



# EXPANDING THE EVIDENCE BASE FOR COLLECTING RACE AND ETHNICITY DATA

Current Health Plan Approaches to Race and Ethnicity Data  
Collection and Recommendations for Future Improvements

**Final Report**

**January 2023**

**Jeni Soucie, MS  
Bryan O. Buckley, DrPH, MPH  
Karri Albanese  
Rachel Harrington, PhD  
Sarah Hudson Scholle, DrPH, MPH**

---

*This project was conducted by the National  
Committee for Quality Assurance (NCQA),  
with support from The Commonwealth Fund.*

# Contents

EXECUTIVE SUMMARY .....	2
BACKGROUND .....	6
METHODS .....	7
Review of Race and Ethnicity Data Collection Toolkits and Related Literature.....	7
Key Informant Interviews .....	8
Advisory Panel.....	9
Limitations .....	9
RESULTS OF HEALTH PLAN INTERVIEWS .....	10
Sources of Race and Ethnicity Data Available to Health Plans .....	10
Key Challenges to Race and Ethnicity Data Completeness.....	12
ADVISORY PANEL FEEDBACK.....	16
Guiding Principles for Addressing Data Challenges .....	16
Key Opportunities for Improving Race and Ethnicity Data.....	18
SYNTHESIS & RECOMMENDATIONS .....	20
Acknowledgments .....	27
Appendix A. Summary and List of Race & Ethnicity Data Collection Resources Consulted.....	28
Appendix B: Interview Guide.....	34
Appendix C: Advisory Panel Members .....	37
Appendix D. Sources for Guiding Principles for Race and Ethnicity Data Collection.....	38

## EXECUTIVE SUMMARY

The COVID-19 pandemic provided a stark reminder that in the United States there are major racial and ethnic gaps in health care access, quality and outcomes. The pandemic also uncovered critical deficits in health care data on race and ethnicity that constrain efforts to advance health equity. Incomplete and inaccurate race and ethnicity data create significant challenges for population health surveillance, measurement of disparities and efforts to promote more equitable outcomes. This project builds on previous work assessing the quality and completeness of race and ethnicity data collected by health plans.

NCQA conducted a targeted environmental scan and key informant interviews to identify current methods for obtaining race and ethnicity data, and strategies to improve these methods. We reviewed findings with a multistakeholder advisory panel. This report summarizes data sources used by a sample of health plans today; key challenges to obtaining complete and accurate data reported by plans; and input from our advisory panel on guiding principles and opportunities for improving race and ethnicity data available to plans. Based on our findings, we make recommendations for key next steps for future work.

### **Data Collection Practices and Challenges for Health Plans**

Health plans use various approaches to gathering data on race and ethnicity, including obtaining enrollment data from government or private payers; collecting data through interactions with plan members; acquiring data from clinical sources serving plan members; and imputing data based on name and place of residence.

A number of themes arose from our interviews, including several that derive largely from new opportunities to share data collected across settings:

- Categories for documenting and reporting race and ethnicity.
- Rationale for data collection.
- Access to employer data.
- Opportunities to improve government enrollment data.
- Consent/transparency of data use.
- Source of truth when there are multiple data sources.
- Frequency of data collection and verification.
- Imputation uses and methods.

### **Multistakeholder Input on Principles and Opportunities**

Our multistakeholder advisory panel recommended that efforts to improve the collection of data on race and ethnicity be grounded in the following principles:

- Prioritize respect for individuals—autonomy, privacy and ability to self-identify using as many categories as needed to accurately represent identity.
- Focus on the goal of achieving equitable outcomes.
- Engage communities.

The advisory panel discussed four areas of opportunity to make the greatest progress toward collecting complete and accurate data.

*1. Categories for documenting and reporting race and ethnicity:* The advisory panel acknowledged the importance of patient autonomy to self-identify race and ethnicity in a meaningful way, but held diverse opinions on the practical implications of how to use data when patients self-identify as multiple races or ethnicities. Panel members also emphasized the importance of providing scripts for staff to explain and answer patient questions about why the information is being collected. The panel noted the need for data collection strategies that allow individuals to self-identify, and data management approaches that make it possible for self-reported information to be converted into a format that allows meaningful analysis.

*2. Access to employer data:* Advisory panel members expressed differing views on the opportunities for obtaining data from employers from employee equal-opportunity records, health risk assessments or open enrollment forms. The panel agreed on the need to protect and respect employees' autonomy. Some panel members were skeptical about employers sharing data with health plans, and expressed concerns about employees' general lack of trust in health plans. Other members felt that employees would find it acceptable if the plan emphasized the goal of advancing equity.

*3. Source of truth when there are multiple data sources:* The advisory panel agreed on the importance of creating a process for identifying a source of truth, or the most valid data source, when there are conflicting data sources of unknown validity. They noted the limited availability of contextual information about race and ethnicity data, such as the date and method of data collection. The panel also emphasized the importance of engaging a diverse group of stakeholders when creating the process for identifying the most valid data source, to understand contextual factors and minimize bias.

*4. Frequency of data collection and verification:* The advisory panel emphasized the importance of creating ways to give patients autonomy over their data, the need to normalize asking patients about their race and ethnicity and updating the data regularly. Panel members had mixed opinions on following up with individuals who decline to respond to questions about their race and ethnicity.

## **Synthesis of Findings**

Making progress toward equity requires more complete and accurate data on race and ethnicity to guide care, quality improvement and national monitoring and benchmarking. The amount of race and ethnicity data available today varies widely, and information about the data's accuracy

is limited. The consensus is that self-reporting is the gold standard for collecting data; thus, health plans are leveraging a range of sources to increase member self-reported race and ethnicity data. Yet, the high level of missing data means that imputation will continue to be used in parallel as the field works to increase the availability of self-reported data.

The critical questions for the field are how to make responsible use of data that are available today, and what steps are needed to improve the accuracy and completeness of data for the future. This means acknowledging the reality—self-reported data are the benchmark—and that improved methods for imputation are needed. Focusing only on self-reporting has the potential to delay or set aside equity efforts; focusing only on imputation methods may blunt the urgency of implementing efforts to obtain self-reported data.

The use of multiple data sources also brings new challenges to synthesis and prioritization of data sources, and raises additional questions about how to support patients' ability to understand and govern sharing and use of their data. The advisory panel advocated for increased focus on ensuring that data collection allows individuals to identify in the most meaningful way, and emphasized that the conversation needs to be normalized so that documenting and verifying race and ethnicity becomes routine, along with efforts to advance equity.

### **Opportunities and Recommendations**

Based on our findings, we identified three opportunities to address the most pressing challenges to improving race and ethnicity data available to health plans: 1.) Define a taxonomy of use cases for collecting, documenting, using and sharing race and ethnicity data; 2.) Create a guide for using interoperability standards to support the collection, documentation, use and sharing of electronic race and ethnicity data; 3.) Improve imputation methods for addressing missing and conflicting data. These recommendations focus on the growing opportunities to collect and share data across health care settings, and address key challenges raised by plans—specifically regarding categories for data collection, establishing a source of truth when multiple data sources are combined, the process for updating and verifying data, and methods for evaluating the accuracy of imputation. Efforts should also use a community-informed process involving patients and community stakeholders, and consider local context, populations and conditions.

**First, we recommend defining a taxonomy of use cases to explain how data should be used and to specify what is permissible and acceptable from the perspective of the health care entity, patients and community members.** Use cases should include care delivery, population management, quality improvement, patient safety and equity reporting and accountability.

**Second, we recommend coordinating a diverse group of stakeholders to develop guidance for implementing interoperability standards that would support the collection, use and sharing of electronic race and ethnicity data for equity reporting.** Fast Healthcare Interoperability Resources (FHIR<sup>®</sup>) or United States Core Data for Interoperability (USCDI) may be suitable interoperability standards to implement widespread adoption. An implementation guide

helps software engineers apply equity reporting consistently across settings and data sources. We propose that future work explore the overlap with use cases described in the first recommendation and in current implementation guides, specifically, the *SDOH Clinical Care FHIR Implementation Guide*. A new or revised implementation guide using a community-focused process could build on that guide and discuss use cases not covered.

**Third, and concurrently, we recommend efforts to improve trust and confidence in imputation methods when such data are needed.** A multistakeholder panel of statistical methodologists, health care decision makers, community members and medical ethicists should identify the technical methods, processes for implementation and guardrails for use that support responsible equity analyses using imputed data.

In this project, we learned that health plans leverage a wide range of sources to increase self-reported race and ethnicity data for members, and that they face a variety of challenges. Plans and other stakeholders expressed the need for guidance on collecting, using and sharing race and ethnicity data. To minimize delays in moving equity efforts forward, NCQA's proposed recommendations will support organizations as they work to increase the availability of self-reported data and suggest ways to increase trust and confidence in imputed data to fill in gaps. as health plans increase their ability to collect, use and share race and ethnicity data, it is critical that they also work with providers and the larger health care ecosystem to ensure that their new efforts complement and are implemented in parallel with established and ongoing efforts.

## BACKGROUND

The COVID-19 pandemic provided a stark reminder that in the United States there are major racial and ethnic gaps in health care access, quality and outcomes. Not only were rates of infection, hospitalization and death greater among racial and ethnic minority groups, but the pandemic also exacerbated preexisting health disparities—for example, individuals with chronic conditions were at increased risk of severe COVID-19 infection, and the prevalence of these conditions was already higher among vulnerable populations. The pandemic also uncovered critical deficits in health care data on race and ethnicity, constraining efforts to advance health equity.

Data from calendar year 2019 show that the completeness of race and ethnicity data varies by type of plan and has changed little from previously published data from 2015.<sup>1</sup> Only about a quarter of Commercial plans reported having race data for most of their members. Medicare plans report more complete data on race and ethnicity - this is because the Centers for Medicare and Medicaid Services (CMS) provide race and ethnicity in enrollment files. The race and ethnicity variable is based on data obtained from Social Security Administration records and a validated imputation approach developed by the Research Triangle Institute uses surname and residence to improve the identification of some ethnic groups<sup>2</sup>. The accuracy and completeness of Medicaid data vary between states. Of the 46 states identified as having usable data in 2020, the DQAtlas (a tool that supports the exploration of the quality and usability of Medicaid data) classified the completeness of data in 16 states as highly concerning, and in another 15 states as of medium concern.<sup>3</sup> Given the level of missing data, imputation is often used to provide vital insights at the population level. Statistical imputation of race and ethnicity can provide vital insights where race and ethnicity data are incomplete since data are unlikely to be missing at random. Many imputation approaches use individuals' name and place of residence to infer race and ethnicity (such as the RTI algorithm for Medicare data) mentioned above, however, there are concerns about the accuracy of proprietary methods without transparent data on validity.<sup>4</sup>

Self-identification of race and ethnicity is strongly preferred, where feasible—this is expressed in both the 1997 Office of Management and Budget (OMB) and 2011 Department of Health and Human Services (HHS) race and ethnicity data collection standards.<sup>5,6</sup> One study found that survey data were more likely to contain complete race and ethnicity data than electronic health records (EHR) for the same patient, perhaps because patients are more likely to provide data when they can record the information themselves.<sup>7</sup>

A number of resources to advance the goal of collecting self-reported race and ethnicity data are available. For example, the Health Educational Trust Disparities Toolkit (HRET) created a toolkit for organizations seeking best practices in race and ethnicity data collection.<sup>8</sup> The toolkit offers strategies and best practices for addressing patient questions about reasons for collecting

race and ethnicity data, but it may need to be updated, following the experiences of the COVID-19 pandemic and the resulting focus on racial justice and health equity.

This project builds on previous work to assess the quality and completeness of race and ethnicity data collected by health sector stakeholders, funded by the Commonwealth Fund and conducted by Grantmakers in Health and the National Committee for Quality Assurance (NCQA).<sup>9</sup> In addition to documenting data deficiencies described above, that project also highlighted a need for technical support among health care organizations that struggle to identify and implement best practices in collecting race and ethnicity data. The project also closely parallels NCQA's current efforts to advance stratified quality reporting by race and ethnicity.

NCQA has begun to introduce scoring incentives for Health Plan Accreditation ratings to incentivize the collection and use of direct, self-reported race and ethnicity data.<sup>10</sup> This is aligned with the requirements of NCQA's Health Equity Accreditation, which advocates for plans to achieve 80% direct data or else supplement with indirect imputation methods.<sup>11,12</sup> Notably, NCQA's definition of direct data includes Medicare enrollment data, which are supplemented with imputed values. Reporting specifications for the Health Effectiveness Data and Information Set (HEDIS®)<sup>1</sup> measure race and ethnicity stratification and note that if both direct and indirect values are available for a patient, the direct self-reported values for race and ethnicity should be prioritized for use.

To advance this goal, NCQA conducted a targeted environmental scan and key informant interviews with health plans to identify current collection methods for race and ethnicity data, and strategies to improve these methods. We reviewed our findings with a multistakeholder advisory panel. This report summarizes the sources of race and ethnicity data used by health plans today; challenges to obtaining complete and accurate data; and advisory panel input on guiding principles and opportunities for improving race and ethnicity data available to plans. In addition, we make recommendations for next steps.

## METHODS

### Review of Race and Ethnicity Data Collection Toolkits and Related Literature

We conducted a targeted environmental scan to identify toolkits, guides and other documents containing best practices for health care entities that collect race and ethnicity data, with a specific interest in resources for health plans. The team used a recently published resource from the Centers for Medicare & Medicaid Services (CMS), *Inventory of Resources for Standardized Demographic and Language Data Collection*, to identify relevant sources.<sup>13</sup> The team then

---

<sup>1</sup> HEDIS® is a registered trademark of the National Committee for Quality Assurance (NCQA).



conducted a general search to identify additional sources not included in CMS’s inventory. Sources submitted by the advisory panel were also included in the scan. In total, 42 sources were included in the scan, with publication dates ranging from 2001–2022.

From the initial review of resources, we sought to identify the key features of collection strategies, including staff and modality involved in data collection, race and ethnicity categories used and information provided to individuals about the rationale for data collection and how the data will be used.

Appendix A includes a summary of findings and a list of sources.

## Key Informant Interviews

To supplement the targeted environmental scan, we conducted nine key informant interviews with a diverse set of health plans. The goal of the interviews was to better understand plans’ current methods of collecting race and ethnicity data, and to identify opportunities to improve these methods. In particular, we asked about plans’ health equity efforts; the availability and quality of race and ethnicity data; their use of publicly available toolkits to suit specific needs of covered populations; and challenges or barriers to achieving complete and accurate data collection. Appendix B includes the full set of interview questions.

We used a purposive sampling approach to identify plans for interviews, with the goal of including different geographic regions and plan types. All organizations interviewed offer plans in more than one market segment, and often in different states or regions. We use the term “plan” to refer to the overall organization and we note when the insights refer to specific market segments for the organization.

Of the nine plans interviewed, five were national health plans, two served the southern region of the United States, one was based in the West and one was based in the Northeast (Table 1). All nine plans operate in the Medicare market, eight plans in Medicaid, seven in Exchange and five in commercial. All plans are NCQA Accredited; four plans have at least one operating unit that previously earned NCQA’s Distinction in Multicultural Health Care, a program that evolved into Health Equity Accreditation in 2021.

**Table 1: Characteristics of Health Plans Participating in Interviews**

Plan Attributes	Health Plans								
	1	2	3	4	5	6	7	8	9
Region	West	Northeast	National	National	South	South	National	National	National
<b>Product Lines Represented</b>									
Medicare (9)	X	X	X	X	X	X	X	X	X
Medicaid (8)	X		X	X	X	X	X	X	X

Exchange (7)	X		X	X	X		X	X	X
Commercial (5)	X	X	X	X		X			X
<b>NCQA Distinction in Multicultural Healthcare (4)*</b>				X	X		X	X	
<p>*NCQA awards Distinction in Multicultural Healthcare to organizations that meet or exceed standards in providing culturally and linguistically appropriate services (CLAS). NCQA evaluates how well an organization complies with standards for collecting race/ethnicity and language data, providing language assistance, cultural responsiveness, quality improvement of CLAS and reduction of health care disparities. As of 2021, this distinction had evolved into NCQA’s Health Equity Accreditation. Not all plan products in national organizations have earned NCQA Distinction.</p>									

## Advisory Panel

The NCQA project team convened a group of diverse stakeholders in July 2022 for a single 2-hour meeting to solicit feedback and prioritization of opportunities to improve the completeness and accuracy of race and ethnicity data available.

We selected stakeholders that represented a variety of perspectives and had expertise in collecting race and ethnicity data and health equity. (Appendix C contains a complete list of advisory panel members.) During the meeting, we shared preliminary findings from the environmental scan and health plan interviews. We invited advisory panel members to prioritize opportunities for making tangible improvements in obtaining complete and accurate race and ethnicity data for reporting and improving equity. Our goal was to understand stakeholder perspectives rather than reach consensus on each topic. We asked the panel to place the discussion in the context of ethical principles commonly used in health care.<sup>14,15,16</sup>

## Limitations

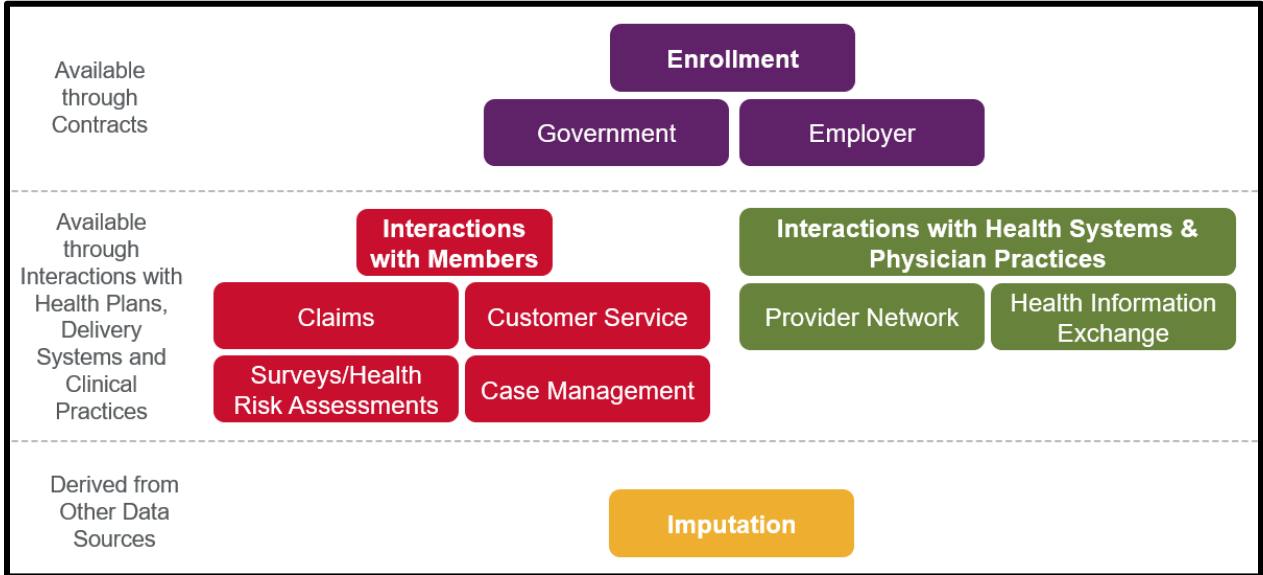
This project included a targeted search of publicly available toolkits for collecting race and ethnicity data, and a limited number of interviews with plans. While we attempted to sample plans from a mix of geographic regions and insurance markets (e.g., Medicare Advantage, Medicaid, commercial), it is not likely that we had a sufficient sample size to achieve exhaustive coverage in each market. As a result, we are unlikely to capture the full range of data collection approaches in the health insurance industry. Additionally, given the breadth of possible modifications to data collection practices, our approach focuses on a set of principles for improvement, supported by examples, rather than on producing a comprehensive, systematic review or exhaustive list of strategies. Although we sought to reflect multiple perspectives of advisory panel participants, the limited size and the single meeting prevented broad participation and deep inquiry into some topics.

# RESULTS OF HEALTH PLAN INTERVIEWS

## Sources of Race and Ethnicity Data Available to Health Plans

The nine health plans in our interview group noted several sources for obtaining race and ethnicity data: 1.) data obtained at enrollment data from the government or private payer; 2.) collecting data through interactions with plan members; 3.) data obtained from clinical sources serving members; and 4.) imputed data based on name and place of residence (including data obtained from vendors) (Figure 1).

**Figure 1. Race and Ethnicity Data Sources Identified by Interviewed Health Plans**



All nine plans interviewed receive most race and ethnicity data through enrollment data provided by state and federal government sponsors. Some plans with commercial contracts reported receiving data from employer sponsors, but this was not consistent and was often for a limited portion of the member population.

Plans also leverage member interactions with the plan or with health care delivery systems. Especially among network model plans, interviewees noted that they take advantage of member interactions with the plan’s web resources, call centers or care management activities to obtain or update race and ethnicity data. Five plans also noted special efforts to field surveys to obtain race and ethnicity data.

Respondents from a plan with an integrated delivery system indicated they rely primarily on data from the delivery setting. In some cases, network plans also had access to data from their provider network’s clinical data systems, but data were limited. Most plans noted a desire to

increase their ability to obtain data either directly from their provider networks or from regional health information exchanges.

All interviewees indicated that they use imputation methods to some extent, given the current incomplete data on race and ethnicity. One plan described extensive efforts to validate and improve imputation methods using data collected through surveys. Other plans indicated they are focused on obtaining data from other sources. Interviewees acknowledged the benefits of imputation for understanding disparities at the population level, and agreed that imputation should not be used for clinical decision making at the individual level.

In general, interviewees reported that plans were exploring multiple avenues for increasing their access to data on race and ethnicity. Their approaches differed widely, however, particularly based on the concentration of government-sponsored contracts and connections to the delivery system (Table 2). For example, a plan with both integrated delivery system and network contracts noted differing availability of data; a national Medicaid plan reported efforts to address the variation in completeness of government enrollment data; and two plans focusing on commercial and Medicare Advantage markets reported limited efforts to use clinical data, and sought other approaches to improve data.

**Table 2: Key Findings from Select Health Plan Interviews**

Health Plan	Key Findings
Integrated Delivery System/Network Plan	<p><b>Current Efforts:</b></p> <ul style="list-style-type: none"> <li>• Markets with integrated delivery systems have race and ethnicity data for over 90% of their members, while markets with network plans have somewhat lower completion rates.</li> <li>• The plan receives race and ethnicity data from selected employer groups.</li> </ul> <p><b>Future Efforts or Considerations:</b></p> <ul style="list-style-type: none"> <li>• Explore additional touchpoints, such as preappointment surveys via mobile app and website, to increase data completeness.</li> <li>• Determine a data hierarchy when there are conflicting data.</li> </ul>
National Medicaid Plan	<p><b>Current Efforts:</b></p> <ul style="list-style-type: none"> <li>• Most race and ethnicity data comes from state enrollment files. Data completeness varies widely, with some states providing data for over 90% of enrollees, while others provide data for less than 50%.</li> <li>• The plan seeks to collect granular data, but is mindful of the need to roll up the data to federal OMB categories for reporting.</li> </ul> <p><b>Future Efforts or Considerations:</b></p> <ul style="list-style-type: none"> <li>• Leverage health information exchanges to increase the amount of available race and ethnicity data.</li> </ul>
National Commercial and Medicare Advantage Plan	<p><b>Current Efforts:</b></p> <ul style="list-style-type: none"> <li>• The plan’s Medicare and Medicaid product lines have race and ethnicity data for between 60% and 70% of enrollees.</li> <li>• The plan’s commercial product line has limited race and ethnicity data available, despite active collection efforts through employers and direct engagement with members.</li> </ul> <p><b>Future Efforts or Considerations:</b></p> <ul style="list-style-type: none"> <li>• Minimize use of imputed data beyond for internal analyses.</li> </ul>

Health Plan	Key Findings
	<ul style="list-style-type: none"> <li>• Encourage employers to share data with health plans.</li> </ul>
Regional Commercial and Medicare Advantage Plan in the Northeast	<p><b>Current Efforts:</b></p> <ul style="list-style-type: none"> <li>• The plan deployed an ongoing survey in its member app to collect race and ethnicity data.</li> <li>• The plan also conducted a randomized control trial to enhance the effectiveness of its paper survey/mailed design to collect self-reported race and ethnicity data.</li> </ul> <p><b>Future Efforts or Considerations:</b></p> <ul style="list-style-type: none"> <li>• Increase self-reported data from essentially 0% in December 2020 to 30% by December 2022.</li> <li>• Use these data to improve imputation efforts.</li> </ul>
OMB = Office of Management and Budget	

## Key Challenges to Race and Ethnicity Data Completeness

The following topics were identified across health plan key informant interviews:

- Categories for documenting and reporting race and ethnicity.
- Rationale for data collection.
- Access to employer data.
- Opportunities to improve government enrollment data.
- Consent/transparency of data use.
- Source of truth when there are multiple data sources.
- Frequency of data collection and verification.
- Imputation uses and methods.

Some topics—for example, race and ethnicity categories, member autonomy and imputation—are addressed in existing toolkits and resources we reviewed (Appendix A). However, some concerns raised by health plans suggest new perspectives (e.g., which categories to use), and some topics—such as “source of truth”—are new concerns developing from the greater opportunity for data collection and sharing through interoperable clinical data. In Table 3, we provide background information on the topic and a summary of concerns noted by plans.

**Table 3: Challenges in Race and Ethnicity Data Collection and Use, Background and Summary of Health Plan Concerns**

Topic	Summary of Health Plan Concerns	Additional Context
<p><i>Categories for documenting and reporting race and ethnicity</i></p>	<p>While some plans use current OMB categories for both data collection and reporting, others noted the importance of having more granular data that would allow them to characterize the needs of their population. Plans also raised questions about how to aggregate categories for people reporting more than one race or ethnicity.</p>	<p>The OMB sets standards for collecting, maintaining and presenting race and ethnicity data for federal programs. There have been a number of suggestions for improving the categories. For example, a 2016 U.S Census Study recommended updating the response categories to include “Middle Eastern or North African” and using the top six disaggregated subcategories for each race and ethnicity</p> <p>The White House recently announced plans to update the standards by Summer 2024. A working group that was convened to guide updates will develop recommendations on topics including:</p> <ul style="list-style-type: none"> <li>• Updates to the minimum reporting categories and how to address detailed race and ethnicity groups.</li> <li>• Whether to update question format, terminology and wording as well as instructions for the questions.</li> <li>• Improvements to the current guidance for collecting and reporting race and ethnicity, including situations where self-identification is not possible.<sup>17</sup></li> </ul> <p>In addition to the OMB categories for race and ethnicity, the CDC developed and maintains a code set of over 900 options for race and ethnicity. The United States Core Data for Interoperability, a standardized set of data classes and elements for nationwide interoperable health information exchange, requires both OMB and CDC race and ethnicity standards in the latest version (V3).<sup>18,19</sup></p> <p>The code set developed by the CDC also is required by the 2015 Office of National Coordinator for Health IT standards for certified electronic health records, and is compatible with the Fast Healthcare Interoperability Resources (FHIR®), a standard for the electronic exchange of health care information that provides a common way to represent data.<sup>20</sup></p>

Topic	Summary of Health Plan Concerns	Additional Context
<i>Rationale for data collection</i>	Plans sought guidance on how to best inform members of the rationale for requesting their race and ethnicity data. One plan conducted a randomized trial to identify the most effective way of collecting the information via survey, and found that a brief explanation (as opposed to a lengthier one) why the information was being requested led to a higher response rate.	Toolkits focused on race and ethnicity data collection often suggest that entities focus on explaining how data will be used to improve quality of care, how data will be protected and that providing the data is voluntary and will not affect the care received. For example, the Institute for Healthcare Improvement created <i>A Roadmap for Race, Ethnicity and Language Data Collection and Use in Connecticut</i> , which provides guidance on framing the importance of collecting the data to patients and training staff on this topic. <sup>21</sup>
<i>Access to employer data</i>	Only one plan stated that it receives race and ethnicity data through employer data. Three plans expressed interest in leveraging employer files to increase their race and ethnicity data.	Employers collect information on employee race and ethnicity for reporting to the Equal Employment Opportunity Commission, but these data are often stored using non-OMB categories. A recent analysis by the American Benefits Council found that there is no federal prohibition on employers collecting, storing, using, disclosing or sharing race and ethnicity data with group health plans. However, the report noted that laws around race and ethnicity data for state insurance plans vary. Some states prohibit the collection of race and ethnicity during enrollment as a safeguard against racial discrimination when offering health insurance, but there is no restriction on collecting the data after enrollment, and some states require it. <sup>22</sup>
<i>Opportunities to improve government enrollment data</i>	Plans noted that government enrollment data are a consistent source of member race and ethnicity data, but the files may have limitations. For example, a recent report from the Department of Health and Human Services, Office of the Inspector General found that the data are less accurate for some groups, especially those who identify as American Indian/ Alaska Native, Asian/Pacific Islander or Hispanic. <sup>23</sup>  Plans also expressed frustration with monthly Medicaid file updates, which override updated enrollee information collected by the plan.	The RTI race and ethnicity variable used in Medicare’s enrollment data is based on are based on Social Security Administration records and a validated imputation approach uses surname and residence to improve the identification of some ethnic groups <sup>24</sup> . Medicare data do not fully align with the OMB categories or guidelines, because race and ethnicity are combined and because the Asian and Pacific Islander categories are combined. <sup>25,26</sup>  Medicaid enrollment data include race and ethnicity information, although data completeness varies across states. Of the 46 states identified as having usable data in 2020, the DQAtlas classified the completeness of data in 16 states as highly concerning. <sup>27</sup>

Topic	Summary of Health Plan Concerns	Additional Context
<b><i>Consent/transparency of data use</i></b>	<p>Plans questioned what permissions would be needed to share race and ethnicity data, and sought guidance on how to navigate various situations; for example:</p> <ul style="list-style-type: none"> <li>• Could race and ethnicity data collected by a provider be shared with the health plan?</li> <li>• Should patients need to give permission for the provider to share the data?</li> <li>• If the patient grants permission, how does the plan ensure that the patient has autonomy over the data and can rescind permission or update the list of entities that can access the data?</li> </ul>	<p>As technology and interoperability improve, entities are more easily able to share information. This has raised questions about how and when individuals should provide consent for information to be shared across entities within the health care space. The Gravity Project created the <i>SDOH Clinical Care FHIR Implementation Guide</i>, which includes processes for obtaining patient consent to share data.<sup>23</sup> Additionally, the Chicago Area Patient Centered Outcomes Research Network (CAPriCORN) created a cross-institutional data infrastructure capable of sharing electronic health records and outcomes data for over one million patients, which required collecting consent from each patient.<sup>28,29</sup> Learnings from both of these efforts could help organizations as they develop their own processes for obtaining patient consent for sharing data.</p>
<b><i>Source of truth when there are multiple data sources</i></b>	<p>Plans frequently ask for guidance on identifying a source of truth when conflicting sources of unknown validity are available. Knowing that members’ responses may be influenced by how the question is asked, who is asking the question, the responses available, and the setting in which they are asked makes it difficult to identify a source of truth. For example, plans questioned whether using the most recent data was most appropriate, or if certain settings should be prioritized over others.</p>	<p>As interoperability improves and health care entities can extract data from multiple sources, it is increasingly important to create a process for determining a source of truth when there are conflicting data sources.</p> <p>Quality organizations have begun to propose guidelines for prioritizing data sources. For example, in HEDIS MY 2023 Volume 2, NCQA offers the following guidance for entities to help prioritize data sources based on anticipated accuracy:</p> <ul style="list-style-type: none"> <li>• Use specific categories over nonspecific categories.</li> <li>• Use the most frequent or consistently reported category.</li> <li>• Select data with clear provenance (source and method of collection) over data without clear provenance.</li> </ul>
<b><i>Frequency of data collection and verification</i></b>	<p>Plans sought guidance on how often members should be asked to provide their race and ethnicity, how often members should update or confirm existing information and how often members who declined to answer or did not previously provide the information should be followed up with to obtain the information.</p>	<p>As health care entities and health plans increasingly look to collect race and ethnicity data, questions have arisen about how often to ask individuals to provide or update their information, and what to do if an individual declines to provide the information.</p>



Topic	Summary of Health Plan Concerns	Additional Context
<i>Imputation uses and methods</i>	All plans indicated that they use imputation to understand populations with missing data. While plans agreed that imputation could be a valuable tool, they sought guidance on when imputation is appropriate.	Data imputation predicts missing values by estimating the probability that patients will self-identify in each race and ethnicity category (typically the OMB-1997 categories) if it is voluntary. <sup>30</sup> Well-known imputation algorithms include the RTI algorithm and the Medicare Bayesian Improved Surname Geocoding, which are both used by Medicare to improve available data. <sup>31,32</sup> Imputation can be helpful for closing information gaps at the population level, but may not be appropriate for decision making at the individual patient level.
<p>OMB = Office of Management and Budget          CDC = Centers for Disease Control and Prevention          RTI = Research Triangle Institute</p>		

# ADVISORY PANEL FEEDBACK

## Guiding Principles for Addressing Data Challenges

Prior to discussing the challenges facing plans, we asked the advisory panel to consider the ethical principles that should govern processes for collecting and using race and ethnicity data. We offered several examples for consideration:

- *Human subject research ethics*, as documented in the 1979 Belmont Report, which govern federal protection of human subjects in biomedical and behavioral research.<sup>33</sup>
- *Data ethics*, as documented in the 2020 Federal Data Strategy Data Ethics Framework, which describe the norms of behavior that promote appropriate judgment and accountability when acquiring, managing or using data, with the goals of protecting civil liberties, minimizing risks to individuals and society and maximizing the public good.<sup>34</sup>
- *Equity principles*, including the principles articulated by the 2022 report of the National Academy of Science, Engineering and Medicine on Measuring Sex, Gender Identity and Sexual Orientation, and principles suggested in the 2021 report by Grantmakers in Health and NCQA for improving data on race and ethnicity.<sup>35</sup>

The panel offered a number of reflections on these ideas. To reflect the ideas offered by the panelists, we created a word cloud to visualize the recommendations of the advisory panel (Figure 2). The size of each word in the figure shows how frequently panelists mentioned this topic.<sup>36</sup>

- ***Prioritize respect for individuals—autonomy, privacy, ability to self-identity***

Respecting individuals' autonomy was a key theme of the panel discussion on principles. The panel noted the importance of addressing the tension between the need for information and individuals' willingness to share it, and emphasized the need to allow individuals the opportunity to self-identify, to ensure that their privacy is respected, and to provide transparency on how the data will be used.

- ***Focus on the goal of achieving equitable outcomes***

The panel encouraged a focus on justice in the collection and use of data on race and ethnicity (e.g., demonstrate how race and ethnicity data are used to identify and close equity gaps) and discussed opportunities for making data available to participants and the community.

- ***Engage Communities***

The panel discussed the importance of engaging with communities throughout the development of methods for data collection and use, noting the importance of allowing flexibility to reflect local needs.

**Figure 3: Principles for Race & Ethnicity Data Collection Identified by Advisory Panel Members**



## Key Opportunities for Improving Race and Ethnicity Data

Panel members prioritized four topics for discussion, based on their chance to have the greatest impact on improving the completeness and accuracy of data:

- Categories for documenting and reporting race and ethnicity.
- Access to employer enrollment data.
- Source of truth when there are multiple data sources.
- Frequency of data collection and verification.

### Categories for documenting and reporting race and ethnicity

The advisory panel acknowledged the importance of patient autonomy to self-identify race and ethnicity in a meaningful way, but some raised concerns about how this would affect data collection and population-level aggregation. They noted that race and ethnicity are not biological categories, and that both social and individual perspectives on these categories have changed since we started collecting these data and are likely to continue to change over time.<sup>37</sup> One member’s health system asks patients to select up to three categories for race and ethnicity, striking a compromise between the patient’s autonomy and the plan’s reporting needs. A member responded that collecting meaningful, accurate data involves allowing patients to answer truthfully and accurately. Another member noted that their health plan encourages members to “select all that apply” as the best approach, if the plan is transparent about how the information is used for reporting and analyses.

Panel members also emphasized the importance of providing scripts for staff to explain and answer patient questions about why the information is being collected. One member noted that their plan designed scripts to explain to patients why the information is collected, and designed a “frequently asked questions” guide to help staff answer patient questions.

The panel noted the need for data collection strategies that allow individuals to self-identify, and data management approaches that make it possible for self-reported information to be converted into a format that allows meaningful analysis. Although categories such as “two or more races” or “multiracial” may be sufficient for reporting, they are not useful for understanding inequities among specific populations. One panel member recommended data management principles that preserve original responses so respondents can “live” in the identity they choose, but have data conversions that identify a person as selecting more than one race, or additional combinations that could be useful for outcomes segmentation. Another panelist wondered if questions about individuals’ experience of discrimination in the health care system should be included along with questions about race and ethnicity. The panel encouraged future research to explore how discrimination plays a role in selecting racial and ethnic categories, and agreed that future updates or changes should be in alignment with the federal government.

### **Access to employer data**

Advisory panel members expressed differing views on the opportunities for obtaining data from employers from employee equal-opportunity records, health risk assessments or open enrollment forms. The panel agreed on the need to protect and respect employees’ autonomy. Some panel members were skeptical about employers sharing data with health plans, and expressed concerns about employees’ general lack of trust in health plans. Other members felt that employees would find it acceptable if the plan emphasized the goal of advancing equity. A panel member noted that employers could play an important role in promoting equity by supporting data collection and sharing. A “how to” guide could be helpful for employers and health plans as they navigate the complexities of employer/health plan data sharing.

### **Source of truth when there are multiple data sources**

The advisory panel agreed on the importance of creating a process for identifying a source of truth when there are conflicting data sources of unknown validity. Current methods do not easily allow collecting and sharing contextual information (metadata), such as whether data are self-reported, who collected the data, and how, or in what setting. These metadata can play a critical role in prioritizing certain sources over another – for example, determining whether to replace older responses with newer data (or data based on self-identification versus unknown source) and to reconcile data across disparate data management systems. One panel member noted that they prioritize data they collected because they can guarantee all aspects of the data collection process. Another member noted the importance of engaging a diverse group of stakeholders when developing a process for identifying a source of truth.

### **Frequency of data collection and verification**

The advisory panel emphasized the importance of creating ways to give patients autonomy over their data, the need to normalize asking patients about their race and ethnicity and updating the data regularly. The advisory panel had mixed opinions about following up with patients who decline to respond to questions about their race and ethnicity. One panel member noted that patients who choose not to respond should be asked again at every visit, because their willingness to respond may change. Another member suggested that a different avenue for data collection could be used (i.e., a provider should still ask, if a patient declines to provide the information at enrollment).

The panel emphasized the importance of creating ways for patients to review, verify and update their information. For example, this information could be confirmed during enrollment, similar to how organizations ask if phone numbers or addresses have changed since the previous year. Normalizing the collection and verification of the information will allow health care entities to move from questions about how to collect the data, to how to use the data to advance health equity.

### **Imputation use and methods**

Although imputation was not included as a formal discussion topic, panel members raised questions and comments about its use throughout the conversation. Some panel members noted that imputation is necessary and will be used for the foreseeable future, and that it can enhance data when used correctly. Others worried that potential inaccuracies in imputed data could hamper efforts to address health inequities, and cautioned that the use of imputation should not lessen the urgency to improve self-reported data collection.

## **SYNTHESIS & RECOMMENDATIONS**

Making progress toward equity requires more accurate and complete data on race and ethnicity to guide care, quality improvement and national monitoring and benchmarking. The amount of race and ethnicity data available today varies widely, and information about the data's accuracy is limited. The consensus is that self-reporting is the gold standard for collecting data; thus, health plans are leveraging a range of sources to increase member self-reported race and ethnicity data. Yet, the high level of missing data means that imputation will continue to be used in parallel as the field works to increase the availability of self-reported data.

The critical questions for the field are how to make responsible use of data that are available today, and what steps are needed to improve the accuracy and completeness of data for the future. This means acknowledging reality—self-reported data are the benchmark—and that improved methods for imputation are needed. Focusing only on self-reporting has the potential to delay or set aside equity efforts; focusing only on imputation methods may blunt the urgency of implementing efforts to obtain self-reported data.

The use of multiple data sources brings new challenges to the synthesis and prioritization of data sources, and raises additional questions about how to support patients' ability to understand and govern sharing and use of their data. The advisory panel advocated for increased focus on ensuring that data collection allows individuals to identify in the most meaningful way, and emphasized that the conversation needs to be normalized so that documenting and verifying race and ethnicity becomes routine, along with efforts to advance equity.

Our findings are consistent with other recent work. Participants from a summit hosted by the Urban Institute, Deloitte and the American Benefits Council echoed the need for updated race and ethnicity categories that encourage individuals to identify race and ethnicity in a way that is most meaningful for them.<sup>38</sup> However, stakeholders had varied opinions on what this looks like in practice. It can be challenging to strike a balance between allowing patients autonomy to self-identify in a meaningful way, but also in a way that can be easily aligned with reporting requirements. Recent recommendations emphasize the need for multistakeholder involvement and standard approaches for data collection, validation and governance.<sup>39</sup>

Based on the panel discussion, we identified three opportunities for national efforts to address the most pressing challenges for improving race and ethnicity data availability for health plans: 1.) define a taxonomy of use cases for collecting, documenting, using and sharing race and ethnicity data; 2.) create a guide for using interoperability standards to support the collection, documentation, use and sharing of race and ethnicity data; and 3.) improve imputation methods for addressing missing and conflicting data. These recommendations focus on the growing opportunities for collecting and sharing data across health care settings, and address key challenges plans raised, specifically about the categories for data collection, data hierarchy when multiple data sources are combined and the process for updating and verifying data. These efforts should be aligned with federal policymaking activities, such as current efforts to update the OMB standards and the current administration's initiatives on equity. Efforts should also use a community-informed process involving patients and community stakeholders, and consider local context, populations and conditions. Data-driven, community-informed intervention and evaluation can better serve communities.

**Recommendation #1: Define a taxonomy of use cases for collecting, documenting, using and sharing race and ethnicity data.**

First, we recommend defining a taxonomy of use cases to explain how data should be used and to specify what is permissible and acceptable from the perspective of the health care entity, patients and community members. We noticed that differences in methods and perspectives presented during our panel discussion seemed to arise from diverse or unclear expectations for how data will be used. Defining expectations and constraints for explicit use cases for race and ethnicity data could provide transparency for patients and communities, and could guide health care organizations in developing infrastructure and business rules that clarify what is (and what is not) acceptable.

Use cases include care delivery, population management, quality improvement, patient safety and equity reporting and accountability. Conversations with health plans and the advisory panel differentiated between using race and ethnicity data to make population insights versus using the data to make inferences about specific individuals. For example, population-level use may include analyses to identify pockets of inequities, understand population needs and report equity results for comparison with regional or national benchmarks.

It is possible that understanding race and ethnicity could be useful in some clinical care cases; for example, to prevent inequities, such as communication differences that have reportedly contributed to inequity in care for people who give birth, or to balance out potentially biased risk-scoring approaches, such as scoring for kidney disease functioning tests that was recently updated.<sup>40</sup>

Population- and individual-level use cases may differ in their requirements for completeness and source data. For efforts that involve contacting individual patients—in direct care or quality improvement, for example—there is low tolerance for inaccurate race and ethnicity data, and there may be a greater opportunity to verify or obtain complete data directly from individuals. For activities designed to understand the population, there may be value in using imputation methods to better estimate population risk or needs where race and ethnicity data are missing. Basing decisions only on available data could lead to faulty decisions if key patient groups are not adequately represented.

For some use cases, like quality reporting, small numbers may make it difficult to draw conclusions about some groups, and may require higher levels of aggregation. Use cases could describe principles for outlining potential processes and alternatives (e.g., options for exploring patterns for anyone who reports multiple race categories vs. for anyone who reports Black/African-American race, whether or not they select additional race or ethnicity categories).

Use cases should be developed through a community-informed process and involve a variety of stakeholders. A use case could identify stakeholders that are involved, describe assumptions and goals, suggest business rules or principles to follow and identify process flow.

**Recommendation #2: Create a guide for using interoperability standards to support collection, documentation, use and sharing, of electronic race and ethnicity data.**

Second, we recommend coordinating a diverse group of stakeholders to develop guidance for implementing interoperability standards that would support the collection, documentation, use and sharing of electronic race and ethnicity data for equity reporting. Both the health plan interviews and the panel discussion indicated the need for clear standards on dealing with the potential for conflicting information resulting from data gathered in different settings. As yet, there is no guidance on how to collect, document, use and share these data for the specific use cases described above and for equity reporting and improvement.

Fast Healthcare Interoperability Resources (FHIR®) or United States Core Data for Interoperability (USCDI) may be suitable interoperability standards to implement widespread adoption. FHIR defines how health care information can be configured to enable its exchange between different health care systems, regardless of how it is originally stored in those systems.<sup>41</sup> USCDI is a standardized set of health data classes and constituent data elements for nationwide, interoperable health information exchange.<sup>42</sup> The Gravity Accelerator Project has also done pioneering work to make it possible to capture and share consistent information on health-related social needs which supports the collection, documentation, use and sharing of race and ethnicity data.<sup>43</sup>

Guidance, such as a FHIR implementation guide, is now needed to utilize the standards defined by the Gravity Project for an equity measurement and reporting use case. A FHIR implementation guide allows software engineers to implement equity reporting consistently across settings and data sources. FHIR standards already include many of the building blocks needed, and the current system has standards for categorizing race and ethnicity. From there, the stakeholder group could develop a guide that would build on and discuss use cases not covered in the *SDOH Clinical Care FHIR Implementation Guide*.<sup>44</sup> The new guidance should be developed in a community-focused process and in collaboration with other equity-focused efforts, such as the Gravity Project and the Gender Harmony Project.

### **Recommendation #3: Improve imputation methods for addressing missing and conflicting data.**

Third, and concurrently, we recommend efforts to improve trust and confidence in imputation methods when such data are needed. The use of imputation will be critical, particularly in the short- and mid-term, as plans work to increase the self-reported data they have available. It is almost certain that plans will never achieve 100% self-reported data, making it important to account for missing information when making decisions at the population level.

Some imputation methods, like the RTI algorithm used in Medicare enrollment data and the Medicare Bayesian Improved Surname Geocoding (MBISG) 2.1 have published data on their validity in comparison with self-reported race and ethnicity. For example, the MBISG was found to have over 95% concordance with self-reported race and ethnicity data for Asian/Pacific Islander, Black, Hispanic and White race/ethnicity, making it a valuable resource for Medicare race and ethnicity data.<sup>45</sup> However, not all imputation methods have been tested with this level of rigor, and their concordance with self-reported data may not be fully understood, which could lead to mistrust in imputed data. To help substantiate the use of imputed data, we recommend convening a group including statistical methodologists, health care decision makers, community members and medical ethicists to identify the technical methods, processes for implementation and guardrails for use that support responsible equity analyses using imputed data.



## SUMMARY AND NEXT STEPS

Making progress toward equity requires having more accurate and complete data on race and ethnicity to guide care, quality improvement and national monitoring and benchmarking. In this project, we learned that health plans leverage a wide range of sources to increase self-reported race and ethnicity data for members, and that they face a variety of challenges. Plans and other stakeholders expressed the need for guidance on collecting, documenting, using and sharing race and ethnicity data. To minimize delays in moving equity efforts forward, NCQA's proposed recommendations will support organizations as they work to increase availability of self-reported data, and suggest ways to increase trust and confidence in imputed data, to help meet short- and mid-term needs. And as health plans increase their ability to collect, document, use and share race and ethnicity data, it is critical that they also work with providers and the larger health care ecosystem to ensure that their new efforts compliment and are implemented in parallel with established and ongoing efforts.

- 
- <sup>1</sup> NCQA (2021, June). *Health Equity and Social Determinants of Health in HEDIS: Data for Measurement*. The National Committee for Quality Assurance (NCQA). [https://www.ncqa.org/wpcontent/uploads/2021/06/20210622\\_NCQA\\_Health\\_Equity\\_Social\\_Determinants\\_of\\_Health\\_in\\_HEDIS.pdf](https://www.ncqa.org/wpcontent/uploads/2021/06/20210622_NCQA_Health_Equity_Social_Determinants_of_Health_in_HEDIS.pdf).
  - <sup>2</sup> Eicheldinger, C., & Bonito, A. (2008). *More accurate racial and ethnic codes for Medicare administrative data*. *Health care financing review*, 29(3), 27–42.
  - <sup>3</sup> *Race and Ethnicity: DQ Assessments*. (n.d.) DQ Atlas. Medicaid.gov: Keeping America Healthy. Retrieved 2021, from <https://www.medicaid.gov/dq-atlas/landing/topics/single/map?topic=g3m16&tafVersionId=25>.
  - <sup>4</sup> Saynisch, P.A., Storfer-Isser A., Elliott, M., Martino, S.C., Dunning S, & Scholle, S.H. *Imputation of Race and Ethnicity for Medicare Advantage Enrollees Varies Substantially by Method*. In preparation.
  - <sup>5</sup> Office of Management and Budget. (1997, October 30). *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*. The White House. Retrieved September 13, 2021, from <https://obamawhitehouse.archives.gov/node/15626>
  - <sup>6</sup> Office of the Assistant Secretary for Planning and Evaluation. (2011, October 30). *HHS Implementation Guidance on Data Collection Standards for Race, Ethnicity, Sex, Primary Language, and Disability Status*. ASPE. Retrieved September 13, 2021, from <https://aspe.hhs.gov/reports/hhs-implementation-guidance-data-collection-standards-race-ethnicity-sex-primary-language-disability-0>
  - <sup>7</sup> Polubriaginof, F.C.G., Ryan, P., Salmasian, H., Shaprio, A.W., Perotte, A., Stafford, M.M., Hripcsak, G., Smith, S., Tatonetti, N.P., & Vawdrey, D.K. *Challenges with quality of race and ethnicity data in observational databases*. *J Am Med Inform Assoc*. (2019);26(8-9):730-736. doi:10.1093/jamia/ocz113
  - <sup>8</sup> Health Research and Educational Trust. (2011, March). *Improving Health Equity Through Data Collection and Use: A Guide for Hospital Leaders*. Chicago: Health Research & Educational Trust. Retrieved September 13, 2021, from <http://www.hret.org/health-equity/index.shtml>
  - <sup>9</sup> Grant Makers in Health and NCQA. (2021, December). *Improving Data on Race and Ethnicity: A Roadmap to Measure and Advance Health Equity*. Grant Makers in Health. The Commonwealth Fund. <https://www.gih.org/wp-content/uploads/2021/12/GIH-Commonwealth-Fund-federal-data-report-part-2.pdf>.
  - <sup>10</sup> NCQA. (2022). *2023 Health Plan Ratings Required HEDIS®, CAHPS® and HOS Measures*. The National Committee for Quality Assurance (NCQA). [https://www.ncqa.org/wp-content/uploads/2022/04/2023-HPR-List-of-Required-Performance-Measures\\_4.25.2022.pdf](https://www.ncqa.org/wp-content/uploads/2022/04/2023-HPR-List-of-Required-Performance-Measures_4.25.2022.pdf)
  - <sup>11</sup> NCQA. (2022). *Stratifying HEDIS Measures by Race & Ethnicity*. The National Committee for Quality

- 
- Assurance (NCQA). <https://www.ncqa.org/health-equity/data-and-measurement/>
- <sup>12</sup> NCQA. (2022). *Health Equity Accreditation*. The National Committee for Quality Assurance (NCQA). <https://www.ncqa.org/programs/health-equity-accreditation/>
- <sup>13</sup> Centers for Medicare & Medicaid Services. (2022, March). *Inventory of Resources for Standardized Demographic and Language Data Collection*.” CMS. <https://www.cms.gov/About-CMS/Agency-Information/OMH>.
- <sup>14</sup> Department of Health, Education, and Welfare, & National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979). The Belmont Report. Ethical principles and guidelines for the protection of human subjects of research. *The Journal of the American College of Dentists*, 81(3), 4–13.
- <sup>15</sup> *Federal Data Strategy Data Ethics Framework*. (2019, June 4). Office of Management and Budget, the General Services Administration, and the Office of Government Information Services <https://resources.data.gov/assets/documents/fds-data-ethics-framework.pdf>
- <sup>16</sup> The National Academies of Sciences, Engineering, and Medicine. (n.d.). *Advancing Anti-Racism, Diversity, Equity, and Inclusion in STEM Organizations: A Consensus Study*. National Academies Sciences Engineering Machine. Retrieved September 26, 2022, from <https://www.nationalacademies.org/our-work/advancing-anti-racism-diversity-equity-and-inclusion-in-stem-organizations-a-consensus-study>.
- <sup>17</sup> Orvis, Dr. Karin. (2022, June 15) *Reviewing and Revising Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity*. The White House. The United States Government. <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/>.
- <sup>18</sup> *United States Core Data for Interoperability (USCDI)*. (n.d.) Interoperability Standards Advisory (ISA). The Office of the National Coordinator for Health Information Technology. Retrieved September 23, 2022, from <https://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi>.
- <sup>19</sup> *Patient Demographics/Information*. (n.d.) Interoperability Standards Advisory (ISA). The Office of the National Coordinator for Health Information Technology. Retrieved September 23, 2022, from <https://www.healthit.gov/isa/uscdi-data-class/patient-demographicsinformation#uscdi-v3>.
- <sup>20</sup> *FHIR Specification - 2.13 FHIR Overview*. (2022, May 28) Overview - FHIR V4.3.0. FHIR HL7. <https://www.hl7.org/fhir/overview.html>.
- <sup>21</sup> Imbeah, K., Howard, P., Brandes, R., Reid, A., Roy, B., & Sivashanker, K. (2021). *Spring 2021 A Roadmap for Race, Ethnicity, and Language Data Collection and Use in Connecticut*. Connecticut Health Foundation Changing Systems, Improving Lives. Connecticut Health Foundation. <https://www.cthealth.org/wp-content/uploads/2021/04/CHF-REL-Report-5-Web.pdf>
- <sup>22</sup> Haley, J.M., Dubay, L., Garrett, B., Caraveo, C.A., Schuman, I., Johnson, K., Hammersla, J., Klein, J., Bhatt, J., Rabinowitz, D., Nelson, H., & DePoy, B. (2022, July). *Collection of Race and Ethnicity Data for Use by Health Plans to Advance Health Equity*. Urban Institute. [https://www.urban.org/sites/default/files/2022-07/Collection%20of%20Race%20and%20Ethnicity%20Data%20for%20Use%20by%20Health%20Plans%20to%20Advance%20Health%20Equity\\_final.pdf](https://www.urban.org/sites/default/files/2022-07/Collection%20of%20Race%20and%20Ethnicity%20Data%20for%20Use%20by%20Health%20Plans%20to%20Advance%20Health%20Equity_final.pdf)
- <sup>23</sup> Nudelman, Jodi, Meridith Seife, & Nancy Harrison. (2022, June 15). *Inaccuracies in Medicare’s Race and Ethnicity Data Hinder the Ability to Assess Health Disparities*. Office of Inspector General Department of Health and Human Services. OEI-02-21-00100 <https://oig.hhs.gov/oei/reports/OEI-02-200100.asp#:~:text=We%20analyzed%20the%20race%20and,applies%20to%20the%20source%20data>
- <sup>24</sup> Eicheldinger C, & Bonito A. (2008). *More Accurate Racial and Ethnic Codes for Medicare Administrative Data*. Health Care Financ Rev. 2008;29(3):27-42.
- <sup>25</sup> Nudelman, Jodi, Meridith Seife, and Nancy Harrison. (2022, June 15). *Inaccuracies in Medicare’s Race and Ethnicity Data Hinder the Ability to Assess Health Disparities*. Office of Inspector General Department of Health and Human Services. OEI-02-21-00100 <https://oig.hhs.gov/oei/reports/OEI-02-200100.asp#:~:text=We%20analyzed%20the%20race%20and,applies%20to%20the%20source%20data>
- <sup>26</sup> Executive Office of the President, Office of Management and Budget (OMB), Office of Information and Regulatory Affairs. (1997, October 30) *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*. Office of Management and Budget. National Archives and Records Administration. [https://obamawhitehouse.archives.gov/omb/fedreg\\_1997standards#main-content](https://obamawhitehouse.archives.gov/omb/fedreg_1997standards#main-content).
- <sup>27</sup> *Race and Ethnicity: DQ Assessments*. (n.d.) DQ Atlas. Medicaid.gov: Keeping America Healthy. Retrieved 2021, from <https://www.medicaid.gov/dq-atlas/landing/topics/single/map?topic=g3m16&tafVersionId=25>.

- <sup>28</sup> Kho, A. N., Hynes, D. M., Goel, S., Solomonides, A. E., Price, R., Hota, B., Sims, S. A., Bahroos, N., Angulo, F., Trick, W. E., Tarlov, E., Rachman, F. D., Hamilton, A., Kaleba, E. O., Badlani, S., Volchenboum, S. L., Silverstein, J. C., Tobin, J. N., Schwartz, M. A., Levine, D., Wong, J.B, Kennedy, R.H, Krishnan J.A., Meltzer, D.O., Collins, J. M., Mazany, T., & CAPriCORN Team (2014). CAPriCORN: Chicago Area Patient-Centered Outcomes Research Network. *Journal of the American Medical Informatics Association : JAMIA*, 21(4), 607–611. <https://doi.org/10.1136/amiajnl-2014-002827>
- <sup>29</sup> Center for Research Informatics. (n.d.). CAPriCORN. <https://cri.uchicago.edu/portfolio/capricorn/>
- <sup>30</sup> Hegde, H., Shimpi, N., Panny, A., Glurich, I., Christie, P., & Acharya, A., (2019). MICE vs PPCA: Missing data imputation in healthcare. *Informatics in Medicine Unlocked*, Volume 17, 100275, ISSN 2352-9148, <https://doi.org/10.1016/j.imu.2019.100275>
- <sup>31</sup> Eicheldinger, C., & Bonito, A. (2008). *More accurate racial and ethnic codes for Medicare administrative data. Health care financing review*, 29(3), 27–42.
- <sup>32</sup> Haas, A., Elliott, M. N., Dembosky, J. W., Adams, J. L., Wilson-Frederick, S. M., Mallett, J. S., Gaillet, S., Haffer, S. C., & Haviland, A. M. (2019). Imputation of race/ethnicity to enable measurement of HEDIS performance by race/ethnicity. *Health services research*, 54(1), 13–23. <https://doi.org/10.1111/1475-6773.13099>
- <sup>33</sup> Department of Health, Education, and Welfare, & National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (2014). The Belmont Report. Ethical principles and guidelines for the protection of human subjects of research. *The Journal of the American College of Dentists*, 81(3), 4–13.
- <sup>34</sup> *Federal Data Strategy Data Ethics Framework*. (2019, June 4). Office of Management and Budget, the General Services Administration, and the Office of Government Information Services <https://resources.data.gov/assets/documents/fds-data-ethics-framework.pdf>
- <sup>35</sup> The National Academies of Sciences, Engineering, and Medicine. (n.d.). *Advancing Anti-Racism, Diversity, Equity, and Inclusion in STEM Organizations: A Consensus Study*. National Academies Sciences Engineering Machine. Retrieved September 26, 2022, from <https://www.nationalacademies.org/our-work/advancing-anti-racism-diversity-equity-and-inclusion-in-stem-organizations-a-consensus-study>. NCQA. (2021, December) *Improving Data on Race and Ethnicity: A Roadmap to Measure and Advance Health Equity*. Grant Makers in Health. The Commonwealth Fund. <https://www.gih.org/wp-content/uploads/2021/12/GIH-Commonwealth-Fund-federal-data-report-part-2.pdf>
- <sup>36</sup> Korab, P. (2021, December 14) *Guide to Using Word Clouds for Applied Research Design*. Towards Data Science. <https://towardsdatascience.com/guide-to-using-word-clouds-for-applied-research-design-2e07a61a513>
- <sup>37</sup> Parker, K., Horowitz, J. M., Morin, R., & Lopez, M. H. (2015, June 11). *Multiracial in America Proud, Diverse and Growing in Numbers*. Pew Research Center. <https://www.pewresearch.org/social-trends/2015/06/11/multiracial-in-america/>
- <sup>38</sup> Haley, J.M., Dubay, L., Garrett, B., Caraveo, C.A., Schuman, I., Johnson, K., Hammersla, J., Klein, J., Bhatt, J., Rabinowitz, D., Nelson, H., & DePoy, B. (2022, July). *Collection of Race and Ethnicity Data for Use by Health Plans to Advance Health Equity*. Urban Institute. [https://www.urban.org/sites/default/files/2022-07/Collection%20of%20Race%20and%20Ethnicity%20Data%20for%20Use%20by%20Health%20Plans%20to%20Advance%20Health%20Equity\\_final.pdf](https://www.urban.org/sites/default/files/2022-07/Collection%20of%20Race%20and%20Ethnicity%20Data%20for%20Use%20by%20Health%20Plans%20to%20Advance%20Health%20Equity_final.pdf)
- <sup>39</sup> Rowen, N.P., Van Stekelenburg, B., Nohria, R., Saunders, R. S., & Whitaker, R. G. (2022, September 14). *How To Improve Race, Ethnicity, And Language Data And Disparities Interventions*. Health Affairs Forefront. <https://www.healthaffairs.org/content/forefront/improve-race-ethnicity-and-language-data-and-disparities-interventions>
- <sup>40</sup> Vyas, D. A., Eisenstein, L. G., & Jones, D. S. (2020, August 27). Hidden in Plain Sight - Reconsidering the Use of Race Correction in Clinical Algorithms. *The New England journal of medicine*, 383(9), 874–882. <https://doi.org/10.1056/NEJMms2004740>
- <sup>41</sup> The Office of the National Coordinator for Health Information Technology. (2019). *What is FHIR?* Health Information technology. <https://www.healthit.gov/sites/default/files/2019-08/ONCFHIRFSWhatIsFHIR.pdf>
- <sup>42</sup> The Office of the National Coordinator for Health Information Technology. (n.d.) United States Core Data for Interoperability. Retrieved from: <https://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi>
- <sup>43</sup> *Gravity Project*. (n.d.) HL7 International. Retrieved September 30, 2022 from <https://www.hl7.org/gravity/>.
- <sup>44</sup> HL7 International – Patient Care WG. (2022, October 25). SDOH Clinical Care. <http://build.fhir.org/ig/HL7/fhir-sdoh-clinicalcare/index.html>

---

<sup>45</sup> Haas, A., Elliott, M. N., Dembosky, J. W., Adams, J. L., Wilson-Frederick, S. M., Mallett, J. S., Gaillet, S., Haffer, S. C., & Haviland, A. M. (2019). Imputation of race/ethnicity to enable measurement of HEDIS performance by race/ethnicity. *Health services research*, 54(1), 13–23. <https://doi.org/10.1111/1475-6773.13099>

## Acknowledgments

The authors would like to thank the health plans who participated in our interviews and our advisory panel members for their insights and suggestions. In addition, we would like to thank individuals who reviewed and commented on the draft report including Phil Saynisch, Lisa R. Sloane, U. Michael Currie, and Ben Hamlin for their feedback and suggestions.

We would also like to thank NCQA's editor Judy Lacourciere.

## Appendix A. Summary and List of Race & Ethnicity Data Collection Resources Consulted

We conducted a targeted environmental scan to identify toolkits, guides and other documents containing best practices for health care entities collecting race and ethnicity data, with a specific interest in resources for health plans. The team used a recently published resource from the Centers for Medicare & Medicaid Services (CMS)—*Inventory of Resources for Standardized Demographic and Language Data Collection*—to identify relevant sources.<sup>19</sup> The team then conducted a general search to identify additional sources not included in the inventory. Sources submitted by advisory panel members were also included in the scan. 42 sources (refer to the list below) were included in the scan, with publication dates ranging from 2001–2022.

Sources focused on race and ethnicity data collection for delivery systems and health plans. The most common topics covered were:

- Training and assigning staff to ask patients/members for their race and ethnicity data.
- Timing and modality for collecting race and ethnicity data.
- Categories for collecting race and ethnicity data.
- Rationale for collecting race and ethnicity data.
- Management and storage of race and ethnicity data.
- Importance of timely and accurate self-reported data.
- Potential barriers to collecting race and ethnicity data.

Although many sources were from the early to mid-2000s, most topics are still relevant and can help inform ongoing race and ethnicity data collection efforts. Sources published after 2020 framed the importance of race and ethnicity data collection, with consideration of the COVID-19 pandemic and its disproportionate burden on communities of color as well as the well-publicized impact of systemic injustice on historically marginalized communities. Table 1 provides an overview of our findings.

**Table 1: Environmental Scan Findings**

Topics	Examples of Content
When should race and ethnicity data be collected?	<ul style="list-style-type: none"> <li>• At enrollment or during registration</li> </ul>
Who should ask patients about race and ethnicity?	<ul style="list-style-type: none"> <li>• Frontline registration staff</li> <li>• Privately by medical staff</li> <li>• Via kiosks/tablets</li> </ul>
What modality should be used to collect race and ethnicity data?	<ul style="list-style-type: none"> <li>• Paper forms, electronic kiosks or tablets, web portals, face to face, over the phone or intake forms/surveys</li> </ul>
How should staff be trained on how to collect race and ethnicity?	<ul style="list-style-type: none"> <li>• Training should be ongoing;</li> <li>• Emphasize respect and consistency in the way the question is asked</li> <li>• Ensure staff understand why they are being asked to collect the data, how data will be used and how to answer patient questions and concerns</li> <li>• Scripts for staff were repeatedly identified as a helpful resource</li> </ul>
What categories should be used to collect race and ethnicity data?	<ul style="list-style-type: none"> <li>• Categories that reflect the community served and that can be rolled up to OMB</li> <li>• Separate race and ethnicity categories</li> <li>• Limited suggestions for inclusion of free text or the option to not respond</li> </ul>
What rationale is given to the patient about data collection?	<ul style="list-style-type: none"> <li>• Important to help improve quality of care for everyone; questions are being asked of everyone</li> </ul>
What assurances are given about how the data will or won't be used?	<ul style="list-style-type: none"> <li>• Scripts to reassure patients that their data will be used to ensure they receive the highest quality of care</li> <li>• Data provided will not be used to discriminate in any way and will be kept confidential</li> <li>• Should be clear that responses are voluntary</li> </ul>
OMB = Office of Management and Budget	

## References:

- A Framework for Stratifying Race, Ethnicity and Language Data Signature Leadership Series* [Review of *A Framework for Stratifying Race, Ethnicity and Language Data*]. (2014). <http://www.hpoe.org/Reports-HPOE/REAL-data-FINAL.pdf>
- A Guide Toward Diversity, Equity, and Inclusion (DEI) in Data Collection*. (n.d.). <https://www.schusterman.org/sites/default/files/DEIDataCollectionGuide.pdf>
- A Roadmap for Race, Ethnicity, and Language Data Collection and Use in Connecticut* | CT Health Foundation. (2021, June). Connecticut Health Foundation. Retrieved June 30, 2022, from <https://www.cthealth.org/publication/a-roadmap-for-race-ethnicity-and-language-data-collection-and-use-in-connecticut/>
- A Vision for Equitable Data Recommendations from the Equitable Data Working Group* [Review of *A Vision for Equitable Data Recommendations from the Equitable Data Working Group*]. 41. (2021, January 20). Equitable Working Group. <https://www.whitehouse.gov/wp-content/uploads/2022/04/eo13985-vision-for-equitable-data.pdf>
- Achieving Progress Toward Health Equity Using Race and Ethnicity Data: State Strategies and Lessons Learned*. (2021, November). Retrieved June 30, 2022, from [https://www.nashp.org/wp-content/uploads/2021/11/NGA\\_HealthEquityData\\_Brief\\_11.5.21.pdf](https://www.nashp.org/wp-content/uploads/2021/11/NGA_HealthEquityData_Brief_11.5.21.pdf)
- Advising Congress on Medicaid and CHIP Policy*. (2022). [https://www.macpac.gov/wp-content/uploads/2022/03/MACPAC-brief\\_Race-and-Ethnicity-Data-Availability.pdf](https://www.macpac.gov/wp-content/uploads/2022/03/MACPAC-brief_Race-and-Ethnicity-Data-Availability.pdf)
- Berg, S. (2020, September 9). *Gathering Race, Ethnicity Info Can Help Fight Inequity. How to Do It Well*. American Medical Association. Retrieved June 30, 2022, from <https://www.ama-assn.org/delivering-care/hypertension/gathering-race-ethnicity-info-can-help-fight-inequity-how-do-it-well>
- Bocchino, C. (2004). *Appendix G: Racial and Ethnic Data Collection by Health Plans* [Review of *Appendix G: Racial and Ethnic Data Collection by Health Plans*]. NCBI; National Academies Press (US). <https://www.ncbi.nlm.nih.gov/books/NBK215748/?report=reader>
- Collection of Race and Ethnicity Data in Clinical Trials Guidance for Industry and Food and Drug Administration Staff* [Review of *Collection of Race and Ethnicity Data in Clinical Trials Guidance for Industry and Food and Drug Administration Staff*]. (2016, October). U.S. Food & Drug Administration. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/collection-race-and-ethnicity-data-clinical-trials>
- Collection of Race, Ethnicity, Language (REL) Data in Medicaid Applications: A 50-State Review of the Current Landscape* [Review of *Collection of Race, Ethnicity, Language (REL) Data in Medicaid Applications: A 50-State Review of the Current Landscape*]. State Health & Value Strategies; State Health Access Data Assistance Center at the University of Minnesota. (2021, May). <https://www.shvs.org/wp-content/uploads/2021/05/SHVS-50-State-Review-EDITED.pdf>
- Collection of Racial and Ethnic Data by Health Plans to Address Disparities*. AAHP Foundation. (2004, July 1). RWJF. <https://www.rwjf.org/en/library/research/2004/07/collection-of-racial-and-ethnic-data-by-health-plans-to-address-.html>
- Create the Data Infrastructure to Improve Health Equity*. (n.d.). www.ihl.org. Retrieved June 30, 2022, from <http://www.ihl.org/communities/blogs/create-the-data-infrastructure-to-improve-health-equity>

- Eliminating Health Disparities Measurement and Data Needs*. (2004). (M. V. Ploeg & E. Perrin, Eds.) [Review of *Eliminating Health Disparities Measurement and Data Needs*]. The National Academies Press.
- Equity Partnership: How Mount Sinai Builds a Foundation for Health Equity*. (2021, January). Info.pressganey.com. Retrieved June 30, 2022, from <https://info.pressganey.com/wistia-initial-videos-to-migrate-to-pg-uberflip-content-hub/equity-partnership-how-mount-sinai-health-system-is-building-a-foundation-for-health-equity-011921>
- Everette, Dwan Tekisha. *Race, Ethnicity and Language Data Collection* (PowerPoint). (2021). [https://www.chcanys.org/sites/default/files/2021-07/REaL%20Data%20Collection\\_Asynchronous%20Sliddeck\\_FINAL\\_07082021.pdf](https://www.chcanys.org/sites/default/files/2021-07/REaL%20Data%20Collection_Asynchronous%20Sliddeck_FINAL_07082021.pdf)
- Gazmararian, J., Carre, R., Olson, N., & Lardy, B. (2012, July). *Exploring Health Plan Perspectives in Collecting and Using Data on Race, Ethnicity, and Language*. The American Journal of Managed Care, 18(7). <https://www.ajmc.com/view/exploring-health-plan-perspectives-in-collecting-and-using-data-on-race-ethnicity-and-language>
- Guide to Demographic Data Collection in Healthcare Settings—A Comprehensive Guide to Planning and Implementing Demographic Data Collection in Healthcare Settings*. (2017). <https://www.colleaga.org/sites/default/files/attachments/Measuring-Health-Equity-Guide-to-Demographic-Data-Collection.pdf>
- Hasnain-Wynia, R., & Baker, D. W. (2006). *Obtaining Data on Patient Race, Ethnicity, and Primary Language in Health Care Organizations: Current Challenges and Proposed Solutions*. Health Services Research, 41(4 Pt 1), 1501–1518. <https://doi.org/10.1111/j.1475-6773.2006.00552.x>
- Health Equity and Race and Ethnicity Data*. The Colorado Trust 1. (2013, September). Retrieved June 30, 2022, from [https://www.coloradotrust.org/wp-content/uploads/2015/03/CT\\_Race\\_EthnicityBrief\\_vFinal2.pdf](https://www.coloradotrust.org/wp-content/uploads/2015/03/CT_Race_EthnicityBrief_vFinal2.pdf)
- HRET HIIN Health Equity Organizational Assessment*. (2018, October). Retrieved June 30, 2022, from [http://www.wsha.org/wp-content/uploads/Health-Equity-Metric-Guidance\\_WSHA.pdf](http://www.wsha.org/wp-content/uploads/Health-Equity-Metric-Guidance_WSHA.pdf)
- Implementing Multicultural Health Care Standards: Ideas and Examples*. (n.d.). <https://theinstitute.umaryland.edu/media/ssw/institute/national-center-documents/Implementing-Multicultural-Health-Care-Standards-Ideas-and-Examples.pdf>
- Improving Health Equity Through Data Collection AND Use: A Guide for Hospital Leaders*. (2011, May). <http://www.hpoe.org/Reports-HPOE/improvinghealthequity3.2011.pdf>
- Lee, W.-C., Sreenivas, Veeranki, P., Serag, H., Eschbach, K., & Smith, K. (2016). *Improving the Collection of Race, Ethnicity, and Language Data to Reduce Healthcare Disparities: A Case Study from an Academic Medical Center* [Review of *Improving the Collection of Race, Ethnicity, and Language Data to Reduce Healthcare Disparities: A Case Study from an Academic Medical Center*]. Retrieved June 30, 2022, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5075235/pdf/phim0013-0001g.pdf>
- LLamas, A., & Joyce, D. (2022, June 3). *Biden Administration Outlines Comprehensive Equity Action Initiatives* [Review of *Biden Administration Outlines Comprehensive Equity Action Initiatives*]. California Health Care Foundation; <https://www.chcf.org/blog/biden-administration-outlines-comprehensive-equity-action-initiatives/>  
<https://www.chcf.org/blog/biden-administration-outlines-comprehensive-equity-action-initiatives/>



- LLanos, K., & Palmer, L. (2006, June). *Using Data on Race and Ethnicity to Improve Health Care Quality for Medicaid Beneficiaries* [Review of *Using Data on Race and Ethnicity to Improve Health Care Quality for Medicaid Beneficiaries*]. Center for Health Care Strategies, Inc.; Center for Health Care Strategies. Retrieved June 2006, from [https://www.chcs.org/media/Using\\_Date\\_to\\_Reduce\\_Health\\_Disparities.pdf](https://www.chcs.org/media/Using_Date_to_Reduce_Health_Disparities.pdf)
- Making CLAS Happen Six Areas for Action -54.* (2013). <https://www.mass.gov/doc/chapter-3-collect-diversity-data-0/download>
- Mcavey, K., Health, M., & Reginal, A. (2021, April). *Unlocking Race and Ethnicity Data to Promote Health Equity in California*. Retrieved June 30, 2022, from [https://www.manatt.com/Manatt/media/Documents/Articles/BSCA\\_Unlocking-Race-and-Ethnicity-Data-to-Promote-Health-Equity-in-CA-April-2021\\_c.pdf](https://www.manatt.com/Manatt/media/Documents/Articles/BSCA_Unlocking-Race-and-Ethnicity-Data-to-Promote-Health-Equity-in-CA-April-2021_c.pdf)
- Melendez, M., Campanella, S., Rosner, S., Pyatt, T., Gustafson, B., Katz, J., & Tchong, E. (2022, February). *The State of the Collection of Race, Ethnicity, and Language Data in Medicaid* [Review of *The State of the Collection of Race, Ethnicity, and Language Data in Medicaid*]. NORC at the University of Chicago. <https://www.norc.org/PDFs/Race%2C%20Ethnicity%2C%20and%20Lanugage%20Data%20collection%20in%20Medicaid/The%20State%20of%20the%20Collection%20of%20Race%2C%20Ethnicity%2C%20and%20Language%20Data%20in%20Medicaid.pdf>
- New York State Toolkit to Reduce Health Care Disparities: Improving Race and Ethnicity Data* [Review of *New York State Toolkit to Reduce Health Care Disparities: Improving Race and Ethnicity Data*]. School of Public Health University at Albany; University at Albany. (2014, February 3). [https://www.albany.edu/cphce/mrt\\_nys\\_toolkit\\_to\\_reduce\\_healthcare\\_disparities.pdf](https://www.albany.edu/cphce/mrt_nys_toolkit_to_reduce_healthcare_disparities.pdf)
- Palanker, D., Clark, J., & Monahan, C. (2022, June 9). *Improving Race and Ethnicity Data Collection: A First Step to Furthering Health Equity Through the State-Based Marketplaces* [Review of *Improving Race and Ethnicity Data Collection: A First Step to Furthering Health Equity Through the State-Based Marketplaces*]. The Common Wealth Fund. [https://www.commonwealthfund.org/blog/2022/improving-race-and-ethnicity-data-collection-first-step-furthering-health-equity-through?utm\\_source=alert&utm\\_medium=email&utm\\_campaign=Advancing+Health+Equity](https://www.commonwealthfund.org/blog/2022/improving-race-and-ethnicity-data-collection-first-step-furthering-health-equity-through?utm_source=alert&utm_medium=email&utm_campaign=Advancing+Health+Equity)
- Pittman, M. A., Pierce, D., & Hasnain-Wynia, R. (2004, May 1). *Who, When, and How: The Current State of Race, Ethnicity, and Primary Language Data Collection in Hospitals* [Review of *Who, When, and How: The Current State of Race, Ethnicity, and Primary Language Data Collection in Hospitals*]. The Commonwealth Fund. [https://www.commonwealthfund.org/publications/fund-reports/2004/may/who-when-and-how-current-state-race-ethnicity-and-primary?redirect\\_source=/publications/fund-reports/2004/may/who--when--and-how--the-current-state-of-race--ethnicity--and-primary-language-data-collection-in-ho](https://www.commonwealthfund.org/publications/fund-reports/2004/may/who-when-and-how-current-state-race-ethnicity-and-primary?redirect_source=/publications/fund-reports/2004/may/who--when--and-how--the-current-state-of-race--ethnicity--and-primary-language-data-collection-in-ho)
- Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement 5. Improving Data Collection across the Health Care System | Agency for Healthcare Research & Quality.* (2012). AHRQ. [Ahrq.gov. https://www.ahrq.gov/research/findings/final-reports/iomracereport/reldata5.html](https://www.ahrq.gov/research/findings/final-reports/iomracereport/reldata5.html)
- Reducing Disparities: The Importance of Collecting Standardized Data on Patient Race, Ethnicity and Language.* (2022). Aligning Forces for Quality National Program Office.

- [https://view.officeapps.live.com/op/view.aspx?src=http%3A%2F%2Fforces4quality.org%2Faf4q%2Fdownload-document%2F4099%2FResource-af4q\\_rel\\_training\\_slides\\_10-14-11.ppt&wdOrigin=BROWSELINK](https://view.officeapps.live.com/op/view.aspx?src=http%3A%2F%2Fforces4quality.org%2Faf4q%2Fdownload-document%2F4099%2FResource-af4q_rel_training_slides_10-14-11.ppt&wdOrigin=BROWSELINK)
- Rubin, V., Ngo, D., Ross, A., Butler, D., & Balam, N. (2018). *Counting a Diverse Nation: Disaggregating Data on Race and Ethnicity to Advance a Culture of Health* [Review of *Counting a Diverse Nation: Disaggregating Data on Race and Ethnicity to Advance a Culture of Health*]. 2. (n.d.).  
[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)
- Shapiro, A., Meyer, D., Riley, L., Kurz, B., & Barchi, D. (2021, September 21). *Building the Foundations for Equitable Care* [Review of *Building the Foundations for Equitable Care*]. NEJM Catalyst; NEJM Catalyst. <https://catalyst.nejm.org/doi/pdf/10.1056/CAT.21.0256>
- Signature Leadership Series Reducing Health Care Disparities: Collection and Use of Race, Ethnicity and Language Data. (2013). [http://www.hpoe.org/Reports-HPOE/Equity\\_Care\\_Report\\_August2013.PDF](http://www.hpoe.org/Reports-HPOE/Equity_Care_Report_August2013.PDF)
- Tan-McGrory, A., Bennett-AbuAyyash, C., Gee, S., Dabney, K., Cowden, J. D., Williams, L., Rafton, S., Nettles, A., Pagura, S., Holmes, L., Goleman, J., Caldwell, L., Page, J., Oceanic, P., McMullen, E. J., Lopera, A., Beiter, S., & López, L. (2018). *A Patient and Family Data Domain Collection Framework for Identifying Disparities in Pediatrics: Results from the Pediatric Health Equity Collaborative*. BMC Pediatrics, 18(1). <https://doi.org/10.1186/s12887-018-0993-2>
- Tools to Address Disparities in Health: Data as Building Blocks for Change. A Data Collection Toolkit for Health Insurance Plans/Health Care Organizations*. (2005, May). Retrieved June 30, 2022, from <https://aapcho.org/wp/wp-content/uploads/2012/02/AHIP-ToolstoAddressDisparitiesinHealth.pdf>
- Ulmer, C., Mcfadden, B., & Nerenz, D. (2009). *Race, ethnicity, and Language Data Standardization for Health Care Quality Improvement Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement Board on Health Care Services*. <https://www.ahrq.gov/sites/default/files/publications/files/iomracereport.pdf>
- Vega Perez, R. D., Hayden, L., Mesa, J., Bickell, N., Abner, P., Richardson, L. D., & Ngai, K. M. (2022). *Improving Patient Race and Ethnicity Data Capture to Address Health Disparities: A Case Study From a Large Urban Health System*. Cureus. <https://doi.org/10.7759/cureus.20973>
- Weissman, J. S., Betancourt, J. R., Green, A. R., Meyer, G. S., Nudel, J. D., Zeidman, J. A., & Carrillo, J. E. (2011, October 4). *Commissioned Paper: Healthcare Disparities Measurement* [Review of *Commissioned Paper: Healthcare Disparities Measurement*]. National Quality Forum. [https://5536401f-20a1-4e61-a28e914fb5dcef51.filesusr.com/ugd/888d39\\_b399d3b7cffa498b821410dab155b401.pdf](https://5536401f-20a1-4e61-a28e914fb5dcef51.filesusr.com/ugd/888d39_b399d3b7cffa498b821410dab155b401.pdf)
- Zambrana, R. E., & Carter-Pokras, O. (2001). *Health Data Issues for Hispanics: Implications for Public Health Research*. Journal of Health Care for the Poor and Underserved, 12(1), 20–34. <https://doi.org/10.1353/hpu.2010.0547>

## Appendix B: Interview Guide

### **Introduction (5 minutes):**

Thank you for taking the time to speak with us today. I am [interviewer] working with NCQA to conduct this project. Also on the call are [introduce other participants]. We've scheduled 60 minutes for this call, does that still work for you?

I wanted to start the conversation by sharing some background information on the project and then we have a series of questions that I wanted to walk through with you. We are recording this conversation for research purposes so that we can refer back to it as we write up the report. Individual interview responses will not be attributed directly to specific people, rather we will refer to the types of individuals interviewed. Only the project team at NCQA and the Commonwealth Fund will have access to the recordings and interview notes. Is that all right with you?

### **Project Background:**

As we indicated in the email to you requesting this interview, we are interested in learning about current methods of collecting race and ethnicity data, emphasizing the approaches used by health plans. We are also seeking opportunities for improving those methods. We hope this work will provide better evidence to guide plans' investment in supporting widespread and accurate reporting on equity. The findings of this project will be summarized in an evidence report.

We have prepared a series of questions to walk through with you today, but before we begin, do you have any questions?

### **Interview Questions:**

**I'd like to start with a quick overview of how your organization approaches health equity efforts, particularly concerning racial/ethnic equity.**

- How would you describe the overall strategy?
- Who is involved?
  - Data/analytic role? Clinical/care delivery? Community outreach?
- Where does discretion/leadership come from? Who or what drives focus on this?
- What are the specific goals?

**Let's turn to the availability of data for health equity efforts, quality improvement, and/or utilization management.**

- Where do the data come from?

- Direct collection from members? Care management functions? 3rd party data collectors or vendors?
- For what populations?
- For what periods?
- What's the provenance of these data? Self-report vs imputed vs community?
- Are these data you could link to quality reporting if coming from an external source?

**Now, I'd like to drill down into how your organization approaches the collection of race and ethnicity data from members.**

Does your organization rely on any standard toolkits or other resources to guide practices for data collection?

- If so, which?
  - How consistently are the practices laid out followed?
- If not, why?
  - Prompts: Lack of awareness of tools? Since the tools available are not tailored for your circumstances?

**We're basing the following questions on themes from existing guidance for collecting race and ethnicity data.**

When and why do you collect self-reported data? What motivates the use of self-report over other sources?

What does the process for data collection look like?

- When are data collected?
  - At enrollment?
  - Updated regularly?
- What collection method(s) are used?
  - Mail or web-based forms?
  - Collection by telephone?
  - Ingesting data from other sources?
- What approaches to categorizing race and ethnicity data do you use?
  - Separate questions for race and ethnicity?
  - Do you use a set typology, like the OMB 1997 categories?
  - Do members have to option to refuse to answer?
    - Is "choose not to say" or similar a stand-alone response?
    - Can members leave the question blank?
- Who collects the information? Do they receive any training?

- Is a rationale given for why the data are collected? If so, what explanation is given?
- Are any assurances about member privacy or the permissible uses of the data given? If so, which?
- What does your data governance process look like?

How are different stakeholders engaged in this process? Members, staff, community, board?

**I want to ask a few questions about process improvement around direct race and ethnicity data collection.**

What is preventing your organization from doing so/doing so more consistently?

- If no/no specific answer, offer prompts
  - Lack of support/prioritization from organization leadership?
  - Legal barriers?
  - Cost of the collection?
  - Technical challenges?
  - Staff reluctant to ask/members reluctant to volunteer information?

What, if any, areas of the collection process have you identified as priorities for improvement?

Would technical support resources on data collection, focused on implementation by health insurance companies, make a meaningful difference in the likelihood that your organization would invest in collecting these data?

- If needed, give examples
  - Training for staff
  - Guidance on the use of race/ethnicity categories
  - Guidance on “help text,” e.g. rationale for collection
  - Guidance on data analysis/QI

How are you assessing/evaluating the success of any changes you’ve made (or plan to make)?

**That covers the questions we’d wanted to ask. Before we conclude, we want to make sure we ask—is there anything we’ve missed? What else, if anything, would you like to share, or do you recommend we ask in future interviews?**

**Thank you again for participating—we appreciate your time and the insights you’ve shared. We’re looking forward to continuing with the interviews and sharing what we’ve learned soon. Is it OK if we reach out with additional questions that might come up?**

## Appendix C: Advisory Panel Members

Type	Name	Title & Organization
Health Plans, Health Systems, and Insurance Exchanges	Mark Friedberg, MD, MPP	Senior Vice President for Performance Measurement and Improvement, Blue Cross Blue Shield of Massachusetts
	Gabriella Silva, PhD	Research Fellow, Blue Cross Blue Shield of Massachusetts
	Taylor Priestly, MSW, MPH	Director for Health Equity and Quality Transformation Division, Covered California’s Health Equity Officer
	Elsie Taveras, MD, MPH	Chief Community Health Equity Officer, Massachusetts General Hospital
	Angela Hagan, MPA, PhD	Director of Health Equity Business Intelligence and Community Engagement, Humana
Employers and Purchasers	Nelly Ganseon, MPH	Executive Director, Community Engagement and Health Equity, Morgan Health
	Cristie Travis, MSHA	CEO, Memphis Business Group on Health
Policy Experts	Cara James, PhD	President & CEO, Grantmakers in Health
	Dan Meuse, MBA	Deputy Director, State Health and Value Strategies, Robert Wood Johnson Foundation
	Ignatius Bau, JD	Independent Consultant
	Katie Keith, JD, MPH	Associate Research Professor, Center on Health Insurance Reforms, Georgetown University
Race/Ethnicity Data and Survey Research	Karen Donelan, ScD	Stuart H. Altman Chair in U.S. Health Policy, Brandeis University, Heller School for Social Policy and Management
	Romana Hasnain-Wynia, MS, PhD	Chief Research Officer, Denver Health
	Deirdre Mylod, PhD	Executive Director, Institute for Innovation and Senior Vice President, Research & Analytics, Press Ganey
Community-Based Organizations	Katherine “Kitty” Bailey, MSW	CEO, Be There San Diego
	Shoshanah Brown, MS, MBA	CEO & Founder, AIRnyc
Advocacy Organizations	Cary Sanders, MPP	Senior Policy Director, California Pan-Ethnic Health Network
	Megan Simmons, JD, MPA	Policy Director, National Birth Equity Collaborative
	Latoshia Rouse	Doula and owner of Birth Sisters Doula Services; Patient and Family Advisor Consultant
Health Information Exchanges	Rachel Eager	Manager of Data Quality, New York eHealth Collaborative

## Appendix D. Sources for Guiding Principles for Race and Ethnicity Data Collection

<b>Guiding Principles</b>	
2022 Measuring Sex, Gender Identity, and Sexual Orientation <sup>46</sup>	<ol style="list-style-type: none"> <li>1. People deserve to count and be counted (inclusiveness).</li> <li>2. Use precise terminology that reflects the constructs of interest (precision).</li> <li>3. Respect identity and autonomy (autonomy).</li> <li>4. Collect only necessary data (parsimony).</li> <li>5. Use data in a manner that benefits respondents and respects their privacy and confidentiality (privacy).</li> </ol>
2021 Improving Data on Race and Ethnicity: A Roadmap to Measure and Advance Health Equity <sup>47</sup>	<ol style="list-style-type: none"> <li>1. Engage with patient and community groups.</li> <li>2. Acknowledge the key role of leadership.</li> <li>3. Use diffusion of innovation frameworks to identify enabling contextual factors and opportunities for peer learning.</li> <li>4. Allow an appropriate level of flexibility in the design of program requirements and incentives.</li> <li>5. Build the business case for greater investment in race and ethnicity data.</li> </ol>
2020 Federal Data Strategy Data Ethics Framework <sup>48</sup>	<ol style="list-style-type: none"> <li>1. Uphold Applicable Statutes, Regulations, Professional Practices, and Ethical Standards</li> <li>2. Respect the Public, Individuals, and Communities</li> <li>3. Respect Privacy and Confidentiality</li> <li>4. Act with Honesty, Integrity, and Humility</li> <li>5. Hold Oneself and Others Accountable</li> <li>6. Promote Transparency</li> <li>7. Stay Informed of Developments in the Fields of Data Management and Data Science</li> </ol>
1979 Ethical Principles and Guidelines for the Protection of Human Subjects of Research <sup>49</sup>	<ol style="list-style-type: none"> <li>1. Respect for Persons</li> <li>2. Beneficence</li> <li>3. Justice</li> </ol>

<sup>46</sup> The National Academies of Sciences, Engineering, and Medicine. (n.d.). *Advancing Anti-Racism, Diversity, Equity, and Inclusion in STEM Organizations: A Consensus Study*. National Academies Sciences Engineering Machine. Retrieved September 26, 2022, from <https://www.nationalacademies.org/our-work/advancing-anti-racism-diversity-equity-and-inclusion-in-stem-organizations-a-consensus-study>

<sup>47</sup> NCQA. (2021, December) *Improving Data on Race and Ethnicity: A Roadmap to Measure and Advance Health Equity*. Grant Makers in Health. The Commonwealth Fund. <https://www.gih.org/wp-content/uploads/2021/12/GIH-Commonwealth-Fund-federal-data-report-part-2.pdf>

<sup>48</sup> *Federal Data Strategy Data Ethics Framework*. (2019, June 4). Office of Management and Budget, the General Services Administration, and the Office of Government Information Services

<sup>49</sup> Department of Health, Education, and Welfare, & National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (2014). The Belmont Report. Ethical principles and guidelines for the protection of human subjects of research. *The Journal of the American College of Dentists*, 81(3), 4–13.