

APPENDIX

Census 2000 Special EEO Tabulation Protecting Privacy

In preparing to publish data from Census 2000, the Census Bureau became extremely concerned that with the emergence of advanced computer technology that decodes information for areas by combining extremely detailed characteristics such as occupations with small geographic units, the possibility of revealing information on individual respondents would be greatly enhanced. To address this problem, the Census Bureau in 1995 created a Disclosure Review Board (DRB), specifically tasked with the responsibility to review specifications for all census data products, and to determine that no product format is approved that contains any degree of disclosure risk.

Previous versions of the Special EEO Tabulation, prepared from the 1980 and 1990 decennial censuses, were used throughout the country by employers and federal agencies to monitor compliance with federal law and regulations prohibiting employment discrimination. These files contained cross-tabulations of over 500 detailed occupational categories by six race/ethnicity groupings, sex, and other characteristics for all states, metropolitan statistical areas (MSAs), and counties irrespective of population size, and for all places with populations of 50,000 or more.

In 1997 the Office of Management and Budget (OMB) rescinded Statistical Policy Directive Number 15 (1977), covering the publication and use of data covering race and ethnicity. Based on OMB's Revised Standards (1997), the Special EEO Tabulation presents data for 12 race/ethnicity groups (15 in Hawaii) instead of six, including several dual-race combinations, such as Black/White and Black/Asian, among others. These expanded race/ethnicity distinctions, combined with the increased ability to electronically match different data files, prompted the DRB to review in detail the potential disclosure risk related to the Special EEO Tabulation format. As a result, the DRB issued two rules in 2000 affecting the Census 2000 Special EEO Tabulation.

1. Any occupational category containing fewer than 10,000 people employed nationwide cannot be shown separately, and must be combined with related occupational categories to create aggregates containing 10,000 or more people.
2. Detailed occupational categories, when cross-tabulated by the 12 race/ethnicity categories (15 in Hawaii), cannot be shown for any geographic area containing fewer than 50,000 people. Geographic areas containing fewer than 50,000 people must be combined in such a way that the resulting aggregation:
 - a. Contains 50,000 or more people, **and**
 - b. The subtraction of any geographic subset from a larger, combined area does not reveal information for any area containing fewer than 50,000 people.

APPENDIX continued...

As a result of the above rules issued by the DRB, the Special EEO Tabulation is now configured in the following manner:

1. The number of detailed civilian occupational categories are reduced from the 509 in the Census 2000 classification system to 472/471.
2. For Datasets No. 1 and 2, containing cross-tabulations of detailed occupational categories (472 in Dataset 1 and 268 in Dataset 2) by the 12 race/ethnicity categories (15 in Hawaii), approximately 70 percent of all U.S. counties are combined to create combinations of counties (referred to as County Sets) containing no fewer than 50,000 people.
3. A further constraint is added to ensure compliance with Rule 2(b), above. That is, no County Set combination is permitted to cross state lines. Otherwise, the risk referenced in Rule 2(b) is unacceptably high that, by subtraction of particular counties across a state boundary, individual respondent information could be revealed.

As further protection, the DRB required that the following additional disclosure avoidance steps be taken with the Special EEO Tabulation:

1. All cells in the Census 2000 Special EEO Tabulation are rounded. The rounding schematic is:
 - 0 remains 0
 - 1-7 rounds to 4
 - 8 or greater rounds to nearest multiple of 5 (i.e., 864 rounds to 865, 982 rounds to 980)
 - Any number that already ends in 5 or 0 stays as is.

Any totals or subtotals are constructed before rounding. This assures that universes remain the same from dataset to dataset, and it is recognized that cells in a dataset will no longer be additive after rounding.

2. If geographic codes are shown, they are Federal Information Processing Standards (FIPS) codes rather than census codes.
3. For Worksite/Residence datasets, no flow is shown from a residence county or residence place unless there are 50 or more unweighted cases for all occupations in that county or place. If there are less than 50 unweighted cases, all cases go into the U.S. Balance flow category. All occupations in the U.S. Balance flow category are shown separately.
4. For Worksite/Residence datasets, no flow is shown from a residence county or worksite for a specific occupation unless there are 3 or more unweighted cases for that occupation in that county or place. If there are less than 3

APPENDIX continued...

unweighted cases for a specific occupation, the cases for that occupation go into the U.S. Balance flow category. All occupations in the U.S. Balance flow category are shown separately.

5. Seven metropolitan areas (MSAs and PMSAs) are not included in datasets 1 and 2 because to identify them in conjunction with identifying County Sets would result in showing data for an area of under 50,000 population. The seven areas are –

Arizona: Flagstaff AZ-UT MSA - FIPS MSA Code #2620
Arkansas: Fort Smith AR-OK MSA - FIPS MSA Code #2720
Maryland: Baltimore MD PMSA - FIPS PMSA Code #0720
New York: Newburgh NY-PA PMSA - FIPS PMSA Code #5660
North Dakota: Grand Forks ND-MN MSA - FIPS MSA Code 2985
Pennsylvania: Pittsburgh PA MSA - FIPS MSA Code #6280
Wisconsin: La Crosse WI-MN MSA - FIPS MSA Code #3870

However, all metropolitan areas (MSAs and PMSAs) are shown in datasets 3, 4, 5, and 7.

6. Sixty places with populations of 50,000 or more are not shown in datasets 14 through 19 because showing them in conjunction with data for the county they are in would result in the identification of a balance of the county that is under 50,000 population. However, all places of 50,000 or more are shown in datasets 3 through 6. The sixty places not shown are:

Alabama:
Dothan city
Montgomery city
Arkansas:
Fort Smith city
Jonesboro city
Pine Bluff city
Colorado:
Pueblo city
Florida:
Jacksonville city

Georgia:
Albany city
Athens-Clarke County (balance)
Augusta-Richmond County (balance)
Columbus city (balance)
Idaho:
Idaho Falls city
Pocatello city
Illinois:
Decatur city
Indiana:
Terre Haute city

Iowa:
Ames city
Council Bluffs city
Dubuque city
Iowa City city
Sioux City city
Kansas:
Kansas City city
Lawrence city
Topeka city
Kentucky:
Owensboro city

APPENDIX continued...

Louisiana: Bossier City city	North Dakota: Bismarck city	Bryon city
Minnesota: Rochester city	Fargo city	Corpus Christi city
Missouri: Kansas City city	Oklahoma: Lawton city	Laredo city
St. Joseph city	Oregon: Gresham city	Longview city
Montana: Billings city	Salem city	Lubbock city
Great Falls city	South Dakota: Rapid City city	Midland city
Missoula city	Sioux Falls city	Odessa city
Nebraska: Lincoln city	Tennessee: Clarksville city	San Angelo city
New Mexico: Rio Rancho city	Jackson city	Victoria city
North Carolina: Durham city	Nashville-Davidson (balance)	Wichita Falls city
Rocky Mount city	Texas: Abilene city	West Virginia: Huntington city
	Amarillo city	Wisconsin: Eau Claire city
		Wyoming: Cheyenne city

7. Thirty-four places with populations of 100,000 or more are not shown in datasets 21 through 24 because showing them in conjunction with data for the county they are in would result in the identification of a balance of the county that is under 100,000 population. The thirty-four places not shown are:

Alabama: Montgomery city	Illinois: Peoria city	Louisiana: Lafayette city
Colorado: Pueblo city	Springfield city	Shreveport city
Florida: Jacksonville city	Indiana: Evansville city	Michigan: Lansing city
Tallahassee city	Indianapolis city (balance)	Mississippi: Jackson city
Georgia: Athens-Clarke County (balance)	Iowa: Cedar Rapids city	Missouri: Kansas City city
Augusta-Richmond County (balance)	Kansas: Kansas City city	Springfield city
Columbus city (balance)	Topeka city	

APPENDIX continued...

Nebraska:

Lincoln city

Omaha city

North Carolina:

Durham city

South Dakota:

Sioux Falls city

Tennessee:

Clarksville city

Nashville-Davidsonville (balance)

Texas:

Abilene city

Amarillo city

Corpus Christi city

Laredo city

Lubbock city

Waco city

Wichita Falls city

Source: U.S. Census Bureau Web Site, December 29, 2003: <http://www.census.gov/hhes/www/eeoindex.html>