## Contents

### Executive Summary

1

### Introduction

4

- How Does CMS Define Health Equity Data? 4
- Why are Health Equity Data Elements Important? 5
- What is CMS’s Commitment to Future Improvement of Health Equity Data? 6

### Current State and Challenges of CMS Health Equity Data

6

- Summary of Health Equity Data Issues Impacting CMS Programs 7
- Sociodemographic Data 8
- Social Drivers/Determinants of Health (SDOH) Data 14

### Progress to Date

16

- Collecting New Elements and Filling Existing Gaps in CMS’s Health Equity Data 17
- Equipping the Industry with New Tools and Capabilities 18
- Providing Access to Disaggregated Data and Insights 19

### Vision for the Future

19

- Actions to Drive the Future of CMS Health Equity Data 20
- Elements Critical to CMS’s Health Equity Data Strategy Success 27

### Conclusion

29

### Appendix A – Data Standards

30

### References

31
Executive Summary

The Centers for Medicare & Medicaid Services (CMS) relies on data to inform planning and execution of policies that affect more than 170 million individuals enrolled in CMS programs nationwide, including but not limited to Medicare, Medicaid & the Children’s Health Insurance Program (CHIP), and the Health Insurance Marketplace®. Through these programs, CMS strives to identify and remedy systemic barriers to health equity so that every person served by its programs has a fair and just opportunity to attain their optimal health regardless of race, ethnicity, tribal sovereignty, disability, sexual orientation, gender identity, socioeconomic status, geography, preferred language, or other factors that affect access to care and health outcomes.

The CMS Administrator’s Six Strategic Pillars demonstrate CMS’s commitment to address health disparities, expand access, and impact Americans’ lives in a sustainable, person-centered, and inclusive way. The first pillar of the CMS Strategic Plan is to advance health equity by addressing the health disparities that underlie our health system. CMS is building this pillar into its core functions. CMS’s strategy will build on the Biden-Harris Administration’s commitment to advancing racial equity and supporting communities who have been historically underserved through the federal government, as described in President Biden’s Executive Order 13985.

The CMS Framework for Health Equity outlines the approach CMS will use to promote health equity for enrollees, mitigate health disparities, and prioritize CMS’s commitment to expanding the collection, reporting, and analysis of standardized data. CMS recognizes that increasing the collection of standardized sociodemographic and social determinants of health (SDOH) data across the health care industry is an important first step towards improving population health.

As the largest payer of health care in the United States, CMS can set the bar for meaningful health equity data collection and use across the health care system. Health equity data elements—i.e., the combination of quantitative and qualitative elements that enable the examination of health differences between populations and their causes—are essential to understanding and addressing health disparities for enrollees across CMS programs.

---

*a Health Insurance Marketplace® is a registered service mark of the U.S. Department of Health & Human Services.*
programs. Accurate and complete data elements support CMS in its efforts to create evidence-based policies and regulations and to assess how well these policies and regulations align with the needs of the communities and individuals that CMS serves.

To strengthen and improve the accuracy of enrollee health equity data, CMS has enhanced the collection of data elements across its programs, including limited collection of SDOH data or Health-Related Social Needs (HRSN) data. For example, the Center for Medicare & Medicaid Innovation’s (CMS Innovation Center’s) Accountable Health Communities (AHC) Model included a mandatory HRSN screener, which includes items such as housing instability, food insecurity, and transportation needs, and CMS has mandated the future collection of new health equity-related data elements in post-acute care (PAC) settings, such as skilled nursing facilities, with information including race, ethnicity, transportation, social isolation, health literacy, and preferred language. CMS has developed tools, such as the Mapping Medicare Disparities Tool,\textsuperscript{4} that allow the public to access and use their health equity data, and makes additional resources available to the public to address health disparities, including the CMS Health Equity Technical Assistance Program\textsuperscript{5} and the use of the CMS Disparity Impact Statement.\textsuperscript{6}

Although progress has been made, gaps in the availability, completeness, and quality of health equity data remain across CMS programs. Historically, some data elements were collected in forms that are not aligned to current standards, such as those established by the U.S. Department of Health and Human Services (HHS) in 2011.\textsuperscript{7} HHS standards include the 2011 HHS data standards and the HHS Office of the National Coordinator for Health Information Technology (ONC) United States Core Data for Interoperability (USCDI).\textsuperscript{8} The USCDI consists of data elements and associated vocabulary standards to support computerized, interoperable use of equity data. The USCDI includes standards for race, ethnicity, preferred language, sex, and disability status aligned with the 2011 HHS data standards that can be used in conjunction with the 2011 HHS data standards and guidance. Where applicable and appropriate, CMS will align with the USCDI standards. The collection of data elements not aligned to HHS data standards can result in inconsistent analyses when using and combining health equity data from multiple sources. Additionally, the lack of consistent data collection at a disaggregated\textsuperscript{b} level, broken down into detailed sub-categories, presents barriers to understanding the needs of specific subgroups, as well as to comparing data across programs and populations as needed or desired.

Efforts to address these health equity-related data issues are already underway and will be prioritized as CMS pursues its future vision for health equity data. Efforts include collecting new health equity data elements across CMS programs to fill existing gaps, aligning health equity data to acceptable standards across all elements, considering health equity measures and health equity scores (e.g., the Health Equity Summary Score), equipping the industry with new tools and capabilities aligned to health equity goals, and providing

\textsuperscript{b} In this context, disaggregated data refers to data that has been divided into smaller units in order to see specific trends of sub-groups within the dataset.
access to disaggregated data insights that the public can use to drive action. This paper, as summarized in the table below, describes the current state of enrollee-focused health equity data across CMS programs, along with the progress made to date and the future actions that will achieve CMS’s vision.

Table 1. Current State and Challenges of Health Equity Data, Progress to Date, and Future Actions

<table>
<thead>
<tr>
<th>Health Equity Data at CMS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current State and Challenges</strong></td>
</tr>
<tr>
<td>- CMS has limited authority to directly collect health equity data elements (e.g., race) from all program enrollees across all settings and control collection standards</td>
</tr>
<tr>
<td>- CMS programs collect limited SDOH data elements in limited settings (e.g., certain Innovation Center Models)</td>
</tr>
<tr>
<td>- Not all elements have adopted data standards; CMS is continuing to explore available standards for each program context</td>
</tr>
<tr>
<td>- Collecting self-reported data in limited settings results in incomplete sociodemographic data</td>
</tr>
<tr>
<td>- Utilization remains low for some standardized data collection methods (e.g., ICD-10-CM Z55-Z65 Codes, CMS Social Needs Screening Tool) many do not adequately capture drivers of health for all populations</td>
</tr>
</tbody>
</table>

| **Progress to Date** |
| - Collecting new data elements across CMS programs to fill gaps (e.g., SDOH data collection in PAC settings) |
| - Collecting race/ethnicity data elements using the 2011 HHS standards in long-term care and PAC assessments, mapping assessments to the USCDI in the data element library, and continuing advancement of implementation guidance for computable assessment data elements including health equity data |
| - Assessing and modifying applications of the Medicare race/ethnicity imputation algorithm across CMS programs to improve data quality in the short term while CMS works to obtain self-reported race/ethnicity data |
| - Equipping the industry with new tools and capabilities (e.g., practical guides on SDOH assessments) aligned to health equity goals |
| - Providing access to certain disaggregated data and insights that the public can use to drive action |

| **Future Actions** |
| - Filling gaps in existing data through new collection venues (Part C/D enrollment), collaborating with partners (e.g., Social Security Administration to collect race and ethnicity data at the time of birth), and allowing broader access to statistical methods to predict race and ethnicity when data elements are missing |
| - Collecting new data elements by increasing utilization of established, reliable collection methods (e.g., Z55-Z65 ICD-10-CM codes for SDOH data) and identifying new venues for collection (e.g., certain CMS Innovation Center Models, the Medicare Current Beneficiary Survey for sexual orientation and gender identity (SOGI)) |
| - Aligning data to standards across existing and emerging data collection practices including to the United States Core Data for Interoperability (USCDI) standards where appropriate and applicable. |
| - Disaggregating data both in collection practices and reporting on outcomes and insights about specific subpopulations when appropriate given the sample size and data availability, following HHS data guidelines to ensure protection of participants’ privacy and identities |
| - Continuing development of equity scores, including refining the Health Equity Summary Score, developing other indices (e.g., Medicare Part C/D Star Ratings health equity index), providing reports to industry stakeholders (e.g., Health Equity Report Card), and considering new applications for equity scores |
| - Releasing resources and training to help health care professionals act on equity data |
| - Addressing bias in tools and methods, including the Medicare Bayesian Improved Surname Geocoding, the Health Equity Summary Score, and other Artificial Intelligence (AI) measurement programs, before release |
| - Acting on feedback received from the public and academic community to further improve health equity data |
Introduction

Since its creation over a decade ago, the CMS Office of Minority Health (CMS OMH) has led efforts across the agency to advance health equity, shaping program policies and producing tools and research that improve health and health care, with a focus on communities and individuals who are underserved. As an update to the previous Medicare-focused CMS Equity Plan to Improve Quality in Medicare, CMS OMH, in collaboration with other CMS Centers and Offices, has launched an enhanced and more comprehensive 10-year approach to further embed health equity within all CMS programs, including Medicare, Medicaid, CHIP, and the Health Insurance Marketplace®: the CMS Framework for Health Equity.

The CMS Framework for Health Equity aligns with the federal government's goal of advancing equity, which is to provide everyone with the opportunity to reach their full potential. Consistent with this aim, the CMS Framework for Health Equity supports CMS's ability to assess whether, and to what extent, its programs and policies perpetuate or exacerbate systemic barriers to opportunities and benefits for the communities referenced above. This includes understanding and addressing the ways in which Medicare, Medicaid, CHIP, and the Health Insurance Marketplace® meet the needs of those we serve, particularly communities and individuals who have been historically underserved.

This white paper supports the CMS Framework for Health Equity’s priorities by outlining how CMS intends to expand the collection, reporting, and analysis of standardized data and use of health equity data to achieve its equity goals. We provide an overview of the current status of health equity data in CMS programs, along with challenges encountered, define actions that CMS has already undertaken, and describe future plans to continue to address challenges in health equity data.

How Does CMS Define Health Equity Data?

Health equity data is the combination of quantitative and qualitative elements that enable the examination of health differences and their causes between populations. This includes sociodemographic data such as race, ethnicity, tribal sovereignty, language, geography, sex, sexual orientation and gender identity (SOGI), language, income, and disability status, as well as SDOH featured in the Healthy People 2030 SDOH Framework across five domains: economic stability, education access and quality, health care access and quality, neighborhood and built environment, and social and community context.

CMS collects and leverages health equity data from a broad range of sources such as: enrollment and eligibility data for Medicare, Medicaid, CHIP, and the Health Insurance Marketplace®; administrative claims...
data from health care professionals; measures aligned to encounter and clinical data from electronic medical records or screening tools; survey data from the U.S. Census Bureau, CMS’s Consumer Assessment of Healthcare Providers & Systems (CAHPS®) surveys, the Medicare Current Beneficiary Survey, PAC assessments, and other self-reported experiences; community-based participatory research, listening sessions, and focus groups; and measurements such as quality metrics, quality reporting programs (e.g., Star Ratings), the Health Equity Summary Score, and indices such as the Centers for Disease Control and Prevention (CDC) Social Vulnerability Index. All of these sources can provide insight into existing disparities and track progress toward achieving health equity.

Why are Health Equity Data Elements Important?

Individuals belonging to specific populations, such as racial and ethnic minority groups, American Indian and Alaska Native tribes or lesbian, gay, bisexual, transgender, queer, and intersex (LGBTQI+) populations, often face greater health and health care challenges than their counterparts in the general population. These challenges are due, in part, to social risk factors, and have included higher rates of chronic conditions, lower access to care, and worse health outcomes (e.g., mortality, hospitalization), often compounded by differential treatment and discrimination in health care settings. Health equity data elements are essential for identifying where disparities exist, directing efforts and resources to address these disparities as they are identified, measuring progress towards improving equity, and establishing accountability for achieving progress.

A growing body of literature suggests that increasing the collection of standardized sociodemographic data and SDOH across health care systems is an important first step towards improving population health. SDOH, a vital portion of health equity data, can significantly affect health care access and outcomes. Medical care alone has been found to cause only 10-15% of preventable mortality in the U.S., with studies showing that factors like educational attainment can be associated with almost half of deaths among working-age adults in the U.S. Data on social risk factors, experience of care, and comprehensive patient sociodemographic data, including race, ethnicity, tribal sovereignty, language, gender identity, sex, sexual orientation, and disability status is a valuable tool for quality improvement.

Increasing the availability of standardized data across settings and programs enables CMS and stakeholders to address changes in populations over time, ensure alignment across HHS and other federal agencies, and leverage information to connect individuals to appropriate and needed social services and supports. Increasing the availability of standardized data encompasses understanding and standardizing data collection across federal agencies serving communities that have been historically underserved, including American Indian and Alaska Native tribes and rural communities, as well as programs with benefits that address individuals’ social risk factors and unmet social needs. Sociodemographic, SDOH, HRSN, and other health equity data can help
drive quality improvement and strengthen CMS’s ability to evaluate programs and policies to assess their influence on health outcomes.

**What is CMS’s Commitment to Future Improvement of Health Equity Data?**

CMS has already made strides to enhance health equity data focused on enrollees, including assessing the current state of sociodemographic and SDOH data collection, targeting collection of new data elements important to understanding disparities across groups who are historically underserved, developing new tools and algorithms to translate individual data points into actionable insights, and working to expand access to health equity data across the health care system.\(^{26,27,28}\) Even with these improvements, there is still much progress to be made to ensure all individuals have access to equitable care and coverage and for CMS, other government agencies, and industry organizations to obtain the health equity data and insights they need to achieve our future vision of health equity. This paper summarizes the actions CMS intends to take to achieve these goals.

**Current State and Challenges of CMS Health Equity Data**

Understanding the current state of health equity data within CMS programs is the first step towards ensuring health equity data quality, accuracy, and completeness in the future. Both socioeconomic data and SDOH data play an important role in understanding health disparities and planning interventions that address them. Improving the quality, completeness, and accuracy of both socioeconomic data and SDOH data requires rigorous collection aligned to appropriate standards. The subsequent sections of the paper describe how CMS is addressing the collection of this critical data.

CMS currently collects and consolidates data under standards established by the Office of Management and Budget (OMB) and HHS, among other authorities. These standards, discussed in greater detail in the Appendix of this report, help establish a common language for data sharing across agencies, programs, and organizations as well as consistency in the quality and completeness of data. However, the degree to which current CMS health equity data elements align to these standards has varied. New standards such as the HHS ONC USCDI standards present technical updates to earlier HHS standards and opportunities for focused alignment and harmonization in some CMS contexts when appropriate and applicable.\(^8\) The next section describes the current state of health equity data across CMS programs, including alignment to standards where appropriate and areas where improvement is needed.

**Summary of Health Equity Data Issues Impacting CMS Programs**

Complete, high-quality, accurate data must form the foundation of efforts to advance health equity. Significant gaps in the availability, structure, and quality of health equity data remain, including the following:
- **Incomplete data reported to CMS on key health equity elements prevents fully data-driven decisions.** Gaps in data challenge CMS’s ability to understand and assess disparities in health care access, quality, and outcomes. For example, CMS does not collect certain SDOH data elements, such as income, for individuals across many of its programs, preventing a complete picture of factors driving health outcomes. Missing health equity data elements may arise when individuals opt out of voluntary self-reported data, CMS has limited authority to collect specific data elements, programs lack policies to require collection of important data elements at all or with a frequency to promote timely capture, or data quality cannot be confirmed. These issues may prevent CMS from interpreting and acting upon health equity data in a timely manner, challenge CMS’s ability to match or combine data elements across CMS programs and data sources, or limit datasets to containing information on only a subset of those served.

- **Health equity data is collected in inconsistent forms, often when it is not aligned to standards or standards do not currently exist.** Lack of data standards and inconsistent use of existing data standards present issues in combining data across CMS’s many sources without losing accuracy. For instance, CMS receives much of its sociodemographic data from the Social Security Administration (SSA) for individuals enrolled in Medicare and from states for individuals enrolled in Medicaid, CHIP, and Health Insurance Marketplace® programs, each with varied alignment to 2011 HHS data standards. The 2011 HHS data standards were originally designed to promote uniform data collection in surveys conducted or sponsored by HHS as required by Section 4302 of the Affordable Care Act and have provided guidelines helpful to CMS programs to track, combine, and compare data over time and across populations. The more recent HHS USCDI consists of data elements to support computerized, interoperable use of equity data aligned with the domains of the 2011 HHS data standards. As a content standard, adoption of health information technology (IT) is not necessary to align with its respective equity related data elements classes and elements. Lack of a common health equity data environment with robust documentation prevents consistent access and use of elements across the CMS enterprise.

- **Lack of disaggregated data can obscure meaningful insights.** Without nuanced, person-centered data elements that support disaggregation across specific subpopulations, issues in access, quality, and outcomes can be masked. For example, disaggregating race and ethnicity data from the Medicare Advantage Health Outcomes Survey (HOS), including that of Asian and Pacific Islander populations, into appropriate subgroups in line with 2011 HHS standards (e.g., Asian Indian, Chinese, Filipina, Japanese, Korean, Vietnamese, Native Hawaiian, Guamanian or Chamorro, Samoan) revealed significantly different cardiovascular disease and stroke burden among Asian subpopulations.

- **Bias in health equity data collection methods challenges CMS’s ability to interpret data.** Implicit bias in how data elements are collected, consolidated, or used can contribute to missing, inconsistent, or hidden factors affecting populations served. For example, limited understanding of how SOGI affects an individual’s health has historically prevented CMS programs from collecting or developing standards around SOGI data that would allow a detailed view of the population’s needs.
These issues, discussed in the context of specific data elements below, have detrimental impacts on CMS programs’ ability to promote health equity. Efforts to address them are already underway and will also be prioritized as CMS pursues our future vision for health equity data.

**Sociodemographic Data**

Sociodemographic data plays a foundational role in understanding health disparities across CMS programs and includes elements like age, race, ethnicity, tribal sovereignty, sex, income, geography, and other factors that provide an objective picture of an individual’s or community’s identity. These data elements can be linked to health care quality, outcome, or other data types to reveal disparities among certain population groups.

Sociodemographic data is often more readily available than SDOH data because these elements are often collected as standard practice for medical records, social security, Census tabulation, and other functions. However, CMS’s alignment with the 2011 HHS standards for sociodemographic data collection across these sources and CMS programs varies. The most common form of sociodemographic data collection is through self-reported means such as surveys with pre-prescribed responses. The composition of the question, the exhaustiveness of response options, and the frequency with which data elements are collected can all contribute to inconsistencies and inaccuracies. Furthermore, accurate survey responses are not always available for all enrollees, and CMS does not survey all of its program participants as a standard practice.

CMS does not collect all of its sociodemographic data elements directly from enrollees via mandatory survey responses or other means, as it has limited statutory authority to do so. To maintain the Medicare Enrollment Database that captures many sociodemographic elements, CMS receives eligibility and enrollment data from the SSA. Due to changes in forms over time, SSA no longer collects some data elements. Other potential options include updating Medicare information when participants contact CMS through their Medicare account or 1-800-Medicare, although this would not apply to all participants. Sociodemographic elements tied to the Medicaid and CHIP programs, along with some elements linked to individuals covered by State-based Exchanges, come from individual states. CMS’s limited control over data collection, along with limitations in what data elements have historically been available, has resulted in several issues in sociodemographic data completeness and alignment to standards. For instance, with the exceptions of biological sex and an individual’s location or address, CMS does not have complete data across any sociodemographic fields noted in the table below. The following table summarizes these issues for all sociodemographic elements described in Executive Order 13985 across CMS programs, although it does not provide the percentage of enrollees with

<table>
<thead>
<tr>
<th><strong>CMS Sociodemographic Data: Key Challenges</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ CMS has limited authority to directly collect health equity data elements (e.g., race) from all enrollees across all settings and control collection standards</td>
</tr>
<tr>
<td>▪ Not all sociodemographic data elements have widely adopted standards, though some are in development</td>
</tr>
<tr>
<td>▪ Collecting self-reported data in limited settings results in incomplete sociodemographic data across all CMS programs</td>
</tr>
</tbody>
</table>

---

[35] Other potential options include updating Medicare information when participants contact CMS through their Medicare account or 1-800-Medicare, although this would not apply to all participants. Sociodemographic elements tied to the Medicaid and CHIP programs, along with some elements linked to individuals covered by State-based Exchanges, come from individual states. CMS’s limited control over data collection, along with limitations in what data elements have historically been available, has resulted in several issues in sociodemographic data completeness and alignment to standards. For instance, with the exceptions of biological sex and an individual’s location or address, CMS does not have complete data across any sociodemographic fields noted in the table below. The following table summarizes these issues for all sociodemographic elements described in Executive Order 13985 across CMS programs, although it does not provide the percentage of enrollees with
available data. Given CMS’s limited control of data collection in State-Based Exchanges, only the Health Insurance Marketplace® (“Marketplace®”) is included in the table for Health Insurance Exchanges programs.

Table 2. Current State of Sociodemographic Data Across CMS Programs

<table>
<thead>
<tr>
<th>Sociodemographic Data Type</th>
<th>Fee-for-Service Medicare™</th>
<th>Medicare Advantage***</th>
<th>Medicaid and CHIP†</th>
<th>Marketplace®‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Geography</td>
<td>◊</td>
<td>◊</td>
<td>○</td>
<td>◊</td>
</tr>
<tr>
<td>Language</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Disability Status</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Income</td>
<td>◊</td>
<td>◊</td>
<td>◊</td>
<td>◊</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td>Sexual Orientation and Gender Identity</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Key: ● Collected aligned to 2011 HHS standards ○ Collected with standards and/or completeness issue(s)
◊ Collected with no major issues, no adopted standard - Not collected

* The data elements included in this table are the same as those prioritized in Executive Order 13985 and the CMS Framework for Health Equity, and do not encompass all data elements that could be collected or improved.1,3 This table does not reflect quality and completeness issues in all cases.

** Data received from SSA and collected via surveys detailed in the sections below.

*** Data collected from Medicare Part C/D enrollment form and various surveys detailed in the sections below, supplemented as needed with SSA data from Fee-for-Service Medicare.

† Data reported from states in the Transformed Medicaid Statistical Information System (T-MSIS).

‡ Data collected from the Marketplace programs using Healthcare.gov platform. Because CMS does not closely regulate data collection on State-Based Exchanges, this table shows data collected on the Federally-Facilitated Exchanges only.

The next section describes specific sociodemographic elements summarized in the table above and concerns about certain categories of socioeconomic data, highlighting the issues CMS will aim to address in the future.

Sex Data

The 2011 HHS data standard for sex utilizes a single question, asking respondents to specify their sex as male or female.7 Medicare, Medicaid, CHIP, and Marketplace® programs collect data on biological sex according to this 2011 HHS data standards but do not yet collect data on gender identity.

Evidence suggests that the methods for collecting data on an individual’s sex can be updated to improve the accuracy of data collected and nuance about different identities.17 New developments, such as the increased uptake of Gender Marker X on U.S. passports, present opportunities to align new data collection practices to updated state and federal practices.36 The HHS USCDI version 2, released by ONC in 2021 for use in certified health IT, includes standards for structured data capture of both “sex assigned at birth”37 and “gender identity.”38
Geography Data

There are no HHS or OMB standards established for the collection of geographic data for individuals. However, CMS collects this information across its programs in the form of individual mailing addresses. Fee-for-Service Medicare, Medicare Advantage plans, Federally-Facilitated Exchanges, Medicaid, and CHIP collect an individual's street address, city, state, and zip code, with the option of including county on enrollment forms and through SSA. Medicaid and CHIP both recently adopted the U.S. Department of Agriculture (USDA) Economic Research Service Rural-Urban Commuting Area Codes (RUCA) standards for geographic data collection. CMS obtains the mailing address used for benefits or billing from SSA for the Medicare program, but in some cases, this may not be the actual address where the person resides. Street addresses and locations are also limited for some populations, such as individuals with housing instability and some tribal members living in remote/frontier settings that are excluded from county and state data collections.

Language Data

Collecting language data is critical to understanding whether programs are delivering linguistically appropriate, patient-centered care. Language barriers may cause individuals to have decreased health care access, reduced quality care, and worse health outcomes. The 2011 HHS data standard for language is a measure of English proficiency and applies to individuals five years old and above. HHS has also released National Standards for Culturally and Linguistically Appropriate Services (CLAS), which can be applied in health care settings.

Beginning with USCDI version 1, adopted by ONC in 2020, all certified health IT includes a data element for “preferred language” in the patient demographics data class. In the Medicaid and CHIP programs, states collect this information on individuals' program application forms in line with standards, but challenges persist for CMS in obtaining these data elements. Many states go beyond the data standard requirements to document both written and spoken language preference. The Marketplace collects language data on preferred written and spoken language once per application and for the application filer only; this data is not collected for all individuals. While Medicare and the Marketplace collect language data, they do not necessarily align with 2011 HHS standards. For example, Medicare Advantage enrollment forms ask if individuals prefer plan information in a language other than English, with a free response field to indicate the preferred language. The CAHPS survey asks what language is mainly spoken at home by the individual.
Disability Status Data

CMS recognizes the importance of collecting accurate, complete disability status data in order to understand and improve the current health status and outcomes of individuals served by CMS programs. The 2011 HHS data standard for disability includes a six-item set of required questions with specified answer options. Although Medicaid and CHIP disability status data collection aligns to standards, completeness issues persist. Fee-for-Service Medicare collects reason for entitlement (i.e., old-age survivors’ insurance, disability, and/or end-stage renal disease) data from SSA, but only for those disabilities that align to SSA disability benefit requirements. Therefore, a full picture of some individuals’ disability status is likely missing in the enrollment data Medicare receives, with individuals who develop a disability after enrollment or those who have a disability but do not receive disability benefits undercounted. The Marketplace® collects limited disability data via two optional questions for applicants rather than the six questions outlined in 2011 HHS data standards.

The USCDI version 3 includes three new data elements for functional status, disability status, and mental/cognitive status, which is a critical step to advancing the standardized capture and exchange of data for people with disabilities. These standards are aligned to the functional status data elements already integrated into CMS PAC assessments and CMS Home- and Community-Based Services Functional Assessment Standardized Items. Functional status, disability status, and mental/cognitive status standards included in the USCDI version 3 may be appropriate for use across certain CMS programs.

Income Data

Income is an eligibility factor for Medicaid and CHIP and is essential to determining subsidies in the Marketplace®. Therefore, data is collected and assessed for quality under all of these programs, but completeness issues persist. Because income is not a criterion of eligibility for Medicare enrollment, it is not consistently collected for enrollees in all programs, and data collected is limited to 1) Medicare enrollees who receive the Low-Income Subsidy under Medicare Part D, are participating in one of the Medicare Savings Programs under Parts A and B, or who have Medicaid partial- or full-benefit status (i.e., dually eligible individuals); 2) higher income enrollees who pay the Income-Related Monthly Adjusted Amount, which is determined by the SSA and applies to both Fee-for-Service Medicare and Medicare Advantage Programs; and 3) a sample of enrollees who report their income via the Medicare Current Beneficiary Survey. Accordingly, the resulting data elements are not inclusive of all individuals covered by the program. The Medicare Advantage program plans do not specifically collect and report income data.
Race and Ethnicity Data

The 2011 HHS data standard for race and ethnicity utilizes a two-question format, with the first question assessing Hispanic ethnicity and a second question assessing race. In both instances, one or more categories can be selected.

Although Medicaid, CHIP, and the Marketplace race and ethnicity data collection adheres to 2011 HHS standards, some race and ethnicity data collected in support of Medicare, particularly enrollment data received from SSA and Qualified Health Plan (QHP) enrollee surveys, use a single field to collect both race and ethnicity. In addition, both Fee-for-Service Medicare and Medicare Advantage plans collect race and ethnicity data for a subset of individuals through national surveys, though only some align to federal standards. These include CAHPS surveys administered in multiple settings, the Medicare Current Beneficiary Survey, and the Medicare HOS—which align to standards—and the Outcome and Assessment Information Set (OASIS) used in home health care—which will not align to standards until January 2023. However, because these surveys are only administered to a sample of enrollees, they do not yield a dataset inclusive of all individuals covered by Medicare programs. To address gaps in their data, Medicare Advantage plans will begin to use the OMB-approved Medicare Advantage enrollment form (CMS-10718) to collect self-reported, voluntary, data from Medicare Advantage enrollees aligned with the 2011 HHS data standards beginning in January 2023. The 2011 HHS data standards align in most instances with the HHS ONC USCDI which consists of granular race and ethnicity data elements to support computerized, interoperable use of this data.

Quality and completeness of race and ethnicity data can vary significantly across specific racial or ethnic groups and over time as standards for data collection change. As detailed in a recent HHS Office of the Inspector General (OIG) report, deficiencies in source data cause inaccuracies and other issues, particularly for individuals who identify as American Indian/Alaska Native, Asian/Pacific Islander, or Hispanic. External studies show that although Medicare enrollment data may be more complete for non-Hispanic White and Black individuals, individuals of other races and ethnicities such as Asian, Pacific Islander, and Hispanic individuals are often misclassified. This is likely due to a number of factors, including the fact that individuals assigned a social security number before 1980 could only specify White, Black, or “other” as their race. As the HHS OIG report also details, SSA stopped routinely collecting these data in 1989, and now collects them only in certain, limited circumstances. The structure in which SSA collects race and ethnicity data can also be too broad to identify specific populations. Hispanic individuals often do not feel they can describe their race and ethnicity with the options provided. Similarly, American Indians and Alaska Natives describe their sovereign status as political and not race-based, which may add to the discrepancies seen in data.

Although CMS currently collects Medicaid and CHIP race and ethnicity data elements that are aligned to 2011 HHS standards, CMS datasets on Medicaid and CHIP enrollees experience quality and completeness issues. Because self-reported race and ethnicity data elements are optional in Medicaid applications due to federal
privacy regulations and not required to determine eligibility, state Medicaid agencies have limited opportunity to collect this data before the individual begins interacting with their health plans. Variations in how state agencies ask questions about race and ethnicity (i.e., combining race and ethnicity into a single question or expanding ethnicity response options separately) and state system issues can impede their ability to accurately transmit data. Health plans have also been shown to have limited race and ethnicity data for individuals covered by Medicaid and CHIP; 50% of race data and 70% of ethnicity data submitted by plans in support of Distinction in Multicultural Health Care applications were incomplete in 2019.

As a result of these limitations, Medicaid’s Data Quality Atlas examination of 2019 Transformed Medicaid Statistical Information System (T-MSIS) Analytic File data nationwide identified five states’ Medicaid and CHIP race and ethnicity data as unusable and 17 states’ data as of high concern. The Marketplace® experiences similar limitations on data collection at enrollment. During the 2017 open enrollment period, 36% of individuals between ages 18 and 64 did not specify their race.

Due to quality concerns with race and ethnicity data, CMS uses statistical methods to impute and improve Medicare race and ethnicity data, including the Research Triangle Institute race code as well as the Medicare Bayesian Improved Surname Geocoding (MBISG) methodology. The Research Triangle Institute race code variable is created by taking the enrollee race code that has historically been used by SSA and applying a name-based algorithm that identifies more enrollees as Hispanic or Asian. The MBISG utilizes an individual’s surname and neighborhood-level characteristics to identify their most likely race and ethnicity. However, these imputation methods are not compliant with 2011 HHS data standards or aligned to Executive Orders 13985 and 14031, which call for more granular race and ethnicity data.

The ONC health IT certification program 2015 Edition certification criteria include both structured data elements and required functionality for race and ethnicity capture. The functionality requires the capability to capture both the OMB standard and the more granular CDC race and ethnicity standards, these functionalities were required for CMS reporting programs that require the use of CEHRT beginning in CY 2019. Both the OMB and the CDC standard are also included in the USCDI version 1 Patient Demographic – “Race” and “Ethnicity” data elements. These standards are included in certified health IT as a capability requirement rather than setting any specific requirements on the users of the technology. In this way, they allow for flexibility as to how health care providers or other technology users implement the level of granularity for structured data capture most appropriate for the use case.

Sexual Orientation and Gender Identity Data

Medicare, Medicaid, CHIP, and Marketplace® programs do not yet collect information on SOGI in a standardized format, although SOGI data standards do exist, such as those included in the following table:

<table>
<thead>
<tr>
<th>SOGI Data Collection Status</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Medicare</strong>: Not currently collected</td>
</tr>
<tr>
<td><strong>Medicaid and CHIP</strong>: Collected in limited settings, not currently shared with CMS</td>
</tr>
<tr>
<td><strong>Marketplace®</strong>: Not collected</td>
</tr>
</tbody>
</table>
in the USCDI version 2, for use when appropriate. Two states, Connecticut and Oregon, collect gender identity information on their Medicaid applications. Some states may choose to collect SOGI data via state-based exchanges, such as the Covered California exchange. However, California collects the data post-enrollment rather than at enrollment, potentially negatively affecting response rates since participants may not be reached or respond after enrollment is complete. Given evidence that consistent health disparities exist among lesbian, gay, bisexual, transgender, queer, and intersex (LGBTQ+) individuals and across different sexual and gender identities, a new look at both sex and SOGI collection standards and processes may be required.

**Social Drivers/Determinants of Health (SDOH) Data**

SDOH refers to the conditions in the environments where people are born, live, learn, work, play, worship, and age that affect a wide range of health, functioning, and quality-of-life outcomes and risks. Healthy People 2030 defines five domains of SDOH: economic stability; educational access and quality; health care access and quality; neighborhood and built environment; and social and community context. SDOH data can include elements on health literacy, transportation, social isolation, housing insecurity, food insecurity, geography, and more. CMS understands the significant role that SDOH data should play in the design and execution of initiatives that promote health equity across all CMS programs.

Unlike sociodemographic data elements which have been collected in some contexts for decades, SDOH data elements are not as widely collected under CMS programs currently, as CMS programs are limited in the SDOH data they are authorized to collect.

There are a number of efforts underway to begin SDOH data collection in specific contexts across several programs. For example, Fee-for-Service Medicare and Medicare Advantage plans collect self-reported SDOH metrics via several means, including surveys like Medicare Current Beneficiary Survey and CAHPS® which collect information on education, full-time student status for applicants aged 18-22 seeking financial assistance, and household size for all applicants seeking financial assistance. In Medicare Advantage, these SDOH indicators are used internally at the plan level and are not reported to CMS. Based on the May 2022 Part C/D final rule, Medicare Advantage special needs plans must meet new requirements to include one or more questions on housing stability, food security, and access to transportation in their health risk assessments beginning in contract year (CY) 2024. CMS also collects metrics for substance use and mobility data elements for individuals experiencing end-stage renal disease and a variety of additional SDOH data elements for individuals attributed under some Innovation Center Models. This includes the AHC Model, which required screening of patients for housing instability, food insecurity, utility needs, interpersonal safety, and other needs, as well as the Integrated Care for Kids Model, which focuses on early identification and treatment
of health-related needs and risk factors.\textsuperscript{72, 73} However, the CMS Innovation Center continues to note that limited or incomplete data hinder the Center’s efforts to design payment models that support SDOH.\textsuperscript{74} Additionally, the low volume of Indian Health Service and Tribal hospitals may limit inclusion into the CMS Innovation Center and Center for Medicare’s programs, which may further skew the data. New proposed rules, cooperative agreements and funding for innovative outreach strategies are being used to promote Indian Health Service and Tribal Hospital inclusion in CMS programs.\textsuperscript{75, 76}

Both Medicare as well as the Medicaid and CHIP programs collect a more expansive set of SDOH with the help of “Z codes” under the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM). Codes Z55-65 (see Table 3) are encounter reason codes that help identify SDOH affecting a patient, such as housing, economic circumstances, and the physical environment.\textsuperscript{77} Capturing SDOH data via these codes can help generate nuanced information that can be aggregated program-wide.\textsuperscript{78, 79, 80} However, quality and relevance issues with these codes prevent them from being a perfect solution for documenting SDOH.\textsuperscript{81} Utilization of these codes has been low; only 1.6% of individuals covered by Medicare Fee-for-Service in 2019 have claims tied to Z codes.\textsuperscript{82}

<table>
<thead>
<tr>
<th>ICD 10 CM Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z55</td>
<td>Problems related to education and literacy</td>
</tr>
<tr>
<td>Z56</td>
<td>Problems related to employment and unemployment</td>
</tr>
<tr>
<td>Z57</td>
<td>Occupational exposure to risk factors</td>
</tr>
<tr>
<td>Z58</td>
<td>Problems related to physical environment</td>
</tr>
<tr>
<td>Z59</td>
<td>Problems related to housing and economic circumstances</td>
</tr>
<tr>
<td>Z60</td>
<td>Problems related to social environment</td>
</tr>
<tr>
<td>Z62</td>
<td>Problems related to upbringing</td>
</tr>
<tr>
<td>Z63</td>
<td>Other problems related to primary support group, including family circumstances</td>
</tr>
<tr>
<td>Z64</td>
<td>Problems related to certain psychosocial circumstances</td>
</tr>
<tr>
<td>Z65</td>
<td>Problems related to other psychosocial circumstances</td>
</tr>
</tbody>
</table>

The Marketplace\textsuperscript{®} QHP Enrollee Experience Survey offers a venue for educational and employment data collection for limited individuals and currently collects metrics on topics such as cost barriers to health care, health literacy, employment, and educational attainment.\textsuperscript{48} Individual and Small Group Markets’ External Data Gathering Environment (EDGE) SDOH Data Collection provides CMS access to data for risk adjustment calculations from issuers of risk adjustment covered plans in the individual and small group markets, including some SDOH elements.\textsuperscript{c} Due to the de-identified nature of EDGE data collection, SDOH elements collected through EDGE servers cannot be linked to identified Marketplace\textsuperscript{®} or State-Based Exchange enrollee data.

\textsuperscript{c} The EDGE server distributed data collection approach provides CMS access to on-exchange and off-exchange data for issuers of risk adjustment covered plans in states where HHS operates the ACA risk adjustment program. For the 2021 benefit year, HHS operated the risk adjustment program in all 50 states and the District of Columbia.
While data standards for SDOH data have been limited to date, recent efforts to advance broader adoption and use of existing standards and to establish new ones are underway. CMS has adopted the ONC finalized version 2 of the USCDI standards in July 2021, with many updates focusing on data elements relevant to SDOH, such as social needs screening tool results. ONC has also published USCDI version 3, which provides further updates on multiple data classes.

Progress to Date

Given the importance of health equity data across its programs, CMS is committed to resolving the major gaps in data quality, accuracy, and completeness discussed above. In subsequent years, CMS will provide updates on the status and availability of health equity data for CMS and external researchers. CMS has engaged in an enterprise-wide review of its data elements and processes, identifying gaps and considering solutions along the way. By prioritizing collection of new data elements, releasing new tools and resources, and collaborating with partners including ONC and industry organizations, CMS has made progress in addressing health equity data gaps for both its own programs and industry organizations. Some of this progress is illustrated in the following timeline and detailed in the sections below.

Figure 1. Illustrative Examples of CMS Progress to Date on Health Equity Data Improvement
Collecting New Elements and Filling Existing Gaps in CMS’s Health Equity Data

The gold standard for collection of sociodemographic and SDOH data remains self-reported data, and CMS is finding new ways to fill gaps in this area. Where self-reported data elements are missing or potentially inaccurate, CMS has applied rigorous data science, updated reporting requirements, and found new venues for enhanced reporting of self-reported data. CMS applies some temporary methods to fill gaps now with the intent to continuously improve self-reported data gathering in the future. These changes to data collection will be executed through rulemaking and with input from individuals and groups in and outside of CMS.

For instance, CMS has added seven Standardized Patient Assessment Data Elements (SPADEs) to required patient assessment tools for PAC settings (see box). These data elements will allow for standardized collection of race and ethnicity aligned to the 2011 HHS standard, preferred language, need for interpreter, health literacy, transportation, and social isolation. Many elements are being collected for the first time and others are being expanded into new settings or in a more standardized way. CMS also developed the Data Element Library for PAC settings to support a centralized resource for assessment instrument data elements and their associated health IT standards including mapping to the USCDI demographic data class. These efforts will give CMS and our partners the ability to tailor programs and policies in PAC settings based on needs and disparities, though their implementation has been delayed due to the COVID-19 Public Health Emergency.

While updating data collection practices will lead to sustained improvements in the completeness of health equity data, implementing those changes may require years to complete appropriate rulemaking steps and integrate input from organizations affected by them. To make more complete data available for our programs and partners that need them today, CMS OMH has supported the development of algorithms designed to indirectly estimate stratification variables, specifically race and ethnicity. This includes the MBISG imputation method discussed above. As indicated, we additionally intend to leverage and align with the ONC USCDI as applicable and appropriate to support data completeness for our programs.

CMS has regularly reassessed the MBISG method and recently made available the MBISG Version 2.1.1, which improves the algorithm’s accuracy in CMS data imputation. Currently, CMS uses the methodology to report on Healthcare Effectiveness Data and Information Set (HEDIS) measures for Medicare Advantage plans. CMS continues to find new opportunities to apply MBISG across its programs and has made this algorithm available across the CMS enterprise. As part of a series of Requests for Information (RFIs) released in 2021, CMS sought comments from health care professionals, payers, industry organizations, and other members of the public on expanding the use of indirect estimation as a temporary measure to better identify health care inequities. This could include applying the methodology to Medicaid and CHIP, for which CMS is developing an imputation model and has issued guidance to states on reporting race and ethnicity as well as the Health Insurance Marketplaces®.
**Equipping the Industry with New Tools and Capabilities**

As the nation’s largest health insurer and a regulator, CMS can help establish and implement an approach supported by the ONC USCDI as applicable and appropriate for how health equity data elements are standardized and collected across the health system. While CMS is working to address health equity data gaps within our own programs, we have also sought opportunities to equip individuals and organizations across the health care field with new tools and capabilities that will help them collect and act on health equity data based on our own lessons learned.

For example, CMS has published the [Inventory of Resources for Standardized Demographic and Language Data Collection](#), which highlights efforts by CMS components, including CMS OMH, and external organizations, such as the American Hospital Association, to address disparities in care through the collection of race, ethnicity, and language data.89 The inventory is updated annually, and the toolkits, webinars, and other resources in it help to explain the challenges with improving data collection methods and offer potential mitigation strategies. Resources developed by CMS OMH to strengthen the use of health equity data include those referenced in the figure below.77, 90, 91

![Figure 2. CMS OMH Resources to Strengthen the Use of Health Equity Data](#)

- An infographic on using **Z codes** to help providers and health care administrators collect SDOH information using ICD-10-CM Z codes and employ this data to improve health outcomes among their patients
- A **practical guide to implementing the National CLAS standards** and building an organizational response to health disparities through improved data collection
- A **Guide to using the AHC Health-Related Social Needs Screening Tool** that provides insight into how to implement robust screening for social determinants of health

CMS has also helped organizations and individuals across the nation develop capacity to act on health equity data. For example, the Center for Medicaid and CHIP Services (CMCS) issued a State Health Official Letter outlining opportunities to address SDOH in their Medicaid and CHIP programs and support officials in designing policies and interventions that can address disparities. The letter also encouraged the use and adoption of SDOH data standards from the Gravity Project.92 Similarly, the CMS Innovation Center’s Health Care Payment Learning & Action Network (HCPLAN) recently launched the Health Equity Advisory Team to help identify and prioritize opportunities to advance health equity through alternative payment models nationwide. The HCPLAN’s efforts will include public-private partnership and regional and national organizations collaborating toward a common goal of advancing health equity.
Providing Access to Disaggregated Data and Insights

In addition to guides and tools, CMS has prepared and released new data that can inform industry planning. This includes releasing stratified reports noting specific disparities across subpopulations in the Fee-for-Service Medicare and Medicare Advantage programs. These reports illustrate the importance of disaggregating data by factors like race, ethnicity, and urbanicity when planning actions to advance health equity.

In 2016, CMS began releasing annual data and reports on the Medicare Advantage population stratified by race and ethnicity. CMS expanded this effort in 2018 and 2021 with the annual publication of its Rural-Urban Disparities in Health Care in Medicare Report for Fee-for-Service Medicare and Medicare Advantage populations and Dual Eligibility or Eligibility For Low-Income Subsidy National Disparities Stratified Report. CMS also recently began sharing Part C and D Star Ratings measures stratified by low-income subsidy, dual-eligibility, and disability status in confidential reports with Medicare Advantage plans and Prescription Drug Plans. CMS has also released the Mapping Medicare Disparities Tool, which allows the public to use CMS data to understand disparities across subpopulations covered by Fee-for-Service Medicare.

Through the release of these tools and resources, CMS has made strides to strengthen health equity data. It will continue to provide leadership across the nation to accelerate the pace of improvement. Future efforts to address gaps in health equity data for both CMS programs and the broader industry are described below.

Vision for the Future

Health care professionals, government agencies, researchers, and health advocates have worked for decades to understand why health disparities exist. Despite advancements in population health, significant disparities persist. Lack of quality health equity data that indicates what drives these differences and provides insight to develop interventions that are the most effective in alleviating them limits progress on eliminating disparities.

CMS is committed to improving the quality, accuracy, and completeness of data that can enable improvements in health equity. This goal is stated clearly within the CMS Framework for Health Equity, which defines the approach to collaborating across government and industry to achieve health equity. The CMS Framework for Health Equity lays out five strategic priorities (see box) to describe system- and community-level approaches to achieve health equity across CMS programs. These priorities were identified through engagement with
organizations across the health care field, review of evidence on health equity drivers, and discussion with federal and non-federal partners. Individually, they address unique drivers of critical health disparities and point to specific actions that can be taken across Medicare, Medicaid, CHIP, the Marketplace®, and the broader small group and individual markets. Together, these priorities plot a path toward creating system-level change through CMS programs.

To achieve the priorities outlined in the CMS Framework for Health Equity, CMS will need to address the quality, completeness, and accuracy issues in health equity data across its programs. The actions described below will help reach the future state of health equity data at CMS.

**Actions to Drive the Future of CMS Health Equity Data**

Though much progress has been made in recent years, significant issues remain in CMS’s sociodemographic and SDOH data. CMS’s partners in government and industry offer deep insight on how we can best fill these gaps. Through a series of RFIs, CMS collected input from the public to understand the challenges hindering the collection of health equity data, policy changes that would lead to overcoming these challenges, and recommendations for filling health equity data gaps.86,96,97 CMS also conducted extensive reviews of data gaps and improvement areas. As a result of these efforts, CMS has identified the following objectives to improve its health equity data landscape:

- Improve completeness of health equity data and alignment to standards
- Expand disaggregated standardized data and stratified reporting for sociodemographic elements
- Gradually implement health equity scores
- Assess and minimize potential bias in data and data algorithms

Each of these actions, and the factors that will make them successful, is further detailed below. Together these actions can move CMS closer to an ideal state of health equity data.

**Improve Completeness of Health Equity Data and Alignment with Standards**

Addressing existing gaps and quality issues for sociodemographic, SDOH, and HRSN elements can provide CMS programs and the broader industry with a more complete picture of needs and disparities, along with how
to address them. CMS is committed to making these improvements through the implementation of new
collection methods, use of imputation where needed, and reevaluation of current practices. To address the
gaps in health equity data, CMS is committed to leveraging health equity data to appropriate standards. This
may involve collecting data elements through new venues that allow for more control over alignment to
standards, accounting for any statutory restrictions. Planned actions for both data types are detailed below.

### Improving Completeness and Standards Alignment

- Identify additional CMS venues to collect new elements (e.g., Innovation Center models) and expand collection of existing ones (e.g., race collection at Part C/D enrollment)
- Collaborate with partners to fill gaps in existing data (e.g., SSA, health plans)
- Allow broader access to statistical methods that can fill existing data gaps
- Increase utilization of established, reliable collection methods (e.g., Z codes)
- Collect data elements previously collected through partners to allow more control over format
- Align new and existing data collection to the HHS ONC USCDI content standard data elements and classes as an update to the HHS 2011 survey-based standards, where appropriate and applicable (e.g., SOGI)

CMS’s efforts to address issues in alignment to data standards across the health equity data landscape will take into account parallel efforts across government and private industry to establish new standards that allow for a more accurate, complete picture of health equity issues. This includes the recently announced effort led by the Chief Statistician of the United States to review and revise standards for maintaining, collecting, and presenting Federal data on race and ethnicity, as well as USCDI updates to promote interoperability, as appropriate.

### Sociodemographic Data

Complete, accurate sociodemographic data can provide CMS programs with a clearer picture of the individuals they serve, the areas of need, and the impacts of their programs. Although CMS has limited statutory authority to collect race and ethnicity data from individuals served by all programs and settings, we have identified venues in which data collection to fill specific data gaps may be possible. For instance, CMS is pursuing collection of voluntary self-reported race and ethnicity data using the 2011 HHS data standard via the Medicare Part C/D enrollment form starting with the 2023 Medicare Advantage open enrollment period in January 2023. ONC is conducting a technical analysis mapping the 2011 HHS standards for race, ethnicity, preferred language, sex, and disability status to the USCDI standards and recommending updates to categories and elements included in the 2011 HHS Data Standards. CMS also finalized the collection and extraction of race and ethnicity data through issuers’ EDGE servers consistent with 2011 HHS data standards beginning with the 2023 benefit year data to assess risk patterns in individual and small group markets, conduct additional policy analysis, and better address discrimination in health care and health disparities.

CMS may explore the collection of new sociodemographic data elements not previously collected by CMS programs. For example, CMS OMH recommended question language for the 2023 Medicare Current Beneficiary Survey to include sex assigned at birth, sexual orientation, and gender identity. CMS may also consider the Medicare Part C/D enrollment form, the Health Insurance Marketplace® QHP Enrollee Survey, the Marketplace® application, and state Medicaid programs enrollment forms as venues for new sociodemographic elements. More robust data on factors like spoken language and religion can advance equity initiatives meant
to address cultural drivers of health disparities and provide more culturally appropriate care for communities who are historically underserved.

CMS is also exploring ways to expand collection of sociodemographic data through cooperation with other government entities and partners to collect these data elements in a more interoperable way. This may involve coordination with SSA to improve collection of sociodemographic data from individuals enrolled in Medicare, yielding more up to date information for the program. Collaboration with the Department of Veterans Affairs or other entities can also improve elements that have consistent gaps, such as veteran status. Additionally, beginning in the 2023 benefit year, individual and small group market issuers will submit to their EDGE servers new data elements including race, ethnicity, and a subsidy indicator, which is intended to indicate whether a particular enrollee is receiving advance payments of the premium tax credit. These data elements will eventually be available in the EDGE Limited Data Set that researchers can request.  

In the CMS Innovation Center 2021 Strategy Refresh, the Center’s stated goals include collecting sociodemographic data from all organizations participating in its Innovation Center Models by 2023. Model participants will be able to submit race, ethnicity, geography, disability and SOGI data via a health equity data collection system. The CMS Innovation Center is currently developing a questionnaire that aligns data collection to OMB standards and the USCDI conceptualizing a process to make the data available across CMS programs via a standards-based Application Programming Interface (API). The CMS Innovation Center is also examining data on existing models to identify methods to improve participation of safety net providers and other venues where care is provided to populations who have been historically medically underserved. CMCS detailed a Strategic Vision for Medicaid & CHIP in which they would work with states to improve collection of a core set of metrics, including race, ethnicity, language, disability status, and other factors, and the Medicaid and CHIP programs have released a proposed rule that would make standardized quality measures aimed at identifying and addressing health disparities mandatory; they anticipate releasing a final rule after reviewing public comments in the coming months.

In addition, CMS is launching efforts to allow broader access to the methods that have filled sociodemographic data gaps to date. CMS OMH is currently allowing CMS program staff access to the MBISG methodology to improve Medicare race and ethnicity data. While acknowledging self-reported data as the gold standard for data collection, these estimation methods provide a means for filling critical data gaps. CMCS is developing an algorithm to impute race and ethnicity for individuals enrolled in Medicaid and CHIP. This effort is in the testing phase across the majority of states and yielding preliminary results for review.

Finally, CMS is considering ways to enhance the accuracy and inclusiveness of SOGI data by assessing how to incorporate newly adopted and updated recommendations on appropriate terminology and data standards from sources like National Academies of Sciences, Engineering, and Medicine, the Health Level 7® Gender Harmony Project, and the USCDI into CMS programs.
and Insurance Oversight (CCIIO) is also exploring means of aligning collection of race and SOGI data to 2011 HHS standards by altering QHP Enrollee Surveys and the Marketplace® application.

**SDOH and HRSN Data**

SDOH data elements are an important factor in advancing health equity for CMS-served populations. Acting on feedback provided by external organizations and individuals, CMS has prioritized the collection of the following SDOH variables: health literacy; social isolation; transportation barriers; food insecurity; and housing insecurity. All of these elements are collected across various settings, but a lack of consensus across the programs and domains that collect them means that this information is not captured in a standardized and accessible way.\(^\text{105}\)

CMS is introducing new venues for SDOH data collection across its programs. For example, the Medicare Shared Savings Program issued an RFI in summer 2022 to engage industry on how to best begin implementing data from SDOH data and measures as they continue to mature and evolve;\(^\text{106}\) and the FY 2023 IPPS Final Rule includes updates to the Hospital Inpatient Quality Reporting (IQR) program that include the adoption of new health equity, and SDOH measures.\(^\text{107}\)

Similar to efforts for sociodemographic data noted above, the CMS Innovation Center also intends to encourage and support the collection and submission of SDOH and demographic data on patients in its models and will begin collecting this data in 2023. These data elements will comply with Fast Healthcare Interoperability Resources (FHIR) standards and USCDI version 2 updates defined by ONC. Participants will be encouraged to collect SDOH data through a validated screening instrument with associated data elements and health IT codes such as the AHC Health-Related Social Needs Screening tool, the Protocol for Responding to and Assessing Patients’ Assets, Risks, and Experiences (PRAPARE) Assessment, or the North Carolina SDOH Screening Tool.\(^\text{108,109}\) Models like Accountable Care Organization Realizing Equity, Access, and Community Health (ACO REACH) will require submission of demographic data annually beginning in 2023, and will likely require SDOH data submission during future model years.\(^\text{110}\)

Medicare Advantage Special Needs Plans are required to conduct an assessment of an enrollee’s physical, psychosocial, and functional needs using a comprehensive risk assessment tool at enrollment and annually.\(^\text{111}\) Starting in 2024, these assessments will include at least one question from a list of screening instruments specified by CMS on each of the categories of housing stability, food security, and access to transportation, providing a more complete picture of risk factors and barriers to access to care. Expanding use of these assessments to other plans may grow the body of data on health risks in CMS programs. CMS is also working to identify and incorporate perspectives from organizations and individuals on what additional elements to collect to better promote health equity. For example, CMS is considering the development of a measure to assess whether Medicare Advantage plans are screening their enrollees for health-related social needs such as food, housing, and transportation.\(^\text{112}\)
ICD-10-CM Z Codes offer a significant opportunity to collect information on the social and structural factors that contribute to each individual’s current health status and future risk. There have been increases in the use of these codes over time, and CMS will continue its efforts to increase use and utility. This includes releasing guidance to health care professionals and administrators on how to use codes and working with partners to expand available Z codes. CMS will continue to assess the quality and completeness of SDOH data and align to federal standards as they are defined.

CMS is working with partners across government and industry to define new ways to share SDOH data and equip other entities, such as states, with that data. Coordination with partners can help both CMS programs and other agencies’ programs to address SDOH. For instance, CMS is collaborating with the Agency for Healthcare Research and Quality (AHRQ) to define opportunities for using and adding to AHRQ’s SDOH database. CMS will also coordinate with the U.S. Department of Agriculture (USDA) to provide data and lessons learned that can help increase both USDA food assistance program enrollment and Medicaid enrollment. This will not only present a fuller picture of populations experiencing food insecurity, but also empower CMS and USDA to identify and help those populations.

HRSN data can also be used to identify ways to address disparities driven by SDOH. A number of CMS programs have started to incorporate standardized ways for collecting and reporting information about HRSNs, including the Innovation Center’s AHC HRSN Screening Tool, Medicaid Waivers and other supports to support HRSN integration, and integration of HRSNs into the quality measurement approach for the Hospital IQR Program. Increasing the use and availability of HRSN data may help CMS and providers reduce disparities by identifying addressable needs that impact health outcomes and supporting the identification of areas where additional incentives or coverage may support disparity reduction efforts.

**Expand Disaggregated, Standardized Data, and Stratified Reporting for Sociodemographic Elements**

Executive Order 13985 notes that a lack of data stratified by “race, ethnicity, gender, disability, income, veteran status, or other key sociodemographic variables” has “cascading effects and impedes efforts to measure and advance equity.” Although the industry has recognized the importance of disaggregating data to drive health equity, commitment has lagged. For example, as of August 2020, 30 states were unable to report COVID-19 data disaggregated to a degree that supported tracking of the disease in Native Hawaiian Pacific Islander individuals. This may have resulted in health disparities being overlooked.

---

**Driving Data Disaggregation and Stratified Insights**

- Collect data in disaggregated forms to understand granular population characteristics
- Continue releasing stratified reports with insights
CMS is identifying new ways to support data collection and categorization that supports disaggregation. This includes potentially overhauling how data elements are collected through certain programs. For instance, the Merit-based Incentive Payment System (MIPS) program currently collects many variables via aggregated clinician data. Identifying means for collecting disaggregated data elements will allow for more nuanced reporting across the program. To effectively provide a nuanced picture of disparities across CMS’s populations, CMS will identify means of applying these principles of disaggregation across all programs.

Distinct from data disaggregation, release of stratified reporting is also important to promoting equity. In 2016, CMS began annually releasing HEDIS and CAHPS® data and reports on the Medicare Advantage population stratified by race and ethnicity. CMS recently began reporting Part C and D Star Ratings data stratified by Low Income Subsidy, dual-eligibility, and disability to Medicare Advantage and PDP plans. CMS is also considering publicly reporting this stratified data by including stratified reporting as part of the display measures on CMS.gov in the future. In December 2021, CMS released a report investigating trends in health disparities in Medicare across a nine-year period using disaggregated data. Public comments suggest support and interest in expanding sociodemographic factors for disaggregation to include gender identity, geographical location, language preference, and dual eligibility status.

CMS intends to use data stratification to promote equity across additional programs. For example, historical benchmarks for ACOs participating in the Medicare Shared Savings Program are stratified based on an individual’s enrollment type (e.g., if they have end-stage renal disease, have a qualifying disability, or are dually eligible for Medicare and Medicaid, or are entitled due to age), allowing for more accurate – benchmarks for ACOs that care for a larger portion of these populations.

**Gradually Implement Health Equity Scores and Other Equity Specific Measures**

Organizations and individuals seeking to advance health equity need to juggle many concepts and data elements when assessing disparities, measuring impact, and tracking progress. A single composite measure that provides a unified view of this complex issue can be invaluable to program planning and execution—not just within CMS, but across the industry. CMS is developing health equity measures, including the Health Equity Summary Score, to meet this need. Developed for Medicare Advantage plans, the Health Equity Summary Score, summarizes the quality of care delivered to patients with specific race, ethnicity, dual eligibility, and low-income status characteristics. To make it easier to translate Health Equity Summary Score, insights into action, CMS is developing a dashboard that will provide Health Equity Summary Score, data to Medicare Advantage organizations, help them identify disparities across their populations, and point to resources that can help

### Implementing Health Equity Scores

- Continue refining the Health Equity Summary Score, and developing other indices (e.g., Medicare Advantage Health Equity Index)
- Provide reports on health equity performance using interim scores or measures (e.g., Health Equity Report Card)
- Consider new applications for equity scores (e.g., Medicaid demonstration approval)
- Release resources and training to help providers use health equity scores
address them. CMS recognizes the value of scores like the Health Equity Summary Score and will prioritize use of these scores across Medicare Advantage and in new settings, tailoring its methods as needed.

As part of the 2021 series of RFIs, CMS received feedback on how a new equity score (such as the Facility Equity Score or Hospital Equity Score) could be applied across various settings and how health care professionals could be assisted in using the score. As a result of feedback around the limitations of the Health Equity Summary Score, CMS is considering adapting other tools as precursors to the Health Equity Summary Score, such as the Health Equity Report Card developed as part of the Elevating Cancer Equity initiative. The report evaluates whether health care professionals met recommendations covering four main areas: community engagement, accessibility of care and SDOH, addressing bias in care delivery, and quality and comprehensiveness of care. Lessons learned from providing these reports could assist CMS in improving health equity scores’ ability to help address disparities.

CMS is also developing a health equity index to enhance existing Part C and D Star Ratings, summarizing performance on Star Ratings measures among individuals with specific social risk factors into a composite score. This index may be used to incentivize improved performance for individuals who are at highest risk based on these social risk factors. This index may replace the current reward factor included in the Part C and D Star Ratings, allowing contracts to have further incentive to reduce health care disparities. Feedback from Medicare Advantage Organizations, Prescription Drug Plans, and other interested parties will be critical to defining and applying this new index.

Furthermore, CMS is considering how to integrate other equity measures into programs. This includes integration into the hospital environment through inpatient prospective payment system (IPPS) rules. One option is creating a commitment to equity measure based on attestation questions and assessing the proportion of adults screened for specific SDOH. CMCS is also considering ways to use health equity metrics to drive better care. This includes possibly tying health equity metrics to waiver, demonstration, and managed care contract approvals (e.g., 1115 and 1915(c).) Additionally, CMS proposed to integrate both the Area Deprivation Index (ADI) and dual eligibility status data into Medicare Shared Savings Program’s methods for allocating advance investment payments, providing upfront capital for investments in accountable care, and adjusting ACO quality performance scores, to recognize high quality performance by ACOs who serve a large number of patients from underserved populations.

Additional resources may be required to assist health care professionals in understanding how scores work and how to apply them to everyday care. CMS may provide technical assistance specific to a type of provider that may use a developed score, such as specific trainings for health care professionals treating patients with end-stage renal disease to understand an end-stage renal disease facility equity score. Technical assistance could help health care professionals interpret report cards and scores to identify patients in need of certain interventions and improve scores overall. Health care professionals could also be encouraged to report
accurate data and provide feedback on the score. CMS is considering options to use relevant equity scores to guide assistance and resources. For instance, CMS could financially incentivize high scoring facilities that are traditionally categorized as safety-net health care professionals (e.g., Federally Qualified Health Centers) and focus technical assistance efforts on the lowest-scoring facilities to encourage improvement. CMS’s Quality Improvement Organizations and its other quality improvement networks are particularly well placed to provide this essential support.

**Assess and Minimize Potential Bias in Data and Data Algorithms**

Equitable health care access and outcomes cannot be achieved without addressing bias. Underlying bias in how data elements are collected, derived, and used can cause numerous issues, including overlooking elements that affect the experience and outcomes of individuals served by CMS programs. This can include failing to collect data that affirms the identities of certain population subgroups, such as the examples of Asian and Hispanic individuals noted previously. Additionally, as the health industry increases its use of AI, machine learning, and similar techniques to impute missing data, predict adverse events, personalize medicine, and plan and monitor care, controlling for bias will become increasingly important. The complex relationships built into these applications are not often understood by the end users at the point of care or program office, increasing the risk of misinformed decisions. For example, algorithmic bias—meaning including instances when the application of an algorithm compounds existing inequities—can significantly impact clinical decision-making, leading to adverse health outcomes for some populations.

CMS is committed to identifying and addressing bias in data collection and AI applications in health care. In line with similar efforts across government such as the HHS Trustworthy AI Playbook and AHRQ RFI on the use of clinical algorithms, as well as input from industry organizations received in the CMS AI Health Outcomes Challenge, CMS will prioritize efforts to help the industry avoid bias in AI, while also assessing its own newly developed tools and processes. CMS will also continue to seek input from the public, including organizations that may represent groups at-risk of being negatively affected by biased tools and methods, through RFIs, public comment periods, and other venues.

**Elements Critical to CMS’s Health Equity Data Strategy Success**

CMS envisions a future where CMS programs have access to complete, reliable data that informs effective decision-making and drives health equity for millions of individuals. Achieving this ambitious goal will require not only concerted effort across the Medicare, Medicaid, CHIP, and the Health Insurance Marketplace programs, but also rigorous processes that are continuously measured to promote improvement. CMS will draw from the following elements to achieve success.
Partners Across Government, States, and Industry

Data collection has never been something that CMS does in a vacuum. Whether through collaboration with SSA and other federal agencies, coordination with other HHS Operating Divisions, regulation of reporting from health care professionals and industry organizations, or collection of feedback from affected groups, CMS’s data landscape has many players. Successfully achieving CMS’s vision for the future of health equity data will similarly depend on partnerships across government and industry.

CMS will continue to collaborate with other federal agencies to receive data, establish standards, and approve program changes to support equity data improvement. This will include continuing to work with SSA to improve collection of race and ethnicity data using the 2011 HHS data standards alignment with the USCDI as applicable and appropriate to support data completeness for CMS programs including for race and ethnicity. Within CMS, resources will be allocated to addressing health equity gaps. For example, CMS has established a data subgroup under the agency’s Advisory Council on Equity to consider the needs and implications of health equity data gaps, track progress in addressing them, and keep leadership apprised.

Feedback from organizations and individuals is vitally important as CMS assesses the feasibility of process changes, determines the metrics useful to its programs, and forecasts the impact of proposed improvements. This may require CMS to compare and select between differing opinions on where to focus efforts. The 2021 RFI responses on how to close CMS’s health equity data gaps show there is an understanding among both external organizations and CMS that more can be done to expand data collection efforts to account for and measure non-clinical factors. However, there are differing views on how best to accomplish this. For example, most responses from the various RFIs support the idea of stratifying collected data by factors such as race, ethnicity, and dual eligibility. Yet some respondents would prefer to expand this list to include additional factors (e.g., language preference, sexual orientation, health literacy). Other respondents would first like to see CMS clearly define factors such as race and ethnicity to prevent confusion during the collection process. CMS will take all of this feedback into account before making final rules or programmatic decisions.

Robust Measurement of Progress

Promoting success as CMS pursues its future vision for health equity data will require thoughtful measurement, allowing CMS to identify risks and course correct as needed. Improvement of health equity data will be an iterative process, with careful consideration of the consequences. To monitor progress toward achieving health equity data strategy goals and aligning to the overall CMS Framework for Health Equity, CMS will continuously monitor how CMS data collection, standardization, and use across CMS programs help achieve the following:

1. Increase understanding and awareness of disparities and their causes
2. Create, test, and implement solutions to advance health equity in CMS programs
3. Lead sustainable actions that advance equity in CMS programs
These goals may be examined by a CMS office (e.g., CMS OMH) as well as working groups and subgroups. Considering these questions will ground CMS’s efforts in the actions that will impact individuals who participate in and administer care under CMS programs.

**Conclusion**

For over a decade, CMS OMH has promoted equitable programs, policies, and impacts across the CMS enterprise, helping CMS place health equity at the center of our work. Though much progress has been made, we know there is work left to be done to help overcome health disparities, eliminate structural barriers that underlie our health system, and advance health equity. CMS has a pivotal role to play across every health care setting in every community, and we are committed to working with partners, organizations across the industry, and the populations we serve to shape the future of health equity data.

CMS is working to advance health equity by designing, implementing, and operationalizing policies and programs that support health for all the people served by our programs, eliminating avoidable differences in health outcomes experienced by people who are disadvantaged or underserved, and providing the care and support that enrollees need to thrive. Together, CMS will work toward a future state where CMS, government agencies, industry organizations, and individuals have the health equity data and insights they need to ensure equal opportunity to health for all individuals and communities that participate in CMS programs.

**Suggested citation:** CMS Office of Minority Health. The Path Forward: Improving Data to Advance Health Equity Solutions. Baltimore, MD: Centers for Medicare & Medicaid Services; November 2022.
Appendix A – Data Standards

Current law requires that data elements, particularly race and ethnicity, align to standards set by the Office of Management and Budget (OMB), which were first published in a 1997 Notice in the Federal Register. Section 4302(a) of the Affordable Care Act also amended the Public Health Service Act to require the Secretary of the HHS to develop data collection standards for sex, primary language, and disability status. The Chief Statistician of the United States announced, in June 2022, a government-wide effort to review and revise standards for maintaining, collecting, and presenting Federal data on race and ethnicity, which will update OMB’s Statistical Policy Directive No. 15 (Directive No. 15): Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity.

In October 2011, the Office of the Assistant Secretary for Planning and Evaluation (ASPE) published guidance to disseminate uniform data collection standards as required by Section 4302 of the Affordable Care Act for race, ethnicity, sex, primary language, and disability status. These standards govern HHS survey data collections and are relevant to many of the data elements collected by and for CMS programs today.

Both CMS and HHS more broadly have continued to identify new standards to improve data across Federal programs and beyond. In 2017, CMS launched the Meaningful Measures Framework to prioritize elements of quality measurement and improvement based on input from patients, health care professionals, payers, and other stakeholders. The framework reduces data reporting burden by focusing on the collection of high impact quality measures across CMS programs. The Meaningful Measures 2.0 Framework is currently being developed to continue this progress, pursuing goals such as aligning measures across value-based programs, prioritizing outcome and patient-reported measures, and transforming measures to be fully digital by 2025.

The HHS ONC is the principal federal coordinating entity advancing health IT for the nation. The ONC USCDI is a standardized set of health data classes and constituent data elements for nationwide interoperable health information exchange and comprises a core set of data needed to support patient care and facilitate patient access using health IT. USCDI establishes a consistent baseline of data for other use cases and includes many data elements related to equity and to SDOH such as race, ethnicity, and SOGI. These capabilities have been widely implemented in health IT certified to ONC’s 2015 Edition certification criteria, are widely used in ambulatory and inpatient settings, and allow for the electronic capture, exchange, and use of elements like race and ethnicity. ONC is conducting a technical analysis comparing the 2011 HHS Data Standards with the USCDI, offering guidance for aligned use of existing computable, data elements and standards as available at this time, and providing recommendations for potential future updates.
References


68 ONC USCDI v1 Patient Demographics – “Race” and “Ethnicity”. The Office of the National Coordination for Health Information Technology (ONC). https://www.healthit.gov/isa/uscdi-data-class/patient-demographicsinformation#uscdi-v1


86 Medicare Program; End-Stage Renal Disease Prospective Payment System, Payment for Renal Dialysis Services Furnished to Individuals With Acute Kidney Injury, End-Stage Renal Disease Quality Incentive Program, and End-Stage Renal Disease Treatment Choices Model. Centers for Medicare & Medicaid Services. November 2021. [https://www.govinfo.gov/content/pkg/FR-2021-11-08/pdf/2021-23907.pdf](https://www.govinfo.gov/content/pkg/FR-2021-11-08/pdf/2021-23907.pdf)


Medicare and Medicaid Programs; CY 2023 Payment Policies Under the Physician Fee Schedule and Other Changes to Part B Payment Policies; Medicare Shared Savings Program Requirements; Medicare and Medicaid Provider
Enrollment Policies, Including for Skilled Nursing Facilities; Conditions of Payment for Suppliers of Durable Medicaid Equipment, Prosthetics, Orthotics, and Supplies (DMEPOS); and Implementing Requirements for Manufacturers of Certain Single-Dose Container or Single-Use Package Drugs To Provide Refunds With Respect to Discarded Amounts.


Request for Information on the Use of Clinical Algorithms That Have the Potential To Introduce Racial/Ethnic Bias Into Healthcare Delivery. Agency for Healthcare Research and Quality. March 2021. [https://www.federalregister.gov/documents/2021/03/05/2021-04509/request-for-information-on-the-use-of-clinical-algorithms-that-have-the-potential-to-introduce](https://www.federalregister.gov/documents/2021/03/05/2021-04509/request-for-information-on-the-use-of-clinical-algorithms-that-have-the-potential-to-introduce)


