Modeling and Imputation Discussion

J. Michael Brick

Westat

March 9, 2016

Suggestions

- Primary user tabulation and files should be simple.

- Data collection should be designed to specific analytic objectives

- Only data collected in the ACS should be labelled ACS data.

Keep It Simple

- Wide variety of users with different levels of sophistication, especially local users, need to be able to implement tables and analyses.

- Mass imputation is possible, but can lead to more difficult inference process that is NOT simple.

- Allow more sophisticated users access in different mechanisms.

Design and Analysis Coordination

- If a design (e.g., matrix sample) is implemented to produce state estimates, then <u>primary</u> analysis mechanisms should support this.

  – Do not mass impute to produce estimates at lower geographies if design for items is for state.

  – This implies tabulation and file structures may need to be re-conceptualized.

Labelling ACS Data

- Administrative and commercial data not collected in the ACS should be labelled as 'linked' data.
  - Data are outside control of survey.
  - Quality of data may vary over time and space.
  - Linking for users should be very simple to support primary analysis. The goal is to keep it simple, but collected and linked data are different.

Example: Phone Service

- Analysis shows commercial data has issues and is not the same as ACS.

- ACS item is not perfect either.

- Uses of admin data:
  - Drop item from ACS and use admin data
  - Impute ACS using admin data
  - Subsample for phone use and impute/substitute with admin data.

Major Implications

- If matrix sampling is used, ensure prime analysis is consistent with this design.

- Imputation retains it current role for filling in for missing data.

  – Admin data has role in imputation

- Allow users to link administrative and commercial data to enhance analysis. But these are <u>not</u> ACS data.

Primary User Analysis Structures

- Existing releases may no longer be appropriate.

- With matrix sampling, releases should be consistent with the design.

- Advanced analyses, that take greater advantage of the resources, need to be encouraged and supported.