

# Is an On-Line Microdata Tool that Uses American Community Survey Transportation Data Feasible?

Tom Krenzke, Westat; Penelope Weinberger, AASHTO;

Jianzhu Li, Westat; and Elaine Murakami, FHWA

November 5, 2013

FCSM Conference, Washington D.C.

# Outline

- Census Transportation Planning Products (CTPP)
- On-line Analytic Systems (OAS)
- Study objectives
- Microdata Analysis System (MAS) development
- Possible scenarios and approaches

# 2006-2010 CTPP

- Pre-specified tables generated from 2006-2010 American Community Survey (ACS) data
- Tables generated are:
  - Part 1 Residence
  - Part 2 Workplace
  - Part 3 Flows
    - e.g., Mean travel time

## 2006-2010 CTPP (2)

- Disclosure risk
- Set A tables
  - Census Bureau Disclosure Review Board (DRB) suppression rules are defined for a very small proportion of the Set A tables
  - Generated tables from original ACS microdata
- Set B tables
  - Census Bureau DRB rules
  - Generated tables from perturbed microdata
  - Rules lifted

# 2006-2010 CTPP (3)

## Challenges/limitations

- Small geographies
- Large undertaking
- Done periodically, after each Census
- What's good for some, may not be good for others
- New analytical interests arise

## Desire for flexibility

# OAS

- Online analytic system (OAS) types
  - Generate from public pre-generated tables – no issues
  - Generate from public microdata – no issues
  - Generate from restricted microdata – disclosure risks
- OAS examples using restricted microdata
  - Developing Microdata Analysis System (MAS) at Census Bureau
  - Developing National Health Interview Survey (NHIS) On-line Analytic Real-time System (OARS) (Gentleman, 2011) at National Center for Health Statistics
  - Australian system (Tam, 2011)
- Table differencing attack

# Study Objectives

## OAS for CTPP Data

- Investigate the feasibility
- Investigate the value added (important features to transportation planners)
  - Flexibility
    - Variables, Categories of variables
  - Updated information
    - Possibly fill the gap between periodic CTPP releases
  - Flows and small geography

# MAS

## Selections from the literature

- Initial paper on advanced queries relating to American Factfinder, and query rules (Rowland and Zayatz, 2001)
- Initial work on disclosure risk in regression models (Reznick, 2003, Reznick and Riggs, 2004)
- Discussion of disclosure issues in designing a system (Steel and Reznek, 2005)
- Status updates (Lucero et al. (2009), Lucero, et al. (2011), Freiman, et al. (2011))
  - Investigates drop Q rule, marginal threshold rule, cutpoints for continuous variables, testing residuals for normality, graphs: Histograms, scatterplots, discussion of other systems



# MAS (2)

## Status – Early 2013

- Test data are from the ACS
- Re-focused toward generating tables
- Beta version to release internally in 2013
- DataFerrett
  - Host system planned for the MAS
  - Table generator system
    - Currently uses public use data for real-time table generation

# MAS (3)

## Meeting transportation needs

- Support team (transportation planners, MAS developers, DRB, ACSO, Web staff, graphics, Security, Systems, Sr Managers)
  - Complex (cuts across staff)
  - Identify ownership of system
- What's the best vehicle for the MAS for transportation needs?
- Interface of MAS/DataFerrett transportation needs -- Focus group
- Performance – Desirable if results returned within 10 seconds
- Policies
  - Security -- Running off restricted use data
  - Consistency with already-published estimates
    - Confidentiality protections, like Drop Q, may add noise

# MAS (4)

## Meeting transportation needs (continued)

- Weighted data – are current disclosure protections sufficient?
  - *Specialized cell-based* subsampling provides protection (Krenzke, et al., 2013)
    - Used random seeds and information from the query, and attain same results between users, and address extreme attacks
- Clustered data – valid variance estimates produced by the MAS?
- An alternative would have to:
  - Be packaged within Census Bureau walls near the restricted data
    - Would Census security policy allow?
  - System protections would need to be approved by the DRB

# Scenarios

## General Setup for Consideration

- Smallest geography
  - Traffic Analysis District (TAD)
    - About 20,000 in population
  - For flows – TAD to TAD
- Types of analysis and tables
  - Same as for CTPP 2006-2010
- Microdata
  - 5 year combined ACS
- Frequency of updates
  - Annual, or every 3 years

# Scenarios (2)

## Option 1






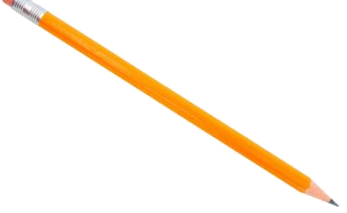
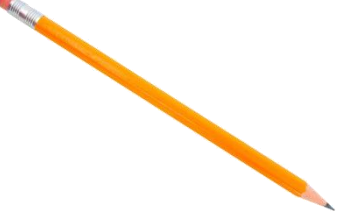

- OAS from microdata, limited to CTPP table shells
- Query restrictions setup to produce the 2006-2010 CTPP tables on demand – no pre-generated tables
- ‘Set A’ microdata file and ‘Set B’ microdata file
  - Use of file triggered by the query submitted

## Option 2 – More Flexibility

- All 2-way tables can be queried within geographic areas, or within TAD-to-TAD flows
- More detailed categories
- Restrictions determine what is accepted after submitting the query

# Scenarios (3)

## Comparison of 2 Options (Illustrative, Not to Scale)

Item	2006-2010 CTPP	Option 1	Option 2
Time, Cost, Staff	\$\$\$\$\$\$\$\$\$\$\$\$	\$\$\$\$	\$\$\$\$
DRB involvement			
Perturbation	<i>Perturbation</i>	<i>Perturbation</i>	<i>Perturbation</i>
Complexity to develop system	NA		
Flexibility			

# Summary

- Transportation planners desire more flexibility
- An OAS may be a solution – options provided
- Feasibility
  - Limited due to...
    - Confidentiality and security
    - Hosting, if alternative to MAS is considered
    - Progress
    - Data prep
  - Optimistic about value added
    - Usefulness
    - Potentially cost effective

# References

- Freiman, M., Lucero, J., Singh, L., You, J., DePersio, M., and Zayatz, L. (2011). The Microdata Analysis System at the U.S. Census Bureau. In *Proceedings of the Joint Statistical Meetings of the American Statistical Association, Section on Survey Research Methods*.
- Gentleman, J. F. (2011). A real-time online system for analyzing restricted data from the U.S. National Center for Health Statistics' National Health Interview Survey. In *Proceedings of the 58th World Statistics Congress of the International Statistical Institute*. Retrieved from: <http://isi2011.congressplanner.eu/pdfs/650208.pdf>.
- Krenzke, T., Gentleman, J., Li, J., and Moriarity, C. (2013). Addressing disclosure concerns and analysis demands in a real-time online analytic system. *Journal of Official Statistics*, 29(1), 99-134.



## References (2)

- Lucero, J., Zayatz, L., Singh, L. (2009). The current state of the microdata analysis system at the Census Bureau. In *Proceedings of the Joint Statistical Meetings of the American Statistical Association, Section on Government Statistics*.
- Lucero, J., Freiman, M, Singh, L., You, J. DePersio, M. and Zayatz, L. (2011b). The Microdata Analysis System at the Census Bureau. *SORT-Statistics and Operations Research Transactions* **35**, 77-98. Special issue on privacy in statistical databases. Privacy in Statistical Databases.
- Reznek, A. (2003). Disclosure risks in cross-section regression models. In *Proceedings of the Joint Statistical Meetings of the American Statistical Association, Section on Government Statistics*.
- Reznek, A., and Riggs, T. (2004). Disclosure risks in regression models: some further results. In *Proceedings of the Joint Statistical Meetings of the American Statistical Association, Section on Government Statistics*.

## References (3)

- Rowland, S., and Zayatz, L. (2001). Automating access with confidentiality protection: The American FactFinder. In *Proceedings of the Section on Government Statistics, American Statistical Association*.
- Steel, P., and Reznek, A. (2005). Issues in designing a confidentiality preserving model server. *Monographs of Official Statistics*, 9:29.
- Tam, S. (2011). On-line access of micro-data at the Australian Bureau of Statistics: Challenges and future directions. In *Proceedings of the 58th World Statistics Congress of the International Statistical Institute*. Retrieved from: [isi2011.congressplanner.eu/pdfs/650030.pdf](http://isi2011.congressplanner.eu/pdfs/650030.pdf).