



UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001
OFFICE OF THE DIRECTOR

SEP 19 2008

Mr. Jack Basso
Director of Management and Business Development
American Association of State Highway and
Transportation Officials
444 North Capitol Street, NW, Suite 249
Washington, DC 20001

Dear Mr. Basso:

On behalf of the U.S. Census Bureau, I want to thank you and the representatives of the American Association of State Highway and Transportation Officials (AASHTO) for sharing your perspectives regarding the Disclosure Review Board's ruling on the three-year American Community Survey (ACS) Census Transportation Planning Package (CTPP) at the August 28 meeting of the Data Stewardship Executive Policy (DSEP) Committee. The Committee appreciated your insights regarding the utility of census data in the context of our increasing concerns about confidentiality.

The DSEP Committee has determined that the Disclosure Review Board's ruling will be upheld. However, I want to assure you that this decision was not made lightly and that we recognize the needs of the transportation community for detailed, small-area data and that we share a common goal to release data that are useful to local communities in making key decisions about transportation.

As explained in the meeting, these needs must be considered in the context of our legal obligation to protect confidentiality—an obligation that the Census Bureau is acutely aware of as we prepare for the 2010 Census. The law is unequivocal, and it establishes a clear standard for the agency; the Census Bureau cannot by law "make any publication whereby the data furnished by any particular establishment or individual...can be identified" (Title 13, United States Code, Section 9). We believe this commitment is essential to the success of our mission and that the public relies on the Census Bureau to maintain this commitment, even when it constrains the agency and limits the data we can release to other agencies and the public.

The Census Bureau has historically erred on the side of caution in releasing public data products. Our approach is based on recognized disclosure avoidance techniques, including the use of data swapping, noise addition, suppressions, thresholds, and rounding. These techniques are shared across the statistical community and were described by the Federal Committee on Statistical Methodology's Confidentiality and Data Access Committee in "Statistical Policy Working Paper 22, Report on Statistical Disclosure Limitation Methodology" (see enclosure).

Now more than ever, with the abundance of technology, coupled with the increasing number of publicly available datasets, as well as the ever-increasing concerns regarding security of data, the Census Bureau has deliberately taken a careful and considered approach to meet the requirements of our law. To that end, we routinely reexamine our data stewardship policies to

U S C E N S U S B U R E A U

Helping You Make Informed Decisions

www.census.gov

ensure that they reflect the most recent advances in protecting and managing data. Since 2002, we have undertaken reidentification studies using historic public data, including the Current Population Survey and the Survey of Income and Program Participation public use microdata (PUMS) files to determine whether we could link variables from these files with other publicly available data, such as vital records files. As a result of these studies, the Census Bureau strengthened the disclosure avoidance procedures for its data products, recognizing the need to adapt in order to protect confidentiality and maintain public trust. These protections apply to all public data products, and are not unique to the ACS. It is important to note that these protections would also have been applied to census long-form data. Research in this area is ongoing, and we plan to conduct these studies with the multiyear estimates from the ACS.

The Census Bureau relies on the Disclosure Review Board to provide guidance in this area—not only to review data products to ensure that the appropriate data protections are used, but also to provide its technical expertise in the area of disclosure avoidance and potential reidentification. The Census Bureau’s efforts, particularly those of the Disclosure Review Board, are recognized within the statistical community, and we are constantly seeking new opportunities, such as the use of synthetic data, that will not only protect individuals’ information, but also lead to better data products.

Therefore, any request or appeal to the DSEP Committee to, in effect, overrule the Disclosure Review Board’s decisions must be seriously considered in light of technical recommendations based on established practices and standards. In this instance, the Disclosure Review Board ruling on the three-year ACS CTPP product required a threshold of three unweighted cases for each category of the “means of transportation” when this variable is crossed with one or more other variables in tables for each geographic area (places of 20,000 population or more). This ruling would suppress tables with marginal totals of one or two unweighted cases for the “means of transportation” variable when that variable is crossed with another variable. It does not suppress all tables with any unweighted cell counts of one or two, provided the marginal total is three or more. It is important to note that there is no threshold for the univariate “means of transportation” table, only for those that are cross-tabulated by other characteristics.

In preparation for the DSEP meeting, the Disclosure Review Board explained its specific concerns with the request to produce tables from the three-year ACS data that violated the threshold of three. With tables containing marginal totals of one or two, data users can recreate microdata records for places of 20,000 or more, a geographic level insufficient to address known disclosure risks and a violation of the Census Bureau’s longstanding guidance for microdata. As the Disclosure Review Board explained, the problem occurs when there is only one person in the ACS with a given means of transportation for the specific place. For example, if this person is the only person who rode a bicycle to work in the given geographic area according to the univariate table, then the table would indicate this person’s weighted count, for instance 30. A data user would then see the weighted count for this single person shown in a single cell for all of the other tables crossed with means of transportation. There would be one occupation category for people who rode a bicycle to work, meaning one cell would contain a weighted count of 30, while all the other cells would contain 0. This phenomenon would be repeated

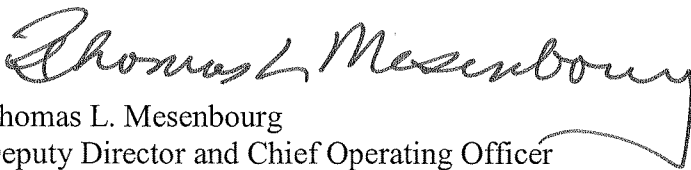
table-by-table for each instance when the means of transportation variable is crossed with another variable ranging from age, to gender, to race, to occupation, to disability, to earnings, to the number of vehicles available. By linking the cells in these tables, the data user could easily recreate a microdata record that could be linked to existing PUMS files or other publicly available data.

It is also important for you to know that the Census Bureau has published standards regarding the release of microdata; we do not publish microdata for geographic areas of less than 100,000 people. The "100,000 rule" is well-established and serves to protect the confidentiality of individual respondents by limiting geographic detail. The Census Bureau has conducted research for geographic areas ranging from 20,000 to 500,000, which demonstrates that "the proportion of records that might be uniquely identified in a microdata file is related to the geographic detail on the file." (See enclosure, "Enhancing the '100,000 rule' on the Variation of the Percent of Unique in a Microdata Sample and the Geographic Area Size Identified on the File.") Geographic detail, such as place of residence as represented in the data tables using the three-year ACS estimates, is especially useful in identifying a particular individual. The creation of microdata based on the means of transportation tables for places of 20,000 violates this standard and is a known and preventable disclosure risk.

The law requires the agency to prevent disclosure, and we cannot discount demonstrated risks. The Census Bureau, however, is committed to working with AASHTO and members of the transportation community to develop alternative data products, such as collapsed versions of the current requested tables or in the development of synthetic data models for the five-year ACS estimates. We have worked successfully together for many decades to identify data products useful for local communities, including 53 means of transportation tables that the Census Bureau will release as a standard data product for the first time this year (see enclosure). As you know, in the past, these tables were released as part of the CTPP special tabulation.

The Census Bureau recognizes the value of AASHTO's partnership to the success of the ACS and that it is important for us to work with stakeholders and data users in anticipation of the release of five-year estimates in 2010, which thereafter replace the decennial census long form with annual data releases. I have asked my staff to continue to meet with your staff to explore options for further research. Recognizing that there is no quick or easy answer, I am confident that we can work together toward a solution that will meet the needs of the transportation community, and by working together we can ensure the ACS will meet the needs of communities throughout the United States.

Sincerely,



Thomas L. Mesenbourg
Deputy Director and Chief Operating Officer
Chair, Data Stewardship Executive Policy Committee

Enclosures