

USDOT meeting with US Census Bureau on Disclosure Issues

July 24, 2006

1:30 - 3:00 p.m.

Hollerith Room (1st floor, behind Auditorium)

U.S. Census Bureau, Building 3

Attendees: Phil Steel, Marie Pees, Laura Zayatz, Celia Boertlein, Phil Salopek, Doug Hillmer, and Lisa Blumberman (Census Bureau), Lee Giesbrecht (BTS), Elaine Murakami, Ed Christopher (FHWA, DOT), and Nanda Srinivasan (Contractor to FHWA), Ed Weiner (OST, DOT)

AGENDA

- 1. Status report on the next CTPP**
- 2. TAZ zone size requirements**
- 3. Collapsing rules**
- 4. Synthetic data tests for small area (TAZ-to-TAZ) flow tabulations**
- 5. ACS PUMS**

SUMMARY OF ACTION ITEMS:

- 1) Doug H. will forward the tables shells for the collapsed tables related to Journey to Work to Nanda to forward to DOT staff.
- 2) Doug will send the handout (powerpoint) explaining the data release rules to Nanda to forward to DOT staff who could not attend the meeting.
- 3) Ed C. will draft a letter requesting examination of rounding rules using ACS person weights.
- 4) Nanda will talk to Frederik Anderson about alternate disclosure proofing methods.
- 5) Nanda will continue to work on the IPF routine for SuperTracts and try use median data for SuperTracts, and income distributions for workers who lived and worked in each supertract.

1. Status report on the next CTPP

Ed Christopher presented an overview of current status of a potential AASHTO pooled fund for data products using ACS.

1a. Workshop on Census Data Needs and Priorities for Transportation

In April 2006, a Workshop on Census Data Needs and Priorities for Transportation was held to identify future census data needs and priorities and scope out a multi-year program of tasks and potential pooled fund contract components

The workshop was sponsored by the AASHTO Standing Committee on Planning (SCOP) Census Data Work Group, supported by the FHWA Planning Capacity Building Program and cosponsored by several Transportation Research Board (TRB) data committees. Representatives from AASHTO, federal agencies, state transportation departments, metropolitan planning organizations (MPOs), the National Association of Regional Councils, and TRB data committees attended the workshop.

There was consensus among workshop participants on the continuing need for census data to support sound transportation planning processes. The participants also agreed that past AASHTO pooled fund projects were a valuable and effective means of providing census data products and tools for state and local transportation agencies.

Ed C. said that most participants at the April workshop expressed strong need to know zone size thresholds for disclosure avoidance, and an assurance that the most important tables for transportation would not be subject to thresholds, as was done for CTPP 2000.

1b. June 14, 2006 AASHTO Standing Committee on Planning (SCOP) meeting

The results of the Census workshop were presented at the SCOP Data Task Force and the full SCOP. At both groups, there was verbal support of a new pooled fund at 1.9 cents per person. This would total to approximately \$5.8 million. The SCOP chair asked Jonette Kreideweis to bring a full proposal to the October meeting.

1c. October, 2006 Full AASHTO meeting

Proposal will be delivered to SCOP, and assuming approval, subsequently forwarded to the full AASHTO Executive Committee.

DISCUSSION:

Laura Zayatz read from a DRB memo to Larry McGinn on December 13, 2006 about DRB rules for workplace tables with mode as a variable, and read from a memo to Alfredo Navarro on August 26, 2003, which stated rules exactly similar to CTPP 2000. Laura said that thresholds would be applied for flows except for total flows, and mode to work. Phil Salopek asked about Table 3-2 (Mode by Vehicles Available).

Ed C. said that rounding 1-7 to 4 resulted in a consistent undercount in CTPP 2000. He recommends that the actual data be examined before selecting the value so that the value with the closest total is selected. Laura asked Ed C. to put the request in writing so DRB can follow-up on this issue with higher level staff at CB. Doug H. said the average person weight in ACS for the 5-year dataset is approximately 14, so values below 7 are less likely anyway.

2. TAZ zone size requirements

As MPOs move toward microsimulation, rather than zone-based travel demand forecasting, MPOs want smaller and smaller geographic units for data. DOT will be talking to CB Geography Division about schedule for TAZ and SuperTAZ boundaries at our CTPP meeting of July 27. The transportation community does not want to run the risk of designing a special product for which all the data is suppressed and collapsed, making it unusable for transportation planning.

- a. Will the CB place restrictions on the population size for reporting data for TAZs (i.e. custom geographic units)?

Marie Pees said that the CB would soon publish a Federal Register Notice asking for minimum population size for BGs/Tracts to be 1,500. Legally defined entities including School Districts are not subject to thresholds. Based on comments to the FR Notice, the DRB will reconsider its decision.

- b. What size population (residential or workplace) would be optimum for getting tables (Residence, workplace, and flow) similar to those released in CTPP2000?

The DRB reiterated the rule of 3 unweighted records for small area data, and pointed to the August 2003 memorandum.

3. Collapsing rules

Doug H. explained the "Data Release Rules" being established to ensure statistical reliability of published ACS data. Unlike decennial Census with a large sample at one point in time, ACS is accumulating data from each year. The collapsing strategy for regular CB data products will ensure that the end user would not make errors in interpreting the data. These rules will not apply for reimbursable or special products. The rules do not apply to 5-year products.

Any table whose median distribution of covariance for individual cell values is greater than 61% would be suppressed for that geography. For example, for a given geography, say County X, for a table with 18 means of transportation, individual covariances are calculated for estimates of workers who drove alone, carpooled etc. If the median of these covariances is greater than 61, then these modes would be collapsed to fewer categories. If the median of covariances for the collapsed table still exceeds 61, the table is suppressed for County X.

Celia has been working with ACS staff on designing the collapsed table structure. She said that the collapsing strategy uses the categories in CTPP. Doug H. said that by beginning of next week, we could have the shells used for collapsing.

ACTION: 1. Doug H. will forward the tables shells for the collapsed tables related to Journey to Work to Nanda to forward to DOT staff.
2. Doug will forward his powerpoint on the data release rules to Nanda to share with DOT staff who could not attend the meeting.

4. Synthetic data tests for small area (TAZ-to-TAZ) flow tabulations

Nanda Srinivasan, Cambridge Systematics

Transportation planners say that they want TAZ-to-TAZ flows stratified by travel mode and household income.

Nanda S. has started testing an IPF routine to create synthetic data at small geography (Tract-Tract) based on Super Tract flows stratified by income and mode to work, and marginal data from Part 1, Part 2, and Part 3 which are not subject to thresholds. One of the issues is that even with the size of a Supertract being about 15,000, roughly 40% of the geography pairs fail the threshold rule of 3. Therefore, a Supertract or SuperTAZ might need to be larger than 15,000.

Laura said that we could continue to pursue the concept of synthetic data for small area flows. However, the threshold of 3 would still apply for any tabulation published from using the original data.

Potential alternatives:

- a. Use “real” data for Supertract-to-Supertract flows and use them in data synthesis process, but do not publish them. Only make the synthetic TAZ-to-TAZ flows available.
- b. Phil Steele suggested to try increasing the size to 25,000 population per Supertract, and examining to see if instead of using flow from a supertract to all supertracts, it would be possible for the IPF routine to just use data for:
 - o Workers who lived and worked in that Supertract
 - o Workers who worked in the Supertract but lived elsewhere.
 - o Workers who lived in the Supertract, but worked outside.
- c. Doug H. suggested using median household income by mode for flows, along with distributions of income for residence and workplace geography.

Is place of work data already synthetic?

One of the issues Nanda found while working on the IPF routine is the high amount of allocation for place of work. Around 25 percent of the records for place of work are already “synthetically” allocated either by the standard POW allocation or the extended POW allocation process. When the size of geography used for tabulation is increased by roughly 3-4 times the size of an average tract, the risk of disclosing using “aggregate” data must be lower. There must be a level of geography where disclosure threshold of the rule of 3 can be relaxed for aggregate data (univariate or bivariate data). We hope this could be SuperTract/TAZ at a population threshold of 25,000.

ACTION: 1. Nanda will meet with Frederik Andersson on synthetic data approaches used in LED.
2. Nanda will try to see if he can implement the ideas from Phil Steele and Doug H for a test area.

5. ACS PUMS

Phil heard that the number of records in the ACS PUMS could not exceed the number of records in the Census 2000 PUMS. Marie and Laura clarified the hearsay information. Marie Pees said that for ACS, CB is collecting the data for 2.5 percent of the population and releasing data for 1 percent of the population. This constitutes 40 percent of the full ACS sample. Laura explained that historically, about one-third of microdata records were released without risking individual disclosure. For the Census 2000 PUMS, 5% from the 16.6 percent sample (33% were included). Top coding and collapsing will be used just as decennial census 2000. Data are topcoded based on state data attributes rather than national data. Ed C. emphasized that transportation planners would like the PUMS data to include a place of work PUMA that was not limited to County level geography. However, for this to occur, the place of work needs to be coded to tract/block to be linked to the correct PUMA.