

RESEARCH ARTICLE

Standardization and accuracy of race and ethnicity data: Equity implications for medical AI

Alexandra Tsalidis¹, Lakshmi Bharadwaj^{2,3}, Francis X. Shen^{6,4,5,6,7}*

1 Future of Life Institute, Brussels, Belgium, 2 Laboratory of Biochemical Pharmacology, Emory University School of Medicine, Atlanta, Georgia, United States of America, 3 Department of Pediatrics, Emory University School of Medicine, Atlanta, Georgia, United States of America, 4 University of Minnesota Law School, Minneapolis, Minnesota, United States of America, 5 Massachusetts General Hospital Center for Law, Brain & Behavior, Boston, Massachusetts, United States of America, 6 Department of Neurosurgery, Dana Foundation Neurotech Justice Accelerator at Mass General Brigham, Boston, Massachusetts, United States of America, 7 Center for Bioethics, Harvard Medical School, Boston, Massachusetts, United States of America

* fxshen@umn.edu

Abstract

The rapid integration of artificial intelligence (AI) into healthcare has raised many concerns about race bias in AI models. Yet, overlooked in this dialogue is the lack of quality control for the accuracy of patient race and ethnicity (r/e) data in electronic health records (EHR). This article critically examines the factors driving inaccurate and unrepresentative r/e datasets. These include conceptual uncertainties about how to categorize races and ethnicity, shortcomings in data collection practices, EHR standards, and the misclassification of patients' race or ethnicity. To address these challenges, we propose a two-pronged action plan. First, we present a set of best practices for healthcare systems and medical AI researchers to improve r/e data accuracy. Second, we call for developers of medical AI models to transparently warrant the quality of their r/e data. Given the ethical and scientific imperatives of ensuring high-quality r/e data in AI-driven healthcare, we argue that these steps should be taken immediately.

Author summary

Healthcare systems are increasingly using artificial intelligence (AI) to improve clinical care in various settings such as hospitals and patient care facilities. In this paper, we discuss how these AI systems may be trained using inaccurate and incomplete patient race and ethnicity data. We identify several key issues underlying this data quality problem: the conceptual challenges in defining race and ethnicity categories, inconsistent data collection practices across health-care facilities, and frequent errors in classifying patients. These problems create



Citation: Tsalidis A, Bharadwaj L, Shen F (2025) Standardization and accuracy of race and ethnicity data: Equity implications for medical AI. PLOS Digit Health 4(5): e0000807. https://doi.org/10.1371/journal.pdig.0000807

Editor: Gloria Hyunjung Kwak, Emory University, UNITED STATES OF AMERICA

Received: August 19, 2024

Accepted: February 27, 2025

Published: May 29, 2025

Copyright: © 2025 Tsalidis et al. This is an open access article distributed under the terms of the <u>Creative Commons Attribution License</u>, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All data are in the manuscript and/or supporting information files.

Funding: This work was supported by the Office of the Director, National Institutes of Health Bridge2AI Program under OT award number 10T2OD032701, Patient-Focused Collaborative Hospital Repository Uniting Standards (CHoRUS) for Equitable AI (to FS),



and by the National Institute of Mental Health of the National Institutes of Health under Award Number R01MH134144, Improving Recruitment, Engagement, and Access for Community Health Equity for BRAIN Next-Generation Human Neuroimaging Research and Beyond (REACH for BRAIN, to FS). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

unreliable training data that undermines efforts to avoid and correct biases within these medical AI tools. To address these challenges, we propose two practical solutions. First, hospitals should adopt best practices for collecting race and ethnicity information, including patient self-reporting, staff training, and transparent processes. Second, developers of medical AI should be required to disclose the quality and sources of the demographic data used to train their models. Our work emphasizes that discussions about fairness in medical AI must include attention to the quality of race and ethnicity data. As these technologies become more widespread in healthcare, ensuring they work effectively for all patients requires addressing these fundamental data issues.

I. Introduction

Artificial intelligence (AI) is being rapidly deployed in many healthcare contexts, with many of these systems relying on individual-level patient data from electronic health records (EHR). There are widespread concerns about potential race bias in such heath care AI models, including critiques raised by many researchers [1–3], the American Civil Liberties Union (ACLU) [4], and the World Health Organization [5]. The E.U. AI Act mandates that the datasets used to train, validate, and test high-risk AI systems be as representative, error-free, and complete as possible (Art. 10.3).

Yet, overlooked in this evolving dialogue is that without accurate and standardized patient race and ethnicity (r/e) data, both the AI models *and* the methods proposed to identify and address r/e bias will fail to deliver on their promise. The road to unrepresentative r/e datasets is paved with procedural mistakes and conceptual uncertainty (see Fig 1 below). As noted in Fig 1, challenges begin with the conceptual question of how, theoretically, to define "race" and "ethnicity". Even once definitions are in place, hospitals are inconsistent in their collection of r/e data. Variation in how this data is entered into EHR systems adds further complexity, as staff may misclassify patient race. In the end, these compounded issues lead to unrepresentative datasets, with the extent of the problem remaining uncertain due to inaccuracy and gaps in r/e data collection and reporting.

An "ethical AI" strategy involving r/e data analysis can only work if (1) the r/e data is being accurately recorded at each site, and (2) the data can be harmonized across the different r/e categories used by each site. Our analysis presented below finds that these remain big "ifs".

Three decades of scholarship in health disparities research reveals both conceptual and practical challenges in ensuring accuracy for r/e data in the EHR [6]. A 2023 systematic review of 56 datasets assessing the availability and accuracy of patient r/e data found that "EHRs often had missing and/or inaccurate data on race/ethnicity" and that these inaccuracies were most acute for non-white populations [7]. The old axiom—"garbage in, garbage out"—applies. If the r/e data on which medical AI is trained cannot be trusted, then neither can the resulting models nor the mitigating strategies that rely on analyses of said data.





https://doi.org/10.1371/journal.pdig.0000807.g001

This raises an even more fundamental question: should biomedical research and clinical practice incorporate r/e variables at all, as distinct from utilizing measures of racism, socioeconomic status, and ancestry [8,9]? We do not advocate for or against using r/e data in AI models: whether to use r/e variables is a complex decision, and we encourage researchers to follow the excellent guidance from the National Academies of Sciences, Engineering, and Medicine on how to weigh the competing factors [9]. In some instances, after careful consideration, researchers may conclude that it is ethical and relevant to use r/e data. For example, in a study on AI models predicting glaucoma progression, incorporating r/e data improved fairness metrics like equalized odds when tested on external populations [10]. Moreover, developers may use r/e data to evaluate models post training by identifying disparities in performance across demographic groups.

Notwithstanding the importance of these broader questions, our focus is on what happens after the decision has been made to utilize r/e data. We propose two solutions to the issues engendered by this stage of the AI development process. First, we provide a quality control (QC) roadmap for healthcare systems to improve the accuracy of their r/e data for medical AI. Second, we argue that all medical AI model developers should warrant the quality (or lack thereof) of the r/e data they use.

II. The challenge of standardizing r/e categories while federal standards are in flux

A prerequisite for standardizing r/e data across sites is clearly defining the r/e categories themselves. This is no simple task given that race and ethnicity are social constructs – they have always been fluid categories and will likely continue to evolve. At present, the U.S. Government is revisiting the most widely used categories in U.S. biomedical research and clinical practice.



Revisiting OMB standards

In 1997, the Office of Management and Budget (OMB) established standardized questions on r/e that federal agencies and recipients of federal funds were obligated to report data on. This move was motivated by a need to enforce civil rights laws, especially by monitoring equal access to housing, education, and employment for "population groups that historically had experienced discrimination and differential treatment" [11]. In March 2024, after nearly two years of consultation and input, the OMB published revisions to Statistical Policy Directive No. 15 [12]. The new directive notes that "the race and ethnicity categories set forth are sociopolitical constructs and are not an attempt to define race and ethnicity biologically or genetically" [12]. It is beyond the scope of this article to examine the new OMB directive in detail, but we address it further below.

Prior to the 2024 revisions, the OMB standards had been comprised of two ethnicity categories ("Hispanic or Latino" and "Not Hispanic or Latino") and five race categories. Section 3101 of the Public Health Service Act also requires any federally conducted or supported healthcare or public health program, activity, or survey to collect and report data using the Department of Health & Human Services (HHS) standards [13]. The HHS standards broke down the OMB's five race categories into 13 subcategories, and the two ethnicity categories into four [14]. Individual hospitals can create more granular r/e categories, but these must roll up into the OMB standards. While hospitals not receiving federal funding will be exempt from these standards, professional associations like the American Hospital Association recommend the collection of r/e data [15].

Others have sought to standardize r/e data collection, while still conforming to the OMB standards. For example, the Observational Outcomes Partnership (OMOP) established the Common Data Model (CDM), "an open community data standard, designed to standardize the structure and content of observational data" [16]. Although there has been significant internal debate on how race and ethnicity should be standardized [17], the CDM currently follows OMB standards, with the "Race" field containing dozens of concepts, while the "Ethnicity" field contains two [18].

The new OMB standards emerged from years of debate. In June of 2022, the Chief Statistician of the U.S. announced that his office would be reviewing and revising current OMB standards on r/e data [19]. This move was in recognition of (1) increasing racial and ethnic diversity; (2) the growing number of people identifying as more than one race or ethnicity; and (3) changing immigration and migration patterns. Following OMB's request for comment on some initial proposals, a debate about r/e data categories has resurfaced, with some stakeholders advocating for the r/e questions to be combined, given evidence that many respondents view race and ethnicity as similar or the same [20]. However, other presenters raised concerns that combining race and ethnicity might lead to erasure within racially and ethnically diverse communities. For example, Afro-Latino patients could feel compelled to choose between the 'Hispanic or Latino' and the 'Black or African American' categories, when they would otherwise have categorized themselves as both [20].

Conceptual challenges

Developing categories for r/e is an epistemically challenging undertaking with broad social consequences for medical research. To take just one example, scholars disagree on whether the 'one drop rule' (the idea that anyone with a Black ancestor is considered Black) is a reductionist approach and ignores the complexities of identity and culture, reinforcing negative stereotypes like that of racial purity [21]. Relatedly, some have advocated for using 'multiracial' as an r/e category [22], while others oppose this designation [23]. Moreover, there is discourse on whether r/e categories should be used at all, or whether they should be replaced by ancestry in biomedical research and clinical practice [24]. R/e categories are essential for understanding health disparities and developing targeted interventions, but there are those who criticize their oversimplification and failure to capture the complexity of identity and culture [25,26].

These historical debates were considered in the National Academies of Sciences, Engineering, and Medicine (NASEM) report on the Use of Race, Ethnicity, and Ancestry as Population Descriptors in Genomics Research [27].



The standardization of r/e categories for use in biomedical AI modeling should be guided by these particularly relevant recommendations:

- "[R]esearchers should avoid typological thinking, including the assumption and implication of hierarchy, homogeneity, distinct categories, or stability over time of the groups."
- "Researchers, as well as those who draw on their findings, should be attentive to the connotations and impacts of the terminology they use to label groups."
- "For each descriptor selected, labels should be applied consistently to all participants."
- "Researchers should disclose the process by which they selected and assigned group labels..."

III. The challenge of accuracy in r/e data collection

Beyond these conceptual issues, the challenge of matching a patient to the right category remains. The same challenges have long applied to sexual orientation and gender identity and expression (SOGIE) data collection [28]. Research on health disparities has repeatedly revealed significant methodological problems and systemic issues in r/e data collection, including clerks determining a patient's race or ethnicity based on observation, often using last name or appearance [29]. One analysis comparing administrative r/e data to patients' self-reported data identified low agreement rates for certain races including Pacific Islanders, Asians, and Native Americans [30]. Another study revealed similar results for patients who were admitted to two different hospitals [31]. If r/e data collection was accurate, these patients should have been categorized as the same race across the two hospitals. Yet, while White and Black patients were usually classified under the same racial category in both hospitals, patients in all other categories had very low reliability coefficients. More recent studies have continued raising doubts about the accuracy of r/e categorization, especially for marginalized populations. A smaller-scale survey of patients at two New York City clinics found that 33% of respondents at Clinic A and 22% of respondents at Clinic B self-identified in a different manner than the race or ethnicity they were recorded as in the clinic registration database [32]. Around a guarter of patient records in two large observational health databases in the U.S. contained "uninformative" r/e data (either categorized as "Unknown" or "Declined to Answer") [33]. Moreover, 57.9% of the 2.5 million patients served by a New York City healthcare system did not have a race or ethnicity identified in their EHRs [33]. At the same time, 66.5% of patients who recorded their own race and ethnicity selected different categories than their EHR [33].

While certain best practices for collecting r/e data have emerged (including the NIH All of Us Research Program's Participant Provided Information (PPI) form), obtaining that data in the biomedical context, especially in critical care situations, is challenging. An additional challenge of accurately capturing patient r/e data is the general reluctance of hospital staff to ask patients to self-identify. It has been found that hospital staff, in particular clerks and administrators, feel discomfort asking these questions which may be interpreted as intrusive or offensive [34,35]. At the same time, one study showed that 28% of respondents were uncomfortable sharing r/e information with a clerk or administrator [34]. Recent scholarship by Owosela et al. (2024) confirms that these issues persist [6]. In their review of the accuracy of r/e data in EHRs, Owosela et al. find that inaccuracies in r/e persist, owing in part to the limitations of self-reporting and a lack of standardization. Their recommendations for improved self-reporting mechanisms and standardization are consistent with our analysis below.

Misclassification of patient r/e is directly related to the standardization and categorization challenges discussed above. Both patients and hospital staff may be confused or frustrated by the options made available to them on intake forms and in EHR drop-down menus [27]. Our experience with Bridge2AI: Patient-Focused Collaborative Hospital Repository Uniting Standards (CHoRUS) for Equitable AI, an NIH project developing an AI-ready dataset from more than 100,000 critically ill patients' EHR data, illuminates the urgency of the problem. While all sites in the project adhered to the two standard OMB



ethnicity options, some provided over 100 subcategories. Some of the sites address the challenge of multiracial patients by having them select a 'multiracial' option, while others have patients check off all applicable r/e options. For medical AI, the implication is that deployment and development across sites may be significantly hampered if the sites are not consistent and accurate in how they classify patient r/e.

These inaccuracies in r/e data collection have potentially pernicious implications for medical AI development. To illustrate, a recent review examining the performance of AI models in treating cardiovascular diseases found racial disparities and concluded that it is essential to identify strategies to limit and mitigate bias during each step of the AI development pipeline [36]. Similar issues have been raised around the deployment of AI models in dermatology [37], radiology [38], and other medical fields [39]. Given these credible concerns about race bias in medical AI systems, it follows that inaccurate r/e demographic data could amplify health inequities by preventing new AI tools from working equally well across all r/e groups.

IV. Toward solutions

To address these issues, we propose two solutions: (1) hospital systems should adopt best practices in r/e data Quality Control (QC), and (2) researchers and developers creating medical AI models should explicitly warrant the quality of their r/e data.

Quality control

Based on two decades of research on optimizing the collection of r/e data in hospital contexts, recommendations for best practices have emerged. In Box 1, we identify several of these best practices that are germane to the issues we raise here (Box 1). The first step is to prioritize this problem on an institutional level and develop a plan for improving r/e data classification accuracy. This plan should be patient-centered, facilitating patient autonomy over how their r/e data is reported and then shared. Patients should also be given a clear rationale for why their r/e information is being requested.

Given the diversity of self-identification preferences and significant cross-cultural and cross-national differences in how r/e is conceptualized, we recognize that it is impossible to create r/e categories that perfectly reflect the sociodemographic groups with which individuals identify. But if the representativeness of AI datasets is to be evaluated, it follows that a categorization scheme must be deployed. Our contention is that in refining those categories institutions should continue to develop more granular categories reflecting the diversity of their unique communities. Precisely because these categories are challenging to define, and vary across cultures, it is important for institutional policymakers to listen to concerns from members of their communities.

Our proposed solution seeks to reconcile the tension between allowing for flexibility in patient identification and facilitating cross-institutional comparisons via standardization. A lack of standardization, both in defining categories and developing procedures for operationalizing them, is at the core of the problems we have identified above (see Section III). In the absence of uniform guidelines for collecting and categorizing r/e, different healthcare institutions may use divergent approaches, leading to incomparable datasets that cannot be meaningfully merged or analyzed at a broader scale. Without a standardized approach to measuring and reporting r/e in AI healthcare training datasets, we would expect to see significant variations across institutions in the quality of r/e data. This variation would contribute to the perennial problem of aggregating across datasets from different hospitals [40]. Using fragmented datasets may also impede efforts to monitor and evaluate AI systems for bias. If standardization across major systems is not practical, then at a minimum, institutions should be transparent about their methods and ensure that however they have defined their categories, they are accurate and consistent in applying them. This transparency and granular data could facilitate standardization across institutions, even when those institutions adopt different categories.



Box 1. Best Practices for Obtaining Patients' Race & Ethnicity Data.

Prioritize and Plan: Make improving self-reported r/e data an organizational priority using intentional policies, procedures, and training [41].

Self-Report, if Possible Before the Visit: Empower patients by inviting them to self-report r/e information, including through pre-visit questionnaires [42,43].

Follow-up: If patient r/e data is not collected before or during the first visit, ensure that there is a personalized follow-up [44].

Provide Rationale: When inviting patients to share their r/e information, explain the rationale, e.g., that the hospital will use this data for quality monitoring and to improve patient care [27].

Improved Training: Ensure that all frontline staff who will interact with patients regarding r/e data collection have undergone appropriate training [42]. The use of scripts can be helpful in ensuring sensitivity when encouraging patients to provide their r/e data [41,43].

Team Effort: If a patient is uncomfortable speaking with the administrator or clerk about their r/e data, involve a doctor or nurse as this may increase comfort levels [34]. Multiple stakeholders should be responsible for the collection of this data, not just one [45].

Electronic Health Record Fields: Self-identifying through a paper-based form can be daunting when there is a long list of r/e labels. Online forms allowing for keystroke recognition can increase the number of r/e options used to populate fields. [27]

Maintain Flexibility: Hospitals should remain flexible given that the sociodemographic makeup of the communities they serve will inevitably change. Having an annual review of current census data can help ensure that an appropriate amount of granularity is included in the r/e options provided to patients [27]. The ELSIhub Collection on race and population identifiers provides a useful resource to stay up to date [46].

Addressing the practical challenges involved in implementing these best practices will need to be a priority if institutions wish to bring about real change in the quality of the r/e data they generate. We identify three practical challenges to prioritize. The first is that we are not aware of a publicly available training program that can be readily imported to address r/e data collection issues. An initial step will be investment in a short training module. The second challenge, related to the first, is that r/e data is likely to be collected by several different types of staff members, including those who admit patients, clinical staff who treat patients, and IT staff who prepare data files and link r/e data from one system to another. It is therefore important for the training module to be completed by all who may be initially collecting, reviewing, and archiving r/e data. The third challenge is that developing a training module and implementing it will face competing priorities in terms of resource allocation. We believe that investing in the implementation of these best practices is likely to yield outsized benefits in the longer term by increasing patient trust and unlocking the full potential of AI systems to improve health outcomes. An incremental, iterative approach can also help alleviate some of these challenges, as pilot programs will enable hospitals to refine workflows, train staff, and measure outcomes before scaling up across hospital systems.



Warranting data quality

Leading AI ethics frameworks stress the importance of transparency and sharing with end users the contents of the black box supporting medical decision-making [47]. One promising strategy to promote transparency is the use of labeling for AI-based medical devices [48]. We advocate for a label on medical AI software which acknowledges data quality limitations and discloses the data collection processes used for the r/e datasets on which the models were created.

We suggest that this disclaimer could take one of the following forms:

The models reported in this analysis and/or implemented in this tool were trained with data that included race and ethnicity (r/e) descriptors for individuals. The individualized r/e data were derived from [*insert data source*].

<u>Option #1 (preferred, if data collection methods are known)</u>: This individualized r/e data were collected in the following way: [describe r/e collection methods].

Option #2 (necessary if data collection methods are not known to those developing and implementing the AI models): We are not aware of how this individualized r/e data were collected, and therefore cannot warranty the quality of the r/e data used in these models.

This disclaimer could be added at minimal time cost. The collection methods that could be specified include patient self-reporting through pre-appointment online forms. Such disclaimers would serve the ethical priority of transparency in AI modeling and the scientific priority of ensuring that AI models are built on high quality data—and if not, making every-one aware of this deficiency.

We recognize that the entities and individuals developing medical AI models will often be distinct from those who are gathering demographic information. Inspired by the longstanding practice of "ethical sourcing" of materials in the field of corporate social responsibility [49], our approach places the onus on AI developers to use reasonable efforts to identify the source(s) and quality of the r/e data they deploy in their models. For example, missing r/e data is regularly imputed in healthcare datasets [50–52]. Developers should be aware of when the r/e data they are using has been imputed and should communicate this to those who use their AI systems. AI developers often work with multiple datasets to train their AI model. In such instances, we believe that developers should disclose the collection methodologies for each dataset that was utilized in the development of the medical AI model, especially when the r/e datasets differ in terms of quality.

Our approach is forward-looking and intended to spur systemic changes in r/e data collection practices. We anticipate that in the near term, many AI developers will need to select Option #2 because they will not know how the r/e data in their models was generated. This would be true, for instance, for those AI developers who are working with historical datasets. However, we also anticipate that precisely because they are initially scarcer, AI models that can provide Option #1 explanations will be more highly valued. For example, a firm could differentiate itself from competitors by pointing to its use of gold-standard r/e data. Relatedly, as medical AI faces increasing regulatory scrutiny, markets may favor those model developers who can offer an Option #1 explanation. Regulatory bodies such as the U.S. Food and Drug Administration and the E.U.'s AI Office will be actively monitoring the potential for bias in medical AI systems. In the case of the E.U. AI Act, high-risk AI systems must be trained, validated, and tested with data that meet quality criteria, which include "data collection processes and the origin of data" (Article 10(2)(b)). The path forward to more accurate r/e data in AI models may be slow, but we believe that without data quality disclaimers progress will be even slower.



V. Conclusion

Our argument is straightforward: if medical AI models are going to utilize r/e data, that data should be of high quality. We have laid out concerns about r/e data quality and proposed high-level guidance for improving and maintaining that quality. Hospitals are the keepers of EHR data and thus are best positioned to implement our recommended QC measures. Nevertheless, researchers and developers who are creating AI models for biomedical use have a parallel ethical responsibility to know where their r/e datasets are coming from and to be transparent about the possible deficiencies of those datasets on which their models are built. For norms to change, professional societies, funders, journal editors, and peer reviewers will need to demand better r/e data accountability. A productive next step would be to convene these stakeholders, along with patient advocates, to further develop the initial solutions we have proposed.

Acknowledgments

The authors thank members of the Shen Neurolaw Lab; Dana Foundation Career Network in Neuroscience & Society; and Patient-Focused Collaborative Hospital Repository Uniting Standards (CHoRUS) for Equitable AI for feedback and assistance on this research project. FS thanks the University of Minnesota Law School for administrative support.

Author contributions

Conceptualization: Francis Shen, Lakshmi Bharadwaj, Alexandra Tsalidis.

Writing – original draft: Alexandra Tsalidis, Lakshmi Bharadwaj, Francis Shen.

Writing - review & editing: Alexandra Tsalidis, Lakshmi Bharadwaj, Francis Shen.

References

- 1. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. JAMA. 2019;322(24):2377–8. <u>https://doi.org/10.1001/jama.2019.18058</u> PMID: <u>31755905</u>
- 2. Christensen D. Medical algorithms are failing communities of color. Health Affairs Blog. Health Affairs; 2021 Sep 9 [cited 2023 Jul 31]. Available from: https://www.healthaffairs.org/content/forefront/medical-algorithms-failing-communities-color
- Sauer CM, Chen L-C, Hyland SL, Girbes A, Elbers P, Celi LA. Leveraging electronic health records for data science: common pitfalls and how to avoid them. Lancet Digit Health. 2022;4(12):e893–8. <u>https://doi.org/10.1016/S2589-7500(22)00154-6</u> PMID: <u>36154811</u>
- 4. Grant C. Algorithms are making decisions about health care, which may only worsen medical racism. ACLU. American Civil Liberties Union; 2022 Oct 3 [cited 2023 Jul 31]. Available from: https://www.aclu.org/news/privacy-technology/algorithms-in-health-care-may-worsen-medical-racism
- 5. World Health Organization Health Ethics and Governance Unit. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021.
- Owosela BO, Steinberg RS, Leslie SL, Celi LA, Purkayastha S, Shiradkar R, et al. Identifying and improving the "ground truth" of race in disparities research through improved EMR data reporting. A systematic review. Int J Med Inform. 2024;182:105303. <u>https://doi.org/10.1016/j.ijmedinf.2023.105303</u> PMID: 38088002
- Johnson JA, Moore B, Hwang EK, Hickner A, Yeo H. The accuracy of race & ethnicity data in US based healthcare databases: a systematic review. Am J Surg. 2023;226(4):463–70. <u>https://doi.org/10.1016/j.amjsurg.2023.05.011</u> PMID: <u>37230870</u>
- Burchard EG, Ziv E, Coyle N, Gomez SL, Tang H, Karter AJ, et al. The importance of race and ethnic background in biomedical research and clinical practice. N Engl J Med. 2003;348(12):1170–5. <u>https://doi.org/10.1056/NEJMsb025007</u> PMID: <u>12646676</u>
- 9. National Academies of Sciences, Engineering, and Medicine. Rethinking race and ethnicity in biomedical research. Washington, DC: The National Academies Press; 2024. https://doi.org/10.17226/27913
- 10. Ravindranath R, Stein JD, Hernandez-Boussard T, Fisher AC, Wang SY. The impact of race, ethnicity, and sex on fairness in artificial intelligence for glaucoma prediction models. Ophthalmol Sci. 2025;5:100596.
- Review of the Racial and Ethnic Standards to the Office of Management and Budget Concerning Changes to the Standards for the Classification of Federal Data on Race and Ethnicity [cited 28 Dec 2024]. The White House. Available from: <u>https://obamawhitehouse.archives.gov/omb/</u> <u>fedreg_directive_15</u>
- 12. Revisions to OMB's Statistical Policy Directive No. 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity. 2024 Mar 29. [cited 2024 Dec 30]. Available from: https://www.federalregister.gov/documents/2024/03/29/2024-06469/ revisions-to-ombs-statistical-policy-directive-no-15-standards-for-maintaining-collecting-and



- 13. Burwell SM. Improving the identification of health care disparities in medicaid and CHIP. Washington, DC: U.S. Department of Health and Human Services; 2014.
- 14. Recommendations from the interagency committee for the review of the racial and ethnic standards to the Office of Management and Budget concerning changes to the standards for the classification of federal data on race and ethnicity. Federal Register. 1997;62(131): 36873–946.
- AHA Institute for Diversity and Health Equity. Addressing health care disparities through race, ethnicity and language (REaL) Data. AHA Institute for Diversity and Health Equity. American Hospital Association; 2020 Jun [cited 2023 Aug 7]. Available from: https://www.aha.org/system/files/media/ file/2020/06/ifdhe_real_data_resource.pdf
- 16. OMOP Common Data Model. OHDSI Coordinating Center at Columbia University [cited 2023 Aug 5]. Available from: <u>https://ohdsi.github.io/</u> CommonDataModel/#:~:text=The%20Observational%20Medical%20Outcomes%20Partnership
- 17. Reich C. Race and Ethnicity in the OMOP CDM. OHDSI Forums. 2019 Nov 19 [cited 2023 Aug 5]. Available from: <u>https://forums.ohdsi.org/t/</u> race-and-ethnicity-in-the-omop-cdm/8700
- Ethnicity Domain and Vocabulary. Observational Health Data Sciences and Informatics. 2016 Mar 12 [cited 2023 Aug 5]. Available from: <u>https://</u> www.ohdsi.org/web/wiki/doku.php?id=documentation:vocabulary:ethnicity
- **19.** The White House. Reviewing and revising standards for maintaining, collecting, and presenting federal data on race and ethnicity. The United States Government. 2022 Jun 15 [cited 2023 Jul 31]. Available from: https://www.whitehouse.gov/omb/briefing-room/2022/06/15/ reviewing-and-revising-standards-for-maintaining-collecting-and-presenting-federal-data-on-race-and-ethnicity/.
- **20.** Management and Budget Office. Initial proposals for updating OMB's race and ethnicity statistical standards. Federal Register. 2023;88(18):5375–84.
- 21. Hickman CB. The devil and the one drop rule: racial categories, African Americans, and the U.S. Census. Michigan Law Rev. 1997;95(5):1161. https://doi.org/10.2307/1290008
- 22. Bijan G. Multiracial minorities: erasing the color line. Law Inequality. 1994;12(1):183-204.
- 23. Phillips A. The multiracial option: a step in the white direction. California Law Rev. 2017;105(6):1853-78. https://doi.org/10.15779/z38h98zd1s
- 24. Nobles M. Shades of citizenship: race and the census in modern politics. California: Stanford University Press; 2000.
- 25. Parrado EA, Rodriguez CE. Changing Race: Latinos, the Census, and the History of Ethnicity in the United States. J Am History. 2001;88(2):744. https://doi.org/10.2307/2675257
- Hochschild JL, Powell BM. Racial Reorganization and the United States Census 1850–1930: Mulattoes, Half-Breeds, Mixed Parentage, Hindoos, and the Mexican Race. Stud Am Pol Dev. 2008;22(1):59–96. <u>https://doi.org/10.1017/s0898588x08000047</u>
- 27. National Academies of Sciences, Engineering, and Medicine. Using population descriptors in genetics and genomics research: a new framework for an evolving field. Washington, DC: The National Academies Press; 2023. https://doi.org/10.17226/26902
- Deb B, Porter K, van Cleeff A, Reardon LC, Cook S. Emphasizing sexual orientation and gender identity data capture for improved cardiovascular care of the LGBTQ+ population. JAMA Cardiol. 2024;9(3):295–302. <u>https://doi.org/10.1001/jamacardio.2023.5267</u> PMID: <u>38265768</u>
- 29. Pittman MA, Pierce D, Hasnain-Wynia R. Who, When, and How: the current state of race, ethnicity, and primary language data collection in hospitals. The Commonwealth Fund; 2004.
- Kressin NR, Chang B-H, Hendricks A, Kazis LE. Agreement between administrative data and patients' self-reports of race/ethnicity. Am J Public Health. 2003;93(10):1734–9. <u>https://doi.org/10.2105/ajph.93.10.1734</u> PMID: <u>14534230</u>
- **31.** Blustein J. The reliability of racial classifications in hospital discharge abstract data. Am J Public Health. 1994;84(6):1018–21. <u>https://doi.org/10.2105/ajph.84.6.1018 PMID: 8203669</u>
- Moscou S, Anderson MR, Kaplan JB, Valencia L. Validity of racial/ethnic classifications in medical records data: an exploratory study. Am J Public Health. 2003;93(7):1084–6. https://doi.org/10.2105/ajph.93.7.1084 PMID: 12835189
- Polubriaginof FCG, Ryan P, Salmasian H, Shapiro AW, Perotte A, Safford MM, et al. Challenges with quality of race and ethnicity data in observational databases. J Am Med Inform Assoc. 2019;26(8–9):730–6. <u>https://doi.org/10.1093/jamia/ocz113</u> PMID: <u>31365089</u>
- 34. Baker DW, Cameron KA, Feinglass J, Georgas P, Foster S, Pierce D, et al. Patients' attitudes toward health care providers collecting information about their race and ethnicity. J Gen Intern Med. 2005;20(10):895–900. <u>https://doi.org/10.1111/j.1525-1497.2005.0195.x</u> PMID: <u>16191134</u>
- Moorthie S, Peacey V, Evans S, Phillips V, Roman-Urrestarazu A, Brayne C, et al. A scoping review of approaches to improving quality of data relating to health inequalities. Int J Environ Res Public Health. 2022;19(23):15874. <u>https://doi.org/10.3390/ijerph192315874</u> PMID: <u>36497947</u>
- Cau R, Pisu F, Suri JS, Saba L. Addressing hidden risks: systematic review of artificial intelligence biases across racial and ethnic groups in cardiovascular diseases. Eur J Radiol. 2025;183:111867. <u>https://doi.org/10.1016/j.ejrad.2024.111867</u> PMID: <u>39637580</u>
- Daneshjou R, Vodrahalli K, Novoa RA, Jenkins M, Liang W, Rotemberg V, et al. Disparities in dermatology AI performance on a diverse, curated clinical image set. Sci Adv. 2022;8(32):eabq6147. <u>https://doi.org/10.1126/sciadv.abq6147</u> PMID: <u>35960806</u>
- Tejani AS, Ng YS, Xi Y, Rayan JC. Understanding and mitigating bias in imaging artificial intelligence. Radiographics. 2024;44(5):e230067. <u>https://</u> doi.org/10.1148/rg.230067 PMID: 38635456
- Chen RJ, Wang JJ, Williamson DFK, Chen TY, Lipkova J, Lu MY, et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. Nat Biomed Eng. 2023;7(6):719–42. <u>https://doi.org/10.1038/s41551-023-01056-8</u> PMID: <u>37380750</u>



- 40. Arora A, Alderman JE, Palmer J, Ganapathi S, Laws E, McCradden MD, et al. The value of standards for health datasets in artificial intelligencebased applications. Nat Med. 2023;29(11):2929–38. <u>https://doi.org/10.1038/s41591-023-02608-w</u> PMID: <u>37884627</u>
- 41. Bullard J. Improving collection of self-reported race & ethnicity data. RTI Health Advance. Research Triangle Institute; 2022 Nov 29 [cited 2023 Aug 6]. Available from: https://healthcare.rti.org/insights/improving-collection-of-self-reported-race-and-ethnicity-data
- 42. 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(4 Pt 1):1501–18. https://doi.org/10.1111/j.1475-6773.2006.00552.x PMID: 16899021
- 43. Berg S. Gathering race, ethnicity info can help fight inequity. How to do it well. American Medical Association. American Medical Association; 2020 Sep 9 [cited 2023 Aug 7]. Available from: <u>https://www.ama-assn.org/delivering-care/hypertension/</u> gathering-race-ethnicity-info-can-help-fight-inequity-how-do-it-well
- 44. Race and Ethnicity Data Improvement Toolkit. Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality; 2014 Sep 30 [cited 2023 Aug 12]. Available from: https://hcup-us.ahrq.gov/datainnovations/raceethnicitytoolkit/home_race.jsp#:~:text=This%20 toolkit%20provides%20practical%20tools,%2FL)%20data%20collection%20efforts
- 45. Yudell M, Gil-Riaño S, Hammonds E. The evolution of race and population identifiers in scientific thought and practice. In: ELSIhub Collections. Center for ELSI Resources and Analysis (CERA). <u>https://doi.org/10.25936/ca3b-rv59</u>
- 46. Manchanda R, Do R, Miles N. A toolkit to advance racial health equity in primary care improvement. California Health Care Foundation. California Improvement Network, California Health Care Foundation, Healthforce Center at UCSF; 2022 Apr [cited 2023 Aug 12]. Available from: <u>https://www.chcf.org/wp-content/uploads/2022/07/ToolkitRacialEquityPrimaryCareImprovement.pdf</u>
- 47. Ursin F, Timmermann C, Steger F. Explicability of artificial intelligence in radiology: Is a fifth bioethical principle conceptually necessary?. Bioethics. 2021;36(2):143–53. https://doi.org/10.1111/bioe.12918
- **48.** Gerke S. "Nutrition Facts Labels" for artificial intelligence/machine learning-based medical devices The urgent need for labeling standards. George Washington Law Rev. 2023;91(1):79–163.
- 49. Blowfield M. Ethical sourcing: a contribution to sustainability or a diversion?. Sust Dev. 2000;8(4):191–200. <u>https://doi.org/10.1002/1099-1719(200011)8:4<191::aid-sd146>3.0.co;2-e</u>
- 50. Xue Y, Harel O, Aseltine RH Jr. Imputing race and ethnic information in administrative health data. Health Services Res. 2019;54(4):957–63. https://doi.org/10.1111/1475-6773.13171
- 51. Filice CE, Joynt KE. Examining race and ethnicity information in medicare administrative data. Med Care. 2017;55(12):e170–6. <u>https://doi.org/10.1097/MLR.000000000000008</u> PMID: <u>29135782</u>
- 52. Grundmeier RW, Song L, Ramos MJ, Fiks AG, Elliott MN, Fremont A, et al. Imputing missing race/ethnicity in pediatric electronic health records: reducing bias with use of U.S. census location and surname data. Health Serv Res. 2015;50(4):946–60. <u>https://doi.org/10.1111/1475-6773.12295</u> PMID: <u>25759144</u>