Improving Data on Race and Ethnicity:

A Roadmap to Measure and Advance Health Equity

GRANT MAKERS IN HEALTH

NCQA
Measuring quality. Improving health care.

DECEMBER 2021

Improving Data on Race and Ethnicity: A Roadmap to Measure and Advance Health Equity

This project, conducted jointly by Grantmakers In Health (GIH) and the National Committee for Quality Assurance (NCQA), with support from The Commonwealth Fund, details how race and ethnicity data are collected across federal health programs. This report builds on an earlier report, Federal Action is Needed to Improve Race and Ethnicity Data in Health Programs (October 2020). The project team included Cara James and Smita Pamar, GIH; Sarah Hudson Scholle, Philip Saynisch, and Jeni Soucie, NCQA; and Barbara Lyons, consultant.

Executive Summary

Racial and ethnic health disparities have been documented in the United States for over a century. In 1985 the Heckler Report provided the first national summary of these disparities, leading to the creation of the Department of Health and Human Services Office of Minority Health. In the 1990s and 2000s the study of gaps in health care access, utilization, and outcomes by race and ethnicity grew rapidly, but confronted a critical limitation of the available data: the lack of standardized, self-identified race and ethnicity.

The COVID-19 pandemic provided stark proof that data limitations are far from being addressed, which has real consequences for the study and practice of public health. Though data have improved since the beginning of the pandemic, the Centers for Disease Control and Prevention can only identify the race and ethnicity of less than 60% of individuals testing positive for COVID-19 or receiving vaccines, significantly limiting the ability of policymakers and health care stakeholders to measure and improve equity in the pandemic's effects and mitigation.

This Roadmap builds on an earlier report, *Federal Action Is Needed to Improve Race and Ethnicity Data in Health Programs,*[1] in three critical respects. First, it expands on that report's summary of the current state of race and ethnicity data in health care programs, offering more detail about whether and how race and ethnicity data are collected across a range of insurance programs, federally administered health systems and public health databases. Second, it summarizes a range of barriers to improving collection and use of race and ethnicity data, and identifies general principles for improving the data. Finally, it expands the range of recommendations for improving the data, considering not only actions the federal government could take, but also identifying actions for states and the private sector.

To advance these goals, the project team carried out a targeted search for information on the completeness and quality of race and ethnicity data; an environmental scan to identify previous reports summarizing challenges to collection and use of race and ethnicity data; and key informant interviews to better define and understand barriers and opportunities.

The environmental scan and informant interviews pointed to a consistent set of barriers faced by health care organizations, including:

- Legal and privacy concerns around collection and use of race and ethnicity data.
- Lack of standardized collection procedures and category definitions.
- Technical barriers to collection and storage of data.
- Cost of collection and lack of financial incentives or program requirements to collect race and ethnicity data.
- Lack of staff and resources in health care organizations to analyze and use data once collected.
- Resistance from patients and clinical providers to collection and use of race and ethnicity data.

Despite these challenges, prior reports and data from the key informant interviews pointed to several opportunities to improve collection and use of race and ethnicity data:

- Highlight early, successful adopters of expanded race and ethnicity data collection.
- Disseminate existing technical support resources and data standards.
- Educate patients and providers about the potential of improved race and ethnicity data to improve outcomes and equity.
- Provide incentives to encourage data collection and finance necessary technology investments and staffing.
- Where incentives fail to produce action, consider mandates (e.g., require collection to meet certain standards as a condition of participating in federal programs or demonstration projects).
- Identify existing resources that could be leveraged to improve analysis of health equity until consistent, complete, self-reported race and ethnicity data are available.

The report concludes with a series of recommendations for federal and state regulators and legislators, health systems and health insurance companies, and a range of other health sector stakeholders. Recommendations are grouped under the following themes:

- Improve data collection, storage and transfer systems.
- Evaluate and expand incentives and requirements to collect.
- Provide updated technical assistance to stakeholders.
- Review, clarify and, if necessary, amend regulations.

I. Background and the Role of This Report

An accumulating body of evidence demonstrates that racial and ethnic minority populations have been disproportionately impacted by the COVID-19 pandemic. These groups are at greater risk of exposure to COVID due to factors like overrepresentation among essential workers and crowded living conditions.[2] Additionally, the pandemic has been marked by racial and ethnic disparities in rates of severe illness leading to hospitalization or death, likely due to differential health care access, preexisting disparities in health status and a range of other contextual factors.[3]

These racial and ethnic health disparities have been documented for well over a century in the United States, though the 1985 Heckler Report was the first national report to summarize health disparities. Since that time, the language used to describe these differences—"health disparities" in the 1990s and 2000s, "health equity" more recently—has changed, but racial and ethnic gaps in health care access, quality, and outcomes have persisted and remain largely unchanged.

At the same time as the pandemic exacerbated existing disparities, it also highlighted major deficiencies in health care data systems. As of July 2021, the Centers for Disease Control and Prevention (CDC) only had race data for 64% of COVID case reports,[4] and eight states had race data on fewer than half of COVID cases as of October 2021.[5] **These examples underscore a key challenge to monitoring and improving health equity: a critical lack of complete, standardized, self-identified race and ethnicity data across federal and state health care and public health programs.** Dating back to the Heckler Report, our ability to monitor and address health disparities has been limited by missing and incomplete data, particularly for small population groups, such as American Indians and Alaska Natives.

This report, a collaboration between Grantmakers In Health and the National Committee for Quality Assurance (NCQA), was funded by the Commonwealth Fund to identify these data limitations, describe barriers to improving data on race and ethnicity across health care systems, and provide recommendations for charting a course forward. In the first project report,1 this group identified the potential scope for short-term actions at the federal level. However, there are barriers to collecting improved race and ethnicity data (and opportunities to use these enhanced data to measure and improve health equity) throughout the health care system. With that expanded perspective in mind, this report builds on the first document, incorporating insights on challenges and opportunities to expand the availability and use of race and ethnicity data from a range of sources. These include prior reports on race and ethnicity data quality and comprehensiveness, a detailed assessment of existing data resources across the health sector and a series of interviews with key health system stakeholders that shared their view of barriers and opportunities.

This report aims to describe the current state of completeness and quality of race and ethnicity data across a variety of health sector settings. Additionally, the work described seeks to identify barriers to improving data and opportunities that federal and state governments, as well as other health sector stakeholders, might leverage to lessen them. The report concludes with recommendations to improve the availability and use of race and ethnicity data as a key tool in measuring and improving health equity.

II. Methods

Data framework and scan of currently available race and ethnicity data

The first goal of this project was to provide an up-to-date overview of the quality and comprehensiveness of data on race and ethnicity across a range of health care settings. In pursuit of this goal, the project team (in consultation with the Commonwealth Fund) identified four key characteristics of race and ethnicity data collected and used by the health sector:

1. The *standards* used to collect the data.
2. The *completeness* of the data.
3. Whether data were *self-reported.*
4. Whether data were *available to researchers or the public.*

In this context, the *collection standards* refer to the categories of race and ethnicity reported in a given setting. One widely used standard was published in 1997 by the Office of Management and Budget (hereafter, OMB 1997). The OMB 1997 standards call for a two-question approach when feasible. In this format, respondents separately indicate which of five race categories (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White) they identify as, with the option to select multiple races. In a second question, respondents indicate whether they are of Hispanic or Latino ethnicity.[6] More recent standards published in 2011 by the Department of Health and Human Services (HHS) Office of the Assistant Secretary for Planning and Evaluation (ASPE) (hereafter, HHS 2011) retain the two-question approach but greatly expand the granularity of the available categories.[7] Notably, even these expanded race and ethnicity standards lack the detail needed to reflect all individuals' self-identification and to understand differences in health care utilization and outcomes. For example, neither standard includes a separate "Middle Eastern or North African" category. Newer, more comprehensive standards exist; for example, standards in the Office of the National Coordinator for Health Information Technology's (ONC) 2015 Edition final rule include over 900 ways of representing race and ethnicity. However, these newer standards are not widely used. Table 1 compares the 2 standards and shows the correspondence between the categories in each.

The completeness of the data refers to the percentage of the population included in a dataset for whom there were usable race and ethnicity data; that is, the percentage with a recorded race/ethnicity category, excluding those labeled as unknown, missing, declined to answer or similar. Additionally, we assessed whether data were *self-reported.* Self-identification is considered the gold standard for race and ethnicity data collection and is recognized as the preferred approach by both the OMB 1997 and HHS 2011 standards.6,7 In most cases, our assessment describes whether the entity responsible for collecting and reporting race and ethnicity data *instructed* staff to record self-reported data. In practice, data collection may deviate from those instructions, but rigorously documenting these practices is beyond the scope of this overview.

Finally, we assessed whether the data described were available to *researchers or the public,* and on what basis. Some data were technically available, but only to employed or affiliated researchers; others are available as a "researcher identifiable file" (RIF, i.e., one with identifying information replaced with pseudo-identifiers that facilitate consistent identification of an individual patient or provider over time and linkage between data sets) on a restricted basis, while others are available as public-use files (PUF).

Table 1: Comparison of OMB 1997 and HHS 2011 Race and Ethnicity Collection Standards

	OMB 1997	HHS 2011
Race*	White	White
	Black or African American	Black or African American
	American Indian or Alaska Native	American Indian or Alaska Native
	Asian	Asian Indian Chinese Filipino Japanese Korean Vietnamese Other Asian
	Native Hawaiian or Other Pacific Islander	Native Hawaiian Guamanian or Chamorro Samoan Other Pacific Islander
Ethnicity*	Not Hispanic or Latino	No, not of Hispanic, Latino/a, or Spanish origin
	Hispanic or Latino	Yes, Mexican, Mexican American, Chicano/a Yes, Puerto Rican Yes, Cuban Yes, Another Hispanic, Latino/a or Spanish origin

* OMB 1997 and HHS 2011 permit the reporting of more than one race; HHS 2011 also permits people to select one or more ethnicity.

We conducted an environmental scan to populate the framework for health care and public health programs that measure and monitor health care quality, primarily focusing on programs at the federal level. This search generally excluded population-based surveys (such as the National Health Interview Survey [NHIS] and the Medical Expenditure Panel Survey [MEPS]) that already have a standard for data collection and cover narrow samples of eligible populations. Where possible, this scan was supplemented with targeted outreach to experts at the relevant agencies (e.g., Veterans Health Administration [VHA], CDC).

One aspect of the data that was not included in these assessments was the accuracy of race and ethnicity data; that is, the concordance between the currently available race and ethnicity data and a reference set of self-reported race and ethnicity data for the same population. Several considerations drove this omission. First, while some efforts to assess accuracy are available (e.g., comparison of administrative data for Medicare beneficiaries to self-reported race and ethnicity data from the Outcome and Assessment Information Set [OASIS][8]), such validations are not consistently available for the data resources we considered. Moreover, in some settings, race and ethnicity data are incomplete or unavailable, rendering such comparisons moot. Nonetheless, future efforts should assess the accuracy of currently available race and ethnicity data, perhaps through original research aiming to validate these data. Additionally, while issues surrounding algorithms and bias in health care are a major topic of concern for health care providers, scholars and policymakers, a detailed treatment of this conversation is beyond the scope of this report. However, this work and the topics covered here are critically linked—without accurate, comprehensive data on self-reported race and ethnicity, we will struggle to know the extent of disparate impact these algorithms may have on health care utilization and outcomes.

Review of prior reports on health equity, with an emphasis on race and ethnicity data

In addition to the review of race and ethnicity data currently collected, the project team also conducted a targeted literature search on the subject that considered reports issued by the Institute of Medicine/National Academy of Medicine (IOM/NAM), the National Quality Forum, NCQA and the HHS ASPE. In particular, the literature search emphasized the sections of these reports that described the state of race and ethnicity data collected by health care stakeholders and the barriers to and opportunities for improving data collection. The review was limited to reports published since 2001, to ensure that commentary on race and ethnicity data was potentially relevant to current challenges. The full list of reviewed reports appears in Appendix Table 2.

Key informant interviews

Finally, in consultation with the Commonwealth Fund, the project team identified a list of 19 key informants representing federal, state, and local health agencies; commercial insurance plans; public and private health systems; and health information technology experts. We conducted 60-minute, semi-structured interviews with each informant to identify barriers to and opportunities for improving race and ethnicity data collection and use that were most relevant to their area of expertise. Additionally, in response to feedback from informants, we conducted 4 targeted, 30-minute interviews with key technical experts. These interviews allowed the project team to confirm and, as needed, update the insights gleaned from the review of prior reports.

III. Findings on Data Completeness and Quality, Barriers and Opportunities

Data completeness and quality

The following sections summarize the project team's findings on the completeness and accuracy of race and ethnicity data in health programs. Table 2 provides a summary of quality and completeness of race and ethnicity data; a more comprehensive treatment appears in Appendix Table 1.

Table 2: Race and Ethnicity Data Collected at Enrollment in Selected Federal Programs

SETTING	DATA COLLECTION STANDARD	COMPLETENESS	SELF-REPORTED?	DATA AVAILABLE FOR RESEARCH?
Medicare	Standards have changed over time	●	✔ [a]	✔
Medicaid	HHS 2011 [b]	◗	✔	✔
Federally-Facilitated and State-Based Marketplaces (FFMs; SBMs)	FFMs HHS 2011; SBMs vary	◗	◗	✔
Commercial Insurance	Unknown	❍	Unknown	✗
Veterans Health Administration	OMB 1997	●	◗	✗
Indian Health Service	Blood Quantum & Tribal Affiliation	◗ [c]	Unknown	✔
Federally Qualified Health Centers	OMB 1997	●	✔ [d]	✔ [d]
Birth Records	HHS 2011	●	✔ [e]	✔
COVID-19 Vaccinations	OMB 1997	◗	Unknown	✔
Pregnancy Risk Assessment Monitoring System	OMB 1997	◗	✔ [f]	✔

LEGEND:

❍ Less complete
◗ Varies (by state, collection method, etc.)
● More complete
✔ Yes
✗ No

NOTES:

a Data are obtained by SSA from the parents at birth, but data are not available for most beneficiaries born after 1989 due to SSA procedure changes. Also includes imputation to improve reporting for Asian and/or Pacific Islanders & Hispanic beneficiaries.

b Data categories roll up to OMB 1997 standards.

c Limited to data from individuals receiving care at IHS providers; 78% American Indians and Alaska Natives live outside tribal statistical areas.

d Data aggregated at center level.

e Based on mother and father's self-report.

f Extracted from birth certificate.

Medicare

Race and ethnicity data are relatively complete for Medicare beneficiaries, though there are well-documented issues with changing data sources and collection standards over time. Before 1989, Medicare's Enrollment Database (EDB) was populated from Social Security Administration (SSA) records derived from form SS-5 (application for Social Security). Prior to 1980, these forms allowed only White, Black or Other as race categories, with all enrollees not identifying coded as "Unknown." In 1980 these categories were expanded to White (non-Hispanic), Black (non-Hispanic), Hispanic, Asian/Asian American/Pacific Islander and American Indian or Alaska Native. However, when SSA switched to enrollment at birth, the race and ethnicity data contained in birth certificates were not recorded in SSA enrollment records. As a result, race and ethnicity data are missing from the EDB for beneficiaries born after 1989.[9] As of 2019, the Centers for Medicare & Medicaid Services (CMS) had race and ethnicity data on roughly 98% of Medicare beneficiaries.[10]

To the extent that health systems collect data on patients' race and ethnicity, these data are not submitted to CMS with Medicare claims. As a result, the race and ethnicity data that appear in Medicare claims databases like those maintained by the Research Data Assistance Center (ResDAC) are drawn from the Medicare EDB, not from data recorded at the point of care. Conversely, race and ethnicity data are collected using the OMB 1997 standard across a variety of quality reporting and assessment data sets, including the Long Term Care Minimum Data Set, the Home Health Outcome and Assessment Information Set and the CMS Hospice Item Set.[11–13] Moreover, these data are generally self-reported.

To address deficits in race and ethnicity information in Medicare data, algorithms such as the Medicare Bayesian Improved Surname Geocoding (MBISG)[14] and Research Triangle Institute (RTI)[9] methods have been developed to impute race and ethnicity from beneficiaries' name and place of residence. In the absence of self-reported data, these imputation approaches may be valuable tools for measuring health equity across populations.

Medicare enrollment, utilization and assessment data are generally available as RIFs from ResDAC, but accessing these data resources can be expensive,[15] which is challenging for researchers without funding for access. Additionally, because of the size of the datasets, analyses can be computationally intensive, requiring either investment in information technology or additional fees for access to computing environments like the Virtual Data Research Center.

Medicaid

Since 2019, Medicaid data have been made available by CMS through the Transformed Medicaid Statistical Information System (T-MSIS), which aims to improve on prior Medicaid data sources by standardizing data, making more timely data available, and other quality enhancements.[12] However, despite these improvements, the quality and completeness of race and ethnicity data varies from state to state. To summarize the state of race and ethnicity data in T-MSIS, the Medicaid Data Quality Atlas uses two criteria: the percentage of beneficiaries missing race and ethnicity data, and the number of race and ethnicity categories for which there was a 10 percentage point or greater difference between the T-MSIS enrollment data and the American Community Survey estimates for that state. On this basis, state data are classified as low concern, medium concern, high concern or unusable. As of 2018, 17 states had "low concern" data quality and 22 states had "high concern" or

"unusable" race and ethnicity data. In the 5 states with "unusable" data, more than 50% of enrollees were missing this information.[17] States report data based on the HHS 2011 standards when race and ethnicity are collected.[18]

Race and ethnicity data are not included in T-MSIS datasets reflecting use of health care services, quality of care and health outcomes or cost of care, but use of services and cost of care datasets can be linked to enrollment data with individual identifiers. As with Medicare data, T-MSIS analytic files are available as RIFs from ResDAC.

Marketplace plans

The latest CMS data indicate that about 12 million Americans are enrolled in insurance coverage purchased via health insurance Marketplaces.[19] As with race and ethnicity data in Medicaid programs, the quality and completeness of data linked to enrollment in Marketplace insurance plans varies considerably from state to state. According to state-level open enrollment PUFs, the percentage of enrollees with missing race data ranges from 11%–59% and ethnicity data are missing for between 4% and 42% of enrollees.[20] Additionally, collection and reporting standards vary depending on whether coverage is offered through state-based or federally facilitated Marketplaces (SBMs or FFMs). SBMs report race and ethnicity as a single variable, while FFMs use the HHS 2011 standards. Colorado does not report any race or ethnicity data to CMS.[21]

With respect to data availability, enrollment by race and ethnicity is publicly reported at the state level, but there is no publicly accessible central repository for Marketplace claims data. Moreover, while quality, outcome and cost of care data are all subject to public reporting, the data are not stratified by race and ethnicity.

Other commercial insurance plans

The majority of working-age adults (61.2% as of 2019) receive health insurance coverage through an employer.[22] Unfortunately, race and ethnicity information on this large population is largely incomplete. Recent research found that as of 2019, 76% of commercial plans had race data for less than 50% of members and 94% had ethnicity data on less than 50% of their membership.[23] Because of the largely incomplete nature of the data, information on collection standards is not centrally available and practices are likely to vary considerably from plan to plan.

Data on enrollment and health care utilization for commercial insurance enrollees are considerably more difficult to access than Medicare and Medicaid data. Claims data are available for a small subset of states through all-payer claims databases (APCD), but availability of race and ethnicity data is highly variable. For example, Minnesota does not include any race or ethnicity data in its APCD.[24] A 2017 analysis of 5 APCDs by the National Association of Health Data Organizations found that only 28% of records had usable race data.[25] Moreover, as of 2018, only 18 states either had legislation creating APCDs or were in the process of establishing an APCD[26] (only 9 had publicly posted rules for releasing data).[27] Data can also be accessed via commercially available claims databases; like APCDs, however, if the databases draw on data available to commercial insurance carriers, they will reproduce the same limitations with respect to race and ethnicity information.

Federal health care delivery systems

The sections above describe how the federal government collects data in its role as a purchaser of health care services, but for several key populations of interest, it also serves as the provider of health care services. This section details race and ethnicity data collected in several federal delivery systems.

The VHA provides health care services to veterans. It collects race and ethnicity data using the OMB 1997 standards, and data have become more complete over time. Prior to 2003, less than 60% of patients had usable race and ethnicity data; since 2015, that figure has climbed above 90%.[29] VA Medical Centers are instructed to collect self-reported race and ethnicity data, but this varies in practice. Enrollment and utilization data are typically not available to researchers outside the VHA system.

The Defense Health Agency (DHA), which provides care to active-duty members of the armed forces, collects eligibility and enrollment data in the Defense Enrollment Eligibility System. Data are collected using more national origin categories than are captured in the OMB 1997 standards but do not match the expanded HHS 2011 standards. Data reflect the race and ethnicity of the enrolled service member and may not match the self-identification of other family members enrolled as dependents. The project team was unable to locate data on the completeness of DHA race and ethnicity information.

Federally Qualified Health Centers (FQHC) are community-based health care providers funded by the Health Resources and Service Administration (HRSA). These centers serve essential safety net functions, including offering sliding scale-based payment structures to patients.[30] FQHCs are required to report center-level data to the Uniform Data System (UDS), which includes race and ethnicity data, using a modification of the OMB 1997 standards. These reflect self-reported data collected at registration,[31] and aggregate statistics from 2019 suggest that race and ethnicity data are available for roughly 85% of FQHC patients.[32]

The Indian Health Service maintains the National Patient Information Reporting System (NPIRS). Because the IHS primarily provides care to members of federally recognized American Indian and Alaska Native (AIAN) tribes, IHS data reflect blood quantum and tribal affiliation rather than the race and ethnicity data used in other databases described here. Additionally, the NPIRS only includes data for individuals receiving care at IHS facilities. Because 78% of AIAN individuals do not reside in tribal statistical areas,[33] the NPIRS only provides data for a subset of AIAN populations.

Federal public health data

In addition to its collection of race and ethnicity data in its role as a payer, the federal government also administers a wide array of public health databases, many of which incorporate data on race and ethnicity. Race and ethnicity data are largely complete in vital statistics records, which include records of births and deaths. Births are recorded with the mother and father's self-reported race and ethnicity using HHS 2011 standards. Deaths are recorded using the OMB 1997 standards and may be reported by the funeral director rather than by the family of the deceased. Race and ethnicity data are largely complete for deaths and for the mother's race and ethnicity in birth records. The father's race and ethnicity are somewhat less complete in birth records, with 82% of records having usable race and 87% having usable ethnicity.[34] Aggregated national-level data are available as PUFs and researchers can apply for access to more granular, restricted-use files as well.[35]

The Pregnancy Risk Assessment Monitoring System (PRAMS) is a surveillance system that samples live births to support improvement in birth outcomes. Race and ethnicity are not included in the PRAMS questionnaire, but the PRAMS data system incorporates maternal and paternal race from birth certificate data, so respondents' race and ethnicity are available to researchers along with the PRAMS files. PRAMS data are also available to researchers as an RIF on a by-application basis.[36]

Another critical data resource is the federal Immunization Information System (IIS), which draws on systems administered by states, territories and local governments. Standards for the federal system call for use of the OMB 1997 standards,[37] but implementation varies from system to system, with major implications for completeness and quality. For example, using COVID vaccines as a case study, as of October 2021, race and ethnicity data were available for only 62% of recipients.[38]

With respect to data sources on communicable diseases, the National Healthcare Safety Network data on health care-acquired infections do not include patient race or ethnicity. Data on community-acquired infections typically use the OMB 1997 standards, but completeness of race data varies by reporting unit. For example, race was available for 57% of reported COVID cases.[39]

Environmental Scan and Key Informant Interviews

To provide critical context for this overview of the current state of race and ethnicity data in health care systems, the project team conducted two further, complementary efforts: a scan of major reports on health equity and a series of semi-structured interviews with key stakeholders. Both efforts were intended to elicit information on *barriers* to improving race and ethnicity data collected and used by health system stakeholders, and *opportunities* to improve the data. The following section describes those barriers. The opportunities, along with the project team's assessment of the policy and practice environment, form the basis of our recommendations.

Barriers

1. Legal and privacy concerns.

One major barrier to collecting race and ethnicity data, particularly in the context of health insurance plans, is the widely shared belief that there are restrictions or even blanket prohibitions on collecting such data. The 2004 National Research Council (NRC) report, *Eliminating Health Disparities: Measurement and Data Needs,* cited four states (California, Maryland, New Hampshire, New Jersey) with a law or insurance regulation restricting collection of race, ethnicity and/or language data by insurers.[40] However, another legal analysis, *Improving Racial and Ethnic Data on Health,* presented at a workshop the same year, indicated that collecting these data is permitted under federal law and—at least under most circumstances—under state law.[41]

While these reports are now almost two decades old, similar concerns were cited in an open letter to the National Association of Insurance Commissioners in 2021, citing the same states and restrictions in some places.[42] In addition to direct prohibitions and restrictions on collecting race and ethnicity data, the same letter highlighted ways in which collection could create liabilities; for example, by being subject to Federal Trade Commission (FTC) action through Unfair and Deceptive Practices laws.

One interview subject raised a related concern about the consequences of *analyzing* race and ethnicity data, once collected. With greater attention focused on health equity (and racial justice more broadly), health plans and health care providers are much more likely to identify when their actions have created or contributed to disparate health outcomes. They may be concerned that *recognizing* the presence of disparities may create liabilities for organizations that fail to subsequently take corrective action.

Our environmental scan and interviews also revealed extensive concerns about ensuring the privacy of race and ethnicity data. These concerns were highlighted in both the 2003 Institute of Medicine report *Unequal Treatment,*[43] as well as in the 2004 report, *Improving Racial and Ethnic Data in Health.*[41] In our interviews, concerns ranged from high-level considerations about the need to protect the data, to specifics about how to construct data use agreements to facilitate data sharing, how to set boundaries around sharing (e.g., to prevent use of data for marketing purposes) and how to balance potential gains from sharing race and ethnicity data with the risks. The financial and administrative costs of privacy protections have also been cited as an issue.

2. Lack of standardized definitions and collection procedures.

Variations in how race and ethnicity data are collected and uncertainty about best practices are another major category of challenges to improving health equity data. The OMB 1997 standards for race and ethnicity categories have been widely—though not universally—adopted. While some competing standards, like the 2011 HHS categories, are compatible with the OMB 1997 approach (Table 1), others are not. For example, race and ethnicity data are collected differently, depending on whether health insurance Marketplace coverage is available through an SBM or FFM. Conflicting standards between administrative data, surveys, electronic health records (EHR) and other sources limit comparability, and in some cases, prevent the use of complementary sources to address missing data issues. Our environmental scan and interviews revealed a lack of clarity about which standards are best, and in some cases, competing guidance from different federal agencies about which standards to use.

One reason for deviations from the standardized sets of race and ethnicity categories is the lack of granularity in these definitions. Several interviewees stressed the challenges created by the coarseness of the OMB 1997 categories. Aggregate categories like "Asian or Pacific Islander" can obscure important differences in experiences, access and outcomes between groups. Moreover, the subgroups that make up these aggregate categories vary by region and health plans or providers often want detailed data on local populations. While the 2011 HHS standards added more granular categories, several groups (particularly people with Middle Eastern or North African ancestry) were not included as separate categories. Several interviews also pointed to the need for better data on the growing population of people with multiracial heritage and better understanding of the factors that lead some individuals to respond with "other" racial or ethnic self-identification.

Uncertainty about best practices in collecting race and ethnicity is another major barrier. In addition to the categories offered as options and whether race and ethnicity are addressed as one question or two, there is considerable variation across plans and providers in when, how often and by whom questions about race and ethnicity are asked. Resources designed to support collection of race and ethnicity data (e.g., the American Hospital Associations Disparities Toolkit) encourage offering patients a rationale for why these data are collected, along with assurances about how the data will (and will not) be used.[44] However, there is little data on specific approaches. One interview subject mentioned the need to improve the evidence base on best practices. Translating guidance into practice may create unexpected challenges, too; for example, legal requirements that *answering* questions about race and ethnicity is voluntary have been interpreted by frontline staff that *asking* the questions is optional, as well.

3. <u>Technological readiness and interoperability concerns.</u>

In addition to concerns about whether different health care stakeholders are collecting data using compatible categories of race and ethnicity, our environmental scan and key informant interviews pointed to major concerns regarding interoperability and data sharing. Several respondents noted that although providers and employers might have data on race and ethnicity, they are often reluctant to share it—the more granular and detailed the data, the greater the perceived pressure to keep the data private.

Other respondents stressed the lack of technical interoperability between EHRs and administrative data, or across systems used by different providers, payers and community-based organizations. While standards like Health Level Seven® (HL7®) exist, not all health IT is required to use them. Moreover, health data system end users may not be aware of all the opportunities to capture data using these standards.

Several interviewees also highlighted the need for clarifying the provenance of race and ethnicity data, or the actors and processes that created a record.[45] Interview subjects also called for clearer procedures for reconciling conflicting records maintained by different entities or collected by a single entity over time. Without clear guidelines, a usable datapoint on race or ethnicity might be overwritten as "missing" by subsequent encounters during which this question is left unanswered.

Critically, contributors to the 2004 NRC report, *Improving Racial and Ethnic Data on Health,* noted that addressing standardization might be precondition for making necessary improvements to health data management systems.[41] In one interview, this was referred to as a need for "semantic interoperability." In other words, providing a common definition of what data to collect may be a precondition to deploying the ideal technology for that purpose.

4. Cost of collection and lack of incentives.

Multiple sources emphasized financial factors that inhibited collection of race and ethnicity data, including both the *costs of expanding data collection* and the *lack of direct financial incentives* to do so. In terms of direct costs, the 2004 NRC report mentioned above highlighted the need to update the process for collecting data, including updating forms and procedures, investing in health IT and retraining staff. At the state- and local-government levels, this can entail substantial investments to upgrade outdated data systems.[41] Several interviews also stressed the opportunity cost of clinical time if billable activities are displaced by additional collection of demographic data.

Reports and interview subjects also pointed to the lack of formal incentives or requirements to collect race and ethnicity data. As a result, these efforts are at risk of being deprioritized, relative to work in performance measurement and other domains with reporting requirements, even if stakeholders view collecting expanded race and ethnicity data is worthwhile. However, while there was broad acceptance of the idea that greater incentives, requirements, and financial support might be necessary, our interviews highlighted a number of potential challenges in designing those incentives. One respondent pointed to the need for caution in designing incentives, to avoiding gaming of reporting requirements. Another noted that even if delivery systems have greater incentives to collect and analyze race and ethnicity data, patients generally lack clear incentives to share their data.

Finally, one respondent highlighted that while identifying and emphasizing the financial return on investment could help build the business case for expanded data collection, this should not be the sole basis for doing so. This caution is well-founded, since securing the support of patients, providers and advocacy organizations could be jeopardized by an overemphasis on economic considerations.

5. Resistance from patients, providers and health care organizations.

The environmental scan and stakeholder interviews also revealed concerns about patients' willingness to share race and ethnicity data, and providers' willingness to collect such information. The 2003 IOM report, *Unequal Treatment,* highlighted providers' potential objections to encroachment on their professional autonomy. They would also be likely to resist a new administrative burden to clinical workflows. Given the additional operational challenges introduced by the COVID pandemic, these concerns warrant consideration.

The *Unequal Treatment* report and our interviews also emphasized patients' well-founded concerns about how race and ethnicity data might be used.[43] One interviewee emphasized fears that data collected by insurers might be used in underwriting, and suggested greater need to clarify rationales for collection. Given past and ongoing discriminatory treatment of vulnerable patients by health care institutions, race and ethnicity data collection procedures need to stress reasons for data collection, patients' privacy protections and other trust-building components.

6. Adequacy of staff and resources.

Even if dedicated funding sources for data collection and related processes are expanded, smaller health care organizations (particularly local public health agencies) may lack the necessary staff and technology to carry out collection and analysis. Additionally, resource constraints in these settings apply to a wide range of functions, not just to health equity efforts. For these less well-resourced organizations, commitment from leadership is needed to ensure that scarce resources are devoted to investments in these data systems and related staffing.

IV. Statement of Principles for Improving Data on Race and Ethnicity

In addition to supporting development of specific recommendations for federal and state policymakers and other health sector stakeholders, our environmental scan and key informant interviews pointed to a range of general principles for improving race and ethnicity data. These principles cut across audiences and topic and can help guide the improvement process.

First, a consistent theme in our interviews was the need to *acknowledge the key role of leadership.* In particular, the federal government can provide a single, consistent vision and discourage states and health care organizations from implementing competing approaches. Interview subjects regularly highlighted leadership's commitment to equity measurement and improvement as a critical predictor of success in collecting and using race and ethnicity data at the organizational level. In health care settings, scarce resources and urgent competing demands can create pressure to deprioritize efforts toward equity, and strong commitment from policy and organizational leaders can help sustain this critical work.

A second theme that emerged was that *diffusion of innovation frameworks*[46] can provide a value lens through which to understand efforts to collect and use race and ethnicity data. Identifying and studying early adopters can provide valuable insights into enabling contextual factors and opportunities for peer learning, and these lessons can be disseminated widely to accelerate innovation and adoption of current best practices. Conversely, this approach can also help identify lagging organizations that may require technical support or financial incentives, or both, to improve.

Third, interviewees emphasized the need to allow an appropriate level of flexibility in the design of program requirements and incentives. For example, standardized coding schemes for race and ethnicity data should support minimum standards and consistent "roll-up" categories but allow health care organizations to capture details appropriate to local demographics. Incentives and financial support should be designed to ensure that not only can states and large health systems access these funding streams, but also smaller, local stakeholders.

Fourth, the process of improving the collection and use of race and ethnicity data must engage with patient and community groups at every step. Health care organizations should identify relevant organizations and solicit input on how to collect race and ethnicity data, as well as how to use the data to advance health equity. Identifying key questions and community goals can serve a vital role in building and preserving trust. When data are used in analyses or quality improvement efforts, the results of these efforts should be reported to patient and community groups as well, reinforcing community consultation at every step in the process.

Finally, advocates for health equity should help build the business case for greater investment in race and ethnicity data. While considerations of health justice should remain paramount, building the business case for investing in equity can help motivate lagging organizations and support sustainable progress. Moreover, the business case should extend beyond purely financial benefits to consider factors like the risk to an organization's reputation for delivering inequitable care.

V. Recommendations for Improvement of Data on Race and Ethnicity

The federal government must play a major role in improving the collection and use of race and ethnicity data, given its control of key incentive systems and regulations and its central position in the health care system. However, a wide range of other actors—state policymakers, health plans and delivery systems, quality measurement and improvement organizations and advocates—have critical roles to play and have varying degrees of ability to influence the completeness and quality of these data. The following sections organize recommendations which follow from the team's review of data resources, the environmental scan and key informant interviews, grouped by theme and identifying relevant actors:

1. Data collection, storage and transfer systems.
2. Incentives and requirements to collect.
3. Technical assistance to stakeholders.
4. Review, clarification and, if necessary, amendment of regulations.

1. Changes in Data Standards and Systems

Data systems and standards refers to the tools and procedures for collecting, storing and sharing race and ethnicity data. Stakeholders identified critical ways to update the categories that describe racial and ethnic self-identification. The federal government should standardize and modernize the collection, analysis and reporting of such data across all programs and provide the resources necessary to accomplish this goal.

Recommendation 1.1:
The federal government should review and update the OMB 1997 Statistical Policy Directive on Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity to more accurately reflect the demographics of the United States population and provide flexibility to state and local governments to capture information representing their communities. In 2016 OMB convened federal agencies and sought public comment on proposed changes, such as including Hispanic as a race option and adding an additional race category for Middle Eastern and North African, but recommendations were never finalized.

Recommendation 1.2:
CMS should include race and ethnicity on the Medicare Parts C & D application. Currently, CMS obtains race and ethnicity data on Medicare beneficiaries at the time of enrollment from the SSA, which gets the data from birth certificates. For most beneficiaries, those categories predate the OMB 1997 Directive. Additionally, beginning in 1990, SSA implemented a new process for issuing Social Security numbers at birth and opted not to include race and ethnicity among the information requested. As a result, it cannot provide race and ethnicity data for beneficiaries born after 1990. The proportion of Medicare beneficiaries enrolled in Medicare Advantage has grown significantly in the past decade to include nearly 40% of all beneficiaries, and more than 75% of Medicare beneficiaries are enrolled in a Part D plan. Collecting the data at enrollment in Parts C and D will help fill the void and allow for data to be collected reflecting the current data standards. Additional recommendations for improving racial and ethnic data in Medicare can be found in a report CMS submitted to Congress in 2017.

Recommendation 1.3:
The Interagency Working Group on Equitable Data (Data Working Group) should work through OMB to standardize collection and reporting of racial, ethnic and other demographic data across the federal government (e.g., USDA, CMS, HRSA reporting requirements), while providing states, local governments and grantees the flexibility to collect data on other populations residing in their area. The federal government should review all data systems to ensure that race and ethnicity data are collected, where appropriate, and that programs and systems allow for more granular data, such as HHS race and ethnicity data categories. Some stakeholders expressed frustration at the inconsistent requirements across federal programs for submission of race and ethnicity data, and at the limited flexibility for more granular categories. For example, the UDS includes data from community health centers and asks respondents to report race data for their patient population using one of eight categories: 1.) Asian, 2.) Native Hawaiian, 3.) Other Pacific Islander, 4.) Black/African American, 5.) American Indian/Alaska Native, 6.) White, 7.) More than one race, or 8.) Unreported/Refused to report race. Meanwhile, the Transformed Medicaid Statistical Information System (T-MSIS) includes more detailed categories for Native Hawaiian and Other Pacific Islander consistent with the HHS standards, and the Women and Infants (WIC) includes Native Hawaiian and Pacific Islander as one racial population, consistent with the OMB 1997 Directive.

Recommendation 1.4:
HHS should edit provider data systems to collect information on provider demographics, and work with the Bureau of Labor Statistics at the Department of Labor to help ensure that the population caring for communities reflects the demographics of those communities. Much of the information related to disparities focuses on the patient or the institution from which they receive care, with considerably less attention paid to the individuals providing care. Two of the National Standards for Culturally and Linguistically Appropriate Services in Health and Health Care (National CLAS Standards), developed by HHS, emphasize the importance of recruiting and developing a diverse workforce and collecting demographic data to monitor disparities and support suggestions about delivery of services. Having a better understanding of who provides care at all levels—from doctors and nurses to nursing home, home health, and community health workers—can help increase the availability of culturally competent care.

Recommendation 1.5:
The ONC should provide funds to support the expansion of health information exchange (HIE) coverage. Patients' data are often stored in multiple records regarding health insurance eligibility, health care utilization in a variety of settings and clinical data. State or regional HIEs can provide a central hub for these data and by combining records from multiple sources, health plans and health care systems can help improve the completeness of race and ethnicity data (compared to what would be available from any one data source. However, the current situation varies significantly from state to state, and considerable technical and financial resources may be needed to promote this function as a "public health utility."

Recommendation 1.6:
Measure developers should assess the feasibility of incorporating the HL7 Fast Healthcare Interoperability Resources® (FHIR®) V3 Code System for race[47] and ethnicity,[48] where appropriate, into quality measure specifications. FHIR is a standard for data formats and elements that allows health information to be exchanged between systems.[49] These specifications are used for required reporting activity such as HEDIS® , MIPS and in other quality programs, and their use could enable sharing of reliable information on race and ethnicity between stakeholders. The ONC Certified Health IT standards require health IT vendors to demonstrate that

HEDIS® is a registered trademark of the National Committee for Quality Assurance (NCQA).

their products *support* data exchange using FHIR standards, but do not mandate that health IT products *include* these functionalities if end users do not request them—and without clear use cases for these features, they are unlikely to be broadly adopted.

Quality measurement and accreditation organizations can play a critical role in advancing interoperability by building FHIR standards for race, ethnicity, and data provenance into measure specifications for required reporting activities and by supporting their deployment through technical assistance and publication of implementation guides, to ensure that implementors can understand and use these requirements.

2. Incentives and Requirements

The completeness and quality of race and ethnicity data varies in federal, state, and local governments; payers; and providers, as does their level of initiative in improving the data. Some stakeholders have invested time and resources, some are interested but not sure how to proceed, and some have other priorities and/or fewer resources to devote to the task. As a result, a one-size-fits-all approach will not work. Therefore, the federal government should use a combination of incentives and requirements to increase racial and ethnic data completeness and quality at the state and local levels. Incentives can encourage data collection and finance necessary technology investments and staffing. Where incentives fail to produce action, mandates (e.g., require data collection to meet certain standards as a condition of participation in federal programs or demonstration projects) may be necessary. For example, in an initial phase, ONC could offer grant funding to support adoption of health IT needed for expanded collection of race and ethnicity data. As data collection practices mature and become more standardized, incentives could be phased out or paired with race and ethnicity data collection as a condition of participation in Center for Medicare & Medicaid Innovation (CMMI) demonstrations or receiving payment from Medicare.

Recommendation 2.1:
The OMB should require all federal program reports to include data stratified by race, ethnicity, and other demographics, where feasible. Some reports (e.g., National Healthcare Quality and Disparities Reports; Health, US; Part C and D Performance Data Stratified by Race, Ethnicity, and Gender) provide information stratified by race, ethnicity, and other demographic characteristics, but many federal program reports do not. In some instances, special reports have been issued, but the lack of routine reporting at leads to reduced availability of timely information and dependence on champions, luck and political or external pressures to get the information needed to identify and monitor health disparities. Making reporting race and ethnicity data standard operating procedure will help ensure that the information is more consistently available and raise awareness about racial and ethnic health disparities, and can help spur improvement in the data. Using the data can also signal to stakeholders that the government is interested in this topic, which may lead to more stakeholders looking at their data.

Recommendation 2.2:
HHS should require state and local health departments, Medicaid programs and public health and human service programs to conduct an audit of their race and ethnicity data to identify information gaps and barriers to completion. This should include an audit of how data are collected (e.g., self-report or another method). Understanding where there are data gaps is critical for guiding efforts to address barriers and solutions and planning for the resources needed to fill data gaps. Smaller agencies, such as local health departments, may need supplemental funding and technical assistance to achieve these goals.

Recommendation 2.3:
Health care quality measurement and improvement agencies should promote in the short term stratification of quality reporting data by race and ethnicity, and require it in the long term. Considerable effort has been devoted to developing and maintaining quality reporting systems for hospitals, health plans, and other entities in the health care system. While these data systems provide vital information for comparison, quality improvement and reimbursement purposes, they are generally limited to providing a "top line" rate for each entity. These aggregates can obscure the fact that entities providing high-quality care *on average* may not provide *equitable care* across groups. Stratified reporting—such as NCQA's new requirement that plans begin reporting certain quality measures by race and ethnicity—can help uncover inequities. In the short term, health plans and health care providers may lack the data on race and ethnicity needed to generate such reports. As a result, a combination of technical assistance now and requirements to report later will ensure development of these capabilities.

3. Education and Technical Assistance

Stakeholders consistently noted three major factors that contribute to incomplete and inaccurate racial and ethnic data: lack of awareness or understanding of best practices regarding the collection of racial and ethnic information; discomfort asking individuals to indicate their race and ethnicity; and lack of consumer awareness about how data are used. Experts noted that in some cases, health plans and systems key informants mistakenly believed that federal or state laws or insurance regulations prohibited collecting data on race and ethnicity. Other interviewees were unaware that standardized data categories exist or were concerned that consumers would not provide the information. Clarifying what is allowed and best practices for data collection are critical steps in improving data completeness and quality.

Interviewees also identified success stories—early adopters of expanded race and ethnicity data collection whose methods and accomplishments could be amplified and shared with peer organizations. Health sector stakeholders would benefit from work to identify existing technical support resources and data standards that would facilitate collection. Outreach to patients and providers about the potential of race and ethnicity data to improve outcomes and equity would also provide critical support to this work.

Recommendation 3.1:
HHS and other departments should provide both technical and financial assistance to improve data systems and interoperability, as well as analytic and reporting capacity. Many departments and agencies work with states to administer their programs. For example, CMS has a long history of working with states through multiple avenues such as state grantee initiatives, State Operations and Technical Assistance calls and informational bulletins. Additionally, the CDC supports state departments of public health, HRSA engages with states on their maternal health programs and the U.S. Department of Agriculture interacts with states through the Supplemental Nutrition Assistance Program (SNAP). As part of its interactions with states, the federal government could raise awareness about the need to improve race, ethnicity and other demographic data and provide tools, resources and best practices. Through these same programs, it could provide additional resources to help states improve their data infrastructure. Medicaid allows states to receive a 90% match from CMS for state administrative activities related to development of core HIE services (e.g., designing and developing a provider directory, privacy and security applications and/or data warehouses), public health infrastructure, electronic Clinical Quality Measurement (eCQM) infrastructure and provider onboarding. Nearly all states already participate in the program, but it is not clear if states use these funds to improve their race and ethnicity data.

Recommendation 3.2:
HHS should develop educational resources on the importance of collecting self-reported race and ethnicity and data, and why the information is important and how it will be used. These resources should provide guidance on informing patients on prohibited uses of the data, to assuage concerns about potentially discriminatory impacts and to provide privacy assurances. In addition, HHS should and work with stakeholder groups (e.g., providers, consumer and community-based organization, public health) to disseminate these resources. Resources like the American Hospital Association (AHA) Health Research and Educational Trust Disparities Toolkit already exist to provide guidance on collecting race and ethnicity data and could be adapted for other stakeholders. With HHS endorsement, resources could be compiled, updated and distributed to a wider audience. Henry Ford Health System's *Why We Ask* page[50] is an example of these principles being translated into a plain-language resource for patients, and offers the concise summary: "We ask because we care!"

HHS should also issue guidance to clarify that federal regulations do not prohibit collecting race and ethnicity data. Doing so would address some perceived barriers to care that have been posited as reasons for not collecting race and ethnicity data. While entities cannot require people to answer these questions, federal law does not prohibit asking for race and ethnicity data.

Recommendation 3.3:
Health plans, provider systems and other health sector entities collecting race and ethnicity data should include community voices in developing and disseminating materials explaining why the information is collected, how it will be used and why it is important. Consumers are often unclear about why they are asked to provide the information and unsure how the information will be used, making them reticent to provide it. The use of "help text" explaining why a question is being asked has been shown to increase response rates. Working with communities to understand their concerns and collaborating with them on the development and dissemination of resources can also help increase response rates and help ensure that response options reflect the diversity of the community.

Information gleaned from data collection efforts should be shared with the communities being measured. Blue Cross Blue Shield of Massachusetts released a *Health Equity Report*[51] documenting disparities in care for its members. Ideally, future efforts would not only include frank documentation of the challenges payers and providers face in delivering equitable health care, but also the remedial actions they take to address them.

Recommendation 3.4:
The National Library of Medicine (NLM) should develop a database of existing resources that provide practical guidance for health plans and health care providers, and should promote their use. Toolkits can provide guidance on how to ask questions about race and ethnicity, how to support and train staff to collect data and how to use the data. There is also the potential to update existing resources; for example, the AHA Institute for Diversity and Health Equity's Disparities Toolkit was developed in 2007. Where best practices for data collection are unknown or disputed, or where updates and improvements to guidance for data collected are needed, the NLM, HHS, and others should fund research to resolve these issues.**Recommendation 3.5:**
Health care quality improvement organizations should leverage insights from diffusion of

innovation frameworks to promote peer learning regarding race and ethnicity data collection. Our interviews indicate that despite significant barriers, some health systems and plans have made significant progress toward complete, self-reported data, and their insights are often replicable.

Health care quality improvement organizations can also provide valuable technical assistance on analysis of race and ethnicity data. Interviews revealed that a lack of guidance can be a critical barrier: sometimes health care organizations resist collecting data they feel they lack the capacity or technical knowledge to productively analyze. Providing this guidance can help create a "virtuous cycle," where data are collected with a clear sense of how they will be used to improve care and promote equity. Clearly defining use cases can also support communicating with patients about the goals of race and ethnicity data collection.

4. Regulation

As noted above, multiple stakeholders expressed beliefs that one or more aspects of federal laws or regulations prevented them from collecting race and ethnicity data. Understanding what federal policies create real or perceived barriers to data collection (including, but not limited to, formal prohibitions) is critical to improving the data available in health programs. Moreover, as the range of stakeholders with interest in (and potential access to) health data grows with the proliferation of health apps, federal policy may need to expand to cover these domains as well.

Recommendation 4.1:
Congress should amend as necessary the Health Information Portability and Accountability Act (HIPAA), the Family Educational Rights and Privacy Act (FERPA), the Patient Protection and Affordable Care Act (ACA), and other data laws to encourage the collection and sharing of information across health, human and social services, including the justice system. Despite recent efforts to improve data sharing, multiple stakeholders noted that this remains a challenge will require congressional action to change.

Recommendation 4.2:
Improve privacy protections of data captured through health apps (e.g., broaden the FTC's authorization to oversee health apps). An increasing number of apps make it easier for consumers to track their health, set goals and monitor progress toward meeting them, but several interviewees suggested that more needs to be done to improve data protections because many apps are not covered by HIPAA. Improving data protections could lead to increased consumer confidence and a greater willingness to share the information.

Recommendation 4.3:
State legislators and regulators should similarly review whether laws or insurance codes prohibit or restrict collection of race and ethnicity data, and amend as necessary. For example, a recent public comment letter from the Blue Cross Blue Shield Association to the National Association of Insurance Commissioners cites several states with these restrictions in place.[42] This echoes a 2004 National Research Council report in which California, Maryland, New Hampshire, and New Jersey were identified as barring health plans from collecting race and ethnicity data.[40]

Recommendation 4.4:
The Department of Labor (DOL) should clarify whether, and when, employers are permitted to share race and ethnicity data collected as part of Equality Employment Opportunity Commission (EEOC) requirements. Under Title VII of the Civil Rights Act of 1964, private-sector companies with 100 or more employees are required to report workforce demographic data to the EEOC.52 While the resulting EEO-1 forms, which describe aggregate firm-level employment data by race and ethnicity, are treated as confidential by the EEOC, these regulations do not specify that employers must treat the disaggregated data they collect as confidential. By sharing workplace race and ethnicity data with the carrier providing health insurance to a company's employees, employers could help address a key data limitation identified in this report. Additionally, the DOL should provide guidance to employers about how to ensure employee privacy in this process and how best to represent the potential uses of the data collected.

Conclusion

The COVID pandemic has provided jarring evidence of racial and ethnic disparities in health care access, quality and outcomes. It has also revealed major weaknesses in health care and public health data resources; for example, as local, state and federal authorities struggled to assess the demographic composition of patients infected with or vaccinated against COVID. Improving these data systems in general, and addressing limitations to data on race and ethnicity in particular, will be a critical step in advancing health equity.

This report provides new details on the quality and comprehensiveness of race and ethnicity data in a wide range of federal health care and public health data bases. It draws on a scan of available literature and interviews with an array of health sector stakeholders on barriers to improving these data—and opportunities to do so. Last, it provides recommendations on how the federal and state governments, as well as other actors, can catalyze improvements in race and ethnicity data.

References

1. James C, Pamar S, Scholle SH, Saynisch P, Soucie J, Lyons B. *Federal Action Is Needed to Improve Race and Ethnicity Data in Health Programs.* Grantmakers in Health and NCQA; 2021. Accessed October 28, 2021. https://www.gih.org/wp-content/uploads/2021/10/GIH-Commonwealth-Fund-federal-data-report-part-1.pdf

2. CDC. Health Equity Considerations and Racial and Ethnic Minority Groups. Centers for Disease Control and Prevention. Published February 11, 2020. Accessed October 28, 2021. https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html

3. CDC. Risk of Severe Illness or Death from COVID-19. Centers for Disease Control and Prevention. Published February 11, 2020. Accessed October 28, 2021. https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/racial-ethnic-disparities/disparities-illness.html

4. CDC. Risk for COVID-19 Infection, Hospitalization, and Death By Race/Ethnicity. Centers for Disease Control and Prevention. Published February 11, 2020. Accessed October 28, 2021. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-race-ethnicity.html

5. COVID-19 Cases by Race/Ethnicity. KFF. Published September 9, 2021. Accessed October 28, 2021. https://www.kff.org/other/state-indicator/covid-19-cases-by-race-ethnicity/

6. Office of Management and Budget. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. The White House. Published October 30, 1997. Accessed September 13, 2021. https://obamawhitehouse.archives.gov/node/15626

7. Office of the Assistant Secretary for Planning and Evaluation. HHS Implementation Guidance on Data Collection Standards for Race, Ethnicity, Sex, Primary Language, and Disability Status. ASPE. Published October 30, 2011. Accessed September 13, 2021. https://aspe.hhs.gov/reports/hhs-implementation-guidance-data-collection-standards-race-ethnicity-sex-primary-language-disability-0

8. Jarrín OF, Nyandege AN, Grafova IB, Dong X, Lin H. Validity of Race and Ethnicity Codes in Medicare Administrative Data Compared With Gold-standard Self-reported Race Collected During Routine Home Health Care Visits. *Med Care.* 2020;58(1):e1. doi:10.1097/MLR.0000000000001216

9. Eicheldinger C, Bonito A. More Accurate Racial and Ethnic Codes for Medicare Administrative Data. *Health Care Financ Rev.* 2008;29(3):27-42.

10. 2019 Medicare Enrollment Section | CMS. Accessed October 29, 2021. https://www.cms.gov/research-statistics-data-systems/cms-program-statistics/2019-medicare-enrollment-section

11. ResDAC. Data Documentation: Long Term Care Minimum Data Set (MDS) 3.0. Accessed October 29, 2021. https://resdac.org/cms-data/files/mds-30/data-documentation

12. ResDAC. Data Documentation: Home Health Outcome and Assessment Information Set. Accessed October 29, 2021. https://resdac.org/cms-data/files/oasis

13. CMS. *Guidance Manual for Completion of the Hospice Item Set (HIS).*; 2021. Accessed October 29, 2021. https://www.cms.gov/files/document/his-manual-v300022321.pdf

14. Haas A, Elliott MN, Dembosky JW, et al. Imputation of race/ethnicity to enable measurement of HEDIS performance by race/ethnicity. *Health Serv Res.* 2019;54(1):13-23. doi:https://doi.org/10.1111/1475-6773.13099

15. ResDAC. CMS Fee List for Research Files. Accessed October 28, 2021. https://resdac.org/sites/datadocumentation.resdac.org/files/2021-08/CMS%20Fee%20List%20for%20Research%20Files_0.pdf

16. Medicaid and CHIP T-MSIS Analytic Files Data Release | CMS. Accessed October 28, 2021. https://www.cms.gov/newsroom/fact-sheets/medicaid-and-chip-t-msis-analytic-files-data-release

17. Single Topic Map View - Race and Ethnicity | DQ Atlas. Accessed October 28, 2021. https://www.medicaid.gov/dq-atlas/landing/topics/single/map?topic=g3m16&tafVersionId=18

18. Medicaid.gov. *DQ ATLAS Background and Methods Resource - Race and Ethnicity.* Accessed October 28, 2021. https://www.medicaid.gov/dq-atlas/downloads/background_and_methods/TAF_DQ_Race_Ethnicity.pdf

19. Marketplace Enrollment Tops 12 Million For 2021; Largest-Ever Funding For Navigators | Health Affairs. Accessed October 28, 2021. https://www.healthaffairs.org/do/10.1377/hblog20210422.65513/full/

20. 2021 Marketplace Open Enrollment Period Public Use Files | CMS. Accessed October 28, 2021. https://www.cms.gov/research-statistics-data-systems/marketplace-products/2021-marketplace-open-enrollment-period-public-use-files

21. CMS. 2021 Marketplace open Enrollment Period Public Use Files Definitions.pdf. Accessed October 28, 2021. https://www.cms.gov/files/document/2021-public-use-files-definitions.pdf

22. Health Insurance Coverage of Adults 19-64. KFF. Published October 23, 2020. Accessed October 28, 2021. https://www.kff.org/other/state-indicator/adults-19-64/

23. NCQA. *Health Equity and Social Determinants of Health in HEDIS: Data for Measurement.* NCQA; 2021. Accessed October 28, 2021. https://www.ncqa.org/wp-content/uploads/2021/06/20210622_NCQA_Health_Equity_Social_Determinants_of_Health_in_HEDIS.pdf

24. MN Department of Health. *Request for Information: MINNESOTA ALL PAYER CLAIMS DATABASE (MN APCD) USES AND PRIORITIES.*; 2017. Accessed October 28, 2021. https://www.health.state.mn.us/data/apcd/docs/apcdrfi.pdf

25. The RAND Corporation. The History, Promise and Challenges of State All Payer Claims Databases. :54.

26. All-Payer Claims Databases. Accessed October 28, 2021. http://www.ahrq.gov/data/apcd/index.html

27. Claims Data Release Rules. APCD Council. Published July 7, 2015. Accessed October 28, 2021. https://www.apcdcouncil.org/claims-data-release-rules

28. 38 U.S. Code § 101 - Definitions | U.S. Code | US Law | LII / Legal Information Institute. Accessed October 28, 2021. https://www.law.cornell.edu/uscode/text/38/101

29. Mor MK. Session #7: Assessing Race and Ethnicity in VA Data. Presented at: Database & Methods Cyberseminar Series; April 1, 2019. https://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/3546-notes.pdf

30. Federally Qualified Health Centers. Official web site of the U.S. Health Resources & Services Administration. Published April 21, 2017. Accessed October 28, 2021. https://www.hrsa.gov/opa/eligibility-and-registration/health-centers/fqhc/index.html

31. HRSA Health Center Program. *2021 UDS Manual.* HRSA; 2021. Accessed October 28, 2021. https://bphc.hrsa.gov/sites/default/files/bphc/datareporting/pdf/2021-uds-manual.pdf

32. data.HRSA.gov. Table 3B: Demographic Characteristics. Published 2019. Accessed October 28, 2021. https://data.hrsa.gov/tools/data-reporting/program-data/national/table?tableName=3B&year=2019

33. American Indian/Alaska Native - The Office of Minority Health. Accessed October 29, 2021. https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=3&lvlid=62

34. CDC WONDER. Natality, 2016-2019. Accessed October 28, 2021. https://wonder.cdc.gov/natality-expanded-current.html

35. Restricted-Use Vital Statistics Data. Published June 25, 2021. Accessed October 28, 2021. https://www.cdc.gov/nchs/nvss/nvss-restricted-data.htm

36. Are PRAMS data available to outside researchers? | CDC. Published June 30, 2021. Accessed October 28, 2021. https://www.cdc.gov/prams/prams-data/researchers.htm

37. Core Data Elements For IIS Functional Standards v4.0 | CDC. Published January 26, 2021. Accessed October 28, 2021. https://www.cdc.gov/vaccines/programs/iis/core-data-elements/iis-func-stds.html

38. Ndugga N, Hill L, Artiga S. Latest Data on COVID-19 Vaccinations by Race/Ethnicity. KFF. Published September 9, 2021. Accessed September 13, 2021. https://www.kff.org/coronavirus-covid-19/issue-brief/latest-data-on-covid-19-vaccinations-race-ethnicity/

39. Gold JAW, DeCuir J, Coyle JP, et al. COVID-19 Case Surveillance: Trends in Person-Level Case Data Completeness, United States, April 5-September 30, 2020. *Public Health Rep Wash DC 1974.* 2021;136(4):466-474. doi:10.1177/0033354921006973

40. National Research Council. *Eliminating Health Disparities: Measurement and Data Needs.* (Ploeg MV, Perrin E, eds.). The National Academies Press; 2004. doi:10.17226/10979

41. Data NRC (US) P on DC of R and E, Melnick D, Perrin E. *IMPROVING RACIAL AND ETHNIC DATA ON HEALTH.* National Academies Press (US); 2003. Accessed October 29, 2021. https://www.ncbi.nlm.nih.gov/books/NBK222062/

42. Comment Letters to Special NAIC (EX) Committee on Race and Insurance 2021. Published online 2021. Accessed October 29, 2021. https://content.naic.org/sites/default/files/call_materials/2021%20Comment%20Letters%20-%20July%201.pdf

43. Institute of Medicine (US) Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care.* (Smedley BD, Stith AY, Nelson AR, eds.). National Academies Press (US); 2003. Accessed March 16, 2021. http://www.ncbi.nlm.nih.gov/books/NBK220358/

44. AHA Disparities Toolkit | IFDHE. Accessed October 29, 2021. https://www.aha.org/hretdisparities/toolkit

45. Provenance - FHIR v4.0.1. Accessed November 22, 2021. https://www.hl7.org/fhir/provenance.html

46. Dearing JW, Cox JG. Diffusion Of Innovations Theory, Principles, And Practice. *Health Aff (Millwood).* 2018;37(2):183-190. doi:10.1377/hlthaff.2017.1104

47. Race - FHIR v4.0.1. Accessed November 10, 2021. https://www.hl7.org/fhir/v3/Race/cs.html

48. Ethnicity - FHIR v4.0.1. Accessed November 10, 2021. https://www.hl7.org/fhir/v3/Ethnicity/cs.html

49. Office of the National Coordinator for Health Information Technology. What is FHIR? Published 2019. Accessed November 22, 2021. https://www.healthit.gov/sites/default/files/2019-08/ONCFHIRFSWhatIsFHIR.pdf

50. Why We Ask. Accessed October 29, 2021. https://www.henryford.com:443/about/culture/diversity/why-we-ask

51. Health Equity Report | Blue Cross Blue Shield of Massachusetts. Accessed October 29, 2021. https://www.bluecrossma.org/myblue/equity-in-health-care/health-equity-report

52. EEO-1 Data Collection | U.S. Equal Employment Opportunity Commission. Accessed November 12, 2021. https://www.eeoc.gov/employers/eeo-1-data-collection

Appendix Table 1: Complete Race and Ethnicity Data Framework

DATA CATEGORY	SOURCE	UNIT OF REPORTING OR ANALYSIS	DATA COLLECTION STANDARD	SELF-REPORTED?	PERCENTAGE OF INDIVIDUALS WITH RACE/ETHNICITY DATA	AVAILABLE FOR ANALYSIS?
Medicare						
Enrollment	Master Beneficiary Summary File (MBSF)	Individual	Varies; relies on Social Security Admin data. Before 1980, limited to White/Black/Other/Unknown; used 6 OMB categories from 1980-1989; since 1989, only collected if a new name or Social Security number requested via SS-5.	Primarily drawn from SSA records; may reflect spouse's race/ethnicity; widely used RTI definitions rely on imputation to improve identification of Asian and Hispanic beneficiaries	1.8% Unknown; 2.6% Unknown or Other (2019)	RIF
Use of Services	Medicare Encounter and FFS Files	Encounter	Merged from MBSF, not collected in claims	Merged from MBSF, not collected in claims	Merged from MBSF, not collected in claims	RIF
Quality of Care and Outcomes	Medicare Part C/D Stratified Reporting	Plan	MBSF data supplemented with MBISG imputation. Only White, Black, Asian or Pacific Islander (API) and Hispanic available	Imputed group membership "For reporting HEDIS data stratified by race and ethnicity, racial and ethnic group membership is estimated using a methodology that combines information from CMS administrative data, surname, and residential location"	Relies on imputed race/ethnicity data	PUF
	CMS Care Compare Provider Data	Provider	NA	NA	0%	PUF
	Long Term Care Minimum Data Set	Individual	OMB 1997	Yes	Unknown	RIF
	Home Health Outcome and Assessment Information Set	Individual	OMB 1997	Yes ("as identified by the patient or caregiver")	Unknown	RIF

Appendix Table 1: Complete Race and Ethnicity Data Framework

DATA CATEGORY	SOURCE	UNIT OF REPORTING OR ANALYSIS	DATA COLLECTION STANDARD	SELF-REPORTED?	PERCENTAGE OF INDIVIDUALS WITH RACE/ETHNICITY DATA	AVAILABLE FOR ANALYSIS?
Medicare						
Quality of Care and Outcomes (continued)	CMS Hospice Item Set	Individual	OMB 1997	Yes	Unknown	Unknown
	CROWNWeb	Individual	OMB 1997	Yes	Unknown	RIF
	Hospital Compare	Facility	Varies; some measure sources have no race/ethnicity data; HCAHPS is similar to OMB 1997 with additional Hispanic subgroups	HCAHPS is self-reported	Unknown	Public reports
Cost of Care	MBSF Cost and Utilization Segment	Individual	NA	NA	0% (linkable to MBSF)	RIF

Appendix Table 1: Complete Race and Ethnicity Data Framework

DATA CATEGORY	SOURCE	UNIT OF REPORTING OR ANALYSIS	DATA COLLECTION STANDARD	SELF-REPORTED?	PERCENTAGE OF INDIVIDUALS WITH RACE/ETHNICITY DATA	AVAILABLE FOR ANALYSIS?
Medicaid/CHIP						
Enrollment	T-MSIS Analytic Files (TAF) Demographic and Eligibility (DE) File	Individual	HHS 2011 plus "unknown" categories; also reported rolled up to OMB 1997	"Although states are expected to report information on both race and ethnicity in T-MSIS, some states may not submit complete information because the data were not collected or technical difficulties arose in reporting. Some states may not have complete data on race and ethnicity because they follow the guidance from the Office of Management and Budget that establishes self-identification as the preferred means of obtaining this information, and not all beneficiaries disclose this information."	Varies by state; 17 with 90+% data, 20 with "high concern" or "unusable" TAF DE data quality	RIF
Use of Services	TAF Utilization Files	Encounter	NA	NA	0% (linkable to TAF DE file)	RIF
Quality of Care and Outcomes	Child and Adult Health Care Core Quality Measures	State	NA	NA	0%	PUF
Cost of Care	TAF Utilization Files	Encounter	NA	NA	0% (linkable to TAF DE file)	RIF

Appendix Table 1: Complete Race and Ethnicity Data Framework

DATA CATEGORY	SOURCE	UNIT OF REPORTING OR ANALYSIS	DATA COLLECTION STANDARD	SELF-REPORTED?	PERCENTAGE OF INDIVIDUALS WITH RACE/ETHNICITY DATA	AVAILABLE FOR ANALYSIS?
Marketplace			PUF			
Enrollment	Marketplace Open Enrollment Period PUF	State	Varies. State-based Marketplaces (SBM) report race and ethnicity as a single variable. HC.gov uses 2011 standards but report rolls up to 6 OMB categories plus multiple races and separate Hispanic, non-Hispanic ethnicity. CO does not report.	Yes	11%-59% unknown race; 4%-42% unknown ethnicity among HC.gov states	PUF
Use of Services	None available	NA	NA	NA	NA	NA
Quality of Care and Outcomes	Child and Adult Health Care Quality Measures	Plan	NA	NA	0%	PUF
Cost of Care	CMS Center for Consumer Information and Oversight (CCIIO) Claims Cost Data	State	NA	NA	0%	PUF

Appendix Table 1: Complete Race and Ethnicity Data Framework

DATA CATEGORY	SOURCE	UNIT OF REPORTING OR ANALYSIS	DATA COLLECTION STANDARD	SELF-REPORTED?	PERCENTAGE OF INDIVIDUALS WITH RACE/ETHNICITY DATA	AVAILABLE FOR ANALYSIS?
Other Commercial Insurance						
Enrollment	None available	NA	NA	NA	NA	NA
Use of Services	All Payer Claims Databases	Encounter	Varies. OMB 1997 most common, some states (e.g. ME, MN) don't include race or ethnicity at all	Varies by state and payer within state.	Varies; one study found 19%-100% missing data for race (2014)	Only 9 states have posted rules for data access
Quality of Care and Outcomes	Healthcare Effectiveness Data and Information Set (HEDIS) Quality Compass	Plan	NA	NA	0%. 24% of plans have race data on 50+% of members, 6% of plans have ethnicity data on 50%+ of members (2019).	Available under license from NCQA
Cost of Care	All Payer Claims Databases (APCD)	Encounter	Varies. OMB 1997 most common, some states (e.g. ME, MN) don't include race or ethnicity at all	Varies by state and payer within state.	Varies;- one study found 19%-100% missing data for race (2014)	Only 9 states have posted rules for data access

Appendix Table 1: Complete Race and Ethnicity Data Framework

DATA CATEGORY	SOURCE	UNIT OF REPORTING OR ANALYSIS	DATA COLLECTION STANDARD	SELF-REPORTED?	PERCENTAGE OF INDIVIDUALS WITH RACE/ETHNICITY DATA	AVAILABLE FOR ANALYSIS?
Public Health Data						
Vital Statistics	Births	Individual	HHS 2011; also reported rolled up to OMB 1997	Instructions say "what the mother/father considers herself/himself to be"	Mother: 100% for race, 99% for Hispanic origin; Father: 82% for race, 87% for Hispanic origin	PUFs for national data; restricted-use files with more granular geographic identifiers
	Deaths	Individual	OMB 1997 (since 1999)	Reported by funeral director/ family of deceased	Over 99%	
Immunizations	Immunization Information Systems (IIS)	Individual	OMB 1997 (CDC functional standards)	Unknown	Implementation varies across state, territorial and local registries. Race/ ethnicity data were available for 57% of COVID vaccine recipients.	Not available
Healthcare Acquired Infections	CDC National Healthcare Safety Network Data	Individual	NA	NA	0%	Public reports available about infection rates
Community-acquired infections (COVID, STIs)	Human infection with Coronavirus Disease 2019 Surveillance Worksheet	Individual	OMB 1997	Instructions ambiguous	Through Sept 2020, 57% complete	RIF with combined race/ethnicity variable

Appendix Table 1: Complete Race and Ethnicity Data Framework

DATA CATEGORY	SOURCE	UNIT OF REPORTING OR ANALYSIS	DATA COLLECTION STANDARD	SELF-REPORTED?	PERCENTAGE OF INDIVIDUALS WITH RACE/ETHNICITY DATA	AVAILABLE FOR ANALYSIS?
Public Health Data						
Community-acquired infections (COVID, STIs) (continued)	Adult HIV Confidential Case Report	Individual	OMB 1997	CDC guidelines refer to HHS 2011 standards, which include self-report	Unknown	Unknown
	National Notifiable Disease Surveillance System/National Electronic Disease Surveillance System	Event	OMB 1997 (as least for WONDER tables)	Recommended but not uniformly applied	Unknown	Public reports available about infection rates
Pregnancy Risk Assessment Monitoring System (PRAMS)	PRAMS	Individual (birth)	Modified OMB 1997; separate Chinese, Filipino, Japanese and Native Hawaiian categories, along with "Asian" roll-up, for mother and father	Extracted from birth certificate records	Unknown	RIF

Appendix Table 1: Complete Race and Ethnicity Data Framework

DATA CATEGORY	SOURCE	UNIT OF REPORTING OR ANALYSIS	DATA COLLECTION STANDARD	SELF-REPORTED?	PERCENTAGE OF INDIVIDUALS WITH RACE/ETHNICITY DATA	AVAILABLE FOR ANALYSIS?
Other Federal Delivery Systems						
Indian Health Service	National Patient Information Reporting System	Individual	Blood Quantum and Tribal Group	Unknown	Only contains records from individuals receiving services from IHS providers. 78% of American Indian/Alaska Native population lives outside tribal statistical areas	Available via IHS General Data Mart (GDM)
Veterans Health Administration	VA Corporate Data Warehouse	Individual	OMB 1997 with "Spanish" as additional, alternate description of "Hispanic" category	Intended to be self-report but source and process varies; 1/3 fill out intake forms online; data on remaining 2/3 collected on intake at VA medical center either by filling out forms themselves or having VA staff read form and record responses	90%+	Need to be VA employee or have "without compensation" affiliate status
Defense Health Agency	Military Health System Mart Defense Enrollment Eligibility Reporting System (MHS M2/ DEERS)	Individual (enrolled service member)	OMB 1997 with additional national origin variable, but not an exact match to 2011 HHS categories	Data reflects enrolled service member ("sponsor"), not dependents	Unknown	Unknown
Federally Qualified Health Centers (FQHCs)	Health Resources and Services Administration (HRSA) Uniform Data System	Center	Modified OMB 1997; separate "Native Hawaiian" and "Other Pacific Islander" categories; separate "More than one race" line item	Yes, instructions are to collect from patients at registration, then annually	85% (2019)	Publicly available at the FQHC level

Appendix Table 2: Reports on Racial and Ethnic Disparities and Health Equity Data

TITLE	YEAR	INSTITUTION
Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care	2003	Institute of Medicine
Improving Racial and Ethnic Data on Health: Report of a Workshop	2004	National Research Council (National Academies of Sciences, Engineering, and Medicine)
Eliminating Health Disparities: Measurement and Data Needs	2004	National Research Council (National Academies of Sciences, Engineering, and Medicine)
Race, Ethnicity and Language Data: Standardization for Health Care Quality Improvement	2009	Institute of Medicine
Accounting for Social Risk Factors in Medicare Payment: Data	2016	National Academies of Sciences, Engineering, and Medicine
First Report to Congress: Social Risk Factors and Performance in Medicare's Value-Based Purchasing Programs	2016	Office of the Assistant Secretary for Planning and Evaluation
Second Report to Congress: Social Risk Factors and Performance in Medicare's Value-Based Care Programs	2020	Office of the Assistant Secretary for Planning and Evaluation
Social Risk Trial Final Report	2021	National Quality Forum
Proposed Changes to Existing Measures for HEDIS® Measures	2021	National Committee for Quality Assurance


GRANT
MAKERS
IN
HEALTH


NCQA
Measuring quality.
Improving health care.