

**US DOT – Census Bureau (CB) meeting on Disclosure Issues  
Suitland, Maryland  
August 23 2007, 10:00 a.m.**

**Attendees:** Ed Christopher, Elaine Murakami, Lee Giesbrecht, Nanda Srinivasan (Contractor) (US DOT), Ron McCready (AASHTO), Celia Boertlein, Michelle Jiles, Chris Henrie, Laura Zayatz, Sam Hawala, Ronaldo Rodriguez, Jennifer Hucket, Marie Pees, Tom Blatt, Jeremy Funk (U.S. Census Bureau)

Ed C. gave an update on the status of CTPP. Ron McCready updated the group on AASHTO's role in the planned suite of products. Laura Zayatz updated the group on the DRB. Currently, the rules planned for ACS include data swapping, rounding, and thresholds. Laura also apprised the group of synthetic data approvals for a linked SIPP/SSA earnings data file, the LEHD Quarterly Workforce Indicators file (which is partially synthetic), and LEHD OnTheMap. The DRB also approved synthetic data to protect GQ population in ACS.

Laura explained that while similar approaches may be taken for both imputation and data synthesis, the purposes are different:

- Purpose of imputation is to fill in missing values, and
- Purpose of data synthesis is to protect confidentiality.

Laura asked hypothetically, if a synthetic ACS CTPP Flow table overlaps with another table, would the CB be “ok” with the release of two different numbers for the same cell. Marie Pees said that the CB has already allowed some cases where there are two values for the same cell. Nanda replied that in a “potential” synthetic data table, the rows and columns will be made to agree (as far as possible) with other published “overlapping” data such as marginals. He thinks that the synthetic data and the real data will be mutually exclusive. It has to be made clear that a synthetic table is appropriately “labeled” as being synthetic. He also believes that there are no problems with non-synthetic data revealing the method used to protect synthetic data.

Elaine reiterated that our preference is to obtain ACS tabulations for CTPP without rounding and thresholds, but that Nandu was working on synthetic data as a concept if it is unavoidable to obtain small area (TAZ) tabulations. Based on recent ACS records, approximately 33 percent of workers have place of work; means of transportation; and/or household income imputed. This high proportion of imputation should reduce the risk of overall individual disclosure. Nanda reviewed his current work on disclosure proofing of flow tables (See attached MS Powerpoint presentation). Laura said that the amount of imputation, aging of data, movement of people, and workers should all be factored into the disclosure proofing method.

For some variables (eg: household income), the imputation procedure is not “strong”, and the imputed records may be treated as 0 real observations. For some variables (eg: sex), there is strong confidence in the imputed records, and these need to be treated as “real records.” Microdata that already have imputed values (for the variable being tabulated)

may not need the same confidentiality protection as microdata records without any imputed values.

Laura said that they have been changing rules and methods as time moves along. For example, the data swapping rules have been changed, such that they are now using multiple years of ACS respondents to find an individual with which to swap, not just in the same ACS survey year.

Laura said that, as in CTPP 2000, the transportation community can continue to get, without any threshold rules, total worker flows, and worker flows stratified by mode and vehicles available.

Ed C. laid out several issues with ACS, and said that the continuous methodology, weighting, and aging of data itself are sufficient disclosure protection.

Laura would like to see some work on how frequently variables changes and/or are missing for the general population over a 5 year period. Ed said that in transit on-board surveys, about 30% of transit users switched travel modes from the previous year. Celia reported that ACS results showed that about 15 percent moved from the previous year, and about 45% over 5 years. Laura said she has been trying to get funding for a project by Michael Larsen (Iowa State) and Jerome Reiter (Duke) on methods to disclosure proof ACS data.