AMERICAN ASSOCIATION OF STATE HIGHWAY AND TRANSPORTATION OFFICIALS
AASHTO
THE VOICE OF TRANSPORTATION

Pete Rahn, President
Director, Missouri Department of Transportation

John Horsley, Executive Director

444 North Capitol Street NW, Suite 249, Washington, DC 20001
(202) 624-5800 Fax: (202) 624-5806 • www.transportation.org

August 20, 2008

Ms. Christa D. Jones
Chief, Office of Analysis & Executive Support
U.S. Census Bureau
4600 Silver Hill Road
Washington, DC 20233

Subject: Census Transportation Planning Products (CTPP) proposal using ACS

Dear Ms. Jones:

Thank you for the opportunity to present information in advance of the August 28, 2008 meeting with the Census Bureau's Data Stewardship Executive Policy committee. We are looking forward to the meeting. We appreciate the opportunity to present our positions regarding the importance of the data we are requesting and the low disclosure risk associated with the ACS three-year cross tabulations on means of transportation characteristics and flows.

In preparation for the meeting, the American Association of State Highway and Transportation Officials (AASHTO) commissioned a data driven disclosure risk analysis from Westat. A copy of the final report is attached. Westat found low levels of disclosure risk in the CTPP tabulations of 3-year ACS data and concluded that the application of the Rule of 3 for the CTPP tabulations is not warranted.

The conclusions and recommendations of their report state:

> *The analysis results noted in this document show the need to balance the low risk of the CTPP tabulations based on the 3-year ACS, with the large amount of data suppressed due to the Rule of 3. The following are recommendations to the AASHTO (where appropriate) and Census Bureau DRB for the 3-year CTPP tables.*
>
> *1. Allow unweighted cell counts of less than 3 for crosstabs in CTPP tabulations from the 3-year ACS. We have demonstrated that the disclosure risk is low due to several components of data protection at the expense of data utility.*

*2. Further investigate the collapsing rules on the CTPP demographic variables (non-MOT variables), most notably industry, and their affect on the suppression of data for tables involving the MOT variables.*

In preparing for the meeting we additionally reviewed the information prepared by Ms. Laura Zayatz, dated August 12, 2008 entitled "The Disclosure Review Board's Decision on the Census Transportation Planning Package Proposal for Special Tabulations of the American Community Survey's Three-Year Estimates". We offer the following comments in response to the information prepared by Ms. Zayatz.

**1. Previous Rulings**

On page 1, Ms. Zayatz states that "None of those previous requests were the same as the current request, which is for CTPP special tabulations for three-year data for areas with a population of 20,000 people or more".

We respectfully disagree with Ms. Zayatz. We have held several meetings over the last 2-3 years to discuss this subject. We believed that the DRB rules on 3-year ACS tabulations would be applied to CTPP using 3-years of accumulated ACS records, with the additional caveats applied to special tabulations (e.g. rounding).

In a memo dated December 13, 2005 and another dated September 13, 2005. Ms. Zayatz wrote "For the 1 year and 3 year estimates, there is no threshold on means of transportation (mode) for residence and workplace tables." Since these memos were written long after the CTPP 2000 was completed, they clearly do not pertain to the CTPP2000. It was understood among all participants that the meetings referred to in the memos were intended to discuss the new ACS and future CTPP data packages. For more information see http://trbcensus.com/drb/dec_13_memo.pdf

**2. The "Threshold of Three" Requirement and the "Threshold of Three" Example**

We respectfully disagree with Ms. Zayatz' comments. We do not believe that the ACS three-year data represent a snapshot. The CTPP tabulations using 3-years of ACS data represent the accumulation of 36 months of surveys for a geographic area with over 20,000 population, and no longer reflect a snapshot of any individual. In fact, Census Bureau ACS staff have stated that the ACS will produce "period estimates", not a snapshot or point-in-time estimate. Our analysis indicates that when 3-years of ACS survey records are accumulated, they do not identify when a specific survey was collected. If someone finds the record for one bicyclist that was surveyed between January 2005 and December 2007, our analysis suggests that it is virtually impossible to identify who that specific person is, and impossible to say when the data was collected in the ACS three-year window.

In addition, statistical analysis undertaken by others, including the National Academy of Sciences Committee on National Statistics, strongly affirms that the information on

the sample records have a high potential for changes. This includes changes in residence, workplace, means of transportation to work, time leaving home to go to work, travel time to work, industry, occupation, and even age and income.

**3. Geographic Requirements for Microdata**

Under this subject heading, Ms. Zayatz suggests that we are requesting the release of micro data records. We respectfully disagree. A "pseudo-microdata" record compiled from cross-tabulations is not the same as a "true" microdata record. Even if several characteristics in collapsed categories can be listed (as both the Census Bureau and our analysis have shown) this is not the same as identifying a specific person. For example, if the CTPP uses an age range of 25-44, then in an ACS three-year accumulation, there are 22 potential years of birth. This is significantly different than a "true" microdata record, where an individual's year of birth is listed.

Our analysis indicates that the "true" microdata record for the ACS three-year data tables has 150 variables and the "pseudo microdata" record using CTPP is limited to less than 20 variables.

All of the analysis we have seen indicates the changes that occur in the data over a three-year period virtually eliminate the ability to link a pseudo-microdata record to a specific individual. Recent statistical analysis performed by Westat substantiate that the risk of re-identification (or the ability to match to a specific individual by name, address) approaches zero. A copy of the Westat position paper is included with this letter.

In regards to the Census Bureau's mandate to minimize the risks of disclosure, we note that OMB Statistical Paper 22 states this mandate is "….concerned with minimizing the risk of disclosure (public identification) of the identity of individual reporting units and information about them." (page 2). Regarding microdata the same OMB report says, "A microdata file consists of individual records, each containing values of variables for a single person, business establishment or other unit' (page 4).

A single record found in the ACS three-year tabulations is not truly a single individual in the population. The single record has an assigned weight, meaning that it could represent an estimated 35, 80 or even 160 people in the population. Having a weigh greater than 1 blurs the identity of the record alone. For further information on this OMB discussion, please refer to
http://www.fcsm.gov/working-papers/SPWP22_rev.pdf

**4. Base Tables, Data Quality, and Options for Collapsing Categories**

We appreciate that the ACS three-year standard tabulations will have workplace tables that were previously only available as part of CTPP and the number of tables crossed with means of transportation. These changes have been very welcome by the transportation community.

We are concerned with Ms. Zayatz' references to data quality and reliability. In all previous discussions, we understood the Census Bureau concerns involved disclosure risk. We have no information to suggest that further data suppression would be applied based on data quality and/or reliability. We are not aware that the role of the DRB is to determine whether or not the results were statistically reliable. The CTPP table request submitted by AASHTO asks that margins of error (MOEs) for each cell be calculated and delivered. We are willing to live with the results as this is a special tabulation.

Further, Ms. Zayatz' comments about potential data reliability issues associated with the number of means of transportation cross tabulated with other variables seems to be in stark contrast to her concerns about the risks of disclosure. If the data are unreliable how can they also be at risk for disclosure of a microdata record?

Ms. Zayatz' comments note the effects of suppression can be reduced to 33 percent if we collapse the number of categories of means of transportation to work. We understand that we may collapse as many tables as we wish. However, it is important to note that 33 percent table suppression is <u>not</u> acceptable. Losing data on means of transportation for one-third or more of the places in the country over 20,000 is not a tenable position. It is important to note again the significance of means of transportation data. There is unprecedented national, state and local interest in single passenger auto, carpooling, motorcycle, transit, bicycle, pedestrian and work at home trends. Understanding these trends, along with other demographic and work place characteristics are essential to a host of transportation policy, planning and investment analyses as well as other related climate change, land use, and evacuation planning efforts.

**5. A Snapshot of a Person at One Point in Time**

In the last paragraph on page 3, Ms. Zayatz indicates that means of transportation is not necessarily one of the variables subject to swapping, imputation and/or other ACS survey design feature intended to minimize disclosure risk.

We wish to note that since the Census Bureau keeps the algorithm for data swapping a secret, then the data users do not know which variables are used, or not used. We assume that when a record is swapped, that all the variables including the means of transportation is moved to the new record. We consider data swapping part of the standard CB practice to reduce disclosure risk.

**6. DRB Efforts to Make Data Available**

On the top of page 4, Ms. Zayatz discusses the process that was followed to identify priorities for the ACS three-year data tabulations.

We would like the record to show that during the entire process, DRB has never shared with us what "too many" cross-tabulations means. We continue to have

questions regarding what number of cross-tabulations would be allowed to avoid creating pseudo-microdata records. In addition, we do not know whether all variables were considered the same way. For example, "poverty status" is a calculated variable based on household size, age of householder, presence of children. We do not understand how "commuting characteristics" such as travel time can be treated as a disclosure risk in the same way as "worker earnings".

Ms. Zayatz correctly notes that we are investigating the feasibility of using synthetic data methods to minimize disclosure risks. We wish to note that these synthetic methods are for the ACS five-year data tabulations. We are undertaking this work because we understood the DRB would require 3-unweighted records per category of means of transportation for the CTPP using 5-years of ACS records. Because the 5-year data product is planned to include SMALL geography, e.g. tracts and Traffic Analysis Zone (similar to Block Groups), we understand why the disclosure risk might be greater when the geographic scale was this small. We did not think that synthetic data was necessary for the ACS three-year CTPP tabulations, because the geographic unit of reporting is limited to places with over 20,000 population. However, it must be noted that further research on the risks associated with any five-year tabulations also needs further study.

As part of the synthetic data discussion on page 4, Ms. Zayatz references work done by Mr. Nanda Srinivasan and his alleged acknowledgement regarding microdata records. We specifically wish the record to note that Mr. Srinivasan's in an e-mail to us says that "…he said none of the CTPP tables can be considered "microdata", it is categorical data that can possibly be used to link to a microdata record from PUMS. The same danger exists whether we used CTPP tables or ACS Standard Tabulations. Moreover, the Westat research shows that the possibility of this happening is close to zero". Further, the reference to Mr. Srinivasan's work is inaccurately described by Ms. Zayatz.

**7. Census Bureau Effort to Make Data Available**

The last points made by Ms. Zayatz on page 4 discuss Census Bureau efforts to make ACS three-year data available and whether we are getting less or more data than in the past. Ms. Zayatz notes that not much was requested or provided from the 1990 census.

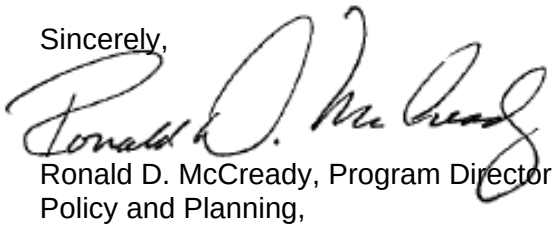
There are expanding uses and applications for the CTPP data. As we have noted in the past, growing concerns over climate change, emergency and security planning, safety, congestion management and other factors are creating unprecedented demands for these data, many of which have been mandated by Congressional action

In respect to disclosure, we would additionally like to note that the smaller sample of the ACS, even after accumulation over 3 years, compared to the decennial Census 2000 "long form" is one of the greatest protections of the ACS. Identifying one

sample record is NOT the same as identifying a specific person in a population.

Thank you again for the opportunity to provide written information prior to our meeting with you on August 28, 2008. We look forward to meeting with you to further discuss these very important issues.

Sincerely,



Ronald D. McCready, Program Director for Policy and Planning,
American Association of State Highway and Transportation Officials
444 North Capitol Street, NW Suite 249
Washington, DC 20001