
Small Area Estimation: New Developments and Directions for Health and Human Services Data

John L. Czajka¹, Amang Sukasih¹ and Susan Queen²

¹Mathematica Policy Research

²Assistant Secretary for Planning and Evaluation, DHHS

2013 FCSM Research Conference

November 4, 2013 (Session C-2)

MATHEMATICA
Policy Research

Panel meeting agenda

- **Overview of SAE applications represented by the panel**
- **Topics in using SAE**
- **Sources of auxiliary data and issues in using them**
- **New developments in methodology**
- **Improving cross-agency collaboration on methods and data**
- **Summary and closing remarks**

Overview of meeting

- **Held June 4, 2013 at the Hubert H. Humphrey building**

- **37 attendees**
 - 19 panelists
 - 11 invited guests
 - 7 ASPE and Mathematica staff

- **Broad representation**
 - 8 operating divisions within HHS
 - 2 other federal departments in person and 2 more through contractors
 - NSF, OMB, CBO, CNSTAT
 - 4 research organizations
 - 3 universities

Principal findings and observations

- **Few applications within HHS produce annual estimates but show a high level of sophistication and cutting edge methods**
- **Development of an SAE program requires significant time and resources, but these are small relative to data collection**
- **Software limitations are an especially challenging part of development and implementation**
- **Auxiliary variables present a number of challenges**
- **Comparative evaluations of competing approaches are rare**
- **Varied approaches to validation have been used**

Principal findings and observations cont'd

- **Interpretation and communication of results require careful attention**
- **The American Community Survey (ACS) changes the landscape for SAE in a number of ways**
- **Collaboration presents both opportunities and challenges; convening a technical group under FCSM was appealing**

Production applications of SAE within HHS

- **State estimates from the National Survey on Drug Use and Health (SAMHSA)**
- **County estimates of diabetes prevalence, incidence, and risk factors (CDC)**
- **State and county estimates of cancer risk factors and screening (NCI, NCHS, Michigan, Pennsylvania)**
- **State estimates of wireless substitution (NCHS)**
- **State estimates from the Medical Expenditure Panel Survey (AHRQ)**
- **Methodological research at NCHS has explored a number of potential applications**

Requirements to establish an SAE program

- **Technically qualified staff—modeling expertise in particular**
- **Strong programming staff**
- **Resources and time for development and evaluation (but these are dwarfed by data collection costs)**
- **Communication skills—for informative presentation and interpretation to users**
- **Hiring qualified staff to do production work can be difficult; alternative options exist:**
 - Collaboration (NCI and NCHS with university faculty)
 - Contracting (SAMHSA with RTI)
 - Assemble an expert panel (Census Bureau)

Software issues

- **Software packages reduce the need for high-level programming staff, but users report issues with convergence and excessive run time**
- **Limited model diagnostics—for example, in detecting over-specification—also an issue with some packages**
- **Some found it necessary to program the entire application in C or R or even FORTRAN**
- **Writing one's own software requires higher level programming staff and greater statistical expertise—not always an option**

Auxiliary variables present challenges

- **Program microdata have played key roles in SAE, but access and use are restricted; aggregate data are more readily available but also more limited**
- **Quality is important, and relying on evaluations performed by the data producer may be insufficient**
- **Potential impact of data anomalies underscores importance of consulting with program staff**
- **Too much time may be spent finding potential covariates when basic variables may work just as well**
- **Variables used in a model are removed from future analyses of area variation**
- **If small area estimates are to be used to measure change, choice of variables may need to reflect this**

Comparative evaluations of methods are rare

- Literature search identified few examples of comparative evaluations of methods
- Typically, agencies explore alternatives in literature and identify an approach that is consistent with their objectives, data and resources
- Effort required to develop and test a competing approach discourages researchers from developing alternatives to the point of empirical evaluation
- Consequently, we know less, empirically, than we might like about the comparative strengths and weaknesses of the major approaches—and variants on those approaches

Varied approaches to validation

- **Validation of estimates can present a significant challenge because reliable estimates are generally limited—if they exist**
- **Simulations using an artificial population have been used in several programs; subsamples are drawn, estimates created, and compared to “truth”**
- **One program used a large survey with related measures and constructed maps for comparison**
- **Preservation of known correlations was another approach**
- **Cross-validation has also been used: removing a subset of areas, re-estimating the model, and applying it to these areas**

Interpreting and communicating results

- **No less than with direct estimates, producers of small area estimates need to consider how best to communicate variability of estimates**
- **Users interested only in point estimates are a particularly difficult challenge**
- **Some producers have developed informative graphics to assist users in making comparisons across areas**
- **Maps can be exceedingly useful in validation, interpretation, and communication**
- **Meeting with stakeholders can be invaluable**

Impact of the ACS

- **Eliminates long-form variables previously used in models and validation**
- **Provides direct estimates for many characteristics and areas that required SAE previously**
- **Provides wide range of new auxiliary variables**
- **Eliminates the need to deal with a variable time lag in incorporating prior census results into models**
- **Expands opportunities for SAE through pairing with other surveys**
- **Shows limits of even large scale surveys**
 - Precision requires 3 or 5-year averages for most substate areas
 - SAE can provide more timely estimates for these areas

Improving cross-agency collaboration

- **Collaboration between agencies can be very useful**
 - Access to broader staff expertise
 - Access to data
- **Challenges to collaboration can be significant—particularly across departments**
- **Common interest in a successful collaboration is critical**
- **Panelists were very receptive to OMB idea of creating a group under the FCSM to communicate on a more regular basis**

Products of the project

- **Literature review focusing on applications (included in final report)**
- **Summary of the meeting presentations and discussion (included in final report)**
- **Slides from panel presentations on SAE methods (to be posted to the ASPE website)**
- **Final report synthesizing background materials and panel discussion (to be posted to the ASPE and Mathematica websites)**

For More Information

- **Please contact**
 - **John L. Czajka**
 - jczajka@mathematica-mpr.com
 - **Susan Queen**
 - susan.queen@hhs.gov