

**U.S DEPARTMENT OF HEALTH AND HUMAN SERVICES
IMPLEMENTATION GUIDANCE ON DATA COLLECTION STANDARDS FOR
RACE, ETHNICITY, SEX, PRIMARY LANGUAGE, AND DISABILITY STATUS**

I. Purpose and Background

Purpose

The purpose of this guidance is to promulgate a set of uniform data collection standards for inclusion in surveys conducted or sponsored by HHS as required by Section 4302 of the Affordable Care Act.

Background

HHS reports, dating back to the landmark 1985 Secretary’s Task Force on Black and Minority Health, emphasize the importance of timely and reliable data to assist in identifying racial and ethnic health disparities, in understanding the causes and correlates of disparities, and in monitoring progress in reducing them. HHS has a long commitment to developing such data, best exemplified by the 1997 HHS Data Inclusion policy, which required the collection of uniform standard data on race and ethnicity in all HHS-sponsored data collection activities. Data improvement efforts enhance the ability of the public health and healthcare systems to identify and track disparities in health and health care, understand their correlates and consequences, and facilitate greater accountability for reducing them. However, the lack of standards related to data collection on vulnerable population subgroups remains a challenge for adequately collecting, reporting, and tracking data on health disparities.

Overview of Section 4302 of the Affordable Care Act

The Affordable Care Act (ACA) includes several provisions aimed at eliminating health disparities in America. Section 4302 (Understanding health disparities: data collection and analysis) of the ACA focuses on the standardization, collection, analysis, and reporting of health disparities data. While data alone will not reduce disparities, it can be foundational to our efforts to understand the causes, design effective responses, and evaluate our progress.

Section 4302 requires the Secretary of DHHS to establish data collection standards for race, ethnicity, sex, primary language, and disability status. The law requires that, once established, these data collection standards be used, to the extent practicable, in all national population health surveys. In response to this statutory requirement, this implementation guidance outlines the new minimum data collection standards for race, ethnicity, sex, primary language and disability status for implementation in HHS, along with a description of the data standards development process, the rationale for each data standard, and instructions for their implementation.

II. Data Standards Development Process

Under the auspices of the ACA Prevention Implementation Workgroup and the Section 4302 Implementation Subgroup, the HHS Data Council was asked to recommend data standards for

race, ethnicity, sex, primary language and disability status to support the implementation of Section 4302 of the Affordable Care Act. The HHS Data Council is the principal, senior internal Departmental forum and advisory body to the Secretary on health and human services data policy and coordinates HHS data collection and analysis activities. A special workgroup within the Data Council, the Section 4302 Standards Workgroup, was formed to lead this task. The Workgroup included representatives from HHS, the Office of Management and Budget (OMB), and the Census Bureau. The Workgroup examined current federal data collection standards, adequacy of prior testing, and quality of the data produced in prior surveys; consulted with statistical agencies and programs; reviewed OMB data collection standards and the Institute of Medicine (IOM) Report *Race, Ethnicity, and Language Data Collection: Standardization for Health Care Quality Improvement*, and built on its members' experience with collecting and analyzing demographic data.ⁱ

The following criteria guided development for data standards for each of the five required variables:

1. Data standards would be evidence-based and demonstrated to have worked well in practice for national survey data collection.
2. Data standards would be framed as minimum data standards, with agencies permitted to include as many additional questions on these topics as desired as long as the minimum standard is included. Agencies would also be permitted to include additional response categories for data standards with as much additional detail and granularity as desired, provided that the additional detail could be aggregated back to the minimum standard and the sample design and sample size support estimates at that level of granularity.
3. The data standards would comply at a minimum, with any standards already mandated by OMB.
4. Data standards are for person-level data collected in population-based health surveys, where subjects either self-report information or a knowledgeable proxy provides information about the subject or responds for all persons in a household.

A draft set of data collection standards were developed, subjected to several levels of internal review, and then published for public comment. This policy guidance reflects the final set of data collection standards and supporting guidance for implementation.

III. Data Collection Standards and Rationale for Selection

A. Race and Ethnicity

The starting point for the race and ethnicity data collection standards is OMB's current government-wide standard, issued in 1997 after a comprehensive public engagement process and extensive field testing. The principles underlying these government-wide standards are described below. The justifications for these principles are described by OMB in detail at http://www.whitehouse.gov/omb/fedreg_1997standards/.

- Self-identification is the preferred means of obtaining information about an individual's race and ethnicity, except in instances where observer identification is more practical. The surveyor should not tell an individual who he or she is, or specify how an individual should classify himself or herself.
- To provide flexibility and ensure data quality, separate questions for race and ethnicity should be used wherever feasible. Specifically, when self-reporting or other self-identification approaches are used, ethnicity is asked first, and then race. The standard acknowledges that this standard might not work in other contexts (e.g., administrative records.)
- The specified race and ethnicity categories provide a minimum set of categories except when the collection involves a sample of such size that the data on the smaller categories would be unreliable, or when the collection effort focuses on a specific racial or ethnic group.
 - The OMB minimum categories for race are: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White.
 - The OMB minimum categories for ethnicity are: Hispanic or Latino and Not Hispanic or Latino.
- When self-reporting or other self-identification approaches are used, respondents who wish to identify their multi-racial heritage may choose more than one race; there is no “multi-racial” category.
- OMB encourages additional granularity where it is supported by sample size and as long as the additional detail can be aggregated back to the minimum standard set of race and ethnicity categories.
- Any other variation will have to be specifically authorized by the OMB through the information collection clearance process. In those cases where the data collection is not subject to the information collection clearance process, a direct request for a variance should be made to OMB.

The categories for HHS data standards for race and ethnicity are based on the disaggregation of the OMB standard used in the American Community Survey (ACS) and the 2000 and 2010 Decennial Census. The data standard for race and ethnicity is listed below. Race and ethnicity data collection applies to survey participants of all ages.

Ethnicity Data Standard

Are you Hispanic, Latino/a, or Spanish origin (One or more categories may be selected)

- a. *No, not of Hispanic, Latino/a, or Spanish origin*
- b. *Yes, Mexican, Mexican American, Chicano/a*
- c. *Yes, Puerto Rican*
- d. *Yes, Cuban*
- e. *Yes, Another Hispanic, Latino/a or Spanish origin*

These categories roll-up to the Hispanic or Latino category of the OMB standard

Race Data Standard

What is your race? (One or more categories may be selected)

- a. *White*
- b. *Black or African American*
- c. *American Indian or Alaska Native*

These categories are part of the current OMB standard

- d. *Asian Indian*
- e. *Chinese*
- f. *Filipino*
- g. *Japanese*
- h. *Korean*
- i. *Vietnamese*
- j. *Other Asian*

These categories roll-up to the Asian category of the OMB standard

- k. *Native Hawaiian*
- l. *Guamanian or Chamorro*
- m. *Samoan*
- n. *Other Pacific Islander*

These categories roll-up to the Native Hawaiian or Other Pacific Islander category of the OMB standard

Rationale for Race and Ethnicity Data Standard

As a result of the 1997 HHS data inclusion policy, the basic OMB standard is already included in most HHS data collection initiatives. The new HHS data standards for race and ethnicity include additional granularity, but all categories roll-up to the OMB standard. However, because additional granularity in the race and ethnicity categories is important for documenting and tracking health disparities, large federal surveys such as the National Health Interview Survey (NHIS), Current Population Survey (CPS), and the ACS have implemented such a more granular strategy, particularly for Hispanic and Asian subpopulations.

Accordingly, the new data standards for race and ethnicity are a slightly modified version of the ACS and Decennial Census questions. These items provide additional granularity for Hispanic (four additional categories) and Asian subpopulations (7 additional categories) beyond the OMB minimum standard categories. The race and ethnicity categories for the ACS and recent Decennial Census have been tested and structured to increase response rates, validity, and reliability.ⁱⁱ The more detailed ACS and recent Decennial Census race categories roll up to the

OMB standard five categories: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White. As with OMB standards, respondents are also instructed to mark all categories that apply (i.e. they may be able to select more than one racial category). The ACS and Decennial Census ethnicity categories roll up to the OMB standard categories: Hispanic or Latino and Not Hispanic or Latino.^{iii,iv} Respondents are also able to select more than one ethnicity category. The recommended standard is in conformance with the methods, logistics, practices and limitations of HHS major surveys, where population estimates are the goal.

HHS agencies may request permission from OMB during the Paperwork Reduction Act clearance process to add a write-in option of “other” to interviewer-administered surveys. This respondent-specified race must then be coded by the agency to the OMB and HHS standards before results are publically reported.

B. Sex

The data standard for sex is male and female. Sex data collection applies to survey participants of all ages.

Sex Data Standard

What is your sex?

- a. _____ *Male*
- b. _____ *Female*

Rationale for Sex Data Standard

For the purpose of this report, the category of sex was defined as biologic sex. Sexual orientation and gender identity were considered as separate concepts. The Department has developed a data progression plan for collecting sexual orientation data and has conducted gender identity data collection listening sessions.

C. Primary Language

The standard for primary language is a measure of English proficiency. The recommended question is based on that used on the ACS. The question applies to survey participants aged five years and above.

Data Standard for Primary Language

How well do you speak English? (5 years old or older)

- a. _____ *Very well*
- b. _____ *Well*
- c. _____ *Not well*
- d. _____ *Not at all*

The primary language data standard represents a minimum standard and the question and answer categories cannot be changed. Additional questions on language may be added to any survey as long as the minimum standard is included.

Optional Granularity

For agencies that wish to collect data on the specific language spoken, the Data Council recommends collecting data on language spoken at home. The recommended survey items are used in the ACS (see below). Collecting this additional information would be optional and at the discretion of the agency, if information on specific language was desired.

- 1. Do you speak a language other than English at home? (5 years old or older)**
 - a. *Yes*
 - b. *No*

For persons speaking a language other than English (answering yes to the question above):

- 2. What is this language? (5 years old or older)**
 - a. *Spanish*
 - b. *Other Language (Identify)*

For agencies that desire to collect information on specific languages beyond Spanish, and have sufficient sample sizes to support such estimates, HHS would publish on the HHS website a list of the ten most prevalent languages spoken in the U.S., as reported by ACS. These would roll up to the “Other Language” category, and provide technical notes to assist in coding. Spanish as a category is reported about 60 percent of the time in the ACS.^v

Rationale for Primary Language Data Standard

The survey item selected for the minimum standard is based on the ACS, which assesses both English proficiency and language spoken other than English, and has been collected by the Census Bureau since 1980.

For statistical, planning, analytical and research purposes, disparities have been associated with English language proficiency rather than specific language spoken. For clinical purposes relating to an individual, specific language and proficiency would both be needed. This recommendation is consistent with language recommendations from the Institute of Medicine report *Race, Ethnicity, and Language Data Collection: Standardization for Health Care Quality Improvement*.

Several HHS surveys currently collect data on language or English proficiency primarily in the preliminary screening phase of in person or telephone interview surveys for administrative purposes in surveys, to determine how or in what language the interview would be administered. It is not the intent of this standard to disrupt those screening practices.

D. Disability Status

The six item set of questions used on ACS and other major surveys to gauge disability is the data standard for survey questions on disability. Note the age thresholds for survey participants for the different disability questions.

Data Standard for Disability Status

1. *Are you deaf or do you have serious difficulty hearing?*
 - a. *Yes*
 - b. *No*
2. *Are you blind or do you have serious difficulty seeing, even when wearing glasses?*
 - a. *Yes*
 - b. *No*
3. *Because of a physical, mental, or emotional condition, do you have serious difficulty concentrating, remembering, or making decisions? (5 years old or older)*
 - a. *Yes*
 - b. *No*
4. *Do you have serious difficulty walking or climbing stairs? (5 years old or older)*
 - a. *Yes*
 - b. *No*
5. *Do you have difficulty dressing or bathing? (5 years old or older)*
 - a. *Yes*
 - b. *No*
6. *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)*
 - a. *Yes*
 - b. *No*

The six-item disability standard represents a minimum standard and the questions and answer categories cannot be changed. Additional questions on disability may be added to any survey as long as the minimum standard is included. If the ACS changes the disability questions in the future, HHS will revisit the standard and modify as necessary.

Rationale for Disability Data Standard

The six item set of questions used on the ACS and other major surveys to measure disability was developed by a federal interagency committee and reflects the change in how disability is conceptualized consistent with the International Classification of Functioning, Disability, and Health. The question set defines disability from a functional perspective and was developed so that disparities between the 'disabled' and 'nondisabled' population can be monitored. The question set went through several rounds of cognitive and field testing and has been adopted in many federal data collection systems. OMB has encouraged the use of this question set by other federal agencies conducting similar population studies due to the extensive testing used in the development of these measures, including the findings that alternative measures did not test as

well. Cognitive testing of these questions revealed that the six questions must be used as a set to assure a meaningful measure of disability.^{vi}

IV. IMPLEMENTATION GUIDANCE

A. Inclusion of Data on Race, Ethnicity, Sex, Primary Language and Disability Status: The minimum data standards described herein on race, ethnicity, sex, primary language, and disability status must be included in all population health surveys conducted or sponsored by HHS.

B. Collecting and Reporting Data on Race, Ethnicity, Sex, Primary Language and Disability Status: Data on race, ethnicity, sex, primary language, and disability status must be collected, analyzed, and reported in an objective, accurate, and useful manner. Both the collection and reporting of the data must be sensitive to constituent concerns about potential misuse or abuse. Only those demographic categories for race, ethnicity, sex, primary language and disability status, with adequate sample sizes to provide statistically reliable data, should be reported. Information on the validity and reliability of the data should be included, whenever possible, to enable the readers to judge the credibility of the findings.

C. Data Collection and Reporting Activities Covered by this Policy: This policy applies to population-based health surveys conducted or sponsored by HHS, in which respondents either self-report information or a knowledgeable proxy provides information about the person or responds for all persons in a household.

For purposes of this guidance, the terms “agency conducted or sponsored” are defined as in the Paperwork Reduction Act (PRA) implementing regulations (5 C.F.R. §1320.3(d)) and would generally include any data collection that would require OMB PRA approval.^{vii} A federal agency is considered to “conduct or sponsor” a collection of information if the agency collects the information, causes another agency to collect the information, contracts or enters into a cooperative agreement with a person to collect the information, or requires a person to provide information to another person, or in similar ways causes another agency, contractor, partner in a cooperative agreement, or person to obtain, solicit, or require the disclosure to third parties or the public of information by or for an agency. Collection of information through investigator initiated grants are generally not subject to OMB review or to this policy, except if 1) the grant recipient is conducting the data collection at the specific request of the agency or 2) the terms and conditions of the grant require specific approval by the agency of the data collection procedures (5 C.F.R. §1320.3(d)).

D. Data Standards Represent Minimum not Maximum: The data standards described herein represent minimum standards and are not intended to limit the collection of needed data. Agencies desiring more granularity or additional data are permitted and encouraged to collect additional data as needed as long as: the minimum standard is included; in the case of race and ethnicity the data can be aggregated up to the minimum standard, and the sample size supports those estimates. For the disability data standard, although survey questions and answer categories included in the standard cannot be changed, additional disability questions may be included.

E. Implementation Schedule: Beginning with the effective date of this guidance, HHS agencies are required to include the data collection standards in all HHS conducted or sponsored person-level data collected in population-based health surveys, where subjects either self-report information or a knowledgeable proxy provides information about the subject or responds for all persons in a household. Implementation will be carried out in accordance with normal agency planning, budgeting and data collection cycles. Any new survey must include the standards, and current surveys must incorporate the data standards no later than the time of the next major revision.

V. EXEMPTIONS FROM POLICY

Exemptions to this inclusion policy for data on race, ethnicity, sex, primary language, and disability status for HHS sponsored data collection activities are as follows:

1. When a data collection activity of an HHS Agency, component, or HHS-funded program is directed to one or a limited number of categories of a specific demographic variable (e.g., women), only that specific demographic variable would be excluded, but other standards would still be required. For example, if a survey specifically focuses on women, it is not necessary to collect data on sex, but data collection for race, ethnicity, primary language, and disability status is required. Similarly, an Indian Health Service survey focusing only on American Indians would not be required to include the full race data standard, but would have to include the standards for ethnicity, sex, primary language and disability status.
2. When consultation with the Agency statistician determines that the data on categories for any particular demographic group- race, ethnicity, sex, primary language, or disability status- are considered statistically unreliable, then such demographic groups should not be reported separately unless accompanied by the appropriate caveats.
3. Special exemptions may be granted on a case-by-case basis by the HHS Secretary or a designee.

VI. EFFECTIVE DATE OF POLICY: (Effective upon Secretary's signature)

VII. RELATED POLICIES

Office of Management and Budget government-wide race and ethnicity data collection standards, originally issued in 1997. http://www.whitehouse.gov/omb/fedreg_1997standards/.

HHS Inclusion Policy for Race and Ethnicity <http://aspe.hhs.gov/datacncl/inclusn.htm>

ⁱ IOM (Institute of Medicine). 2009. *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement*. Washington, DC: The National Academies Press.

ⁱⁱ Alberti, N. (2006) *The 2005 National Census Test: Analysis of the Race and Ethnicity Questions*. Final Report, 2005 National Census Test Analysis. U.S. Census Bureau

-
- ⁱⁱⁱ Office of Management and Budget. (1997a) *Recommendation from the Interagency Committee for the Review of the Racial and Ethnic Standards to the Office of Management and Budget Concerning Changes to the Standards for Classification of Federal Data on Race and Ethnicity*, Federal Register: 62: 36873-36946, July 9.
- ^{iv} Office of Management and Budget. (1997b) *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*, Federal Register: 62: No.210, October 30.
- ^v Shin, Hyon B. and R. Kominski. (2010). *Language Use in the United States: 2007*, American Community Survey Reports, ACS-12. U.S. Census Bureau, Washington, DC.
- ^{vi} Brault, M, S. Stern, D. Raglin. (2007). *Evaluation Report Covering Disability*, American Community Survey Content Test Report P.4. U.S.Census Bureau, Washington, DC.
- ^{vii} Controlling Paperwork Burdens on the Public. Code of Federal Regulations. 5 CFR Section 1320.