Resource of Health Equity-Related Data Definitions, Specifications, & Stratification Practices

November 2024

CENTERS FOR MEDICARE & MEDICAID SERVICES'

Resource of health equity-related data definitions, specifications, and stratification practices

Advancing health equity and improving health equity-related data collection are priorities for the Biden-Harris Administration, as detailed in Executive Order 13985 (Advancing Racial Equity and Support for Underserved Communities Through the Federal Government), released in January 2021, and subsequent equity-focused Executive Orders, including Executive Order 14031 (Advancing Equity, Justice, and Opportunity for Asian Americans, Native Hawaiians, and Pacific Islanders, released in May 2021, Executive Order 14075 (Advancing Equality for Lesbian, Gay, Bisexual, Transgender, Queer, and Intersex Individuals), released in June 2022, and Executive Order 14091 (Further Advancing Racial Equity and Support for Underserved Communities Through the Federal Government), released in February 2023. In alignment with these Executive Orders, *Advance Equity* is the first of the Centers for Medicare & Medicaid Services (CMS)'s six Strategic Pillars, highlighting the importance of addressing the health disparities that underlie the United States' health care system. Health equity-related data collection is critical for advancing equity as the lack of data impedes measurement of disparities and the identification of opportunities to drive action. Standardized data collection and stratification allow for comprehensive analyses that can be combined or compared across multiple programs or initiatives; and in turn, offer an important first step towards improving population health.

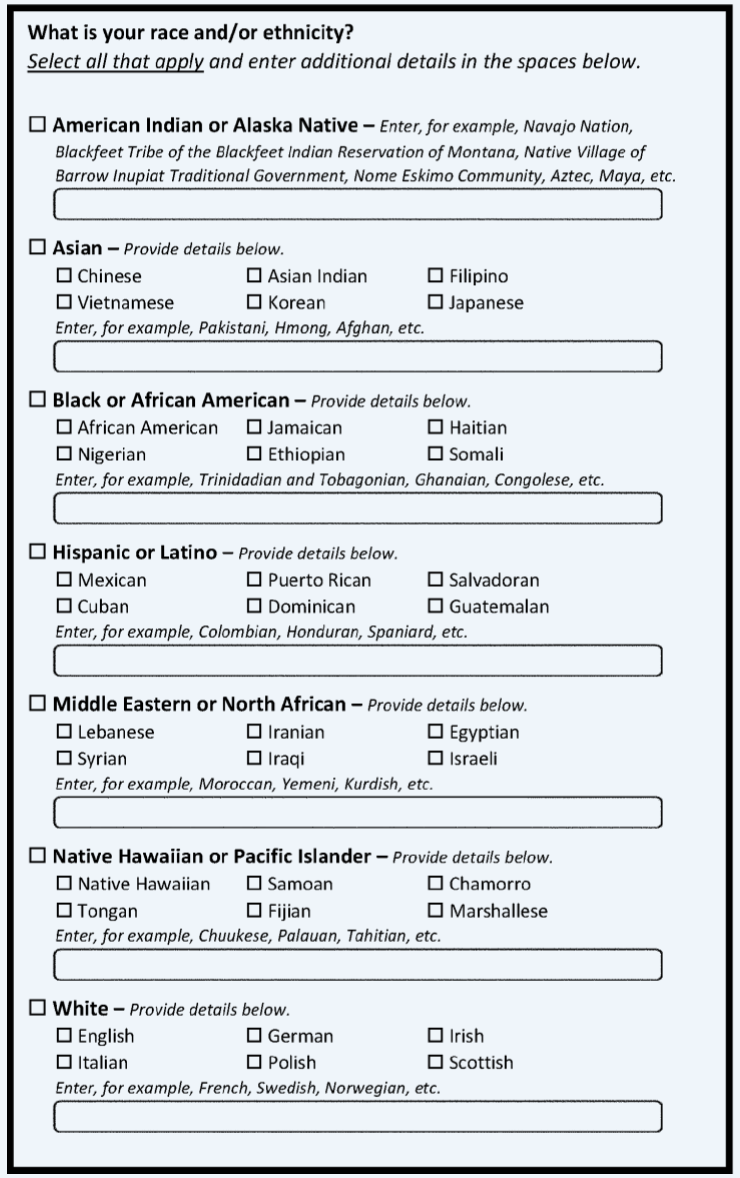
In 2022, CMS released its *Framework for Health Equity 2022-2032*, which contains five health equity priorities for reducing disparities in health. Priority 1, "*Expand the Collection, Reporting, and Analysis of Standardized Data,*" reflects the importance of ensuring that CMS's collection of health equity-related data, including but not limited to sociodemographic and social determinants of health (SDOH) information, is comprehensive, consistent, and interoperable. Efforts toward achieving this priority – i.e., addressing current gaps in relevant data collection, reporting, and analysis – will not only enhance overall CMS data integrity by better capturing the experiences of people marginalized by the healthcare system, but will drive meaningful transformation in policies and programs for the individuals that CMS serves.

This document serves as a technical resource that can be used by organizations and entities, such as providers, states, community organizations, and others, that wish to harmonize with CMS when collecting, stratifying, and/or analyzing health equity-related data. It may also clarify differences in results that may arise when different data standards and definitions are used. This document includes suggested definitions, specifications, and stratification practices for the following sociodemographic elements:

- Race and Ethnicity
- Sex (assigned at birth)
- Gender identity
- Sexual orientation
- Disability status
- Primary Language
- English language proficiency
- Rurality/urbanicity or residence

NOTE:

CMS has prepared the following table as a technical resource for organizations and entities, such as providers, states, community organizations, and others. We hope these standards can be useful as different organizations and entities work to improve their data collection. Where applicable to a given organization or entity and appropriate, CMS recommends harmonization with the United States Core Data for Interoperability (USCDI) standards. We may update this resource in the future to reflect advancements in sociodemographic data collection. We note that this resource does not create or override any data collection or stratification requirements.

Race and Ethnicity				Notes
Standards	**Definition**	**Question specified**	**Categories**	
OMB's 2024 Statistical Policy Directive No. 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity Minimum categories, multiple detailed checkboxes, and write -in response with example groups – Default (shown) When an agency receives an OIRA exemption from collecting detailed data, it may use a format that includes only the minimum categories: Minimum categories and examples Minimum Categories only Refer to spd15revision.gov for exact guidance.	N/A	“What is your race and/or ethnicity? Select all that apply and enter additional details in the spaces below.	**Figure 1. Race and Ethnicity Question with Minimum Categories, Multiple Detailed Checkboxes, and Write-In Response Areas with Example Groups**  **What is your race and/or ethnicity?** *Select all that apply and enter additional details in the spaces below.* ☐ **American Indian or Alaska Native –** *Enter, for example, Navajo Nation, Blackfeet Tribe of the Blackfeet Indian Reservation of Montana, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, Aztec, Maya, etc.* ☐ **Asian –** *Provide details below.* ☐ Chinese ☐ Asian Indian ☐ Filipino ☐ Vietnamese ☐ Korean ☐ Japanese *Enter, for example, Pakistani, Hmong, Afghan, etc.* ☐ **Black or African American –** *Provide details below.* ☐ African American ☐ Jamaican ☐ Haitian ☐ Nigerian ☐ Ethiopian ☐ Somali *Enter, for example, Trinidadian and Tobagonian, Ghanaian, Congolese, etc.* ☐ **Hispanic or Latino –** *Provide details below.* ☐ Mexican ☐ Puerto Rican ☐ Salvadoran ☐ Cuban ☐ Dominican ☐ Guatemalan *Enter, for example, Colombian, Honduran, Spaniard, etc.* ☐ **Middle Eastern or North African –** *Provide details below.* ☐ Lebanese ☐ Iranian ☐ Egyptian ☐ Syrian ☐ Iraqi ☐ Israeli *Enter, for example, Moroccan, Yemeni, Kurdish, etc.* ☐ **Native Hawaiian or Pacific Islander –** *Provide details below.* ☐ Native Hawaiian ☐ Samoan ☐ Chamorro ☐ Tongan ☐ Fijian ☐ Marshallese *Enter, for example, Chuukese, Palauan, Tahitian, etc.* ☐ **White –** *Provide details below.* ☐ English ☐ German ☐ Irish ☐ Italian ☐ Polish ☐ Scottish *Enter, for example, French, Swedish, Norwegian, etc.*	• Note this standard combines the separate questions on race and ethnicity into a single combined race and ethnicity question. • Recommend sufficient sample size to produce reliable estimates, as noted in OMB’s SPD 15. • Recommend using high quality, self-reported data when available as self-report is the preferred method. • When data are not self-reported, agencies are encouraged to transparently describe how the data were collected or generated, and how nonresponse or other missing data were assigned or allocated. For statistical survey reporting, agencies are **required**, rather than encouraged, to transparently describe whether race and ethnicity data are self- reported or collected by proxy, along with any imputation or coding procedures. • When data quality / missingness are a concern, recommend using imputed race (Medicare Bayesian Improved Surname Geocoding (MBISG) for Medicare, which are more updated than the RTI Race Code; Bayesian Improved First Name Surname Geocoding (BIFSG) for Medicaid and CHIP). • Consider focus group testing or formal “cognitive testing” if translating to a different language to ensure the adequacy of translations.

Data Element		Sexual Orientation and Gender Identity (SOGI)				Notes
		Specifications	Example Definition	Example Question Specified	Example Response Options	
Two-Step Questions	Gender Identity (Question 1 of 2)	Recommendations on the Best Practices for the Collection of Sexual Orientation and Gender Identity Data on Federal Statistical Surveys Federal Evidence Agenda on LGBTQI+ Equity Modified NASEM Best Practices	Gender identity is a core element of a person's individual identity	What is your gender?	- Woman - Man - Non-binary - [Shown only if respondent has previously identified themselves as American Indian or Alaska Native:] Two-Spirit - I use a different term: [free text] - Prefer not to answer [If this is an option for other demographic questions]	- There is no single, best practice set of questions for soliciting information about a person's sexual orientation or gender identity (SOGI) and there is evidence continuing to emerge from the field that may continue the evolution and improvement of SOGI data collection methods. - The NASEM report laid out important steps for research to pursue next, and the Federal Evidence Agenda on LGBTQI+ Equity stated that SOGI data are demographic data and data collection must start immediately even as research continues. - To minimize burden and privacy risk to respondents, only the minimum amount of information needed to meet the planned uses should be collected. - Gender identity and sex assigned at birth should not be conflated. - Gender identity should be self-reported. - Using sex assigned at birth together with these gender identity codes is necessary for mapping to the latest USCDI standards. - Where data are available, stratify by sex assigned at birth as well as gender identity. - Consider limits to generalizability if the distribution is unexpected. - In alignment with NASEM, AI/AN response option of "Two-Spirit" should be reserved for those data vehicles that can enable skip logic and ensure that the response option appears only for those who have already self-identified as AI/AN. - Consider focus group testing or formal "cognitive testing" if translating to a different language to ensure the adequacy of translations.
	Sex (as Assigned at Birth) (Question 2 of 2)	Modified 2023 Recommendations on the Best Practices for the Collection of Sexual Orientation and Gender Identity Data on Federal Statistical Surveys Federal Evidence Agenda on LGBTQI+ Equity Modified NASEM Best Practices	A multidimensional construct based on a cluster of anatomical and physiological traits (sex traits)	What is your sex assigned at birth, for example on your original birth certificate?	- Female - Male - I don't know	- Sex assigned at birth and Gender identity should not be conflated. - If statutory requirement mandates the collection of sex, then it may not be necessary to collect sex assigned at birth. - Where data are available, stratify by sex assigned at birth as well as gender identity. - Consider limits to generalizability if the distribution is unexpected. - Using sex assigned at birth together with gender identity codes is necessary for mapping to the latest USCDI standards. - Consider including "Intersex" for new data collection as feasible, consistent with EO 14075. - Consider focus group testing or formal "cognitive testing" if translating to a different language to ensure the adequacy of translations.

Data Element		Sexual Orientation and Gender Identity (SOGI)				Notes
		Example Specifications	Example Definition	Example Question Specified	Example Categories	
Sexual Orientation and Gender Identity (SOGI)	Sexual Orientation	Modified 2023 Recommendations on the Best Practices for the Collection of Sexual Orientation and Gender Identity Data on Federal Statistical Surveys Federal Evidence Agenda on LGBTQI+ Equity Modified NASEM Best Practices	Sexual orientation is a multidimensional construct encompassing emotional, romantic, and sexual attraction, identity, and behavior	Which of the following best represents how you think of yourself? (Select one	• Lesbian or gay • Straight, that is, not gay or lesbian • Bisexual • [Shown only if respondent has previously identified themselves as American Indian or Alaska Native:] Two-Spirit • I use a different term [free- text] • Prefer not to answer • (Don't know)	• There is no single, best practice set of questions for soliciting information about a person's sexual orientation or gender identity (SOGI) and there is evidence continuing to emerge from the field that may continue the evolution and improvement of SOGI data collection methods. • The NASEM report laid out important steps for research to pursue next, and the Federal Evidence Agenda on LGBTQI+ Equity stated that SOGI data are demographic data and data collection must start immediately even as research continues. • To minimize burden and privacy risk to respondents, only the minimum amount of information needed to meet the planned uses should be collected. • Sexual orientation should be self-reported. • Consideration should be given if the question is asked in languages other than English. • In alignment with NASEM, AI/AN response option of “Two-Spirit” should be reserved for those data vehicles that can enable skip logic and ensure that the response option appears only for those who have already self-identified as AI/AN. • Consider focus group testing or formal “cognitive testing” if translating to a different language to ensure the adequacy of translations.

Disability				Notes
Example Specifications	Example Definition	Example Question Specified	Example Categories	
HHS Implementation Guidance on Data Collection Standards for Race, Ethnicity, Sex, Primary Language, and Disability Status	N/A	• Are you deaf or do you have serious difficulty hearing? • Are you blind or do you have serious difficulty seeing, even when wearing glasses? • Because of a physical, mental, or emotional condition, do you have serious difficulty concentrating, remembering, or making decisions? (5 years old or older) • Do you have serious difficulty walking or climbing stairs? (5 years old or older) • Do you have difficulty dressing or bathing? (5 years old or older) • Because of a physical, mental, or emotional condition, do you have difficulty doing errands alone such as visiting a doctor's office or shopping? (15 years old or older)	Responses per question: • Yes • No	• Ideally, as is the case with many other health equity-related sociodemographic data elements, disability status is self-reported. Implementation of this approach will likely require attention to whether the question is asked in languages other than English. • CMS recognizes that many federal and state programs use eligibility-based definitions of disability and that the 2011 HHS Standards may be challenging. Consider using the 2011 HHS Standard as data collection efforts evolve to the extent permitted by statute, regulation, or guidance. • Consider focus group testing or formal “cognitive testing” if translating to a different language to ensure the adequacy of translations.

Data Element	Language				Notes
	Example Specifications	Example Definition	Example Question Specified	Example Categories	
English Language Proficiency	[HHS Implementation Guidance on Data Collection Standards for Race, Ethnicity, Sex, Primary Language, and Disability Status]()	N/A	How well do you speak English? (5 years old or older)	• Very well • Well • Not well • Not at all	• Ideally, as is the case with many other health equity-related sociodemographic data elements, English Language Proficiency is self-reported. Implementation of this approach will likely require attention to whether the question is asked in languages other than English. • Consider focus group testing or formal “cognitive testing” if translating to a different language to ensure the adequacy of translations.
Primary Language	[HHS Implementation Guidance on Data Collection Standards for Race, Ethnicity, Sex, Primary Language, and Disability Status]()		Do you speak a language other than English at home?	• Yes • No (If Yes) What is this language? • Spanish • Another language	• Ideally, as is the case with many other health equity-related sociodemographic data elements, Primary Language is self-reported. Implementation of this approach will likely require attention to whether the question is asked in languages other than English. • You may wish to expand the list of response options as appropriate, but depending on subgroup sizes (i.e., if they are especially small), you can collapse or "roll-up" categories to meet the 2011 HHS Data Standards. • Consider focus group testing or formal “cognitive testing” if translating to a different language to ensure the adequacy of translations.

Data Element	Specification for Metropolitan/Micropolitan Level Analysis				Specification for Granular/Zip Code Level Analysis				Notes
	Specification	Definition	Question specified	Categories	Specification	Definition	Question specified	Categories	
Geography (Rurality)	U.S. Census Housing Patterns and Core-Based Statistical Areas	Metropolitan and Micropolitan Statistical Areas are collectively referred to as Core-Based Statistical Areas. Metropolitan and micropolitan statistical areas are defined in terms of whole counties or county equivalents, including the six New England states.	N/A	• Metropolitan Statistical Areas: Have at least one urbanized area of 50,000 or more population, plus adjacent territory that has a high degree of social and economic integration with the core as measured by commuting ties. (e.g., New York-Newark-Jersey City, Washington-Arlington-Alexandria) = **Urban** • Micropolitan Statistical Areas: Have at least one urban cluster of at least 10,000 but less than 50,000 population, plus adjacent territory that has a high degree of social and economic integration with the core as measured by commuting ties. (e.g., Tupelo, MS; Augusta- Waterville, ME) = **Rural** • Non-CBSA: Do not have at least one urban cluster of at least 10,000 (e.g., Millington, MD) = **Rural** • NOTE: Metropolitan and micropolitan statistical areas are defined in terms of whole counties or county equivalents, including the six New England states. As of June 6, 2003, there are 362 metropolitan statistical areas and 560 micropolitan statistical areas in the United States.	USDA Rural-Urban Commuting Area Codes	The rural- urban commuting area (RUCA) codes classify U.S. census tracts using measures of population density, urbanization, and daily commuting.	N/A	• 1 = Metropolitan area core: primary flow with an urbanized area (UA) = **Urban** (e.g., Baltimore County, MD; Dallas County, TX) • 2 = Metropolitan area high commuting: primary flow 30% or more to a UA = **Urban** (e.g., Montgomery County, MD; Galveston County, TX) • 3 = Metropolitan area low commuting: primary flow 10% to 30% to a UA = **Urban** (e.g., Calvert County, MD; Jefferson County, TX) • 4 = Micropolitan area core: primary flow within an urban cluster (UC) of 10,000 to 49,999 (large UC) = **Rural** (e.g., Dorchester County, MD; Anderson County, TX) • 5 = Micropolitan high commuting: primary flow 30% or more to a large UC = **Rural** (e.g., Talbot County, MD; Calhoun County, TX) • 6 = Micropolitan low commuting: primary flow 10% to 30% to a large UC = **Rural** (Worcester County, MD; Cooke County, TX) • 7 = Small town core: primary flow within an urban cluster of 2,500 to 9,999 (small UC) = **Rural** (e.g., Garrett County, MD; Falls County, TX) • 8 = Small town high commuting: primary flow 30% or more to a small UC = **Rural** (e.g., Kent County, MD; Hunt County, TX) • 9 = Small town low commuting: primary flow 10% to 30% to a small UC = **Rural** (e.g., Queen Anne's County, MD; Limestone County, TX) • 10 = Rural areas: primary flow to a tract outside of a UA or UC =**Rural** (e.g., St Mary's County, MD; Shelby County, TX)	• At a minimum, response options or groupings should include "rural" and "urban," but could be more granular when and where appropriate and possible (e.g., "metropolitan," "micropolitan," "rural," and "non-core- based statistical area (CBSA)"). • You may consider collapsing or "rolling up" categories to meet your sample size needs for reporting, but the data used, ideally, should have been collected in as granular a manner as possible (and reported in as granular a manner as possible, while still ensuring validity). • When applicable, consider more granular rural categories such as Frontier. • The Rural-Urban Continuum Codes (RUCC) are another standard used to define rural and urban.

Frequently Asked Questions

1. What is this resource document and why is CMS releasing it?

This document is a technical resource of definitions, specifications, and stratification practices for health equity-related data elements that can be used by organizations/entities such as providers, states, community organizations, and others that wish to harmonize with CMS when collecting, stratifying, and/or analyzing health equity-related data. It may also clarify differences in results that may arise when different data standards and definitions are used.

CMS's Framework for Health Equity 2022-2032 contains five priorities for reducing health disparities. Priority 1, "Expand the Collection, Reporting, and Analysis of Standardized Data," reflects the importance of ensuring that CMS's health equity-related data and social determinants of health (SDOH) data are comprehensive, consistent, and interoperable. Standardized data collection and stratification allow for comprehensive analyses that can be combined or compared across multiple programs or initiatives; and in turn, offer an important first step towards improving population health.

2. What data elements are included in the resource document?

This document includes suggested definitions, standards, and stratification practices to address populations and areas identified in Executive Order 13985 "Advancing Racial Equity and Support for Underserved Communities Through the Federal Government"[1] and Executive Order 14091 "Further Advancing Racial Equity and Support for Underserved Communities Through the Federal Government." [2] The included data elements are race, ethnicity, sex, disability status, gender identity, sexual orientation, primary language, English language proficiency, and rurality/urbanicity of residence.

3. How were these standards selected?

We selected the standards based upon broadest applicability and feasibility for the various data collection methods and systems, as well as usage of the data to help harmonize assessment of the included data elements across CMS.

---

[1] https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/

[2] https://www.whitehouse.gov/briefing-room/presidential-actions/2023/02/16/executive-order-on-further-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/

4. Will all CMS data collection efforts align with the information in the resource document?

Under certain situations, some of the standards listed in this resource document are binding on CMS as a federal agency (e.g., OMB standards for race and ethnicity, HHS standards for the collection of demographic data in population health surveys). Other standards listed in this document are not binding but CMS programs will begin moving towards the voluntary adoption of these standards where practicable and legally permissible to facilitate consistency on the "front end" collection of demographic data and its analysis.

Regardless of the standard adopted for the "front end" collection of demographic data, CMS recommends where applicable and appropriate harmonization with the Office of the National Coordinator for Health Information United States Core Data for Interoperability (USCDI) standards on the "back end" to facilitate the sharing of interoperable data.

5. Why do some CMS data collection efforts use different standards and will that change?

Many CMS programs adopted standards prior to the development of this resource document. Where practicable and legally permissible, some collection will align with these listed standards over time. Additionally, taking into consideration the program and the enrollees whom we ask the questions, we perform cognitive and/or consumer testing on the question and response options to ensure our enrollees understand and are able to respond appropriately. CMS is continuing to explore how to implement questions and response options for different use cases such as benefits applications (eligibility/enrollment), surveys, and model patient forms in care settings, and consider how the method of consumer engagement may require different approaches to best support the interaction channel, for example a self-service online form vs. a phone interaction (live representative vs. an automated telephone system).

6. Will this resource be updated as standards are revised?

Standards continuously evolve and may be subject to change. We will update this resource as applicable.

7. Are the listed data standards now required of CMS partners for alignment?

This document is solely a technical resource of suggested data definitions, standards, and stratification practices. Unless and until otherwise noted in statute, regulation, guidance, or formal agreement (including, but not limited to, CMS terms and conditions included in contracts and other binding agreements), these standards are non-binding for CMS partners.

8. What does this mean for state Medicaid and CHIP agencies?

As noted earlier, this document is a technical resource. Any actual data collection requirements for Medicaid and CHIP agencies would be described in statute or regulation as interpreted in guidance, or state-specific documents like the State's Medicaid State plan, section 1915 waivers, or section 1115 demonstration special terms and conditions.

There are many different ways state Medicaid and CHIP agencies could collect this information, such as standalone surveys, surveys fielded as addenda to state Medicaid and CHIP applications, and by leveraging demographic data collected through health information exchanges and electronic health records.

As required in regulations at 42 CFR 435.907(b) and 457.330, the single, streamlined application that individuals use to apply for Medicaid or CHIP benefits must be either the model application, or an alternative developed by the state and approved by CMS through a state plan amendment (SPA). In addition to the single, streamlined application, many states have applications or supplemental forms for determining eligibility on a basis other than MAGI, or other alternative applications not subject to CMS approval.

States that wish to update their Medicaid or CHIP alternative, single streamlined application to include these standards may need to work with CMS to determine whether CMS approval and a SPA is needed. States may need to submit a SPA if they want to add questions that are not needed for an eligibility determination or are not for a purpose directly connected to the administration of the state plan, or if they want to make substantive changes that significantly deviate from the model application.

More detailed information on the development, review, and approval of alternative applications is available in the June 18, 2013, joint Center for Consumer Information and Insurance Oversight (CCIIO) and Center for Medicaid and CHIP Services (CMCS) guidance entitled "Guidance on State Alternative Applications for Health Coverage"[3] and in the November 9, 2023, CMCS informational bulletin entitled "Guidance on Adding Sexual Orientation and Gender Identity Questions to State Medicaid and CHIP Applications for Health Coverage."[4]

9. Are there any additional considerations when it comes to data collection and stratification for American Indian/Alaska Native (AI/AN) populations?

States are encouraged to use Tribal consultation in developing Tribal data collection, analysis, and data distribution approaches, and may be required to conduct Tribal consultation in certain circumstances (such as prior to submitting a Medicaid SPA that is likely to have a direct effect on Indians and Indian health programs). CMS and Tribes share the goals of eliminating health disparities for AI/ANs and ensuring access to critical health services, including those made available through Medicare, Medicaid, CHIP, and the Health Insurance Marketplace,®[5] is maximized to advance or enhance Tribal members' physical, mental, and economic status. To achieve these goals, and to the extent practicable and permitted by law, it is important that CMS, States, and Indian Tribes engage in open, continuous, and meaningful consultation.

---

[3] https://www.cms.gov/CCIIO/Resources/Regulations-and-Guidance/Downloads/state-alt-app-guidance-6-18-2013.pdf

[4] https://www.medicaid.gov/sites/default/files/2023-11/cib11092023.pdf

[5] Health Insurance Marketplace® is a registered service mark of the U.S. Department of Health & Human Services.

More information on the CMS consultation policy can be found here:

https://www.cms.gov/Outreach-and-Education/American-Indian-Alaska-Native/AIAN/Downloads/CMSTribalConsultationPolicy2015.pdf

10. What is the recommended solution when it is not feasible to align with the standards in this resource?

This document is a technical resource of suggested definitions, standards, and stratification practices. Alignment with these standards may not be feasible under certain circumstances, necessitating deviations. Regardless of which data standards are adopted for “front end” collection of demographic data, CMS recommends harmonization with the ONC USCDI standards to facilitate “back-end” data interoperability.

11. How can one receive assistance in implementing these standards?

Each of the listed standards provide documentation and guidance for implementation that can be reviewed in detail. Where assistance is needed with regard to CMS usage of these standards and stratification practices, the Health Equity Technical Assistance program can be contacted at HealthEquityTA@cms.hhs.gov.

12. Where can I find more information on OMB’s 2024 Statistical Policy Directive No. 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity (SPD 15)?

Please refer to spdrevision.gov including its Contact Us page and page for Federal Employees for exact guidance on SPD 15.

Contact Us: https://spd15revision.gov/content/spd15revision/en/contact.html

Federal Employees: https://spd15revision.gov/content/spd15revision/en/federal-government.html