

# Race and Ethnicity Code Set Version 1.0

Prepared by the  
Centers for Disease Control and Prevention  
March 2000

## *Background and Purpose*

The U.S. Centers for Disease Control and Prevention (CDC) has prepared a code set for use in coding race and ethnicity data. This code set is based on current federal standards for classifying data on race and ethnicity, specifically the minimum race and ethnicity categories defined by the U.S. Office of Management and Budget (OMB) and a more detailed set of race and ethnicity categories maintained by the U.S. Bureau of the Census (BC). The main purpose of the code set is to facilitate use of federal standards for classifying data on race and ethnicity when these data are exchanged, stored, retrieved, or analyzed in electronic form. At the same time, the code set can be applied to paper-based record systems to the extent that these systems are used to collect, maintain, and report data on race and ethnicity in accordance with current federal standards.

### *Race and Ethnicity Concepts*

The code set consists of two tables: (1) Race and (2) Ethnicity. Concepts in the Race and Ethnicity tables include the OMB minimum categories, 5 races and 2 ethnicities, along with a sixth race category, Other race, and a more detailed set of race and ethnicity categories used by the BC. The OMB minimum categories for data on race are American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White. There are two OMB minimum categories for data on ethnicity: Hispanic or Latino, and Not Hispanic or Latino. More information on the OMB minimum categories and federal standards for classifying data on race and ethnicity is available at <http://www.whitehouse.gov/omb/fedreg/ombdir15.html>

The categories used by the BC adhere closely to the basic classification framework provided by the OMB standards, with the notable exception of a sixth race category, Other Race, used to collect and tabulate decennial census data. The BC also collects more detailed race and ethnicity data in its decennial census by asking respondents to self-identify their race and ethnicity in a series of categorical questions and free text data fields. The BC's summarization and categorization of these responses provide a detailed set of race and ethnicity concepts, grouped in accordance with the OMB standards. The detailed concepts in the Race and Ethnicity tables are based on the BC's summarization and categorization of responses to the March 1998 dress rehearsal for the year 2000 decennial census.

### *Code Set Specifications*

Within each table, discrete concepts, e.g., Asian, Hispanic or Latino, Arapaho, are specified by a unique identifier, hierarchical code, synonym (if any), date added to version (i.e., code set version), and date removed from version. The unique identifier is the permanently assigned numeric code value intended for use in electronic interchange of coded race and ethnicity data. A check digit, calculated using the Mod 10 check digit scheme (see appendix), is appended at the right of the unique identifier; it is a single digit separated from the preceding digits by a dash. The hierarchical code is an alphanumeric code that places each discrete concept in a hierarchical

position vis a vis other related concepts, e.g., Costa Rican, Guatemalan, and Honduran are each ethnicity concepts whose hierarchical codes place them at the same level vis a vis the concept Central American, which in turn is at the same hierarchical level as Spaniard vis a vis the broader concept Hispanic or Latino. The letter in the first position of the hierarchical code, R or E, indicates the location of the concept in one of two hierarchies: Race or Ethnicity. The length of the hierarchical code also conveys information about the hierarchical position of the associated concept vis a vis other concepts.

In contrast to the unique identifier, the hierarchical code can change over time to accommodate the insertion of new concepts or to represent changes in the hierarchical organization of concepts. When a new concept is inserted in the code set, the date of insertion is specified as the date added to version. When a decision is made that a concept is obsolete, this date is noted as the date removed from version. However, obsolete concepts and their unique identifiers will remain visible in the code set tables to allow accurate interpretation of coded data.

#### *Code Set Maintenance*

The Race and Ethnicity Code Set introduced here as Version 1.0 will require maintenance over time. The specified race and ethnicity concepts do not provide comprehensive coverage of all race and ethnicity categories. Additions are needed. These additions need to be made in accordance with the basic framework and guiding principles used to create the initial version of the code set. Further, as race and ethnicity concepts evolve over time, changes will be needed to reflect contemporary understanding, needs, and usage. The initial version of the Race and Ethnicity code set is intended to serve as a starting place. The next version will include changes based on the BC's summarization and categorization of race and ethnicity responses to the 2000 decennial census.

## *Appendix - Calculating Mod 10 Check Digits*

The algorithm for calculating Mod 10 check digits, using the identifier 1234 as an example, is as follows:

<i>Steps</i>	<i>Example</i>
(1) Take the odd digit positions of the identifier counting from the right	42
(2) Multiply the number from step (1) by 2	84
(3) Take the even digit positions of the identifier counting from right	31
(4) Prepend the number from step (3) to the number from step (2)	3184
(5) Add all digits of the number from step (4) together	16
(6) Using the number from step (5), find the next highest multiple of 10	20
(7) The check digit is the number from step (6) minus the number from step (5)	4

Thus, in the example, 4 is the Mod 10 check digit for 1234