

# Best Practices for Collecting and Analyzing Race and Ethnicity Data

## What is this guide?

A comprehensive overview of best practices for collecting and analyzing race and ethnicity data in a way that promotes equity and improves health outcomes.

## Who is this guide intended for?

Public health professionals and researchers who work to

- Inform action by leadership and clinical teams to reduce disparities in care.
- Engage community members in data-driven decision making.
- Effectively pinpoint and allocate resources to address gaps in access or quality care.
- Meet state and federal reporting requirements.

## What questions does this guide seek to answer?

- What do functional race and ethnicity data collection forms look like?
- What resources and training promote comfortable and effective race and ethnicity data collection?
- How can race and ethnicity data be analyzed and applied to promote equity in public health services?

## What is race? Ethnicity?

Race is a social construct based on perceived physical traits. It does not determine health outcomes, but racism contributes to disparities. Ethnicity refers to identity tied to shared culture, language, or history.

## Who determines how race and ethnicity are reported?

In 1977, OMB issued the Race and Ethnic Standards for Federal Statistics and Administrative Reporting that are set forth in Statistical Policy Directive No. 15. The standards in this Directive have been used for almost two decades throughout the Federal government for recordkeeping, collection, and presentation of data on race and Hispanic origin. The standards have been used in two decennial censuses and in surveys of the population, data collections necessary for meeting statutory requirements associated with civil rights monitoring and enforcement, and in other administrative program reporting.

On March 18, 2024, the OMB published its review of Statistical Policy Directive No. 15 (SPD 15) and issued updated standards for collecting, maintaining, and presenting race and ethnicity data across federal agencies. <sup>1</sup> The updates are as follows:

- Using one combined question for race and ethnicity, and encouraging respondents to select as many options as they identify with.
- Adding “Middle Eastern or North African” as a new minimum reporting category.
- Requiring the collection of detailed race and ethnicity responses that go beyond the required minimum categories for most situations. <sup>2</sup>

The updated standards also aim to provide balance across the race/ethnicity definitions and to remove any outdated and offensive terminology. <sup>1</sup> OMB expect that updates will be made as expeditiously as possible for all its data collections that include race and ethnicity, such as the American Community Survey and the 2030 Census.

Minimum Race/Ethnicity Reporting Category	Definition
American Indian or Alaska Native	Individuals with origins in any of the original peoples of North, Central, and South America, including, for example, Navajo Nation, Blackfeet Tribe of the Blackfeet Indian Reservation of Montana, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, Aztec, and Maya.
Asian	Individuals with origins in any of the original peoples of Central or East Asia, Southeast Asia, or South Asia, including, for example, Chinese, Asian Indian, Filipino, Vietnamese, Korean, and Japanese.
Black or African American	Individuals with origins in any of the Black racial groups of Africa, including, for example, African American, Jamaican, Haitian, Nigerian, Ethiopian, and Somali.
Hispanic or Latino	Includes individuals of Mexican, Puerto Rican, Salvadoran, Cuban, Dominican, Guatemalan, and other Central or South American or Spanish culture or origin.
Middle Eastern or North African	Individuals with origins in any of the original peoples of the Middle East or North Africa, including, for example, Lebanese, Iranian, Egyptian, Syrian, Iraqi, and Israeli.
Native Hawaiian or Pacific Islander	Individuals with origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands, including, for example, Native Hawaiian, Samoan, Chamorro, Tongan, Fijian, and Marshallese.
White	Individuals with origins in any of the original peoples of Europe, including, for example, English, German, Irish, Italian, Polish, and Scottish.

## Revised Definitions for Minimum Race/Ethnicity Reporting Categories in 2024 SPD 15 **1**

### Best Practices for Data Collection

#### Creating a standardized data collection process

- *When*: Collect information early in the encounter, preferably during registration.
- *Who*: Front-line staff, such as receptionists. If a patient is uncomfortable providing this information to a receptionist, a provider should collect the data.
- *How*: Self-reporting is the preferred method. Use clear, standardized forms and scripts to explain

the purpose and confidentiality of the data collected.

#### Addressing patient concerns

Before asking for any information, tell patients:

- Data will *NOT* be used to discriminate against patients.
- Data *WILL* be used to identify inconsistencies in service.
- Data *WILL* be used to ensure that all patients receive high-quality care and services.

- Information *WILL* be kept confidential, and participation is voluntary. <sup>3</sup>

Justify race and ethnicity data collection as ‘Quality Monitoring.’ Compared to other explanations, the following introductory statement was found to improve patient comfort levels the most:

*“We want to ensure that all our patients get the best care possible, regardless of race or ethnic background. We would like you to tell us your race or ethnic background so that we can review all patients’ treatment and ensure that everyone gets the highest quality of care.”* <sup>4</sup>

Assure patients that data will be kept confidential by HIPAA laws and that a limited number of people can access the data.

### Wording instruction questions

The instructions “*Mark/select all that apply*” were found to perform better than the instructions “*Mark [x]/select one or more boxes*” for individuals reporting multiple race and ethnicity groups and yielded more consistent reporting for major race and ethnicity groups. <sup>4</sup>

### Staff Education and Training

- Provide scripts explaining the rationale for collecting race and ethnicity data, its importance, and how it will be used.
- Use case examples to guide staff on answering patient questions or concerns.
- Include instructions on how to record data in new or existing systems.

### Appropriate Hispanic and Latino/a/e/x Terminology

- *Hispanic* refers to individuals with ancestry from Spanish-speaking countries, while *Latino/a* refers to those from Latin America. These terms should remain distinct, even though they are sometimes used interchangeably in surveys.
- *Latinx* is a gender-neutral term that has gained popularity since 2018. However, only a small

percentage of the Hispanic and Latino population identifies with it. <sup>5</sup> It’s recommended to use “Hispanic and Latino/a” for consistency with federal categories, and allow individuals to specify further if needed (e.g., Mexican-American, Spanish). <sup>6,7</sup>

### Inappropriate Techniques with Multiracial Data

- Avoid excluding multiracial responses or aggregating them into a single category.
- Do not count a single individual multiple times across different racial categories. The current standard requires individuals to identify once, even if they have multiracial identities.

### Tools for Collection and Analysis

#### Forms and surveys

- Ensure that forms and surveys capture a comprehensive range of race and ethnicity categories. This means not only including broad groups but also offering options for individuals to self-identify if they do not see their group represented.
- Provide translations to accommodate non-English speakers, ensuring all patients have access to the same information regardless of language proficiency.
- Clearly explain why the data is being collected, how it will be used, and the benefits it brings—such as improving healthcare services and addressing disparities. This builds trust and encourages accurate reporting.

### Information Systems

- Choose electronic systems that allow for detailed, nuanced entries of race, ethnicity, and language data rather than just broad categories.
- Ensure these systems seamlessly integrate with existing healthcare databases and electronic health records (EHRs). This integration facilitates efficient data sharing, comprehensive patient profiles, and more robust data analysis.
- Opt for systems that can update in real-time, allowing for dynamic tracking of patient demographics and prompt identification of

identification of disparities or emerging trends.

## **Scripts and Training**

- Develop clear, standardized scripts for staff to use when asking for race, ethnicity, and language information. This consistency helps ensure that patients receive the same information and guidance, regardless of whom they speak with.
- Provide detailed training sessions for staff on how to use these scripts, the importance of collecting this data, and how to handle sensitive questions with cultural competence and respect.
- Regularly update training materials and provide refresher courses to keep staff informed of any changes in data collection protocols and best practices, ensuring sustained consistency and quality in data collection.

## **Integrating Data Collection into Existing Systems**

Consider incorporating new categories into existing forms and systems instead of creating a new form entirely. Also, ask patients to report their race, ethnicity, and language preferences as part of the standard registration process. As employees become familiar with new systems, use reminders, such as flags or checkboxes in forms.

## **Collapsing race and ethnicity data**

Collapsing data refers to aggregating more specific data into broader categories. This typically occurs when larger numbers are needed for data analysis, or data are merged across databases with different race and ethnicity categories. The main benefit of this is to reduce the complexity of data, making it easier to analyze. Comprehensive justification regarding groupings (and if groups are excluded) is critical to explain why these data handling methods do not compromise data quality.

## **Key Materials for Effective Data Collection**

- A script for staff to use when asking for data and answering questions in a uniform manner.
- Paper and electronic versions of the collection form in all relevant languages.

- Standardized instructions for recording/coding data into the information system.
- A section in patient electronic records for storing race and ethnicity data.
- Staff training plan.
- Community outreach plan.

## **Analysis and Application of Data**

### **Using race and ethnicity data to identify health disparities**

Clinical outcomes can be arranged by race and ethnicity to corroborate known existing disparities or reveal previously unknown disparities in larger data sets. These analyses can then be leveraged to draw associations with barriers to care, such as language, transportation, or access to technology for telehealth.

### **Stratification**

Stratifying data uses expansive race and ethnicity categories to reveal hidden disparities in public health outcomes within broader groups. Specific race and ethnicity data can be tied to health outcomes data to understand how contextual factors (such as cultural beliefs, access, and attitudes toward health) affect a person's health risks in diverse communities.

### **Intersectional Analysis**

If the sample is large enough, data can be stratified by two or more categories, such as race AND gender. This helps provide a nuanced insight into how the multidimensionality of identity affects health outcomes.

### **Evaluating Interventions**

Stratifying health intervention data by race and ethnicity can point toward specific targets for improving intervention efficacy – such as culturally tailoring programs to meet patients' needs, navigation, and engagement of interdisciplinary care teams for intervention delivery.

As an example, studies found that while opioid death rates flattened between 2018-19, there was a 38% increase in opioid overdose death rates among non-Hispanic Black individuals. Generalized interventions

to reduce opioid overdose deaths had little effect on or worsened disparities between racially minoritized and white populations. <sup>8,9</sup> Based on this research, faith-based rehabilitative treatments were piloted and found to be more effective than traditional interventions for minoritized communities.

## Improving Clinical Tools

Current data collection practices privilege certain people over others and use data predominantly taken from homogenous populations as the medical standard. Historically, this practice has resulted in the development of clinical tools that are not generalizable and further marginalize people with identities that were excluded when the medical “standard” was developed.

## Planning for Diversity in Services and Staff

Up-to-date race, ethnicity, and language data can help inform how you allocate resources and funds to different programs and interpreter services to ensure your agency can meet a diverse population's needs. Data can also help determine how well staff diversity and skills match the diversity of the people you serve.

## Best Practices for Reporting & Privacy

Sharing race and ethnicity data is a great way to justify research conclusions and meet state-required benchmarks. All race and ethnicity data *MUST* be deidentified when being shared with the public or other organizations unless you have received explicit permission from the patient to share their data.

This is particularly important for communities with large undocumented populations. Extra care should be taken to prevent data from being accessed by immigration and deportation services.

## Recommendations for Researching on Structural Racism

### Do not use race as a proxy for racism

- Use variables that capture multiple dimensions of structural racism.
- Reconsider your model; include in the discussion that significant differences between racialized

groups do not account biological or cultural differences.

## Race and social determinants

- If necessary to use race, intersect race with another variable to capture intersectionality (e.g., race × socioeconomic status, race × redlining).
- Become aware of the potential variables race can be confounded with.

## Interdisciplinary variables and data sets

- Review interdisciplinary literature on racialized group(s); consult with colleagues outside of discipline; partner with marginalized communities.
- Become aware of biases embedded in funding and publishing that may influence access to scholarship on racism.

## Mixed methods

- Collect qualitative and quantitative data.
- Become aware of your biases toward other methodologies.

## Life course, time, and history

- Incorporate life-course pathways into models; explore impact of time on exposure and health outcomes; engage interdisciplinary colleagues and literature to build in the role of history.
- Become aware of your lack of knowledge on history; acknowledge historic abuses of your discipline and others; counter embedded biases that would create or reproduce harm.

## Mixed data

- Engage with disciplines and practices that are advancing the field of mixed data integration; engage with discourse from the humanities and social sciences; practice researcher reflexivity.
- Disentangle your biases toward other methodologies; become aware of how your biases are embedded in the study.

## Multilevel and multidimensional models

- Utilize an index of variables or a latent construct;



take care when selecting variables.

- Use culturally responsive assessments that are valid for the population of interest; acknowledge limitations of model. 10

### Primary beneficiary: the community

A key part of just data practices is reciprocity, the idea that the community whose data is being collected should be the primary benefactor of research done with that data.

This is rooted in frameworks developed by Indigenous health researchers, which strive to move beyond producing research on communities to conducting research for and with communities.

Community members should be directly engaged in decision-making about collecting and using data. While this will look different for every community, data should be put toward community-identified goals and ultimately stretch beyond academic/corporate research interests to develop communities' autonomy and educational, economic, and government institutions. 11

### Right to decline participation

Communities may not want to be involved in data collection or research, as it has been historically used for surveillance. The right for communities to refuse involvement without being harmed must be protected.

### Ensuring community benefits

Historically, data collection has been used to discriminate or tokenize marginalized communities, perpetuating oppressive power dynamics between health systems and people. Researchers must use data to eliminate systemic barriers and discrimination rather than reinforce them. Ultimately, this requires engaging community members in the collection, ownership, and use of data to allow them to support community autonomy.

## Cheatsheet

### 1. Identify Populations Served

- Use up-to-date demographic data collected from diverse populations.
- Include community members in decisions.

### 2. Create a Standardized Process

- Ensure clear processes for data collection, confidentiality, and staff training.
- Reflect categories for race and ethnicity for self-reported identities that can be rolled into minimum state and federal categories.

### 3. Assemble Tools for Collection and Analysis

- Update forms, records, and methodologies to include relevant questions.
- Provide data collection forms in multiple languages.

### 4. Analyze and Apply Data

- Regularly review data to identify disparities in access and utilization of care in an intersectional analysis.
- Use data to inform interventions and plan funding.

### 5. Best Practices

- Engage community in decision-making around the collection and use of data.
- Ensure data are used to address community-identified problems and that the community benefits most from the research.

1. Marks, R., Jones, N., & Battle, K. (2024, April 8). Updates to OMB's Race/Ethnicity Standards. United States Census Bureau.
2. Orvis, K. (2024, March 28). OMB Publishes Revisions to Statistical Policy Directive No. 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity | OMB. The White House.
3. Office of Health Equity, MDPH, Making CLAS Happen: Ch 3. Collecting Diversity Data. 54–84 (2013). Massachusetts Department of Public Health.
4. Mathews, K., Phelan, J., Jones, N. A., & al., e. (2015). 2015 National Content Test Race and Ethnicity Analysis Report. In: United States Census Bureau.
5. Greenwood, S. (2021, March 15). About one-in-four U.S. hispanics have heard of Latinx, but just 3% use it. Pew Research Center's Hispanic Trends Project.
6. Borrell, L. N., & Echeverria, S. E. (2022). The use of latinx in public health research when referencing Hispanic or Latino populations. *Social Science & Medicine*, 302, 114977.
7. María del Río-González, A. (2021). To Latinx or not to latinx: A question of gender inclusivity versus gender neutrality. *American Journal of Public Health*, 111(6), 1018–1021.
8. Larochelle, M. R., Slavova, S., Root, E. D., Feaster, D. J., Ward, P. J., Selk, S. C., Knott, C., Villani, J., & Samet, J. H. (2021). Disparities in opioid overdose death trends by Race/ethnicity, 2018–2019, from the Healing Communities Study. *American Journal of Public Health*, 111(10), 1851–1854.
9. James, K., & Jordan, A. (2018). The Opioid Crisis in Black Communities. *The Journal of law, medicine & ethics : a journal of the American Society of*