

# **NCHS Procedures for Multiple-Race and Hispanic Origin Data: Collection, Coding, Editing, and Transmitting**

Division of Vital Statistics  
National Center for Health Statistics  
Centers for Disease Control and Prevention  
May 7, 2004

As of January 1, 2003, Federal programs were required by the U.S. Office of Management and Budget to adopt revised standards for collecting and reporting racial and ethnic status. These standards were published in the Federal Register on October 30, 1997, as “Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity.” The notice is posted on the OMB web-site at:  
<http://www.whitehouse.gov/omb/fedreg/ombdir15.html>

The U.S. Census Bureau was one of the first federal agencies to implement the revised standards, incorporating in the 2000 Decennial Census a format for the race question that included 15 checkbox items and 3 write-in lines, plus the instruction to “Mark one or more races to indicate what this person considers himself/herself to be.” Subsequently, the Panel to Evaluate the U.S. Standard Certificates recommended that the revised certificates should have race and Hispanic origin questions nearly identical to those in the 2000 census in order to maintain comparability of the data collected in census and vital statistics. The Panel’s report can be found at the following web-site:  
[http://www.cdc.gov/nchs/data/dvs/panelreport\\_acc.pdf](http://www.cdc.gov/nchs/data/dvs/panelreport_acc.pdf)

The revised standard certificates, with the revised race and Hispanic origin formats, may be found by going to the following web-site, where the data collection, transmission, edit, and file layout specifications are also posted:  
[http://www.cdc.gov/nchs/vital\\_certs\\_rev.htm](http://www.cdc.gov/nchs/vital_certs_rev.htm).

To facilitate coding and processing of multiple-race/Hispanic-origin data in a uniform manner for all vital statistics jurisdictions, NCHS has developed a computer system to code (with minimal manual intervention) and edit reported data. For any jurisdiction that collects and transmits multiple-race/Hispanic-origin data to NCHS, those data will be coded, edited, and the results returned to the jurisdiction for its use. The system is fairly flexible and has the ability to receive and process data from race question formats that differ from the standard. Jurisdictions with variants of the race/origin questions must contact NCHS staff to work out the necessary details for transmission of data.

The system receives multiple-race and Hispanic-origin data in a file layout that reflects the question format specified in the revised U.S. Standard Certificates, which is essentially the same format that was used in the 2000 decennial census forms (see Edit Specifications in the web-site: [http://www.cdc.gov/nchs/vital\\_certs\\_rev.htm](http://www.cdc.gov/nchs/vital_certs_rev.htm)). The system embodies a 3-digit code structure and an edit process analogous to those used by the U.S. Census Bureau to code and edit the 2000 decennial census data. The NCHS code lists for race and origin are accessible in PDF at the following web-sites:  
<http://www.cdc.gov/nchs/data/dvs/RaceCodeList.pdf> for race and  
<http://www.cdc.gov/nchs/data/dvs/HispanicCodeTitles.pdf> for Hispanic origin.

Using these codes the edit program does some sorting (eliminating redundancies and resolving inconsistencies) and produces a cohesive set of race codes (up to 8), which may be used to tally multiple-race data after some further recoding. The system also will bridge multiple-race data into the single race format of the prior OMB race standard (four races only). A description of bridged race data from the 2000 census for counties, states, and nation is accessible at the following NCHS web site:

<http://www.cdc.gov/nchs/about/major/dvs/popbridge/popbridge.htm>,

which includes a report describing the bridging algorithm, including its development and characteristics. Essentially this same bridging process is being applied to the vital records race data within the NCHS edit program. Appended to this document is a PowerPoint presentation describing the bridging concept in more detail.

The NCHS procedure is to receive multiple-race and Hispanic-origin source data from a jurisdiction and then return the coded and edited data back to the jurisdiction. NCHS is building a coding dictionary around the NCHS 3-digit code list by inserting various misspellings, abbreviations, etc. to automate the coding of literals as much as possible. Our goal is to eventually make the coding and editing algorithm available interactively on the web, so states can submit race and Hispanic origin data to the program and receive back the edited results on a real-time basis.

As for presenting race data in NCHS publications in the near future, most national tabulations will use bridged race in place of multiple race as long as some jurisdictions are using the old race standard. Moreover, NCHS will continue to use bridged-race population estimates for denominators to calculate rates.

Appended to this document are the record formats of coded data that will be returned to each jurisdiction; there is a listing for each certificate type. The record formats include both the old standard and new standard layouts. By default, the returning coded data will be placed in the jurisdiction's download area on the Secure Data Network. The returned file will have a new file extension: Mortality (MRE), Natality (NRE), and Fetal Deaths (FRE). The file name will follow a format similar to the NCHS standard data file name. The first eight characters will define the State, data year, a unique process control number (PCN), and the letter 'S'. As an example, the 12<sup>th</sup> mortality file received from California in 2003 would be named CA03012S.MRE.

As can be seen in the file layouts for returning data appended to this document, the multiple-race file format has been expanded to include fields for both the NCHS and FIPS geographic codes. Regarding the Hispanic origin question, the Hispanic checkbox fields (\_ETHNIC1...5) will be blank for jurisdictions collecting multiple-race data using certificate formats with the old Hispanic origin format. Instead, the old standard Hispanic coded response will be shown in \_ETHNICO, a corresponding 3-digit code from the new Hispanic code list will be shown in the \_ETHNICE field, and the \_ETHNIC5C coded literal field will be blank.

The 2-digit Bridged Multiple Race code (RACEBRG) is a recode indicating either the single race reported (codes 01 to 15) or the bridged race—specific to the old race standard—for multiple races reported (21 to 24), as follows:

Bridged Multiple Race code structure (RACEBRG):

Single race specified:

- 01 ... White
- 02 ... Black
- 03 ... American Indian or Alaskan Native
- 04 ... Asian Indian
- 05 ... Chinese
- 06 ... Filipino
- 07 ... Japanese
- 08 ... Korean
- 09 ... Vietnamese
- 10 ... Other Asian
- 11 ... Native Hawaiian
- 12 ... Guamanian or Chamorro
- 13 ... Samoan
- 14 ... Other Pacific Islander
- 15 ... Other

Bridged multiple specified race:

- 21 ... White
- 22 ... Black
- 23 ... American Indian or Alaskan Native
- 24 ... Asian or Pacific Islander

The following three slides summarize the 3-digit race code structure employed by NCHS to code and then edit multiple-race data:

## NCHS/DVS Race Code List...

- 100 = White Checkbox
- 101-199 = Specific White responses
  - e.g., 107 = Italian
- 200 = Black Checkbox
- 201-299 = Specific Black responses
  - e.g., 219 = Nigerian
- 300 = American Indian & Alaska Native Checkbox
- A01-R96 = Specific Indian tribes (~900)
  - e.g. E31 = Oklahoma Kiowa,
  - N69 = Greenland Eskimo

## NCHS/DVS Race Code List, cont.

400 = Asian Indian Checkbox  
401-409 = Spec. origin (e.g. 402 = Bangladeshi)  
410 = Chinese Checkbox  
411-419 = Spec. origin (e.g. 412 = Taiwanese)  
420 = Filipino Checkbox  
421-429 = Spec. origin (e.g. 422 = Hmong)  
430 = Japanese Checkbox  
440 = Korean Checkbox  
450 = Vietnamese Checkbox

## NCHS/DVS Race Code List, concluded

460 = Other Asian Checkbox  
461-499 = Specified origin (e.g. 472 = Nepalese)  
500 = Native Hawaiian Checkbox  
510 = Samoan Checkbox  
520 = Guamanian Checkbox  
530 = Other Pacific Islander Checkbox  
531-599 = Spec. origin (e.g. 531 = Mariana  
Islander)  
600 = Other Race Checkbox  
601-995 = Spec. origin, e.g. 616 = Mestizo, 617 =  
Mexican,

The following slides display the various steps in the NCHS coding, editing, and bridging procedure for multiple race data, as applied to four examples. In the first slide of Example 1, source data are displayed from the race question for an actual case. This case had three checkboxes checked, with two literal race entries written on each of two lines. The initial coding is displayed for the various source items.

## Example 1:

### Data from Certificate                      Initial coding

- *Checkboxes checked:*

– White	Y	100
– Native Hawaiian	Y	500
– Other Pacific Islander	Y	530
- *Other Pacific Islander write-in line:*

– Part Hawaiian		503
– Fijian		542
- *“Other” write-in line:*

– Irish		106
– Italian		107

In the second slide for Example 1, two steps in the edit procedure applied to this case are high-lighted, and the edited multiple race data are displayed along with the NCHS 3-digit codes.

## Example 1, cont.

### Editing

- Checkbox codes removed when more specific write-in information present.
- Multiple white races collapsed to code 199 (Multiple white responses)
- *Edited Multiple Race Data*

– Multiple White Responses	199
– Part Hawaiian	503
– Fijian	542

In the concluding slide for Example 1, the bridging process is illustrated. First, the edited codes from the previous slide are collapsed into the four race groups specified in the 1977 OMB Standards for Race and Ethnicity; in this case we have only White and Asian or Pacific Islander (API)—a two-race combination. Using the bridging algorithm model for API/White, along with data on county of residence and other covariates for this case, we find that the probability is 60% that White would be the main race and 40% that API would be selected as main race. Next, these percentages are converted to adjacent intervals in the range from 0 to 100. The next step is to select a random number between 0 and 100. To do this, the program takes the last three digits of the certificate number and inverts them into a number with one decimal place. In this particular example, the last three digits are 433, which inverted yield 33.4. Since 33.4 lies between 0 and 60, White is chosen as the main race for this case.

## Example 1, concluded:

### **Bridging**

- *Edited Codes Collapsed to Four Race Groups*
  - White
  - Asian or Pacific Islander
- *Select Proportions (based on race combination and county of residence)*

▪ White	60%	(0-60)
▪ Asian or Pacific Islander	40%	(61-100)
- *Random Number (last three digits of certificate number inverted)*
  - Certificate Number: 0245433
  - Random Number: 33.4
- *Bridged Race*
  - White

Similar procedures are displayed in Examples 2, 3, and 4, below.

## Example 2:

<u>Data from Certificate</u>		<u>Initial Coding</u>
White checkbox	Y	100
Black checkbox	Y	200
American Indian or Alaska Native box	Y	500
<u>Write-in: Apache</u>		A09
<u>Arikara</u>		A31
Other Asian checkbox	Y	460
<u>Write-in: Taiwanese</u>		412
<u>Indonesian</u>		423
Other Pacific Islander checkbox	Y	530
<u>Write-in: Tahitian</u>		512
<u>Chamorro</u>		522

## Example 2, cont.

### Editing

- Checkbox codes removed when more specific write-in information present.
- Reduce stacks of more than 8 codes per individual (specific rules as to which codes will be collapsed or dropped and in what order to get to 8 or fewer codes).

## Example 2, cont.

### Edited Multiple Race Data

White (Checkbox)	100
Black (Checkbox)	200
Apache	A09
Arikara	A31
Taiwanese	412
Indonesian	423
Tahitian	512
Chamorro	522

## Example 2, concluded:

### **Bridging**

#### Edited Codes Collapsed to Four Primary Race Groups:

White  
Black  
American Indian or Alaska Native  
Asian or Pacific Islander

#### Select Proportions:

White	38%	(0-38)
Black	11%	(39-49)
American Indian or Alaska Native	21%	(50-70)
Asian or Pacific Islander	30%	(71-100)

#### Random Number (invert last three digits of cert. num.)

Certificate Number: 649256 => Random Number: 65.2

Bridged Race: American Indian or Alaska Native



In Example 3, no checkboxes are checked, and six literal entries are written in on the “Other race” line. (Note: This is the method NCHS uses to process multiple-race data from jurisdictions that do not use checkbox format but do accept multiple write-in entries.)

### Example 3:

**Data from Certificate**      **Initial coding**

***No checkboxes checked;***

***“Other” write-in:***

JPSE	(Japanese)	431
OKINAWAN	(Okinawan)	444
HAWN	(Hawaiian)	502
ITAL	(Italian)	107
GER	(German)	105
IRISH	(Irish)	106

### Example 3, cont.

**Editing**

- Multiple white races collapsed to code 199 (Multiple white responses)

- **Edited Multiple Race Data:**

Multiple White Responses	199
Japanese	431
Okinawan	444
Hawaiian	502

## Example 3, concluded:

### **Bridging**

#### Edited Codes Collapsed to Four Primary Race Groups

White

Asian or Pacific Islander

#### Select Proportions

White	12%	(0-12)
Asian or Pacific Islander	88%	(13-100)

Random Number (invert last three digits of cert. num.)

Certificate Number: 258754 => Random Number: 45.7

Bridged Race: Asian or Pacific Islander

## Example 4:

### **Data from Certificate**                      **Initial coding**

- *Checkboxes checked:*

– White	Y	100
– Black	Y	200

- *“Other” write-in line:*

– Mexican		617
-----------	--	-----

## Example 4, cont.

### Editing:

- No edits needed
- Edited Multiple Race Data

White (Checkbox)	100
Black (Checkbox)	200
Mexican	617

## Example 4, concluded:

### Bridging

**Edited Race Codes collapsed to four primary race groups (codes in the range 600-699 dropped):**

White  
Black

### **Select proportions**

White	75%	(0-75)
Black	25%	(76-100)

**Random number (invert last three digits of cert.num.)**

Certificate number: 024568 => Random number: 86.5

**Bridged race:** Black