicity categories that should be explicitly listed as answer options or regrouped.

Finalized Race and Ethnicity Standard

In November 2024, the PDI released a race and ethnicity data collection and reporting standard,²⁴ including a shortened version that includes race and ethnicity categories and an expanded version that includes both categories and their associated subcategories (TABLE). The PDI also developed a list of guiding principles, including the voluntary nature of the questions, the role of write-in responses, the importance of alphabetizing response options, and guidance on data reporting.²⁵ The final standard's comments on data reporting guide institutions to clearly identify their preferred strategy for reporting. Data reporting should be sufficiently detailed such that it is clear to the user how and why groupings were created.

Implications for GME and Future Work

The race and ethnicity data collection and reporting standard represents an important advancement with the potential to enhance our understanding of the applicant pool and workforce of physicians. There are several key implications for GME. First, we expect to have an improved ability to conduct collaborative

research across the continuum of medical education. If the same categories of race and ethnicity data are collected, then it will be possible to more easily track individuals as they progress through their medical training and career. Fewer discrepancies in race and ethnicity data across organizations will also result in improved accuracy in understanding representation in GME.

Second, the shared standard presents a model that could be used as guidance for collection of race and ethnicity data at individual GME programs and sponsoring institutions. This increases opportunities for meaningful multisite research as well as the possibility of tracking learners' race and ethnicity self-reporting over time despite changing institutional affiliation.

Third, by enhancing the collection of subcategories of race and ethnicity, the proposed standard may enable more individuals to be counted using race and ethnicity descriptors that are not only statistically but also personally meaningful. Finding one's identity on a questionnaire's response choices can serve as an important signal of belonging, particularly for groups that have been historically marginalized.²⁶ The AAMC, ACGME, and AMA are medical professional organizations at the forefront of education and practice and may be viewed as gatekeepers to the profession. It is important for all individuals to feel seen and have the opportunity to share their identity, if they wish to do so.

The process of developing a shared race and ethnicity standard has limitations. First, this work is, by definition, dynamic. This means that standards will need to be modified over time as social constructs related to race and ethnicity also evolve. The term Hispanic or Latino, for example, aligns with the US Census and the OMB. However, some medical education leaders who identify with this group favor more inclusive terminology, such as Hispanic/Latinx²⁷ or Latino, Latina, Latinx, Latine, Hispanic or of Spanish Origin+ (LHS+) with the plus signifying the varied additional descriptors (eg, Chicano/a, Boricua) or nationalities (eg, Cuban, Mexican) with which individuals in this group may identify.²⁸

Individuals' willingness to disclose race and ethnicity may also change over time. For example, some may feel that disclosing their race or ethnicity makes them vulnerable to discrimination or disadvantages them in some way. Finally, while the PDI made efforts to consult with experts in multiple race and ethnicity categories to develop the current standards, we recognize that a standard is unlikely to satisfy all perspectives. In particular, some groupings may be small in number but important to make visible, such as Indigenous communities worldwide whose identities have been largely erased due to centuries of colonialism.^{29,30} Individuals who identify with these

communities may not be able to document a tribal affiliation due to the substantive erasure of their history, limiting how they can report their identity in the currently available race and ethnicity standard. Qualitative research may be well-suited to explore the perspectives of small communities to inform the race and ethnicity categories that would best allow such populations to accurately self-identify.

Next steps include implementing and evaluating the shared standard across the 3 institutions—AAMC, ACGME, and AMA—and establishing a periodic process for updates longitudinally. Future work should evaluate the effectiveness of the standard in medical education and health care settings.

References

1. Marrast LM, Zallman L, Woolhandler S, Bor DH, McCormick D. Minority physicians' role in the care of underserved patients: diversifying the physician workforce may be key in addressing health disparities. *JAMA Intern Med.* 2014;174(2):289-291. doi:10.1001/jamainternmed.2013.12756
2. Jetty A, Jabbarpour Y, Pollack J, Huerto R, Woo S, Petterson S. Patient-physician racial concordance associated with improved healthcare use and lower healthcare expenditures in minority populations. *J Racial Ethn Health Disparities.* 2022;9(1):68-81. doi:10.1007/s40615-020-00930-4
3. Alsan M, Campbell RA, Leister L, Ojo A. Investigator racial diversity and clinical trial participation. *National Bureau of Economic Research.* Published October 2, 2023. Accessed March 25, 2025. doi:10.3386/w31732
4. Snyder JE, Upton RD, Hassett TC, Lee H, Nouri Z, Dill M. Black representation in the primary care physician workforce and its association with population life expectancy and mortality rates in the US. *JAMA Netw Open.* 2023;6(4):e236687. doi:10.1001/jamanetworkopen.2023.6687
5. Title VII of the Civil Rights Act of 1964. Accessed January 20, 2025. <https://www.eeoc.gov/statutes/title-vii-civil-rights-act-1964>
6. US Equal Employment Opportunity Commission. Accessed January 20, 2025. <https://www.eeoc.gov/>
7. Strmic-Pawl HV, Jackson BA, Garner S. Race counts: racial and ethnic data on the US Census and the implications for tracking inequality. *Soc Race Ethn.* 2018;4(1):1-3. doi:10.1177/2332649217742869
8. Hu YY, Ellis RJ, Hewitt DB, et al. Discrimination, abuse, harassment, and burnout in surgical residency training. *N Engl J Med.* 2019;381(18):1741-1752. doi:10.1056/NEJMsa1903759
9. US Equal Employment Opportunity Commission. Race/color discrimination. Accessed January 20, 2025. <https://www.eeoc.gov/racecolor-discrimination>

10. Ford ME, Kelly PA. Conceptualizing and categorizing race and ethnicity in health services research. *Health Serv Res.* 2005;40(5 Pt 2):1658-1675. doi:10.1111/j.1475-6773.2005.00449.x
11. Chakkalakal RJ, Green JC, Krumholz HM, Nallamothu BK. Standardized data collection practices and the racial/ethnic distribution of hospitalized patients. *Med Care.* 2015;53(8):666-672. doi:10.1097/MLR.0000000000000392
12. Hernandez SE, Sylling PW, Mor MK, et al. Developing an algorithm for combining race and ethnicity data sources in the veterans health administration. *Mil Med.* 2020;185(3-4):e495-e500. doi:10.1093/milmed/usz322
13. Jia H, Zheng YE, Cowper DC, et al. Race/ethnicity: who is counting what? *J Rehabil Res Dev.* 2006;43(4):475-484. doi:10.1682/jrrd.2005.05.0086
14. Federal Register. Revisions to OMB's Statistical Policy Directive No. 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity. Accessed March 25, 2025. <https://www.federalregister.gov/documents/2024/03/29/2024-06469/revisions-to-ombs-statistical-policy-directive-no-15-standards-for-maintaining-collecting-and-presenting-federal-data-on-race-and-ethnicity>
15. Association of American Medical Colleges. FACTS Glossary. Accessed January 20, 2025. <https://www.aamc.org/data-reports/students-residents/data/facts-glossary>
16. Accreditation Council for Graduate Medical Education. About the ACGME. Accessed January 20, 2025. <https://www.acgme.org/about/overview/>
17. Association of American Medical Colleges. Applying to residencies with the ERAS system. Accessed January 20, 2025. <https://students-residents.aamc.org/applying-residencies-eras/applying-residencies-eras-system>
18. American Medical Association. Physician Professional Data. Accessed January 20, 2025. <https://www.ama-assn.org/about/physician-professional-data/ama-physician-professional-data>
19. Atekruse SF, Cosgrove C, Cronin K, Yu M. Comparing cancer registry abstracted and self-reported data on race and ethnicity. *J Registry Manag.* 2017;44(1):30-33.
20. Accreditation Council for Graduate Medical Education. ACGME releases 2021-2022 statistics on graduate medical education programs and resident physicians. Accessed January 20, 2025. <https://www.acgme.org/newsroom/2022/10/acgme-releases-2021-2022-statistics-on-graduate-medical-education-programs-and-resident-physicians/>
21. National Center for Education Statistics. Collecting race and ethnicity data from students and staff using the new categories. Accessed January 20, 2025. <https://nces.ed.gov/ipeds/report-your-data/race-ethnicity-collecting-data-for-reporting-purposes>
22. Andrews JS, Mathews C. US graduate medical education, 2023-2024. *JAMA.* 2024;332(24):2127-2153. doi:10.1001/jama.2024.21756
23. Kader F, Chebli P. Disaggregation of race and ethnicity group data: research-to-practice issues in clinical environments. *JAMA.* 2022;328(14):1395-1396. doi:10.1001/jama.2022.17194
24. Accreditation Council for Graduate Medical Education. Physician Data Initiative. Accessed January 20, 2025. <https://www.acgme.org/physiciandatainitiative/>
25. Accreditation Council for Graduate Medical Education. Physician Data Initiative, Race and Ethnicity Data Collection and Reporting Standard. Accessed January 20, 2025. <https://www.acgme.org/globalassets/pdfs/pdi/physiciandatainitiativraceandethnicitystandardfinal.pdf>
26. Anjorin O, Busari JO. Unpacking the social constructs of discrimination, othering, and belonging in medical schools. *Teach Learn Med.* 2024;36(5):660-668. doi:10.1080/10401334.2023.2230211
27. Ortega P, Martínez G, Alemán MA, Zapién-Hidalgo A, Shin TM. Recognizing and dismantling raciolinguistic hierarchies in Latinx health. *AMA J Ethics.* 2022;24(4):e296-e304. doi:10.1001/amajethics.2022.296
28. Sánchez JP, Rodriguez D. *Latino, Hispanic, or of Spanish Origin+ Identified Student Leaders in Medicine: Recognizing More Than 50 years of Presence, Activism, and Leadership.* Springer Nature; 2024.
29. Weaver HN. Indigenous identity: what is it, and who really has it? *Am Indian Quart.* 2001;25(2):240-255.
30. Lock MJ, McMillan F, Warne D, et al. Indigenous Cultural Identity of Research Authors Standard: research and reconciliation with Indigenous Peoples in rural health journals. *Rural Remote Health.* 2022;22(3):7646. doi:10.22605/RRH7646



Pilar Ortega, MD, MGM, is a Clinical Associate Professor, Departments of Medical Education and Emergency Medicine, University of Illinois College of Medicine, and at the time of submission was Vice President, Diversity, Equity, and Inclusion, Accreditation Council for Graduate Medical Education (ACGME), Chicago, Illinois, USA; **Renee Bauschke, MA, PMP**, is Manager, Data Analytics, Quality, and Reporting, Information Services, ACGME, Chicago, Illinois, USA; and **William McDade, MD, PhD**, is Chief Diversity, Equity, and Inclusion Officer, ACGME, Chicago, Illinois, USA.

The authors would like to thank all participants of the Physician Data Initiative and expert consultants for their contributions.

Corresponding author: William McDade, MD, PhD, Accreditation Council for Graduate Medical Education, Chicago, Illinois, USA, wmcdade@acgme.org