



The Truth can be Nebulous

Clara Reschovsky
Metropolitan Washington
COG

September 25, 2013

Data User Perspective

- Transportation planner for a Metropolitan Planning Organization (MPO)
 - MPOs assume planning responsibilities for areas of 50,000 population or more
 - Over 300 MPOs in the US
- First time receiving data with perturbed values
 - Values had to be changed otherwise not much data would be provided
 - The ACS reduced sample size is a significant source of the uncertainty surrounding the CTPP Traffic Analysis Zone (TAZ) estimates

Data User Perspective (2)

- First time receiving data with perturbed values (continued)
 - An extensive evaluation was conducted using transportation model inputs (NCHRP 08-79)
 - Perturbed values followed closely to the original ACS estimates
 - Users have questions and need to be 'educated' about SDC treatments

MPO Responsibilities

- Ensure compliance with federal laws and requirements
- Provides technical resources for decision-making
- Develop and maintain travel forecasting model for predictions about future travel conditions
 - Requires accurate data on current conditions

Which Census Data Sources MPOs Use

- Decennial Census population/household /GQ counts
 - Can roll up block geography to custom (TAZ) geography
- Census Transportation Planning Products
 - Household Size, Vehicles Available, Workers in Households for model validation

How an MPO uses Census Data

- CB data is a travel forecasting model input
- Modelers view CB data as “truth”
 - Learning that it contains error is unsettling
 - Generally data analysts have gone over data for reasonableness before it goes to modelers
- Data is used for historical trends with decennial data
- Generates base Pop/HH/Employment counts which are used to generate forecasts
- Federal requirements to measure income, means of transportation, etc. in formal grant applications

Margins of Error

- When do I give 'weight' or credence to MOE, when do I ignore it?
 - I have to determine what level of geography is appropriate and reliable
 - Depends on how CB data is being used – often it is a comparison with a different data source or a sniff test
 - Will the data when presented take on a life as its own – will it be 'truth'? (particularly with elected officials)
- Its helpful that the CTPP MOEs account for perturbation error and ACS sampling error so we have a good measure of the uncertainty of the estimate

The CTPP is Coming!

- Adjust geography (Census TAZ to COG TAZ)
- Jurisdictional checks for reasonableness with base year from Forecasts (applying local knowledge)
- Make sure all the big lumps are in the right places (proper distributions of the data)
- Determine which tables are essential in my organization and at what geographic levels
 - Are any of these B-Tables?

How do I assess SDC data?

- My first step when I get the SDC data will be:
 - Compare rim totals with non-SDC data
 - More closely examine county level geography in depth for reasonableness
 - Determine if data can be delivered to modelers or not be used
 - If used, document the differences in detail
- If I do not use SDC data, alternatives include Census 2000 data or Household Travel Survey data (2007/8)

Concerns with SDC Data

- If SDC data is different, how to I determine when it is too different to be useful for my users?
- How do I maintain confidence among my users when numbers do not match or if trend lines diverge?
- How much effort do I put into making sure data users know what they are using?

Explaining SDC data

- How many of my users do I educate about SDC?
- I need vocabulary words that I (and others) can remember and fully understand
- I also need to be able to justify my usage or non-usage of all the data (including non-SDC data)
 - For instance, I generally do not recommend using Block Group data for anything unless BGs are combined into custom geography and MOE recalculated

Where Does the User Community Go From Here?

- First, we need data to analyze (CTPP will be released this fall)
- Next, comprehensive analysis needs to be done for usability by multiple organizations
- Depending on the outcome, changes may need to be made to future tabulations

Summary

- ◉ With restrictions on releases and growing concern about disclosure risk, SDC treatments are more prevalent
- ◉ Provided a glimpse of how government data is prepared for public use
- ◉ Continuing to work on language to communicate between data producers and data users, and developing diagnostics to determine what is an acceptable amount of uncertainty to tolerate
- ◉ Questions?