General Data Issues

The purpose of this document is to provide background information on a wide range of methodological and statistical issues related to Healthy People 2020 (HP2020) data. References and links to related webpages are included under each topic.

Contents

Acronyms

AAR, Age-adjusted rate

ACS, American Community Survey

AHS, Access to Health Services

BRFSS, Behavioral Risk Factor Surveillance System

BMI, Body mass index

CDC, Centers for Disease Control and Prevention

DoD, Department of Defense

DSU, Data are suppressed

GED, General Educational Development

HHS, Department of Health and Human Services

HP2020, Healthy People 2020

ICD, International Classification of Diseases

ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification

ICD-10-CM, International Classification of Diseases, Tenth Revision, Clinical Modification

LGBT, Lesbian, gay, bisexual, transgender

MSA, Metropolitan statistical area

NAMCS, National Ambulatory Medical Care Survey

NCHS, National Center for Health Statistics

NHAMCS, National Hospital Ambulatory Medical Care Survey

NHANES, National Health and Nutrition Examination Survey

NHDS, National Hospital Discharge Survey

NHIS, National Health Interview Survey

NNDSS, National Notifiable Disease Surveillance System

NSDUH, National Survey on Drug Use and Health

NSFG, National Survey of Family Growth

NVSS, National Vital Statistics System

OMB, Office of Management and Budget

SES, Socioeconomic status

USRDS, United States Renal Data System

WHO, World Health Organization

A. HP2020 Objectives: Definition of Types

HP2020 included 1,198 objectives when it launched in December 2010. Over the decade, objectives were added and archived based on emerging needs and public health recommendations. The Healthy People 2020 Midcourse Review included 1,271 objectives (1). Of these, data were available for 1,054 objectives; these objectives are said to be measurable since they have at least one data point (baseline data). 135 objectives did not yet have baseline data, even though a data source may have been identified; these objectives are said to be developmental. The remaining 82 objectives included in the Healthy People 2020 Midcourse Review have been archived.

1. Measurable Objectives

Measurable objectives have a national baseline value. The baselines use valid and reliable data generally derived from currently established and, where possible, nationally representative data systems.

a. Trackable Objectives

Trackable objectives are a subset of measurable objectives that have a target and at least two data points (i.e., data for the baseline year(s) and at least one additional point in time). Progress toward the 2020 target can be assessed only for trackable objectives.

b. Baseline-only Objectives

Baseline-only objectives are a subset of measurable objectives for which progress toward target attainment cannot be assessed because there is only a single data point.

c. Informational Objectives

Informational objectives are a subset of measurable objectives with baselines for which targets have not been established, meaning progress toward target attainment cannot be assessed. These objectives remain of public health importance and are maintained for informational purposes.

2. Developmental Objectives

Developmental objectives lack baseline data and, therefore, do not have targets and cannot be tracked for progress. However, most developmental objectives included in HP2020 have a potential data source that ultimately could provide baseline data and at least one follow-up data point.

3. **Archived Objectives**

Archived objectives are no longer being monitored due to changes in science, lack of data source, or replacement with other objectives.

B. HP2020 Objectives: Selection of Targets

All trackable objectives in HP2020 have national targets for the decade. For example, the 2020 target for the Access to Health Services (AHS) objective AHS-1.1 [Increase the proportion of persons with medical insurance] is 100 percent. HP2020 targets are proposed by the topic area workgroups, and then reviewed and approved by the Federal Interagency Workgroup for HP2020.

When sufficient data, knowledge, and expertise exist, HP2020 targets have been set through the application of science- or evidence-based methods, such as modeling, projection, or trend analysis. When circumstances do not permit the use of science- or evidence-based methods, targets generally have been set using a projected 10% improvement relative to baseline values, which was based on an analysis done by the National Center for Health Statistics (NCHS) of hypothetical Healthy People targets. The other methods used for setting HP2020 targets have included:

- Maintaining baseline level
- Maintaining consistency with national programs, regulations, policies, or laws
- Establishing minimum statistical significance
- Retaining Healthy People 2010 targets
- Achieving fixed percentage point improvement
- Achieving total elimination or total coverage

Applicable information on each HP2020 objective's baseline, target, target-setting method, data source, and revision history can be found at: https://www.healthypeople.gov/2020/topics-objectives.

C. HP2020 Data: Methodological and Statistical Issues

1. HP2020 Data Systems

DATA2020 is the HP2020 database and can be found online at:
https://www.healthypeople.gov/2020/data-search/. About 200 data sources are used in
DATA2020. A comprehensive list of HP2020 data systems, with links to online technical details, is
available on healthypeople.gov: https://www.healthypeople.gov/2020/data-search/Data-Sources

2. Measuring Variability in HP2020 Data Systems

Two main types of data systems are used to track HP2020 objectives: sample surveys and
population-count systems (also called surveillance systems). Evaluating data from both types of
systems requires consideration of variability. For sample surveys, sampling error (also called
sampling variability) is of primary interest, whereas for population-count systems, random
variation is of primary interest. Issues of data quality (e.g., item non-response, bias, and non-
representativeness) can affect data from both types of sources.

a. Measuring Variability in Sample Surveys

Much of the data used to monitor the HP2020 objectives are derived from sample surveys—
such as the National Health Interview Survey (NHIS), National Health and Nutrition
Examination Survey (NHANES), or the National Survey on Drug Use and Health (NSDUH) —
that make estimates for a population from a representative sample of respondents. These
estimates are subject to sampling error. One commonly used measure of sampling error is the
standard error, which represents the variation in an estimate that can occur by chance since
only a sample of the population is surveyed rather than the entire population. Assuming a
normal distribution of events, the chances are about 95 in 100 that the difference between
the sample estimate and the true population value would be slightly less than twice the
standard error. This concept is often referred to as a 95% confidence interval, where the
estimate is expressed as a range of the observed rates, approximately ±1.96 × standard error.

To interpret differences between rates for different population groups or changes over time
in data derived from sample surveys properly, it is important to consider the variation
associated with each rate. HP2020 data tables display detailed race and ethnicity and
socioeconomic categories for population-based objectives, when available. However, some
data are associated with relatively large standard errors. Thus, point-estimate differences
between population subgroups or between a population subgroup and the total population

may be within the expected sampling error, meaning they are not statistically significant. When available, standard errors and 95% confidence intervals are displayed in the HP2020 data tables and should be considered when evaluating progress over time or comparing population groups for objectives using survey data.

More information on the sample design and variance estimation for some of the major data systems used to monitor the HP2020 objectives can be found in related publications (2–9). Information on additional data systems can be accessed through: https://www.healthypeople.gov/2020/data-search/Data-Sources.

b.   Measuring Variability in Population-Count Systems

Some of the data systems used to track the HP2020 objectives are based on complete counts of events occurring to the population (e.g. the National Vital Statistics System [NVSS] or the United States Renal Data System [USRDS]). Therefore, these data are not subject to sampling error, although they are subject to errors in the registration process. However, when the estimates are used for analytical purposes, such as the comparison of rates over time or among groups or areas, the number of events that actually occurred may be considered as one of a large series of possible results that could have arisen under the same circumstances. This is known as random variation.

Therefore, it is desirable to compute the standard error of the rates and use that computation in the comparison of interest. Standard errors and associated confidence intervals for rates derived from population-count systems are included in HP2020 data tables when available.

More information on random variation can be found in the "technical notes" section of the annual National Vital Statistics Report (10, 11).

3.   Data Suppression

When data are not shown, one of the following symbols is shown in place of a data value:

DSU   Data are suppressed (because they do not meet the criteria for statistical reliability, data quality, or confidentiality).

---   Data are unavailable (because data for the specific population are not collected or because data have been collected but have not yet been analyzed).

There are three main reasons for why a HP2020 data value is suppressed (shown as DSU):

- The number of events is too small to produce a reliable estimate or may violate confidentiality requirements.

- The sample design does not produce representative estimates for a particular subgroup.

- There is a high item non-response rate or a large number of unknown entries.

For most objectives, an estimate or count based on a single year of data can be shown. However, for some objectives that are based on relatively few events, multiple years of data are used to produce reliable estimates. For example, objectives that are tracked using NHANES use estimates based on 2, 4, or 6 years of data. Similarly, for subpopulations that are small or that experience relatively few events, multiple years of data and/or population categories may be combined to produce reliable estimates.

D. HP2020 Data: Population Data Issues

Mortality and some HP2020 estimates of morbidity are calculated using population counts from the U.S. Census Bureau as denominators. The U.S. Census Bureau provides estimates of the total U.S. population by age, sex, race/ethnicity, state of residence, and urbanicity of county of residence. The methodologies used by the Census to collect and estimate these population counts have changed over time. Detailed information on historical, current, and future population data collection methodologies can be found at http://www.census.gov.

1. Postcensal Population Estimates

Postcensal population estimates are estimates made for the years following a census, before the next census has been taken. National population estimates for the years after the 2010 decennial census (postcensal estimates) are calculated using the 2010 census as the base population and adjusting those counts using the following measures of population change: births and deaths (provided by NCHS), immigration estimates (provided by the U.S. Citizenship and Immigration Services), data on the movement of Armed Forces personnel (from the U.S. Department of Defense [DoD]), movement between Puerto Rico and the U.S. mainland (from the Puerto Rico Planning Board), and movement of Federal employees abroad (from the Office of Personnel Management and DoD). These estimates reflect the U.S. population as of July 1 of each year. Postcensal estimates for state and county populations are also calculated using these data, as well as data

from the Internal Revenue Service and state departments of education. Each year following the decennial census, the U.S. Census Bureau produces a series of postcensal estimates that includes estimates for the current vintage data year and revised estimates for earlier years. The vintage year refers to the final year in the time series. Postcensal estimates become less accurate as the date of the estimates moves farther from the date of the census (13).

2. Intercensal Population Estimates

Intercensal population estimates are estimates made for the years between two completed censuses and take into account both censuses. After each decennial population census, intercensal estimates for the preceding decade are calculated to replace postcensal estimates. Intercensal estimates reflect the population as of July 1 of a given year and are more accurate than postcensal estimates because they include data from the enumerations at the start and end of the decade. Once data from the 2020 decennial census are released, intercensal estimates for the period 2011–2019 may be used to revise HP2020 data that is used for trend purposes within the 2030 iteration of Healthy People.

3. Population Undercounts

Some subgroups of the population (including some racial, ethnic, and age groups) are less likely than others to be completely enumerated in the decennial census. The undercounts of these groups decrease the denominators and result in overestimates of morbidity and mortality rates (14, 15). The U.S. Census Bureau makes estimates of net census undercount for the total, white, and black populations by age, and these estimates are then used to weight the populations used by most of the national health surveys, including the NHIS, NHANES, the National Survey of Family Growth (NSFG), and the National Health Care Surveys. However, the NVSS—Mortality and Natality statistics use population denominators that are not adjusted for net census undercount.

4. Population Coverage

Data for HP2020 objectives cover different "universes" or definitions of the national population, depending on the data source(s) for the individual objective. For example:

a. Resident Population
The resident population includes all persons whose usual place of residence is in one of the 50 states or the District of Columbia, including Armed Forces personnel stationed in the United States (U.S.) and institutionalized persons (e.g. prisoners). The resident population is generally the denominator used for calculating death rates from NVSS, for example.

b. Civilian Population

The civilian population is the resident population, excluding members of the U.S. Armed Forces (although their family members are included). Some HP2020 data sources, such as the National Hospital Discharge Survey (NHDS), use the civilian population as the denominator for calculating rates or proportions.

c. Civilian, Noninstitutionalized Population

The civilian, noninstitutionalized population is the civilian population not residing in institutions (e.g. correctional facilities, psychiatric hospitals, and nursing homes). This population is the denominator for rates from HP2020 data sources such as the National Ambulatory Medical Care Survey (NAMCS) and the National Hospital Ambulatory Medical Care Survey (NHAMCS). This population is also used in the weighting procedure to produce national estimates from health surveys such as the NHIS, NHANES, and NSDUH.

E. HP2020 Data: Vital Records Data Issues

The U.S. standard certificates of birth and death were revised in 2003 to improve the quality of data reported, to facilitate the collection of data needed to address coding changes resulting from the implementation of the International Classification of Diseases (ICD)-10 in 1999, and to address the modification of questions on race and Hispanic origin to reflect standards established by the Office of Management and Budget (OMB) in 1997. For additional information on ICD-10, please see Section G.

Some of the revisions to the death certificate included:

- Provision of additional space in the cause of death section to give the attending physicians more opportunity to list contributing conditions;

- Addition of a question regarding tobacco use to help reduce underreporting of tobacco use as a contributing factor to death;

- Addition of a question regarding pregnancy status of female decedents to improve the reporting of maternal deaths;

- Addition of a question to gather information about the decedent's role in transportation accidents;

- Revision of the education item from an open-ended item about years of education to check boxes of educational degree categories, consistent with U.S. Census Bureau data collection methods;

- Revision of the race and Hispanic origin questions from open-ended items to check boxes to replicate the format of the race and Hispanic origin items on the Census questionnaire (16).

Some of the revisions to the birth certificate included:

- Development of standard worksheets for the mother and the facility to enhance the collection of quality, reliable data;

- Recording cigarette usage both before and during certain intervals of pregnancy to document level of smoking;

- Addition of new items on fertility therapy, use of WIC for food during pregnancy, infections during pregnancy, breastfeeding, and principal source of payment for delivery;

- Revision of the education item from an open-ended item about years of education to check boxes of educational degree categories, consistent with U.S. Census Bureau data collection methods;

- Revision of the race and Hispanic origin questions from open-ended items to check boxes to replicate the format of the race and Hispanic origin items on the Census questionnaire (16, 17).

The transition to the 2003 standard vital certificates occurred over several years. By 2016, all states and the District of Columbia had adopted the new birth certificate. By 2017, 49 states and the District of Columbia had adopted the new death certificate and by 2018 the final state also completed the transition.

F. HP2020 Data: Age-adjustment

Age adjustment is a technique to control for differences in age composition between populations, thereby permitting comparison of their rates. In HP2020, age-adjusted rates are usually computed by the direct method, which consists of applying the age-specific rates in a population of interest to a standardized age distribution to eliminate differences in observed rates that result from age differences in the population composition (18).

1. **Overview of Methodological Issues with Age-adjustment**

Age-adjusted rates are useful for comparing two or more populations (such as racial/ethnic groups) at a point in time or a single population at two or more points in time. The actual numerical value of an age-adjusted rate is dependent upon the standard population used and should be viewed as a construct or index rather than a direct or actual measure. It is also important to note that age-adjusted rates may only be compared to rates adjusted to the *same* standard population (18).

Similarly, the actual numerical value of HP2020 targets for objectives that are measured using age-adjusted rates is dependent upon the standardization used for the target population. Consequently, HP2020 targets for age-adjusted rates do not apply to crude (unadjusted) rates such as those that are displayed for specific age groups in DATA2020.

In HP2020, many of the mortality objectives are age-adjusted, as are many of the objectives that measure health outcomes and risk factors. Age-adjusted data may be shown for objectives that target either the total population or a subgroup of the population with a large age range. Analyses of population subgroups or objectives that target groups with relatively small age ranges (generally less than 40 years) are usually not adjusted (18). Data for older adult age groups (e.g., 50 and over, 65 and over, etc.) are generally age-adjusted.

For some population groups, the age-adjusted rates are considerably different from crude (unadjusted) rates. This happens because the age distribution of the group is quite different from the age distribution of the standard population. For example, the Hispanic population has a younger age distribution than the standard population. Consequently, the age-adjusted rates for Hispanics for outcomes and behaviors that are more frequent among older people are higher than the corresponding crude (unadjusted) rates for Hispanics.

Almost all age-adjusted rates in HP2020 are adjusted using the direct method and are based on the year 2000 standard population, which was derived from the U.S. projected 2000 population. Exceptions may include data provided by Topic Area Workgroups. For example, data provided by the USRDS are adjusted by age, gender, and race based on the contemporaneous national population (19).

Objectives tracked with age-adjusted data are identified as such in the Data Details, DATA2020 data tables, and DATA2020 charts on the Healthy People 2020 website which can be found at https://healthypeople.gov/2020. Otherwise, data should be considered crude (unadjusted).

2.   **Method for Direct Age-adjustment to the 2000 U.S. Standard Population**

The age-adjusted rate (AAR) is a weighted average of age-specific rates where the age-specific weights represent the relative age distribution of a standard population. The AAR is calculated by the direct method using the following formula (19):

$$AAR = \sum_{i=1}^{I} w_{si} \times R_i ,$$

where R_i is the age-specific rate for age interval i ($i=1,...,I$), w_{si} denotes the standard weight, given by:

$$w_{si} = \frac{P_{si}}{\sum P_{si}} ,$$

P_{si} denotes population in age interval i in the standard population, $0 < w_{si} < 1$, and the weights (w_{si}) sum to one.

The 2000 U.S. standard population tables used for age-adjustment are available online at: https://www.cdc.gov/nchs/data/nvsr/nvsr68/nvsr68_09-508.pdf

The age groups used to adjust HP2020 data are shown in the Data Details on the Healthy People website (https://healthypeople.gov/2020) for each objective. The specific age groups used depend on the data system and the population targeted by the objective (21).

A number of HP2020 objectives use age-adjusted data from national health surveys such as the NHIS, the NHAMCS, the Behavioral Risk Factor Surveillance System (BRFSS), and NHANES. Data for these objectives are age-adjusted using the above equations with R_i as the age-specific rate for the health status, health behavior, or health care utilization variable, as appropriate.

In some cases, the applicable age range for the objective may not be the total population. For example, an objective may refer to persons aged 18 years and over, females aged 40 years and over, or persons aged 45–74 years. Similarly, population categories may include a different age range than the total population for that objective. For example, educational attainment for most objectives is reported only for persons aged 25 years and over. In these cases, the weights are based on the population or subpopulation in the specified age range selected for adjustment.

The age distribution of persons with chronic diseases tends to differ considerably from the standard population used for age adjustment. Using the standard age groups to age-adjust an objective with a chronic disease denominator places relatively large weights on the younger age

groups. The relatively small numbers of people with those conditions in these age groups may result in highly variable rates. When the denominator of an objective is the population of persons with a chronic disease (specifically, arthritis, cancer, chronic kidney disease, coronary heart disease, diabetes, or hypertension), the data for those under 45 years (or, in some cases, 40 years) are aggregated into a single group to stabilize the age-adjusted rates. Combining the younger age groups increases stability and reliability of estimates.

For more on age adjustment of survey data for Healthy People objectives, see (21).

G. Disease Definitions and Classification

For every measurable HP2020 objective, the Data Details page on the Healthy People website (https://healthypeople.gov/2020) provides information about how the specific disease of interest is measured and classified, including specific codes based on the data source, as described below.

1. Mortality Data

Data for HP2020 objectives that monitor specific causes of death are classified and coded according to the World Health Organization's (WHO) Tenth Revision of the International Classification of Diseases (ICD-10) (22). The ICD for mortality includes disease, injury, and poisoning categories, as well as the rules used to select the single underlying cause of death from the several diagnoses that may be reported on the death certificate. The ICD for mortality also includes definitions, tabulation lists indicating cause-of-death groupings used to present mortality data, and the format of the medical certificate of death. Use of the ICD for the classification and coding of mortality statistics is required under an agreement between the U.S. and the WHO (10).

Since 1900, the ICD for mortality has been revised approximately every 10 years, with the exception of the 20-year interval between the Ninth and Tenth revisions. The revisions are essential to ensure that disease classifications are consistent with advances in medical science and changes in diagnostic practice. The ICD-10 for mortality was implemented in the U.S. effective with deaths occurring in 1999 (23).

Data for most HP2020 mortality objectives are based on the underlying cause of death. The underlying cause of death is defined as the disease or injury that initiated the sequence of events leading directly to death or as the circumstances of the violence or accident that produced the fatal injury (22). It is selected from the conditions entered by the physician in the cause of death section on the death certificate. When more than one cause is entered by the physician, the

underlying cause is determined by the sequence of conditions on the certificate, provisions of the ICD, and associated selection rules and modifications (23). Generally, more information is reported on the death certificate than is directly reflected in the underlying cause of death. Several objectives use all mentions of a cause (or "multiple" cause) on the death certificate.

2. Morbidity Data

Baseline data for cause-specific morbidity objectives are coded to the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) (24). Specific ICD-9-CM codes used for morbidity objectives are shown on the Data Details page on the Healthy People website.

ICD-9-CM is a clinical modification of ICD-9 that includes a fifth digit, thus providing greater specificity and detail than ICD-9. ICD-9-CM is intended to serve as a tool in the classification of morbidity data for indexing of hospital medical records, medical care review, and ambulatory and other medical care programs, as well as for basic health statistics. It is used to code and classify morbidity data from inpatient and outpatient records, physicians' offices, long term care facilities and most health surveys (24).

Code assignment using ICD-9-CM is based on national coding guidelines. The guidelines for selecting the "first-listed" or principal diagnosis for morbidity records differ from those used to select the underlying cause of death on death records. Under morbidity coding rules, the principal diagnosis is that condition established after study to be chiefly responsible for the admission to the hospital or the encounter with the health care provider for care. Thus, the principal diagnosis may sometimes be an effect of the disease rather than the underlying cause. For example, if a patient with a primary malignant neoplasm with metastasis is admitted and receives treatment directed toward a secondary site, the latter could be the principal diagnosis on the hospital discharge form (25).

In general, the HP2020 morbidity objectives are tracked using the principal diagnosis. However, for some objectives "all-listed" diagnoses, which include the principal and all other (secondary) diagnoses appearing on the medical record, are used to identify cases of a particular condition.

Additional codes have been added and code changes have been made to the ICD-9-CM since its implementation in 1986. A conversion table for diagnosis and procedure code changes between 1986 and the current data year is available to assist users in data retrieval. The table shows the date the new code became effective and its previously assigned code equivalent. The latest additions to the classification appear in bold print (24).

On October 1, 2015, the U.S. Department of Health and Human Services (HHS) replaced the ICD-9-CM with a clinical modification of ICD-10 (International Classification of Diseases, Tenth Revision, Clinical Modification [ICD-10-CM]) to be used in all healthcare transactions covered by the Health Insurance Portability and Accountability Act (26).

H. Population Group Definitions

1. Age

Specific age groups are identified as population categories in the DATA2020 national data tables for some HP2020 objectives. The specific age ranges for which data are available varies by the requirements of the objective or the data source. In addition, many objectives are limited to specific age ranges. For example, some objectives focus only on adolescents and others focus only on the elderly.

Several population categories may be limited to persons of a minimum age (e.g., 25 years for educational attainment), or a maximum age (e.g. 64 years for health insurance status). The age groups that apply to these categories are specific to each objective and may vary by data source. Caution must be exercised when comparing results for age-limited groups with results for the all-ages population.

2. Sex

Data for males and females are reported separately in HP2020 when available and applicable. Sex is self-reported by respondents to surveys and other interview-based data collection systems. In many surveillance systems (e.g., NVSS), sex is usually recorded by health care professionals based on observation or written records. Sex may also be reported by proxies (e.g. the adult head of family or household may report sex of children).

In addition, in HP2020, there are a few objectives that are sex-specific, meaning they track progress toward a 2020 target only for males or females. Examples include C-7 (Reduce the prostate cancer death rate), which applies only to males; and C-4 (Reduce the death rate from cancer of the uterine cervix), which only applies to females.

3. **Race and Ethnicity**

    a. **Racial Categories**

    In 1997, OMB revised the 1977 Race and Ethnic Standards for Federal Statistics and
    Administrative Reporting (OMB Directive 15) and introduced new standards for record-
    keeping, data collection, and data presentation for race and ethnicity in federal programs (27,
    28). The 1977 standards specified four single-race categories: American Indian or Alaska
    Native, Asian or Pacific Islander, black, and white. The revised 1997 standards required that
    Federal data collection programs allow respondents to select one or more race categories
    when responding to a query on their racial identity. This provision means that there are
    potentially 31 race groups, depending on whether an individual selects one, two, three, four,
    or all five of the race categories.

    Federal agencies are required to offer respondents the option of selecting one or more of the
    following five racial categories:

    > **American Indian or Alaska Native**. A person having origins in any of the original
    > peoples of North and South America (including Central America), and who maintains
    > tribal affiliation or community attachment.

    > **Asian**. A person having origins in any of the original peoples of the Far East, Southeast
    > Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan,
    > Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

    > **Black or African American**. A person having origins in any of the black racial groups of
    > Africa.

    > **Native Hawaiian or Other Pacific Islander**. A person having origins in any of the
    > original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

    > **White**. A person having origins in any of the original peoples of Europe, the Middle East,
    > or North Africa.

    These five categories are the minimum set for data on race for Federal statistics, program
    administrative reporting, and civil rights compliance reporting. The 1997 standards were
    implemented by the U.S. Census Bureau starting with the 2000 decennial census. Other
    Federal programs were required to incorporate the new standards into household surveys,
    administrative forms and records, and other data collections by January 1, 2003.

    The 1997 standards explicitly do not include an "other race" category for data collection;
    however, an "other" category may be used for tabulating and data reporting (28). In the 2010

census, of persons self-identifying their race as "some other race," 95% also identify themselves as Hispanic; 36.7% of Hispanics selected "some other race" as their sole racial identification (29).

Collection of additional detail on race or ethnicity is permitted so long as the additional categories can be aggregated into the minimum categories. The OMB standards require that at a minimum, the total number of persons identifying with more than one race be reported when data are available. Note that the presentation of detailed information on specific racial combinations is subject to constraints of data reliability and confidentiality standards. In the 2010 census, 2.9% identified with two or more races (29).

b. Hispanic Origin

The standards regarding Hispanic origin provide for the collection of data on whether or not a person is of "Hispanic or Latino" culture or origin. This category is defined as follows:

Hispanic or Latino—A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. The term, "Spanish origin" can be used in addition to "Hispanic or Latino." Persons of Hispanic origin may be of any race and persons in the various race groups may be of any origin.

To provide flexibility and to assure data quality, the 1997 OMB guidelines recommend that a two-question format (separate race and Hispanic ethnicity questions) be used, especially when respondents can self-identify.

- When race and ethnicity are collected separately, ethnicity should be collected first. Most HP2020 data systems that use self-identification, such as the NHIS and NHANES, use the two-question format.

- When self-identification is not feasible (e.g. the National Notifiable Disease Surveillance System [NNDSS]) or when there are overriding data collection considerations (e.g., the Youth Risk Behavior Surveillance System prior to 2007), a combined race and ethnicity question can be used that includes a separate Hispanic category co-equal with the other (racial) categories. When a combined question is used, more than one entry (race and ethnicity or multiple races) is possible.

c.   Transition from the 1977 to the 1997 Standards

Over time, data systems have been revising their collection and tabulation procedures to comply with the 1997 standards. Some data systems quickly implemented the 1997 standards between 1999 and 2003, whereas others were delayed in the process of planning for or implementing the 1997 standards. Therefore, the display of race and ethnicity data in DATA2020 data tables varies across the population-based HP2020 objectives.

During the transition to full implementation of the 1997 OMB standards, two different methods for the collection of race/ethnicity data were used, creating incomparability across data systems. Further, within a given data system, the change in the race standards results in incomparability across time, thus making it difficult to perform trend analyses. The OMB recognized that approaches to make data collected under the 1997 standards comparable to data collected under the 1977 standards would be needed. Therefore, the OMB issued "Provisional Guidance on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity" (30, 31). This guidance document contains a detailed discussion of "bridging" methods. (Bridging is a mechanism for collapsing the multiple-race group population counts into single-race group counts.)

NVSS is one of the major HP2020 data sources that was delayed in implementing the 1997 OMB standards. This is because vital rates are based on information obtained from vital records collected through the state-based Vital Statistics Cooperative Program. Implementation of the 1997 OMB standards within the Vital Statistics Cooperative System started in 2003, on a state-by-state basis. As of 2016 and 2018, respectively, the race question has been revised on the birth certificates and death certificates of all states and the District of Columbia (see Section E, above). Other major data systems (e.g., the NSFG) have implemented the 1997 OMB standards, but are not reporting data for all the domains, due to lack of reliability of data for some race categories.

d. Bridged-Race Population Estimates

Starting in the 2000 decennial census, the U.S Census Bureau has used the 1997 OMB revised standards for the classification of Federal data on race and ethnicity. Thus, race data from the 2000 and 2010 census are not comparable with race data from data systems that continued to collect data using the 1977 OMB standards. For comparability, NCHS, in collaboration with the U.S. Census Bureau, developed methodology to bridge the race groups in Census 2000 and 2010 to the four single-race categories specified under the 1977 standards. Each year, NCHS has released bridged-race population estimates of the July 1 resident population of the U.S.,

for use in calculating vital rates. Even though Federal programs were to fully implement the revised standards by January 1, 2003, the transition from the 1977 to the 1997 OMB standards has been uneven. Federal systems that rely on information obtained from vital records through state-based programs, such as NVSS, did not fully implement the 1997 standards until data year 2018.

Bridged-race estimates of the July 1 resident population from the latest vintage postcensal series, and the revised 2000-2009 intercensal series are available at: https://wonder.cdc.gov/bridged-race-population.html

The bridging methodology was developed using information from the 1997–2000 NHIS. The NHIS provides a unique opportunity to investigate multiple-race groups because, since 1982, it has allowed respondents to choose more than one race but has also asked respondents reporting multiple races to choose a primary race. The bridging methodology developed by NCHS involved the application of regression models relating person-level and county-level covariates to the selection of a particular primary race by the multiple-race respondents (32). Bridging proportions derived from these models were applied by the U.S. Census Bureau to the Census 2000 Modified Race Data Summary file. This application resulted in bridged counts of the April 1, 2000 and April 1, 2010, resident single-race populations for the four racial groups specified in the 1977 OMB standards.

e. Misreporting Race/Ethnicity Data

Most health surveys and population-count systems obtain the self-reported race of the respondent, which is considered to be the most accurate representation of a person's racial or ethnic background. However, some data systems cannot collect self-reported race or ethnicity. For example, the NVSS mortality component (NVSS—M) collects information about the decedent from an informant, usually a family member. In other systems, such as those derived from hospital/patient care records, it is often unclear whether the information is self-reported. In these cases, race and ethnicity may be entered by clerical staff, hospital personnel, or someone else based on observation or the report of proxy respondents.

National Vital Statistics System—Mortality
Death rates by race and Hispanic origin may be biased from misreporting of race and Hispanic origin in the numerator of the rates and misreporting and under-coverage in the denominator of the rates (15). Numerator data are from the death certificate as reported by the funeral director based on information from an informant, usually a family member (14).

Denominator data, from surveys or the Census, are either self-reported or reported by a member of the household. Past studies comparing death certificate information with that from independent sources such as the Current Population Survey indicate that the reporting of race on the death certificate is relatively reliable for the white and black populations; however, the reporting of race for other groups and for groups with Hispanic origin may be seriously under-stated (31, 15). Additional problems, such as population undercount (see Section D.3, above), affect population censuses and estimates (33–35). For more on the validity of race and Hispanic origin reporting on death certificates, see https://www.cdc.gov/nchs/data/series/sr_02/sr02_172.pdf.

For HP2020, infant mortality rates for races and ethnic populations are based on linked files of infant deaths and live births (1, 36). These rates use the race of mother as self-reported on the birth certificate and, therefore, are not affected by the misreporting of race on the death certificate.

Patient care data

Systems that collect data from patient records, such as the NHDS, NAMCS, NHAMCS, NNDSS may also misreport the race of some individuals. It is often unclear how race and ethnicity are reported in these systems. The race and ethnicity of the patient may be reported by hospital or other medical care personnel by observation, by proxy report, or by the patient. Therefore, one must use information on race and ethnicity from these systems with caution (37).

f. Missing Data

In addition to the problems of misreporting race and ethnicity, the information on race reported by some data systems are often missing or incomplete. Some of these systems are described below.

National Ambulatory Medical Care Surveys

Race is not reported for about 10% to 32% of records in NAMCS and NHAMCS. However, missing values are imputed for both surveys (38–43).

National Notifiable Disease Surveillance System

Although staff in state health departments and the Centers for Disease Control and Prevention (CDC) attempt to obtain complete demographic information associated with nationally notifiable cases of disease, some data (particularly for the variables of race and ethnicity) are not available for some cases of disease. Laws, regulations, and mandates for public health reporting (including specific data items that are reported) fall under the

authority of individual states, and in some states, race and ethnicity may not be approved for reporting to the national level. Race and ethnicity data may also be unknown when cases are reported from a laboratory or when cases are reported as aggregate disease totals.

4. Education

Educational attainment is typically measured either by the number of years of education an adult has completed or by the highest credential received. The mutually exclusive categories for educational attainment that are generally used in HP2020 are:

Less than high school (adults with less than 12 years of schooling or no high school diploma),

High school graduate [adults with either 12 years of schooling, a high school diploma, or Certificate of General Educational Development (GED)];

At least some college (adults with a high school diploma or GED and more than 12 years of schooling, but without a college degree);

Associate's Degree (AA, etc.);

4-year college degree (BA, BS, etc.); and

Advanced degree (Master's Degree, PhD, MD).

In general, data on educational attainment are presented for adults 25 years and over, consistent with guidance given by the U.S. Census Bureau. However, the ages used to report educational attainment vary by data system, as shown in Table 1. Because of the requirements of the different data systems, the age groups used to calculate educational attainment for an objective may differ from the age groups used to report the data for other select populations and the overall measure of the same objective, as shown in the DATA2020 data tables for applicable objectives. Because of possible differences in age ranges, caution must be used in comparing the data by educational attainment with data for the main objective and other select populations.

Table 1. Ages used to report educational attainment for some HP2020 data systems

Data system	Ages used to report educational attainment
Behavioral Risk Factor Surveillance System (BRFSS)	25 years and over
National Health and Nutrition Examination Survey (NHANES)	25 years and over
National Health Interview Survey (NHIS)	25 years and over
National Survey of Family Growth (NSFG)	20–44 years
National Vital Statistics System—Natality (NVSS-N)	Of mother, 20 years and over
National Vital Statistics System—Mortality and Linked (NVSS-M and NVSS-L)	Currently suppressed (through data year 2017)
National Survey on Drug Use and Health (NSDUH)	18 years and over

Source: DATA2020, Centers for Disease Control and Prevention, National Center for Health Statistics (CDC/NCHS).

5. Income

Income is one of the most common measures of socioeconomic status (SES). Current income provides an indicator of the quality of food, housing, leisure-time amenities, and health care an individual is able to acquire, as well as reflecting the relative position in society. The use of income as a measure of SES involves more practical difficulties. In many health surveys, a substantial number of persons either do not know or refuse to report their incomes (44). As a result, statistical techniques may be used to impute missing values, as described for the 2018 NHIS in https://www.cdc.gov/nchs/data/nhis/tecdoc18.pdf.

a. Family Income

For most health surveys, income is defined as income before taxes and does not include the value of non-cash benefits such as food stamps, Medicare, Medicaid, public housing, and employer-provided fringe benefits.

Family income includes wages, salaries, rental income, interest, dividends, profits, and fees from their own businesses, pensions, and help from relatives. The definition of family income depends on a working definition of 'family'. For example, for the purpose of assessing family income, the top two major data systems in HP2020, namely NHIS and NHANES, consider that all persons within a household who are related to each other by blood, marriage, or adoption constitute a family. Thus, each member of a family is classified according to the total income of the family. Unrelated individuals are classified according to their own individual incomes. Family income data are used in the computation of the poverty threshold that is used for many HP2020 objectives.

b. Poverty Status

Converting family income to poverty status adjusts for family size and composition as well as inflation, facilitating comparisons between groups and over time. Poverty status measures family income relative to family size and composition (e.g. single parent, married couple with children) using the poverty thresholds developed by the U.S. Census Bureau, based on definitions originally developed by the Social Security Administration. These thresholds vary by family size and composition and are updated annually to reflect changes in the Consumer Price Index for all urban consumers. Multi- or single-person families with incomes below their appropriate thresholds are classified as below poverty.

For example, for a family of four, the average poverty threshold weighted for family composition was $22,025 in 2008 (the baseline year for many HP2020 objectives). Detailed poverty thresholds by year, size of family, and number of children are available on the U.S. Census Bureau website at:
http://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html

Focusing simply on the dichotomy of "above" versus "below" poverty, however, obscures the full gradient of inequalities in income distribution and the impact on health. In HP2020, additional categories are used to further separate the population above the poverty threshold. For example, the most commonly used set of family income categories in HP2020 are:

- Below the poverty threshold

- At 100–199% of the poverty threshold

- At 200–399% of the poverty threshold

- At 400–599% of the poverty threshold

- At or above 600% of the poverty threshold.

In addition to the limitations discussed for income, converting income to poverty status introduces other issues that need to be considered. If income data are collected by selecting an appropriate income category, rather than giving the actual dollar amount, then the conversion to poverty status must be performed using category means or medians and will thus result in some misclassification.

6. Disability

Disability is operationally defined in a number of different ways for program and research purposes, depending on the data collected by the data systems. In HP2020, many data sources use the American Community Survey (ACS) disability questions, also known simply as the ACS questions, to determine a person's disability status. The ACS questions were endorsed by the HHS in 2011 as a standard definition of disability to be included on HHS national surveys.

The six ACS questions used to gauge disability are:

For all persons:

- *Is this person deaf or does he/she have serious difficulty hearing?* [Yes/No]
- *Is this person blind or does he/she have serious difficulty seeing even when wearing glasses?* [Yes/No]

For persons aged 5 years or over:

- *Because of a physical, mental, or emotional condition, does this person have serious difficulty concentrating, remembering or making decisions?* [Yes/No]
- *Does this person have serious difficulty walking or climbing stairs?* [Yes/No]
- *Does this person have difficulty dressing or bathing?* [Yes/No]

For persons aged 15 years old or over:

- *Because of a physical, mental, or emotional condition does this person have difficulty doing errands alone such as visiting a doctor's office or shopping?* [Yes/No]

Respondents who answer "yes" to one or more of the questions are classified as being "at risk" of disability. Used in this way, these questions can identify the majority of persons at risk for disability, depending on their environmental context, and can be used as a stub variable to assess program participation, access to services, or other policy issues.

For some data systems, Healthy People 2020 reports limitation of activity instead of disability. Sub-categories include basic actions difficulty (a composite measure designed to capture any limitations of difficulties in movement, emotional, sensory, or cognitive functioning at the person level) and complex activity limitation (a composite measure that measures any inability to

function successfully in certain social roles). Data systems that use limitation of activity in Healthy People are labeled or footnoted appropriately in Healthy People data tables.

7. **Geographic Location**

Urban residence in HP2020 (labeled as "urban" in the DATA2020 data tables) is defined as living either within or outside a metropolitan statistical area (MSA) or within or outside the boundaries of an urbanized area and the urban portion of places outside an urbanized area that have a decennial census population of 2,500 or more.

The U.S. Census Bureau (1) uses published criteria to determine the qualification and boundaries of urbanized areas. An urbanized area is an area consisting of a central place(s) and adjacent urban fringe that together have a minimum residential population of at least 50,000 people and generally an overall population density of at least 1,000 people per square mile of land area (45).

The U.S. Office of Management and Budget revises the MSA standards before each decennial census. When census data become available, the standards are applied to define the actual MSAs. An MSA is a county or group of contiguous counties that contains at least one urbanized area of 50,000 or more population. In addition to the county or counties that contain all or part of the urbanized area, an MSA may contain other counties that are metropolitan in character and that are economically and socially integrated with the main city. In New England, cities and towns, rather than counties, are used to define MSAs. Counties that are not within an MSA are considered nonmetropolitan. For current and historical definitions of MSAs, see (45).

8. **Health Insurance**

The health insurance status population category applies only to persons under age 65 years, because most persons 65 years and over are eligible for Medicare. In the NHIS, respondents are identified as having health insurance if they are covered by either private or public health plans. Private insurance includes fee-for-service plans, single service hospital plans, and coverage by health maintenance organizations. Public insurance includes Medicare, Medicaid, Children's Health Insurance Program, state-sponsored or other government-sponsored health plan, or military plans. Respondents are defined as being uninsured if they had only Indian Health Service coverage or had only a private plan that paid for one type of service, such as accidents or dental care (46). HP2020 data systems other than NHIS define health insurance status similarly.

9.   **Sexual Orientation**

HP2020 includes a Topic Area on lesbian, gay, bisexual, and transgender (LGBT) population health, with an overall goal of improving the health, safety, and well-being of LGBT individuals. See the Healthy People website for an overview.

Data by sexual orientation are currently available in DATA2020 for objectives from the NHIS across multiple topic areas, and NSFG for objectives in the Family Planning; Maternal, Infant, and, Child Health; and Sexually Transmitted Diseases topic areas. NCHS continues to investigate cognitively meaningful and statistically reliable techniques for collecting data on sexual orientation and gender identity (47). For a comparison of sexual orientation estimates based on NHIS, NHANES, and NSFG data, see https://www.cdc.gov/nchs/data/series/sr_02/sr02_169.pdf Development, testing, and implementation of questions on gender identity in the NHIS are ongoing (48). New or revised questions aim to reduce misclassification, misinterpretation, and nonresponse. For more information, see https://www.cdc.gov/nchs/nhis/sexual_orientation/faqs.htm

10. **Veteran Status**

Veteran status is specified as Veteran and non-Veteran. Veterans include those who identify themselves as ever being honorably discharged from active duty in the U.S. Army, Navy, Air Force, Marine Corps, or Coast Guard. Veteran status data are available for selected objectives in the following HP2020 Topic Areas: Arthritis, Osteoporosis, and Chronic Back Conditions; Chronic Kidney Disease; Diabetes; Heart Disease and Stroke; Mental Health and Mental Disorders; Nutrition and Weight Status; Older Adults; Oral Health; and Substance Abuse.

11. **Obesity Status**

Some HP2020 objectives may specifically track populations of persons who are overweight or obese. Alternatively, categories representing overweight or obese subgroups may be included in the DATA2020 national data tables for more general population objectives. For adults, obesity is determined by using weight and height to calculate a body mass index (BMI). BMI is a measure that adjusts body weight for height. The BMI is calculated as weight in kilograms divided by height in meters squared. Obesity is defined as a BMI of 30 or more. For youth, obesity is defined as BMI at or above the sex-and age-specific 95th percentile of the 2000 CDC Growth Charts. Depending on the data system, obesity status may be based either on directly measured or self-reported height and weight. BMI estimates derived from self-reported height and weight tend to be lower than those derived from measured height and weight due to underreporting of weight and

overreporting of height (49). However, self-reported data remain useful in making comparisons by obesity status.

I. Geographic Coverage

1. National Data

Generally, HP2020 data are based on events occurring in the 50 states and the District of Columbia. Unless specifically noted, data for U.S. territories are not included. Most population-based data in HP2020 are from: a national census of events, such as the NVSS or NNDSS; a nationally representative sample survey, such as the NHIS or NSDUH; or a nationally representative catchment area, such as the National Antimicrobial Resistance Monitoring System.

For some objectives, however, complete national data are not available, and data for selected states and/or areas are used to monitor the objectives in lieu of nationally representative data. In these cases, the coverage area (e.g. number of states in the reporting area) is described with the data for the objective and in the Data Details. Examples of these data systems include the Surveillance, Epidemiology, and End Results Program and the Healthcare Cost and Utilization Project.

For some national data systems that cover the entire U.S., such as the NVSS and BRFSS, some data are not available for all states. This is either because data for a specific variable are not collected by some states or because the quality of data for some states is not sufficient to produce reliable estimates for some variables. For example, in the BRFSS, the number of reporting states for some variables can vary from year to year. For example, the data for HP2020 Diabetes objective D-13 (to increase the proportion of adults with diabetes who perform self-blood glucose-monitoring at least once daily) were based on 33 states in 2012, the baseline year, whereas the 2015 data for that objective were based on 38 states. This information is included in the Data Details for all relevant objectives.

2. State and Local Data

For three decades, Healthy People has established benchmarks and monitored progress over time to encourage collaborations across communities and sectors, empower individuals toward making informed health decisions, and measure the impact of prevention activities. For more about the impact of Healthy People at the state, territorial, and local level, see https://healthypeople.gov/2020.

Important data issues arise when comparing Healthy People objectives' progress between states, or between a state and the nation. To ensure that federal, state, tribal, and local health agencies have the necessary infrastructure to effectively provide essential public health services, HP2020 includes a separate Topic Area, Public Health Infrastructure, aimed at improving infrastructure and surveillance capability; see https://healthypeople.gov/2020.

a.   Objective Wording—Definitions

Many state and local agencies and organizations have tailored the HP2020 objectives to better focus on specific concerns of their constituents. These modifications may more effectively address the health concerns of the state or local population, but they also reduce comparability when evaluating objective progress relative to the nation, other states, or localities. Examples include EH-14 (increase the proportion of homes with an operating radon mitigation system for persons living in homes at risk for radon exposure) and EH-15 (increase the proportion of single-family homes constructed with radon-reducing features, especially in high-radon-potential areas). Some states have adopted these objectives verbatim, whereas others have included schools or day care centers within the same objective. In addition, the technical definitions and data sources for the same or similar national and state or local objectives may vary considerably.

b.   Population Data—Race and Ethnicity Reporting

Many Healthy People objectives are population based and are expressed as rates (e.g. C-2, to reduce lung cancer deaths per 100,000), where the denominator is a population estimate produced by the U.S. Census Bureau. The Census Bureau provides population counts from the decennial census for the nation, states, counties, and large municipalities. It also produces annual postcensal estimates for the years following the census. These estimates are available by sex, age, and race and ethnicity. However, the sizes of some racial groups (e.g. American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander) are relatively small, even at the national level, and are distributed unevenly across state and local areas. This precludes many jurisdictions from producing reliable rates for objectives that focus on these populations.

c.   Data Sources for State and Local Data

The availability and comparability of data for national, state, and local monitoring of Healthy People objectives vary considerably. Some data, especially vital statistics, are readily

available at all geographic levels. The standardization of vital statistics data contributes to their comparability across jurisdictions.

Many Healthy People objectives are monitored using data from the NHIS. Some of these objectives are also monitored at state and local levels using data from the BRFSS. In general, it should be noted that differences in the data collection methods (e.g. in-person vs. telephone interview), differences in response rates, and wording of questions used to monitor the same objectives at the national versus the state level can affect the comparability of the information collected. In addition, even objectives that are monitored using identical questions in the NHIS and the BRFSS may be included only periodically, either in specific rotating modules of the BRFSS (e.g. D-13, to increase the proportion of adults with diabetes who perform self-blood glucose-monitoring at least once daily) or in supplements to the NHIS (e.g. HDS-6, to increase the proportion of adults who have had their blood cholesterol checked within the preceding 5 years). Not all states use these BRFSS modules in any given year. Further, the year of the module rotation may not coincide with national data from the NHIS, which limits comparability between national and state data.

d.   Confidentiality of State and Local Data

Some HP2020 objectives address important and sensitive health issues (e.g. deaths from HIV) that are relatively rare events. Reporting event counts or rates by certain characteristics or geocoding and displaying maps of the distribution of sensitive or rare events may jeopardize privacy of affected individuals or confidentiality of the source data. It is often necessary to aggregate data from smaller to larger geographic areas, demographic characteristics, and/or data years to protect confidentiality. In some cases, state or local data may be suppressed to protect confidentiality.

References

1. National Center for Health Statistics. Healthy People 2020 Midcourse Review. Hyattsville, MD: National Center for Health Statistics. 2016. Available from: https://www.cdc.gov/nchs/healthy_people/hp2020/hp2020_midcourse_review.htm

2. Chowdhury SR, Machlin SR, Gwet KL. Sample designs of the Medical Expenditure Panel Survey Household Component, 1996–2006 and 2007–2016. Agency for Healthcare Research and Quality. MEPS Methodology Report (33). 2019. Available from: https://meps.ahrq.gov/data_files/publications/mr33/mr33.pdf

3. Johnson CL, Dohrmann SM, Burt VL, Mohadjer LK. National Health and Nutrition Examination Survey: Sample design, 2011–2014. National Center for Health Statistics. Vital Health Stat 2(162). 2014. Available from: https://www.cdc.gov/nchs/data/series/sr_02/sr02_162.pdf

4. Parsons VL, Moriarity C, Jonas K, Moore TF, Davis KE, Tompkins L. Design and estimation for the National Health Interview Survey, 2006–2015. National Center for Health Statistics. Vital Health Stat 2(165). 2014. Available from: https://www.cdc.gov/nchs/data/series/sr_02/sr02_165.pdf

5. Dennison C, Pokras R. Design and operation of the National Hospital Discharge Survey: 1988 redesign. National Center for Health Statistics. Vital Health Stat 1(39). 2000. Available from: https://www.cdc.gov/nchs/data/series/sr_01/sr01_039.pdf

6. Wolter KM, Smith PJ, Khare M, Welch B, Copeland KR, Pineau VJ, Davis N. Statistical methodology of the National Immunization Survey, 2005–2014. National Center for Health Statistics. Vital Health Stat 1(61). 2017. Available from: https://www.cdc.gov/nchs/data/series/sr_01/sr01_061.pdf

7. Lepkowski JM, Mosher WD, Groves RM, West BT, Wagner J, Gu H. Responsive design, weighting, and variance estimation in the 2006–2010 National Survey of Family Growth. National Center for Health Statistics. Vital Health Stat 2(158). 2013. Available from: https://www.cdc.gov/nchs/data/series/sr_02/sr02_158.pdf

8. Kyle TM, Brener ND, Kann L, Ross JG, Roberts AM, Iachan R, et al. Methods: School Health Policies and Programs Study 2006. J Sch Health 77(8):398-407. 2007.

9. Centers for Disease Control and Prevention. Methodology of the Youth Risk Behavior Surveillance System – 2013. MMWR Morb Mortal Wkly Rep 2013; 62 (No. RR—1). Available from: https://www.cdc.gov/mmwr/pdf/rr/rr6201.pdf

10. Kochanek KD, Murphy SL, Xu JQ, Arias E. Deaths: Final data for 2017. National Vital Statistics Reports; vol 68 no 9. Hyattsville, MD: National Center for Health Statistics. 2019. Available from: https://www.cdc.gov/nchs/data/nvsr/nvsr68/nvsr68_09-508.pdf

11. Martin JA, Hamilton BE, Osterman MJK, Driscoll AK. Births: Final data for 2018. National Vital Statistics Reports; vol 68 no 13. Hyattsville, MD: National Center for Health Statistics. 2019. Available from: https://www.cdc.gov/nchs/data/nvsr/nvsr68/nvsr68_13-508.pdf

12. Parker JD, Talih M, Malec DJ, et al. National Center for Health Statistics data presentation standards for proportions. National Center for Health Statistics. Vital Health Stat 2(175). 2017. Available from: https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf

13. U.S. Census Bureau. 1990-2000 Intercensal Estimates: Methodology. Washington, DC. 2016. Available from: https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/intercensal/intercensal-nat-meth.pdf

14. National Center for Health Statistics. Vital statistics of the United States: Mortality, 1999. Technical appendix. Hyattsville, MD. 2004. Available from: https://www.cdc.gov/nchs/data/statab/techap99.pdf

15. Rosenberg HM, Maurer JD, Sorlie PD, MacDorman MF, Hoyert DL, Spitler JF, Scott C. Quality of death rates by race and Hispanic origin: A summary of current research, 1999. National Center for Health Statistics. Vital Health Stat 2(128). 1999. Available from: https://www.cdc.gov/nchs/data/series/sr_02/sr02_128.pdf

16. National Center for Health Statistics. Report of the Panel to Evaluate the U.S. Standard Certificates. Hyattsville, MD: National Center for Health Statistics, 2000. Available from: https://www.cdc.gov/nchs/data/dvs/panelreport_acc.pdf

17. National Research Council (US) Committee on National Statistics. 4. Methodological Issues and the 2003 Revision of Standard Instruments. In: Vital statistics: Summary of a workshop. Washington, DC: National Academies Press; 2009. Available from: https://www.ncbi.nlm.nih.gov/books/NBK219875/

18. Siegel JS, Swanson DA. The methods and materials of demography, 2nd ed. San Diego, CA: Elsevier Academic Press, 2004.

19. United States Renal Data System. 2020 USRDS annual data report: Epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD. 2018. Available from: https://adr.usrds.org/2020

20. Curtin LR, Klein, RJ. Direct standardization (age-adjusted death rates). Healthy People Statistical Notes, no. 6. Hyattsville, MD: National Center for Health Statistics. 1995. Available from: https://www.cdc.gov/nchs/data/statnt/statnt06rv.pdf

21. Klein RJ, Schoenborn CA. Age adjustment using the 2000 projected U.S. population. Healthy People Statistical Notes, no 20. Hyattsville, MD: National Center for Health Statistics. January 2001. Available from: https://www.cdc.gov/nchs/data/statnt/statnt20.pdf

22. World Health Organization. International statistical classification of diseases and related health problems, tenth revision (ICD–10). 2nd ed. Geneva, Switzerland. 2004. Available from: https://apps.who.int/iris/handle/10665/42980

23. Hoyert DL, Arias E, Smith BL, Murphy SL, Kockanek KD. Deaths: Final Data for 1999. National Vital Statistics Reports; vol 49 no 8. Hyattsville, MD: National Center for Health Statistics. 2001. Available from: https://www.cdc.gov/nchs/data/nvsr/nvsr49/nvsr49_08.pdf

24. U.S. Department of Health and Human Services. International classification of diseases, 9th revision, clinical modifications. 6th ed. Washington, D.C.: Health Care Financing Administration, 1997. DHHS publication No. (PHS) 96-1260.

25. National Center for Health Statistics. Conversion table of new ICD-9-CM codes. Hyattsville, MD: National Center for Health Statistics, 2013. Available from: https://www.cdc.gov/nchs/data/icd/icd-9-cm_fy14_cnvtbl_final.pdf

26. National Center for Health Statistics. International classification of diseases, 10th revision, clinical modification (ICD–10–CM). Hyattsville, MD: National Center for Health Statistics, 2011. Available from: https://www.cdc.gov/nchs/icd/icd10cm.htm

27. Office of Management and Budget. Revisions to the standards for the classification of federal data on race and ethnicity. Fed Regist 62(210):58782–90. 1997.

28. Office of Management and Budget. Race and ethnic standards for Federal statistics and administrative reporting. Statistical Policy Directive 15. 1997. Available from: https://wonder.cdc.gov/WONDER/help/populations/bridged-race/Directive15.html

29. Humes KR, Jones NA, Ramirez RR. Overview of race and Hispanic origin: 2010. Washington, D.C.: U.S. Census Bureau, 2011. 2010 Census Briefs. Available from: https://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf

30. Office of Management and Budget. Provisional guidance on the implementation of the 1997 Standards for Federal Data on Race and Ethnicity. Washington, D.C.: U.S. Office of Management and Budget, 2000. Available from: https://www.ofm.wa.gov/sites/default/files/public/dataresearch/pop/asr/re_guidance2000update.pdf

31. Sorlie PD, Rogot E, Johnson NJ. Validity of demographic characteristics on the death certificate. Epidemiol 3(2):181–184. 1992.

32. Ingram DD, Parker JD, Schenker N, Weed JA, Hamilton B, Arias E, Madams JH. United States Census 2000 population with bridged race categories. National Center for Health Statistics. Vital Health Stat 2(135). 2003. Available from: https://wonder.cdc.gov/wonder/help/populations/bridged-race/VitalHealthStatistics-Series2No135.pdf

33. Mule T. 2010 Census coverage measurement estimation report: Summary of estimates of coverage for persons in the United States. DSSD 2010 census coverage measurement memorandum series #2010–G–01. Washington, DC: U.S. Census Bureau. 2012 Available from: https://www2.census.gov/programs-surveys/decennial/2010/technical-documentation/methodology/g-series/g01.pdf

34. Arias E, Schauman WS, Eschbach K, Sorlie, PD, Backlund E. The validity of race and Hispanic origin reporting on death certificates in the United States. National Center for Health Statistics. Vital Health Stat 2(148). 2008. Available from: https://www.cdc.gov/nchs/data/series/sr_02/sr02_148.pdf

35. Arias E, Heron M, Hakes JK. The validity of race and Hispanic-origin reporting on death certificates in the United States: An update. National Center for Health Statistics. Vital Health Stat 2(172). 2016. Available from: https://www.cdc.gov/nchs/data/series/sr_02/sr02_172.pdf

36. MacDorman MF, Atkinson JO. Infant mortality statistics from the 1996 period linked birth/infant death dataset. 1998. Monthly Vital Statistics Report, Vol. 46, No. 12. supp. Available from: https://www.cdc.gov/nchs/data/mvsr/supp/mv46_12s.pdf

37. National Research Council. Eliminating health disparities: Measurement and data needs. Washington, D.C.: The National Academies Press, 2004. Committee on National Statistics, Division of Behavioral and Social Sciences and Education.

38. McCaig LF, Nawar EW. National Hospital Ambulatory Medical Care Survey: 2004 emergency department Summary. Hyattsville, MD: National Center for Health Statistics, 2006. Advance data from vital and health statistics, No. 372. Available from: https://www.cdc.gov/nchs/data/ad/ad372.pdf

39. Middleton KR, Hing E. National Hospital Ambulatory Medical Care Survey: 2004 outpatient department summary. Hyattsville, MD: National Center for Health Statistics, 2006. Advance data from vital and health statistics, No. 373. https://www.cdc.gov/nchs/data/ad/ad373.pdf

40. Hing E, Cherry DK, Woodwell DA. National Ambulatory Medical Care Survey: 2004 summary. Hyattsville, MD: National Center for Health Statistics, 2006. Advance data from vital and health statistics, No. 374. Available from: https://www.cdc.gov/nchs/data/ad/ad374.pdf

41. Niska R, Bhuiya F, Xu J. National Hospital Ambulatory Medical Care Survey: 2007 emergency department summary. National health statistics reports; no 26. Hyattsville, MD: National Center for Health Statistics. 2010. Available from: https://www.cdc.gov/nchs/data/nhsr/nhsr026.pdf

42. Hsiao CJ, Cherry DK, Beatty PC, Rechtsteiner EA. National Ambulatory Medical Care Survey: 2007 summary. National health statistics reports; no 27. Hyattsville, MD: National Center for Health Statistics. 2010. Available from: https://www.cdc.gov/nchs/data/nhsr/nhsr027.pdf

43. Hing E, Hall MJ, Ashman JJ, Xu J. National Hospital Ambulatory Medical Care Survey: 2007 Outpatient Department Summary. National health statistics reports; no 28. Hyattsville, MD: National Center for Health Statistics. 2010. Available from: https://www.cdc.gov/nchs/data/nhsr/nhsr028.pdf

44. Pamuk E, Makuc, D, Heck K, Reuben C, Lochner K. Socioeconomic Status and Health Chartbook. Health, United States, 1998. Hyattsville, MD. National Center for Health Statistics. 1998. Available from: https://www.cdc.gov/nchs/data/hus/hus98cht.pdf

45. U.S. Census Bureau. Metropolitan and Micropolitan. Washington, DC. Available from: https://www.census.gov/programs-surveys/metro-micro/about.html

46. Cohen RA, Terlizzi EP, Martinez ME. Health insurance coverage: Early release of estimates from the National Health Interview Survey, 2018. National Center for Health Statistics. May 2019. Available from: https://www.cdc.gov/nchs/data/nhis/earlyrelease/insur201905.pdf

47. Dahlhamer JM, Galinsky AM, Joestl SS, Ward BW. Sexual orientation in the 2013 National Health Interview Survey: A quality assessment. Vital Health Stat 2(169). 2014. Available from: https://www.cdc.gov/nchs/data/series/sr_02/sr02_169.pdf

48. Elliott MN, Dahlhamer JM, MacCarthy S, Beckett MK, Orr N, Guerino P, Agniel D, Saunders CL, Schuster MA, Ng JH, Martino SC. Using ancillary sociodemographic data to identify sexual minority adults among those responding "Something Else" or "Don't Know" to sexual orientation questions. Med Care. 2019 Dec;57(12):e87-e95. Available from: https://journals.lww.com/lww-medicalcare/Fulltext/2019/12000/Using_Ancillary_Sociodemographic_Data_to_Identify.17.aspx

49. Merrill RM, Richardson JS. Validity of self-reported height, weight, and body mass index: Findings from the National Health and Nutrition Examination Survey 2001–2006. Prev Chronic Dis 2009;6(4):A121. Available from: https://www.cdc.gov/pcd/issues/2009/oct/pdf/08_0229.pdf