Small Area Estimation: New Developments and Directions for HHS Data

Final Report

March 31, 2014

John L. Czajka
Amang Sukasih
Alyssa Maccarone

MATHEMATICA
Policy Research

This page has been left blank for double-sided copying.

Small Area Estimation: New Developments and Directions for HHS Data

Final Report

March 31, 2014

John L. Czajka
Amang Sukasih
Alyssa Maccarone

Contract Number:
HHSP23320095642WC/HHSP23337038T

Mathematica Reference Number:
40143.301

Submitted to:
Office of the Secretary
Dept. of Health and Human Services
200 Independence Avenue, SW
Washington, DC 20201
Project Officer: Susan Queen

Submitted by:
Mathematica Policy Research
1100 1st Street, NE
12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763
Project Director: John L. Czajka

MATHEMATICA
Policy Research

This page has been left blank for double-sided copying.

ACKNOWLEDGMENTS

The authors would like to acknowledge the contributions of several individuals to this report. We especially want to thank the panelists—named in Appendix B—who provided descriptions of their respective agencies' production and use of small area estimates, and we want to express our appreciation to all who attended the meeting—named in Appendix A. Their thoughtful discussion of the issues underlies the most important parts of this report. We are grateful as well to our colleagues Scott McCracken, the late Sameena Salvucci, and Frank Yoon for their contributions to the literature review; Allen Schirm for his careful review of earlier drafts; and Alfreda Holmes for preparing the final manuscript. Lastly, we want to thank our project officer, Susan Queen, in the Office of the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services, the sponsoring agency, for her experienced guidance throughout the project, and both Susan and her colleague Joan Turek for their helpful comments on the draft report.

This page has been left blank for double-sided copying.

CONTENTS

IV *(continued)*

EXECUTIVE SUMMARY

Household surveys provide timely estimates of a wide range of population characteristics and play a critical role in policy analysis and development in the federal government, but the sample designs of most household surveys do not support estimates below the national level. Program administrative data avoid the limits of sample size when they include the entire population of participants in a program, but often their contents are narrowly focused, the universes that they cover are narrowly defined, and they do not include program-eligibles who are nonparticipants.

Small area estimation (SAE) methods were developed to address the need for estimates of characteristics for geographic areas or other domains (population subgroups, for example) that cannot be obtained directly from survey or administrative data. Many of the techniques used in SAE combine data from surveys and administrative sources. With advances in data processing, it's now feasible to apply estimation methods that were impractical not that long ago. Applications of SAE by federal agencies have multiplied in the past two decades.

Greater sharing of information about methods, data sources, and how to present small area estimates would be useful to all who are involved in preparing such estimates—but especially to those who are new to this endeavor, those who have more limited resources than the major statistical agencies, or those who continue to use less effective methods.

Seeing the opportunity for a significant step forward, the Office of the Assistant Secretary for Planning and Evaluation (ASPE) in the U.S. Department of Health and Human Services (HHS) undertook the planning of an expert panel meeting to bring together practitioners of SAE to share their insights and jointly explore possibilities for improving their methods and outcomes. ASPE contracted with Mathematica Policy Research to (1) identify prospective panelists and other attendees and develop an agenda, (2) prepare background materials for the meeting participants, (3) facilitate the meeting and prepare a summary of the panel discussion, and (4) synthesize the background materials and the panel discussion into a final report. The broadly written objective of the project was to assist agencies within HHS and the broader federal statistical community in making more effective use of SAE to address data gaps for small geographic areas and populations. This final report represents the culmination of the project but not the broader effort.

Panel Meeting

The panel meeting was held on June 4, 2013, at the Hubert H. Humphrey Building with 37 attendees, including ASPE and Mathematica project staff. Seven of the 11 operating divisions of HHS were represented, along with seven other federal agencies and the Committee on National Statistics. Panelists included faculty from three universities and members of four research organizations. The agenda included the following topics:

- 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 interagency collaboration on methods and data
- Summary and closing remarks

The meeting was convened at 9:00 a.m. and concluded at 4:00 p.m.

Small Area Methods

A concept central to SAE is that of "borrowing strength," which means using data from other areas or time periods to increase the effective sample size for an estimate. For example, a common method of borrowing strength to improve the precision of an estimate from an annual survey is to calculate an average of the survey estimates over multiple years. If the samples in consecutive years were independent, then combining two years of data would double the effective sample size and combining three years of data would triple the effective sample size.

Estimates from other areas or time periods are generally not identical in expectation to the area and time period of interest. Consequently, the estimates obtained by pooling areas or time periods are biased. If the bias is small enough, however, the reduction in variance due to pooling may be great enough to produce a net reduction in mean squared error—a measure of total error equal to the sum of the variance of an item and the square of its bias. This is the goal of SAE—to reduce the variance of a small area estimate by more than enough to compensate for the bias that is introduced by drawing sample observations from areas or times that are not identical.

Most of the small area methods widely used today incorporate two features. First, they borrow strength through the use of predictive models that express the relationship between the small area characteristic and a set of auxiliary variables measured for each area of interest. The models are estimated by regressing an unbiased survey estimate of the characteristic on the auxiliary variables—which often come from administrative records or census data and, therefore, are measured with little or no sampling error. Second, the small area estimate is calculated as a weighted sum of the direct survey estimate and the prediction from the model. The weights for a given area depend upon the relative accuracy of the two estimates. Generally, the less precise the survey estimate, the more weight is given to the model-based estimate. If an area has no sample observations, all of the weight goes to the model-based estimate—that is, the small area estimate equals the model prediction. In this situation, which is common for substate estimates, an estimate for the small area would not be possible without some form of small area methodology.

Findings

Within HHS the use of SAE methods to produce small area estimates on a regular basis is limited, but these applications show a high level of sophistication and cutting-edge methods. Small area estimates are produced currently by the Centers for Disease Control and Prevention (CDC), the National Center for Health Statistics (NCHS), the National Institutes of Health (NIH), and the Substance Abuse and Mental Health Services Administration (SAMHSA).

The development of a program of small area estimates requires significant expertise, time, and resources. Requisite skills include statistics or econometrics with an emphasis on modeling, high-level programming, and strong communication ability. Development and implementation can take two or more years. Nevertheless, these costs are dwarfed by the expense of undertaking new or expanded data collection to enhance the available set of direct estimates.

Most of the formal SAE methods used by federal agencies and their contractors are computationally intensive. As such, they require specialized software or specialized routines within larger software packages. Computational speed is a particular issue for users, as is the extent of diagnostic information that the software can provide. Some of the most sophisticated users of SAE methods have written their own software to address these limitations.

Auxiliary variables play a critical role in the production of small area estimates. Administrative records are a valuable source of auxiliary variables because they can provide very precise estimates for small domains. In addition, the quality of measurement in administrative records can exceed that of comparable survey estimates—although, quality may vary widely across items, across areas, and over time. This underscores the importance of consulting with agency staff when using an agency's administrative data.

Decennial census long form data have been an important source of auxiliary variables as well, but the replacement of the census long form by the American Community Survey (ACS) changes the landscape for small area estimates in a number of ways. For states and large substate areas, annual ACS estimates have sufficient precision to use either as direct estimates alone or as predictors in small area models. For smaller areas, however, only three-year or five-year averages are published and their precision may be insufficient for many applications. So far, the use of ACS data in SAE applications outside the Census Bureau has been limited, but it is likely to grow.

Comparative evaluations of alternative approaches are not as common as prospective users might wish. As a result, too little is known about how different SAE approaches compare when applied to the same problem. Agency staff developing an application may not be able to determine from existing research what method is best for their needs. Instead, they may learn what methods have been applied to similar problems and, if they choose one of these methods, what issues they can expect to encounter and, if the method has been widely used, how they can address those issues.

How a set of small area estimates is likely to be used may have implications for how they are created and how they are evaluated. For example, small area estimates produced by federal agencies may be used for funding allocation, where the quality of the estimates has direct implications for how well federal funds are targeted and how efficiently they are distributed. If the estimates for individual areas vary too much from year to year, program administrators or legislators may respond by establishing hold-harmless provisions or other administrative remedies that undermine the use of the estimates. Statistical solutions would be preferable but may be difficult to implement once the estimates are in use.

Validation of small area estimates remains a significant challenge. Given the critical role of modeling, careful examination of model diagnostics—that is, internal validation—is common. Evaluations of the end results apply a variety of methods. Simulations using an artificial population are becoming more popular. Comparisons using the most precise direct estimates (for example, the largest states) are a long-standing approach. Estimates have also been evaluated on how well they preserve known correlations and whether adjacent areas on maps show similar results. Cross-validation has also been applied.

Producers of small area estimates may have to explain why their estimates are better than direct estimates and communicate an intuitive sense of how they were derived. Good graphics and maps can be helpful in demonstrating the implications of statistical uncertainty. When there are explicit stakeholders who will use the estimates, meetings with these groups can be helpful in ensuring that the underlying concepts are understood and that the stakeholders' concerns are addressed.

Collaboration presents both opportunities and challenges, yet the sharing of technical resources across agency divisions and even across agencies may be necessary to more fully exploit the potential of small area methods. Models of successful collaboration—such as the one among the National Cancer Institute (NCI), NCHS, and two universities—can be studied for ideas on how to make such ventures effective. Interagency working groups can be helpful for sharing ideas, keeping members apprised of ongoing activities, and discussing new ventures that are under consideration. Panelists responded enthusiastically to an offer to create a working group on SAE under the Federal Committee on Statistical Methodology.

Recommendations

One of the goals of this project was to help ASPE communicate to others in HHS what small area methods can and cannot do to address data needs involving states, substate areas, and other small domains. The successful SAE programs at CDC, NCHS, NCI, and SAMHSA—along with the ongoing work at the Agency for Healthcare Research and Quality (AHRQ) to develop county estimates for its quality indicators program—illustrate (1) the potential of these methods to fill data gaps, (2) the amount of effort that may be involved, and (3) the types of data that must be assembled. We recommend that ASPE use these examples to convey a realistic appreciation of SAE.

Other agencies in HHS that routinely use small area statistics recognize the value of such data but may not have the requisite staff or resources to develop their own small area estimates. We suspect that each of these agencies may be able to identify needs for expanded small area data in the future. Should that occur, we recommend that each agency explore potential collaboration with NCHS, which has the deepest staff with significant experience with SAE techniques.

Administrative data collected by the Centers for Medicare & Medicaid Services (CMS) could provide a potential source of new auxiliary variables for applications of SAE within HHS. We recommend that ASPE monitor applications within the operating divisions to identify needs that could be addressed by CMS data and, where such needs exist, encourage collaboration. We also recommend that ASPE explore interest in state estimates from the Medicare Current Beneficiary Survey (MCBS). If interest is identified, then we suggest that ASPE facilitate a collaborative program between CMS and NCHS to produce such estimates.

The Division of Program Statistics in the Indian Health Service (IHS) produces a major report, *Regional Differences in Indian Health*, that relies heavily on decennial census data. With the elimination of the census long form, many of the estimates in this report will have to be obtained from the ACS. This opens up the possibility of producing estimates more often than once a decade to better document trends, but sample size becomes an issue. We recommend that IHS assess the merits of using SAE methods to produce more frequent statistics on American Indians and Alaska Natives.

Although the Health Resources and Services Administration (HRSA) does not require SAE methods in order to identify Health Professional Shortage Areas, the application of small area methods could enable HRSA to broaden the way they define these areas. We recommend that HRSA explore whether there might be value in this approach.

Finally, some mechanism for sharing experiences with SAE would serve a broad array of current and prospective users within the federal government. We encourage the Office of Management and Budget (OMB) to pursue the establishment of a working group on small area methods and applications. We also encourage ASPE to participate in this effort and inform the operating divisions of HHS of developments that are communicated through this group.

This page has been left blank for double-sided copying.

I. INTRODUCTION

Household surveys provide timely estimates of a wide range of population characteristics and play a critical role in policy analysis and development in the federal government, but the sample designs of most household surveys do not support direct estimates below the national level. Those that do can provide only limited estimates below the state level, with the exception of the American Community Survey (ACS). Even the ACS requires that sample data for most counties and nearly all smaller areas be aggregated over three to five years to provide sufficient sample size for statistically reliable estimates. This greatly limits the value of such estimates for monitoring trends. The cost of expanding most household surveys to provide statistically precise annual estimates at the state level—much less for geographic areas or other domains below the state level—is generally prohibitive. The ACS, with an annual sample of two million interviewed households, demonstrates that even a sample of that size encounters serious limits in the precision of its estimates for small areas.

Program administrative data avoid the limits of sample size when they include the entire population of program participants, but their contents are often narrowly focused, and the universes that they cover may be very limited as well. Moreover, persons who are eligible for the program but do not participate are excluded. Tax records depart from most administrative record systems in the breadth of their coverage of the population, but the thousands of items that they collect span an exceedingly limited range. For example, tax records lack even the most basic information on the demographic characteristics of taxpayers and their dependents.

Small area estimation (SAE) methods were developed to address the need for estimates of characteristics for geographic areas or other domains (population subgroups, for example) that cannot be obtained directly from survey or administrative data. Many of the techniques used in SAE combine data from multiple sources and often include both a survey component and an administrative component. With advances in data processing capacity and computational algorithms, it has become feasible to apply estimation methods that were impractical not long ago. Applications of SAE by federal agencies have multiplied in the past two decades. The ACS, by extending the range of survey estimates below the national level, has also increased the potential reach—and quality—of small area estimates.

Greater sharing of information about methods, data sources, and how to present small area estimates would be useful to all who are involved in preparing such estimates—but especially to those who are new to this endeavor, those who have more limited resources than the major statistical agencies, or those who continue to use less effective methods.

Seeing the opportunity for a significant step forward, the Office of the Assistant Secretary for Planning and Evaluation (ASPE) in the U.S. Department of Health and Human Services (HHS) undertook the planning of an expert panel meeting to bring together practitioners of SAE to share their insights and jointly explore possibilities for improving their methods and outcomes. ASPE contracted with Mathematica Policy Research to (1) identify prospective panelists and other attendees and develop an agenda; (2) prepare background materials for the meeting participants, facilitate the meeting, and prepare a summary of the panel discussion; and (3) synthesize the background materials and the panel discussion in a final report. The broadly written objective of the project was to assist agencies within HHS and the broader federal statistical community in making more effective use of SAE to address data gaps for small geographic areas and populations. There are significant challenges in applying such methods, to be sure, but ASPE concluded that a broader sharing of information on agency activities in this area would aid and encourage greater and more effective use of the available techniques and data. This final report represents the culmination of the project but not the broader effort.

This report is organized as follows. Chapter II provides a brief overview of SAE methods. Chapter III describes current programs of small area estimates in the federal government, with a focus on HHS. Chapter IV discusses issues in the application of small area methods, as identified during the panel discussion and our literature search. Chapter V presents our conclusions and recommendations. Appendix A contains the panel meeting agenda and a list of attendees and their affiliations. Appendix B summarizes the formal presentations at the panel meeting, and Appendix C summarizes the discussion sessions. Finally, Appendix D contains a literature review that provides a more extensive overview of small area methods and detailed synopses of applications.

II. STATISTICS FOR SMALL AREAS: METHODS OF ESTIMATION

This chapter provides a broad overview of methods used to derive statistical estimates for small areas when direct estimation is not viable. An extended discussion, from which this material was drawn, is presented in Appendix D.

A. Early Approaches to Indirect Estimation for Small Areas

Federal agencies have been producing estimates for states and substate areas for decades. These efforts include county crop estimates from the U.S. Department of Agriculture (USDA) and state and county personal income estimates from the Bureau of Economic Analysis (BEA).[1] The methodologies for producing these estimates have evolved over time and can be expected to continue to do so.

The National Agricultural Statistics Service (NASS) of USDA began producing state and county estimates of crop acreage and production more than 90 years ago, using the Census of Agriculture, which is conducted every five years, and state surveys of farmers as the basic data source—applying adjustments to refine the estimates and ensure additivity to higher levels of geography where desired. As a pioneer in survey sampling, NASS was well positioned to manage this effort. Probability sampling of area frames was introduced in the 1960s to supplement the list frame samples used exclusively in earlier years. Satellite photos were added as a data source when they became available. Because the program is annual, current estimates can be informed by prior year data. Data from the Census of Agriculture enable validation of the estimates every five years, as well as corrections when necessary. Strategies for dealing with sample zeroes and small sample counts generally have been modified over time, taking advantage of some of the methodological developments discussed below.

BEA initiated the publication of aggregate personal income data below the national level in 1939 with the release of the first estimates of personal income by state (FCSM 1993). Counties were added later. Personal income represents one of the five national income and product accounts. The state and county estimates of aggregate personal income are produced by disaggregating national totals. Methods of disaggregation have changed over time as other data have become available. For example, wages paid by employer were an important early source for disaggregating one component of personal income—although the employer data were recorded by place of work while the goal was to report income by place of residence. This necessitated an adjustment, which has grown more sophisticated over time. Some sources of income data actually originate at the state or local level, and they can be estimated directly (in some cases with adjustments). Other sources must be allocated in proportion to related sources for which there is state or local information.

B. Small Area Methods

A concept central to small area methods is that of "borrowing strength," which means using data from related areas or time periods to increase the effective sample size for an estimate. For

[1] These and several other examples were discussed in a report by the Federal Committee on Statistical Methodology (1993).

example, a common method of borrowing strength to improve the precision of an estimate from an annual survey is to calculate an average of the survey estimates over multiple years. If the samples in consecutive years were independent, then combining two years of data would double the effective sample size, and combining three years of data would triple the effective sample size. Likewise, methods of spatial smoothing, which average the estimate for an area with the estimates for surrounding areas, can increase effective sample size several fold.

For a given characteristic, the true values for related areas or time periods may be similar, but they are generally not identical. Consequently, the estimates obtained by pooling areas or time periods are biased. If the bias is small enough, however, the reduction in variance due to pooling may be great enough to produce a net reduction in mean squared error (MSE).[2] This is the goal of SAE—to enhance overall accuracy by reducing the variance of a small area estimate by more than enough to compensate for the bias introduced by using data from areas or times that are not identical.

A limitation of the most simple forms of borrowing strength, which rely on pooling estimates over time or space, is that they require that there be direct estimates for the areas of interest and direct estimates for nearby areas or time periods as well. National surveys may not contain representative samples of all 50 states, and their samples tend to be distributed among just a small proportion of counties, leaving most counties with no representation at all. Furthermore, some important national surveys are not conducted annually, which diminishes the usefulness of multi-year averages by increasing their bias.

Rao (2003)—a standard reference on SAE methods—describes borrowing strength in more general terms, asserting that borrowing strength is accomplished through models that link a small area to other areas or time periods. Models can be either implicit or explicit. Both types of models make use of supplemental data, but explicit models take account of the variation between areas that is not explained by the additional data. Most of the examples of SAE in the federal government discussed in this report involve explicit models, but methods using implicit models remain in use today. Rao describes demographic methods of population estimation and synthetic estimation as using implicit linking models. Composite estimation is presented by Rao as using implicit models as well, although many of the estimators that use explicit models can be expressed as composites as well. Next we discuss these three approaches followed by the methods based on explicit models. We conclude with a brief overview of the new developments in SAE methods since Rao's book was published.

1. Demographic Estimation

Demographers have relied on a variety of methods to develop population estimates for small (and large) areas for the years following decennial censuses. These include the use of demographic accounting, where population change in an area is equal to the base (census) population plus births minus deaths plus net migration (in-migration minus out-migration). Vital statistics provide estimates of births and deaths at the national, state, and local levels, but in the United States there are no administrative data that provide equally complete coverage of migration. Enrollment records for Medicare, which covers more than 90 percent of the population 65 and older, are an important source of information about where part of the

[2] MSE is a measure of accuracy that equals the sum of the variance of an item and the square of its bias.

population lives. Estimates of both gross and net migration have been developed from the addresses reported on tax returns—which cover a very substantial fraction of the population but which systematically exclude those with low incomes. Demographers have developed procedures—including regression models—to estimate net migration from "symptomatic indicators," which include changes in housing stock and school enrollment, for example. These approaches differ from most of the other small area methods discussed in this report in that they generally do not involve survey data.

2. Synthetic Estimation

Synthetic estimation encompasses a class of estimators that use relationships measured reliably for a large area to estimate the characteristics of small areas within the large area. In its simplest form, with no auxiliary variables, synthetic estimation uses the estimated mean or proportion for a large area—for example, the unemployment rate—to estimate the mean or proportion for a small area within that larger area. If information on the composition of the area population is available, that information can be incorporated into the synthetic estimator. Continuing with the unemployment example, estimates of the unemployment rate for subgroups that can be identified at the small area level (for example, subgroups defined by age, sex, race, or ethnicity) can be applied to the composition of the population (the auxiliary information) at the small area level. The small area estimate of the unemployment rate is thus a weighted sum of the subgroup unemployment rates for the larger area, where the weights reflect the subgroup composition of the small area. Synthetic estimation was used in a number of federal agencies in the 1970s, but its use has diminished with the development of better estimators—and the computational power to apply them. A disadvantage of synthetic estimation is that it understates variability across the small areas because of its assumption that the subgroup rates are identical within the larger area.

3. Composite Estimation

A composite estimator combines two estimates. As a general methodology, composite estimation may be the most widely used approach in the federal government today to producing small area estimates. Typically, the composite estimator of a small area characteristic is a weighted average of a direct estimate and an alternative estimate, where the weights are ωi and $(1 - \omega i)$ with ωi bounded by 0 and 1. The general idea is to select weights that minimize the error of the composite estimate, given the error associated with each of the components, but this can be approached in different ways. The simplest approach assigns a uniform value to ωi for all areas. More generally, the weights reflect estimates of the relative variability of the direct estimate and the alternative estimate, with ωi approaching 1 as the direct estimate becomes more precise. If there is no direct estimate for area i because no sample observations fall into that area, then ω_i will be zero.

4. Methods with Explicit Models

Predictive models borrow strength by using data from multiple areas or multiple times—and often from multiple data sources—to estimate the model parameters. For example a predictive model estimated from state-level data uses information from all the states to make predictions for each individual state based on the values of the predictors for that state.

The models used for SAE often include both fixed and random effects, with the latter being used to represent variance contributions that are area-specific. When linear models involve combinations of fixed and random effects, they can be estimated by a type of estimator called best linear unbiased prediction (BLUP). This is notable because the BLUP estimator can be expressed as a weighted sum of a direct estimate and a regression synthetic estimate, which gives it the form of a composite estimator (Rao 2003). A limitation of BLUP estimation is that it depends upon variance parameters that are typically unknown, but empirical best linear unbiased prediction (EBLUP) provides a way to estimate a BLUP model.

The EBLUP estimator can be used with linear mixed models, which encompass a wide range of SAE applications, but such models are inappropriate for binary data or count data (Rao 2003). Moreover, accurate estimates of MSE for EBLUP estimators generally require that the errors be normally distributed—an assumption that is often not satisfied. Empirical Bayes and hierarchical Bayes methods can handle both binary data and count data and do not depend upon normality. With a linear mixed model with normally distributed errors, empirical Bayes and EBLUP are identical.

Empirical Bayes and hierarchical Bayes methods typically use simulations to generate the parameter estimates of the models. Because of that, they require extensive computations. A fully Bayesian approach, such as hierarchical Bayes, begins with a prior distribution of the parameters to be estimated. This prior distribution may be based on earlier research (an "informative" prior) or not (a "diffuse" or "noninformative" prior). The final, "posterior" distribution is derived from the prior distribution, given the data. Bayesian methods offer greater flexibility than non-Bayesian methods, but estimation is computationally more difficult because Bayesian estimators typically capture more sources of error. Different approaches have been developed to estimate such models. Markov Chain Monte Carlo (MCMC) methods are popular currently and are found in the most widely used software. It should be noted, however, that with sophisticated models and estimation techniques the distinction between empirical Bayes and fully Bayesian approaches blurs, such that it may be difficult to characterize a particular application unambiguously as one or the other.

5. New Developments in SAE

Pfeffermann (2013) reviewed developments in SAE in the years since the publication of Rao's book. Calibration of the survey base weights to known totals of auxiliary variables can improve the precision of design-based estimates of small areas. Model-dependent and model-based direct estimates are another development that may be useful when sufficient sample data for direct estimation are available. Developments in model-based SAE include (1) new ways of estimating the MSE of the prediction model, (2) computation of prediction intervals, (3) new approaches to benchmarking, (4) ways of accounting for measurement error in the auxiliary variables, (5) identification of outlying estimates, (6) methods to enhance the robustness of models, and (7) new approaches to the problem of predicting ordered means.

III. SMALL AREA ESTIMATES PROGRAMS IN THE FEDERAL GOVERNMENT

Small area estimates at the state and substate level and small domain estimates for population subgroups are produced by a number of federal agencies. In addition, several agencies have active research and development programs exploring the application of small area methods to new data series. This chapter reviews all programs in HHS, selected programs in other federal agencies, and—to illustrate their growing reach—one state government.

A. Programs in HHS

- HHS has 11 operating divisions, some of which include major subdivisions:
- Administration for Children and Families (ACF)
- Administration for Community Living (ACL)
- Agency for Healthcare Research and Quality (AHRQ)
- Agency for Toxic Substances and Disease Registry (ATSDR)
- Centers for Disease Control and Prevention (CDC), which includes the National Center for Health Statistics (NCHS)
- Centers for Medicare & Medicaid Services (CMS)
- Food and Drug Administration (FDA)
- Health Resources and Services Administration (HRSA)
- Indian Health Service (IHS)
- National Institutes of Health (NIH), which includes several institutes, among them the National Cancer Institute (NCI)
- Substance Abuse and Mental Health Services Administration (SAMHSA)

Among these, CDC, NCHS, NCI, and SAMHSA have small area estimates programs that produce state-level or county-level estimates on a more or less regular basis. These established programs and ongoing research into potential future applications are discussed below.

1. Established Small Area Estimates Programs

Small area and small domain estimates produced by SAMHSA, CDC, NCI, and NCHS are discussed below. More detailed descriptions of particular programs of estimates are provided in Appendices B and D.

a. State Estimates from the National Survey on Drug Use and Health

SAMHSA conducts the National Survey on Drug Use and Health (NSDUH) and publishes age-group-specific state prevalence estimates for 25 behaviors, including past-month or past-year use of a number of illicit drugs, alcohol, and tobacco products (SAMHSA 2012). The estimates are derived from a survey-weighted hierarchical Bayes methodology (Folsom et al. 1999) that has been used and refined since the early 1990s, when it was first applied to a predecessor of the NSDUH. Predictors used in the regression models are drawn from a number of sources, including the Census Bureau, the Federal Bureau of Investigation, HRSA, the Bureau of Labor Statistics, BEA, NCHS, SAMHSA, and the firm Claritas. Because the modeling is done at the person level, most of the predictors represent substate geography. Recently, SAMHSA has begun to use two-year moving averages to produce the state estimates. Estimates from the 2011 and 2012 surveys can be retrieved from the following web page, which also contains links to extensive documentation and estimates from other years: http://www.samhsa.gov/data/NSDUH/2k12State/NSDUHsae2012/Index.aspx .

b. County Estimates of Diabetes Prevalence, Incidence, and Risk Factors

CDC produces estimates of the prevalence of self-reported diabetes, the incidence of diagnosed diabetes, and the prevalence of selected risk factors for states and all U.S. counties.[3] All of these estimates use data from the Behavioral Risk Factor Surveillance System (BRFSS), a telephone survey of noninstitutionalized adults designed to produce state estimates of health status and factors affecting health. The approach to estimating diabetes prevalence treats the survey data as observations collected from a larger set of complete data, most of which are unobserved (Cadwell et al. 2010). The unobserved prevalence data are predicted using an expansion of a Bayesian model proposed by Malec et al. (1997), which is estimated from the observed BRFSS data and demographic data from the Census Bureau's population estimates program. To evaluate the model-based estimates during development, the final estimates were compared with direct survey estimates for large counties. Diabetes incidence is estimated using a variation on this methodology (Barker et al. 2013). Because incidence is so much lower than prevalence, the estimates of incidence were evaluated by comparisons with direct estimates at the state level, as direct estimates for even the large counties are too imprecise. This work has been extended to the estimation of selected risk factors—obesity and leisure-time physical inactivity—at the county level. Currently, state-level and county-level estimates of all four sets of indicators for the years 2004 through 2009 are available from the CDC website at: http://www.cdc.gov/diabetes/atlas/countydata/atlas.html.

c. State and County Estimates of Cancer Risk Factors and Screening

Working in collaboration, NCI, NCHS, the University of Michigan, and the University of Pennsylvania developed a model-based methodology and produced state and county-level prevalence estimates of (1) current and past smoking for adult males and females, (2) mammography screening for women 40 and older, and (3) pap smear tests for adult women for the periods 1997-1999 and 2000-2003 (Raghunathan et al. 2007). The methodology combines estimates from two surveys—the BRFSS and the National Health Interview Survey (NHIS)—with county-level demographic and socio-economic information by means of a hierarchical Bayesian model. The BRFSS is a state-based survey with substantially larger samples for small states than the NHIS and a national sample that reached 500,000 in 2011, but its sample frame includes only households with telephones. The NHIS is conducted in person and, therefore, includes households without telephones. The NHIS also has a higher response rate than the BRFSS. Including the NHIS in the estimates provides a means to correct for undercoverage and nonresponse bias. In addition, estimates of smoking behavior and cancer screening from the two surveys for households with telephones differ, providing another reason to combine the two surveys. Regression models predicting county-level outcomes utilize 26 county-level variables obtained from the 2000 census and other sources. Because the outcomes of interest are known to vary by socio-economic status, the county-level variables include measures of per capita income, percentage below poverty, median home value, and the percentage of the population graduating

[3] Prevalence refers to the number of cases or proportion of the population with a specific characteristic. Incidence refers to the number of cases or proportion of the population acquiring or diagnosed with this characteristic over a specified period of time (often a year).

from college. Estimates for states, counties, and health service areas (combinations of counties) can be obtained from an NCI website along with documentation of the methodology: http://www.sae.cancer.gov. Currently, only estimates for the two earlier time periods are available, but estimates for 2004-2007 and 2008-2010 are in production.

d. State Estimates of Households with Only Wireless Telephone Service

The NHIS is unique among national household surveys because it collects data on the type of telephone service used by households in its sample. Because NHIS interviews are conducted in person, the survey is also able to identify households with no telephone service at all. Such households are rare nationally, although the frequency differs among states and by other geographic as well as personal characteristics. Households with only wireless phone service (those that have substituted wireless for landline service) have grown dramatically—to the point where telephone surveys that sample solely from landline phones exclude a substantial proportion of the population with a decidedly different demographic composition than households with landlines (including those with both wireless and landline service). In the past four years NCHS has collaborated with the University of Minnesota and NORC at the University of Chicago to publish four sets of state-level estimates of telephone usage by type. The most recent estimates cover mid-2011 through the end of 2012 and provide estimates for 93 nonoverlapping areas consisting of individual states and county groups (Blumberg et al. 2013). In addition to the NHIS, the estimation methodology uses data from the ACS and information on listed telephone lines per capita for geographic areas. The principal users of these estimates are survey organizations—including federal agencies—that require such information to improve the weighting of their surveys or enhance their sample designs. The estimates are particularly useful for weighting mixed-frame telephone surveys (such as the National Immunization Survey), but organizations have also used the estimates to develop weighting adjustments to compensate for their survey frames' exclusion of households lacking telephones of any kind and the exclusion of households with only wireless service.

e. Model-based Estimates for Small Domains in Vital Statistics

NCHS has also used models that borrow strength over time or space in order to produce estimates of vital statistics for minority race groups in state or county areas.

Each year the Division of Vital Statistics publishes national life tables by race (white and black) and sex. At 10-year intervals the Division also publishes state-specific life tables. In some states the black populations are too small to support direct estimates of the mortality rates needed to construct life tables. In tables published after the 1990 decennial census, life tables for blacks had to be omitted in 18 states. In the most recent publication, NCHS used model-based methods to assist in the development of life tables for blacks in states with 300 to 700 black deaths, reducing (to 11) the number of states without life tables for blacks (Wei et al. 2012). A model was used to estimate the one-year probability of dying by single year of age and sex. The model drew on historical mortality data for the black populations in these states to improve the precision of the estimates.

Another example of borrowing strength involves estimates of population by race at the county level. The 2000 decennial census used the new racial classification issued by the Office of Management and Budget (OMB) in 1997. This new standard for federal data collection allows respondents to report multiple races, rather than the five single-race categories recognized in the

1977 OMB standard. State vital statistics agencies have continued to collect race data using the single-race standard. To calculate birth and death rates by race, it is necessary to bridge the two standards. NCHS has elected to do so by converting the multi-racial population estimates from the Census Bureau into the five single-race categories rather than decomposing births and deaths into the 31 multi-racial categories defined by OMB. Following OMB guidelines, single-race responses in the census are converted directly to the corresponding single-race categories, so the problem reduces to one of converting the responses for multi-racial categories to single-race responses that respondents would have given if only a single race were permitted. The NCHS methodology uses the NHIS, which asks respondents who report multiple races the race with which they identify most closely. NCHS estimated a model predicting the response to this question using a combination of person-level and county level-variables, as bridged data are required at that level (see Ingram et al. 2003 for a complete description of the methodology). The model borrows strength across areas and across time, pooling three years of NHIS data. The results have been applied to both decennial census data and annual postcensal population estimates, which are used as denominators in the calculation of postcensal vital rates. New county-level population estimates for the five single races are published each year.

2. Research and Development

In addition to those programs that are producing small area estimates for public release, some agencies within HHS have active research programs that are exploring potential applications of small area methods or are engaged in the development of estimates that they expect to release in the near future.

NCHS has been a leader in the development of small area estimates to address a range of needs and to demonstrate the value of SAE methods. In addition to the applications discussed in the preceding section, NCHS has conducted research in the following areas:

- The prevalence of diabetes among 11 small domains, including Native Hawaiians and Other Pacific Islanders; this involved the application of RAND's Modified Kalman Filter model to NHIS data
- Fast screening to identify health outcomes that vary by small area; this research focuses on developing a procedure to determine which outcomes have sufficiently large variation across small areas to warrant the application of small area methods
- Small area estimates from the NHIS utilizing block-linked ACS data; this joint work with the Census Bureau is using ACS estimates as covariates to estimate health insurance and access-to-care outcomes for selected states and counties
- Small areas estimates of the adoption of electronic medical record systems by office-based physicians; this work is using data from the Electronic Medical Records supplement to the National Ambulatory Care Survey, with covariates drawn from the Area Health Resources Files

These efforts are discussed further in Appendix B.

At AHRQ, estimates for state and substate areas have been limited by small sample sizes in the agency's major surveys. The agency is exploring ways to expand the number of states and areas for which estimates can be produced. Three such efforts are discussed below.

1. The Household Component (HC) of the Medical Expenditure Panel Survey (MEPS), cosponsored by AHRQ and NCHS, is the premier source of household-level data on the use of and expenditures for health care services in the United States. The survey also collects extensive additional information on health status, health conditions, health insurance coverage, and household income. The sample sizes of individual panels are small, however, and not designed to be state representative. Direct state estimates with acceptable precision can be produced for the largest states for characteristics that are relatively common—AHRQ has published reports with estimates for the 10 largest states. An empirical analysis concluded that the use of SAE techniques would be essential to expand the number of states and extend the estimation to less common variables (Sommers 2005).
2. The Insurance Component (IC) of MEPS is an independent survey of employers that collects detailed information on the health insurance coverage that firms provide. With its large sample size, the MEPS-IC supports extensive tabulations for all states and the 20 largest metropolitan statistical areas (MSAs). AHRQ has explored the possibility of applying SAE in order to expand the number of MSAs for which estimates from the survey can be produced (Baskin and Sommers 2008).
3. AHRQ is also developing experimental county-level estimates of selected health conditions that can be used in conjunction with hospital inpatient discharge data to measure the quality of care for conditions where good outpatient care or early intervention can reduce the risk of more serious illness later. A fuller discussion of this research is presented in Appendix B.

Finally, NCI has three small area projects underway. First, NCI is exploring potential new outcomes for the BRFSS/NHIS modeling discussed earlier. Second, NCI is collaborating with the Census Bureau to develop small area estimates of several tobacco-related measures utilizing the Tobacco-Use Supplement to the Current Population Survey (CPS). Third, NCI is developing state-level estimates of cancer-related knowledge variables using data from the agency's Health Information National Trends Survey (HINTS), which collects nationally representative data on how the American public uses information about cancer. [4]

3. HHS Agencies without Small Area Estimates Programs or Research

Of the operating divisions at HHS that do not currently produce small area estimates, three were represented at the panel meeting. Their representatives provided an overview of how small area estimates might figure into future plans (see Appendix B):

1. The ACL, which was created in 2012 and includes the Administration on Aging, the Administration on Intellectual and Developmental Disabilities, and the Center for Disability and Aging Policy, makes use of data from the decennial census and a number of national surveys, including NHIS and the Health and Retirement Study (HRS). The ACL also compiles state-level and county-level data from other sources. As a consumer of small area estimates, the ACL recognizes the value of such statistics but is not staffed to assume the role of producer.

[4] Conducted by mail, HINTS collects data from between 3,000 and 4,000 respondents. Recent data collection has been annual. Previous surveys were conducted at two to three year intervals.

2. CMS collects and aggregates extensive data on medical claims to produce summary measures of spending, utilization, and quality for regional healthcare markets. CMS also provides access to the individual claims data for research purposes. With direct access to the microdata, CMS has the ability to construct estimates for alternative small domains without employing sophisticated techniques to borrow strength. CMS also administers the Medicare Current Beneficiary Survey (MCBS), which collects survey and administrative data for a sample of Medicare beneficiaries. A longitudinal survey that follows a small sample cohort over four years, the MCBS was not designed to produce subnational estimates. Annual files that combine four cohorts (a new panel is started every year) contain between 12,000 and 16,000 enrollees—too few for direct estimates of all but the largest states but large enough that CMS could consider the application of SAE techniques if there were a demand for state estimates.
3. HRSA produces a major source of auxiliary variables for small area analysis as well as SAE—the Area Health Resources Files (see Chapter IV). The data are compiled from more than 50 sources and currently contain over 6,000 variables at the county level. HRSA has added extensive geographic codes that make it easier for users to link the data to areas in their analyses. HRSA also generates statistics on the healthcare labor force and has developed criteria for defining Health Professional Shortage Areas. Producing the statistics that make it possible to identify shortage areas does not require SAE methods, however, as HRSA has access to population-level data on the geographic distribution and professional characteristics of the health labor force. Whether SAE techniques could enable enhancements to the way that shortage areas are defined is a question for HRSA to consider as it weighs options for future development of these important resources.

In correspondence regarding the panel meeting, representatives at three of the remaining operating divisions presented varying assessments of the potential value of small area estimates in their work.

1. FDA does not produce statistics for geographic areas or other domains, whether small or large. Of the 11 operating divisions, FDA has the least potential use for SAE methods.
2. IHS serves a small minority population and produces periodic statistics to document the socio-economic characteristics of that population as well as to compare members residing in different service areas. The source of those data in the past—the decennial census long form—has been replaced by the smaller but continuous ACS. As IHS determines how to deal with this change in data source, SAE methods may become more relevant to the agency. Agency research staff already have experience with methods of analysis for small area data that share some elements with SAE.
3. ACF compiles and publishes extensive state-level statistics, which the states produce from their administrative records on the Temporary Assistance for Needy Families (TANF) program, child support enforcement, foster care, and adoption. The state data are based primarily on statistics drawn from the full population of records. The states are allowed to base some of their reports on samples, but ACF must approve all sample designs in advance. One of the criteria that the sample designs must satisfy is adequate precision. ACF also uses state estimates from the ACS in some of its reports (for example, estimates of children in poverty). Given the nature of the statistics that

ACF collects from the states and publishes, there is no apparent need for small area estimates to either replace some of these statistics or extend the ACF data into new areas.

B. Programs in Other Federal Agencies

In the previous chapter we discussed long-running programs of state and substate estimates operated by USDA and BEA. Here we focus on much newer programs operated or administered by the Census Bureau, the Food and Nutrition Service (FNS), BLS, the Treasury Department, and the National Center for Education Statistics (NCES). Then we will discuss research and development programs in NCES and other federal agencies.

1. Programs Using Small Area Methods

The first five programs discussed below use empirical or hierarchical Bayes methods to combine direct survey and model-based estimates. The remaining programs employ alternative methods developed specifically for the applications for which they are used.

a. Small Area Income and Poverty Estimates

Under the Small Area Income and Poverty Estimates (SAIPE) program, the Census Bureau produces annual estimates of the following characteristics for states and counties: total number of people in poverty, number of children under 5 in poverty (states only), number of related children ages 5 to 17 in families in poverty, number of children under 18 in poverty, and median household income. To implement certain provisions of the No Child Left Behind Act of 2001, the SAIPE program also produces annual estimates of three characteristics for school districts: (1) total population, (2) the number of children ages 5 to 17, and (3) the number of related children 5 to 17 in families in poverty. The most recent estimates, for 2012, were released in December 2013.

When the SAIPE program was conceived, the decennial census was the only source of household income and poverty statistics below the state level. To produce estimates for postcensal years, the SAIPE program applies a Bayesian methodology that combines direct survey estimates and predictions of these direct estimates from a regression model using predictors drawn from (1) administrative records (federal tax returns and recipient counts from the Supplemental Nutrition Assistance Program (SNAP)—formerly the Food Stamp Program), (2) Census Bureau population estimates, (3) BEA estimates of personal income, and (4) income and poverty statistics from the most recent decennial census. Prior to 2005, the source of the direct survey estimates was the CPS Annual Social and Economic Supplement (ASEC), which is the official source of household income and poverty statistics for the United States. When the ACS reached full scale, in 2005, the ACS replaced the CPS ASEC as the source of direct survey estimates in SAIPE. Under a congressional mandate, the U.S. Department of Education commissioned a study by the Committee on National Statistics of the National Academy of Sciences (NAS) to review the SAIPE methodology. The NAS panel provided extensive recommendations in a series of reports (see, for example, National Research Council 2000). An extended description of the SAIPE methodology is presented in Appendix D. Detailed information on the SAIPE methodology along with the annual estimates can be found on the SAIPE website: http://www.census.gov/did/www/saipe/index.html.

With the ACS, which replaced the decennial census long form in 2010, estimates of income and poverty for geographic areas with at least 65,000 in population are available annually. But estimates for the many counties below this size require three-year or five-year averages, and the ACS does not produce estimates for school districts. For counties below 65,000 in population, SAIPE staff note that the model-based estimates are likely to be more reflective of current conditions than the multi-year ACS estimates.

b. Small Area Health Insurance Estimates

Since 2005, the Census Bureau's Small Area Health Insurance Estimates (SAHIE) program has produced annual state and county estimates of health insurance coverage among persons under the age of 65. The program is partially funded by the CDC National Breast and Cervical Cancer Early Detection Program, which requires some of the estimates that SAIHE produces. The first SAHIE estimates included the number insured, the number uninsured, and the percent uninsured by age-group, race or Hispanic origin (state only), sex, and income level (including either under 200 percent or 250 percent of poverty, depending on the state). For both states and counties there was an additional estimate of children under the age of 19 in families under 200 percent (or 250 percent) of poverty.

As with the income and poverty estimates, the health insurance estimates are produced using a Bayesian methodology that combines direct survey estimates with regression predictions using covariates drawn from administrative records, Census Bureau population estimates, and the 2000 decennial census. The administrative records include aggregated federal tax returns; participation records from Medicaid, the Children's Health Insurance Program, and SNAP; and County Business Patterns. Beginning with the 2008 estimates the ACS replaced the CPS ASEC as the source of the survey estimates. This change enables the production of estimates with greater precision and for an expanded set of income categories (adding under 138 percent and under 400 percent of poverty). These income levels have meaning under the Affordable Care Act, but with a somewhat different definition of income. The most recent SAHIE estimates, for 2012, were released in March 2014. The estimates and extensive documentation can be obtained from the SAHIE web page: http://www.census.gov/did/www/sahie/index.html.

c. SNAP State Participation Rates

Participation in SNAP is underreported in surveys, and underreporting of SNAP participation in the CPS ASEC has grown substantially over time. Consequently, rather than using survey reports of participation, SNAP participation rates are estimated as the ratio of state administrative counts of participants to state estimates of eligible persons derived from a microsimulation model that applies SNAP eligibility rules to sample households in the CPS ASEC. These direct estimates lack sufficient precision in most states, so FNS' contractor, Mathematica, uses a Bayesian shrinkage estimator to derive more precise estimates by combining the direct estimates with regression predictions of participation rates (Cunnyngham et al. 2013). Predictors in the regression model, which is estimated at the state level, include the percentage of the state population receiving SNAP benefits, the percentage of school-age children certified to receive a free lunch, a poverty rate calculated from tax data, and four measures derived from the ACS. The equations are estimated over multiple years. Thus, the estimates borrow strength not only across states but over time as well. A more detailed description of the methodology is presented in Appendix D. The latest estimates and the full methodological report can be found at: http://www.fns.usda.gov/ops/supplemental-nutrition-assistance-program-snap-research. [5]

d. State Estimates of WIC-Eligible Children for Funding Allocation

Also for FNS, Mathematica has estimated the number and percentage of infants and children ages 1 to 4 in families with incomes at or below 185 percent of the federal poverty level and therefore income-eligible for the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), which FNS administers. Program regulations specify that such estimates be used to allocate federal WIC funds to the states (Cunnyngham 2012). The estimates are derived using a Bayesian shrinkage estimator. Direct sample estimates from the CPS ASEC are combined with model-based estimates of eligible percentages obtained from a state-level regression of the direct estimates on predictors drawn from aggregate tax statistics and estimates from the ACS. The methodology is very similar to what is used to estimate the SNAP-eligible population, borrowing strength over time as well as across states. A fuller description of the methodology is presented in Appendix D.

e. State-Level and County-Level Estimates of Adult Literacy

In 2009, for the first time, NCES published indirect estimates of adults at the lowest literacy level for individual states and counties for 1992 and 2003 using SAE techniques that included a single area-level hierarchical regression model and a hierarchical Bayes model that used the MCMC method to derive the estimates (Mohadjer et al. 2009). The models predicted the percentage of adults lacking basic prose literacy, which included those considered nonliterate in English as well as those with prose literacy below the basic level. Such individuals lack the skills required to successfully complete basic everyday English prose literacy tasks, such as comprehending a news story and using that information to accomplish daily goals. Direct estimates of these variables were drawn from the 1992 National Adult Literacy Survey (NALS) and the 2003 National Assessment of Adult Literacy (NAAL). Auxiliary variables used as

[5] Of note, a separate brochure not available on the website includes a graph that shows for each state which states' participation rates were significantly lower or significantly higher than that state's own participation rate.

predictors were obtained from the 1990 census (for the 1992 model) and the 2000 census (for the 2003 model) and other sources and included county-level measures of educational attainment and race and ethnicity distributions, census division, and indicators from state assessments. The 1992 model included the proportion of county residents who were native English speakers while the 2003 model included the poverty rate and the percentage foreign-born instead. State estimates were calculated as weighted averages of the county estimates. NCES notes that the predictive ability of the auxiliary variables was not high and cautions users that the confidence intervals around the estimates are large. Indirect county and state estimates of low literacy are presented on the web page: http://nces.ed.gov/naal/estimates/index.aspx. The state indirect estimates are also provided in appendices of Mohadjer et al. (2009), which can be found on the NAAL website: http://nces.ed.gov/NAAL.

f. Monthly Estimates of State and Local Area Employment and Unemployment

Through its Local Area Unemployment Statistics program, BLS produces monthly estimates of total employment and unemployment for approximately 7,300 areas—including census regions and divisions, states, counties, and cities. Estimates are based on data from several sources, including the CPS, the Current Employment Statistics (CES) program, state unemployment insurance (UI) systems, and the decennial census. Estimates for all substate areas (except seven large areas and their respective state balances) are generated through the "handbook method," which employs a building-block approach that dates back to the 1950s but has been modified a number of times. State UI statistics provide monthly counts of persons receiving UI benefits down to small levels of geography, but the unemployed at any one time include persons who have exhausted their benefits, do not qualify to receive them, or have not applied. None of these additional unemployed persons are included in the state UI statistics and must be estimated indirectly. Different segments of the uninsured unemployed are estimated in different ways. For example, the number of unemployed who have exhausted their benefits is projected by applying survival rates to counts of prior exhaustions by time period.

State-level estimates of employment and unemployment, which rely on the CPS, have been improved in stages through a research program that has focused on progressive enhancements. The state estimates utilize time series methods that take account of autocorrelation and other features in the monthly survey estimates. State-specific models include predictors from state UI data and the CES program. The state estimates are benchmarked to CPS national estimates. Both the state and substate methods are discussed more fully in Appendix D.

g. State Estimates of Potential Program Changes from Microsimulation Modeling of SNAP

In policy analysis, the outcomes of interest and the policy changes under consideration are typically numerous and changing. Although state-level estimates of the impacts of program changes may be desired, it is not feasible to develop small area models for every outcome—or even selected outcomes. To address this problem, Schirm and Zaslavsky (1998) developed a methodology for estimating a vector of state weights that could be assigned to every individual or household in a policy microsimulation model (see also Schirm et al. 1999). Each sample observation receives a nonzero weight for every state, where the state weight is the product of the observation's sample weight and an estimated probability of residing in that state, given the observation's characteristics. Using a predictive model to assign the probabilities gives each observation a relatively greater weight for states with residents that the observation more closely resembles. At the same time, using the entire sample to estimate outcomes for each state increases the precision of the state estimates. Mathematica staff have made use of this methodology in a microsimulation model based on the Survey of Income and Program Participation and used for policy analysis of SNAP.

h. Reweighting a Sample to Support State-level Tax Analysis

The Office of Tax Analysis (OTA) in the Treasury Department uses a microsimulation model to estimate the impact of hypothetical changes to individual tax law. The database for the model is a sample of individual tax returns drawn from the population of tax returns filed in a given year. The sample—which is edited, enhanced with modeled tax variables, and extrapolated to a future tax year—is large enough to support estimates for the nation and the largest states. Estimates for the remaining states, however, lack sufficient precision. To address this, OTA staff applied the methodology of Schirm and Zaslavsky (1998) and generated for each observation in the tax model database a vector of state probabilities. These probabilities were calculated from a set of tabulations produced from unedited data for the entire population of returns. Each element in the vector represents an estimate of the probability that the taxpayer filed in a given state. These probabilities sum to 1 for each record. When applied to an observation's sample weight, these probabilities yield a vector of state weights. Across observations, a given state weight sums to the total returns filed in that state. The state weights can be used to estimate state totals of a wide variety of tax variables. To evaluate the results, state estimates from the model were compared to published tabulations of tax items by state. An OTA staff paper describing the approach to deriving and evaluating the state weights is forthcoming.

2. Research and Development

The Bureau of Justice Statistics (BJS), USDA, and NCES, among others, are engaged in research to develop small area estimates programs for selected data series. BJS has contracted with Westat to develop state and substate estimates of crime victimization using the National Crime Victimization Survey. USDA staff are conducting research to evaluate an SAE methodology for producing county agricultural estimates to expand or replace the traditional estimates program. NCES has contracted with the American Institutes for Research to develop state and some substate estimates from the Trends in International Mathematics and Science Study—TIMSS, which has been conducted periodically since 1995.

C. Programs at the State Level

Many states conduct their own periodic surveys on varied topics. For example, substate variation in health and other conditions is of interest to state policymakers, but even these state surveys lack the sample size to support direct estimates for all but the largest substate areas. The University of Minnesota's State Health Access Data Assistance Center has produced substate estimates of health insurance coverage for Oklahoma and Minnesota. The Minnesota effort, described here, differed from typical SAE modeling applications in that the state survey estimate of the characteristic of interest was supplemented by estimates of the same characteristic from two additional sources from the Census Bureau: the ACS and SAHIE (Graven and Turner 2011). First, county-level health insurance coverage was modeled from the state survey using demographic measures from the ACS and additional auxiliary variables drawn from administrative records and other sources using an hierarchical Bayes methodology. Next—because ACS estimates of health insurance coverage were not yet available for all counties—

ACS estimates for higher levels of aggregation were used to derive county-level estimates, where direct estimates did not exist. Lastly, a simultaneous equations model was used to derive the final county estimates from the outputs of the first two steps and the published SAHIE estimates. All of the county estimates satisfied standards of precision commonly used by the Census Bureau and NCHS.

IV. ISSUES IN THE APPLICATION OF SMALL AREA METHODS

In preparation for the panel meeting, project staff at Mathematica and ASPE developed a set of discussion questions for the panel to address. Each topic in the meeting agenda was given between two and seven questions. [6] The questions were distributed to the panel in advance of the meeting. Several of the panel members responded to some of the questions in their prepared presentations (see Appendix B). Additional responses were obtained during the discussion portion of the meeting (see Appendix C) when the questions were used to frame the issues under each topic. The literature review (Appendix D) provided further information on many of the questions. This chapter summarizes what we learned about each of the topics through the panelists' presentations, the discussion, and our preparation of the literature review. This chapter's structure follows the agenda of the panel meeting.

A. Topics in Using SAE

Six topics related to the use of small area methods were included in the panel meeting. These topics were (1) data needs that SAE can address, (2) implementation issues, (3) choosing a method, (4) identifying suitable auxiliary data, (5) validation, and (6) interpreting and communicating results. These topics are discussed in turn below.

1. Data Needs That SAE Can Address

Historically, one of the major uses of small area estimates has been for funding allocation. The requirements for specific statistics are often written into the law, so in producing such statistics federal agencies are responding to a congressional mandate. SAMHSA's state estimates of drug use were developed to address a requirement in the law. The Census Bureau's SAIPE program had barely started when the Improving America's Schools Act of 1994 specified the use of SAIPE statistics for allocating Title I funds to benefit educationally disadvantaged students—although the law also stipulated that a panel of the NAS review the methodology to determine if it was suitable for this use (National Research Council 2000). Sometimes legislation goes so far as to define the statistics to be calculated and how they are to be used. For example, the law that established CHIP specified that the allotment of funds to the states would be based on state estimates of uninsured low-income children calculated from three-year moving averages of CPS ASEC data (Czajka and Jabine 2002). Later legislation provided funds for a significant expansion of the CPS ASEC sample to increase the precision of these estimates. [7] Even without a legislative mandate, the use of small area statistics for funding allocation has at times provided an incentive to improve the quality of the estimates (National Research Council 2003).

In general, small area methods are needed in two common situations: (1) when direct estimates do not exist, or (2) when direct estimates exist but are too imprecise. The former situation occurs when areas are too small to be allocated sample observations consistently in

[6] The full set of questions is reproduced at the beginning of Appendix C.

[7] The name for the annual supplement—that is, the ASEC—that measures income, poverty, health insurance coverage and a number of demographic characteristics was assigned in response to the sample expansion, which added interviews in February and April to those conducted in March. Prior to the sample expansion, the supplement was known most commonly as the March supplement.

national surveys, or when the quantity to be estimated is derived from a source that does not identify locations below the state level.

Although substate estimates of income and employment have been produced for some time, state estimates still account for many of the examples of small area estimates cited in the previous chapter. One reason is that most national surveys are not designed to produce precise estimates for all of the states, much less substate areas. Obtaining more reliable state estimates is the first priority. Yet the need for substate estimates is growing. For example, better targeting of healthcare resources requires estimates of specific disease prevalence at the local level.

Small area methods often combine direct estimates and model-based estimates. But for many substate areas the surveys that are the source of the direct estimates provide no survey data or too little data to support a direct estimate that could be used in a composite estimator. Two strategies for producing estimates for such areas have been used. One is to rely entirely on the model for those areas. This is often done in producing county-level estimates from national surveys, where the samples provide direct estimates for some counties but not all. The other strategy is to apportion the estimates for larger areas among their component areas—often based on shares of relevant variables. For example, in producing estimates of poverty for school districts, the Census Bureau disaggregates estimates that were generated at the county level. A school district model cannot be estimated because the key predictors used in the county-level models are not available for school districts. The county is a common geographic unit for substate data produced by administrative agencies, but smaller areas are not.

It may also turn out that the most useful substate areas for examining geographic patterns in the distribution of a particular characteristic do not coincide with political divisions. For example, diseases or medical conditions may show more variation among areas separated by geographic barriers than by political boundaries. Determining the most appropriate geographic unit may require extensive analysis, which may be hampered by a lack of suitable area-level data.

2. Implementation Issues

To establish an SAE program, a federal agency must address a number of issues. The implementation issues that panelists were asked to discuss included personnel needs, use of external resources (contractors, expert panels, collaborations with other agencies), and development time. Panelists brought up additional issues related to software and benchmarking.

a. Requirements for Establishing an SAE Program

To establish and maintain an SAE program requires specific kinds of personnel resources. It also requires sufficient financial resources and time to support a development effort that is likely to span multiple years. The required financial resources, however, are dwarfed by the data collection costs that would be incurred in expanding sample size in order to produce direct estimates of adequate precision. This is particularly true of substate estimates. But even expanding a national sample to produce improved state estimates is likely to require a substantial sample increase. Thus, establishing an SAE program is likely a cost-effective strategy for producing subnational estimates.

The personnel resources required to develop and maintain an SAE program include staff with specific technical expertise, strong programming ability, and communication skills sufficient to persuade users that the small area estimates are sound. Although it may not be necessary that staff have direct experience with SAE, strong statistical skills and especially expertise in modeling are crucial. Most of the small area estimates in production in the federal government rely heavily on modeling. New applications are not likely to be any different. Software packages that can perform different types of SAE are available, and these may be sufficient if the required number of estimates is not too extensive. If so, the programming resources needed to develop and implement an SAE methodology will be more modest. But as noted below, staff in several agencies have found it necessary to write their own code in a programming language. That increases the requirements for both the programming staff and the technical staff, who must specify and review the code. Lastly, users who are unfamiliar with SAE are likely to require a detailed explanation of why the estimates represent an improvement over direct estimates and a non-mathematical explanation of how they are created. Depending upon the users, experience in explaining technical topics to a non-technical audience may be an important skill—and one that might be essential if the estimates are to gain broad acceptance among users.

Hiring qualified technical staff can be difficult. Individuals with advanced research degrees may not be attracted by the production work that will eventually follow the development of a small area methodology. If hiring proves difficult, there are alternative ways to obtain access to qualified technical staff. Collaboration with another agency that has such staff is one way, although this is more of an option for the developmental phase of a project than the production effort that will follow. For example, NCI staff teamed with NCHS staff (and with university faculty) to develop the methods for producing small area estimates of cancer incidence, which NCI has put into production. Contracting can address both development and production, and a contractor may also be able to assist with the effort to communicate the methodology and explain the results to users. As described above, SAMHSA and FNS have had long and fruitful relationships with contractors in developing, producing, and disseminating small area estimates. Another option for obtaining short-term expertise is to assemble an expert panel or technical advisory group (TAG). When there is a well-defined estimation problem, such groups can be very efficient in coming up with recommendations. For a more extensive effort, an agency might consider funding a panel of the NAS, which will meet multiple times and produce a monograph with recommendations. The Census Bureau made important use of an NAS panel's recommendations in its development of the SAIPE program. Such panels are a more costly option than convening a TAG but are more suitable if the goal is to develop something new or to obtain a fully independent review.

b. Software

The more sophisticated SAE methods used by federal agencies and their contractors are computationally intensive. Application of these methods—especially at the county level—may involve tens of thousands of parameters that have to be estimated and re-estimated many thousands of times. Given both the volume of the calculations and the complexity of the models, these methods require specialized software or specialized routines within larger software packages. Computational speed is a particular issue for users, as is the extent of the diagnostic information that the software can provide. With regard to the latter, over-specification is a significant concern—especially with state models, where the number of predictor variables can be large relative to the degrees of freedom. With as few as 51 observations, adding a variable may appear to improve the model's fit, but the improvement may be due to a reduction in error in a single state. Packaged routines often provide too little information that could help the user identify over-fitting or other common problems with model specification.

Because of the limitations of existing software packages and the demands of the applications, some of the most sophisticated users of SAE methods have chosen to write their own software. For such programming the language C has been popular. But some applications were programmed in R while at least one used FORTRAN. Writing software may not be an option for many agencies, however.

c. Benchmarking

In general, it is desirable that substate estimates sum to state totals and that state estimates, in turn, sum to national totals. Often such benchmarking is done outside of the formal estimation process as a separate step at the end. When done in this way, however, it raises several issues. Should all areas be adjusted proportionately? If the direct estimates for some areas are very strong, will the adjustment be detrimental? How can the impact of such adjustments be reflected in the estimated variances? There are ways to incorporate the benchmarking into the modeling process itself. Staff in some agencies have done so. This strategy adds complexity but, handled in this way, the impact of benchmarking can be incorporated directly into the variance estimates.

3. Choosing a Method

The literature search and a discussion at the panel meeting identified few examples of comparative evaluations of SAE methods, and fewer still involving current state-of-the art methods. Typically, agencies explore alternative approaches in the literature and identify a methodology that is consistent with the agency's objectives, the available data, and the agency's resources—both personnel and funding. The effort required to develop and test a competing approach discourages researchers from carrying their exploration of alternative methodologies to the point of empirical evaluation. As a result, we have less empirical data than we might like on the comparative strengths and weaknesses of the major, alternative approaches to SAE—and variations on those approaches. One notable empirical observation is that the more computationally complex estimators can make run time an issue when the number of domains to be estimated is large. For some applications, this was a consideration in the choice of method.

When there is a possibility that a set of small area estimates under development may be used for funding allocation, this may influence the choice of method and the choice of auxiliary variables. Exceedingly complex methods may lack the transparency needed to persuade state and local agencies that a method is fair. Considerations of timeliness may rule out particular variables or dictate that they be used with a lag, which may reduce their effectiveness. Other variables, such as those influenced by state and local policies and administrative practices, may be deemed inappropriate as predictors in models that will ultimately influence how funds are distributed. It is also important to understand how the properties of an estimator may interact with features of the funding formula in which it will be used. For example, funding formulas that utilize thresholds that award or cut off funds when estimates either exceed or fall below specified values may exaggerate the impact of bias or variability in the estimates (Zaslavsky and Schirm 2002). Choosing between older census data and more current but also more variable estimates from the ACS illustrates the type of decision that might be influenced by knowledge of exactly how the estimates will be used.

The use of small area estimates for funding allocation also gives rise to issues in maintaining or upgrading the estimates over time. Decisions to add or drop variables from the predictive model could have implications for the allocation, which the producer of the statistics may want to take into account—although it may be difficult to assess the full implications. In general, improvements to the estimates are desirable, but they do have consequences—potentially—for the allocation of funds and may create a discontinuity when the changes are first introduced.

4. Identifying Suitable Auxiliary Data

For many applications, the set of relevant auxiliary variables may be small. With few choices, some variables may dominate others as predictors. This has been the experience of the Census Bureau with its SAIPE program. A poverty measure constructed from tax data is consistently the strongest predictor of small area poverty rates.

A potentially important consideration in selecting auxiliary variables is how the small area estimates will be used. If a major purpose of the estimates is to measure year-to-year change, then it may be important to include auxiliary variables that capture change at the small area level. In particular, variables from an earlier time period may be less effective in this case than in situations where the measurement of change is not a principal use of the estimates.

A reason to be cautious in the selection of auxiliary variables is that the variables used in an SAE model to predict a characteristic are effectively removed from consideration for future analyses of variation in the small area estimates of that characteristic. This is because a relationship between a predictor, X, and the small area estimate of characteristic Y will be created by the modeling process. Subsequent analysis of the relationship between X and Y may yield biased information on their true relationship.

5. Validation

Varied approaches to validation have been used with applications of SAE. The principal challenge in evaluating small area estimates is that there may be no observations of the "truth" for any of the areas for which a set of estimates is prepared. This characterizes a lot of substate estimates.

Prior to the ACS, when a small area estimate corresponded to something measured on the decennial census long form (and the area was not so small in population that even the long form sample was inadequate), the estimates could be evaluated by constructing estimates for the census year and comparing them to the census results. Often such estimates used census long form data as covariates, but the covariates were from the prior census. Although comparing an estimate constructed in part from 2000 census data to a 2010 census estimate was entirely appropriate, this represented close to a worst-case scenario in that the census data used as covariates were 10 years old. How well the estimator performed after, say, five years could not be determined. Even same-year comparisons with decennial census estimates were not perfect in that the estimators of interest did not always correspond exactly to what was measured in the decennial census. For example, although labor force status could be determined from the data collected in the census, the long form questions did not replicate the questions used to measure labor force status in the CPS, which is the source of the official measure of the unemployment rate. Consequently, a census measure of a small area unemployment rate was not identical, conceptually, to a CPS-based small area estimate. Nevertheless, the approximation was close enough to enable effective use of the decennial census data for validation, and comparisons at higher levels of aggregation—such as states—could be used to understand the bias in the decennial census estimates and provide a basis, in turn, for assessing bias in the small area estimates.

For small area estimates of characteristics measured in the ACS, direct comparisons to the ACS estimates take the place of comparisons to decennial census data. In one respect, such comparisons are an improvement upon validation based on decennial census data because they can be done in any year. But this method cannot be used for small area estimates that incorporate ACS data. Furthermore, precise estimates from the ACS for most substate areas require pooling of either three or five years of data. Thus, the ACS estimates are multiyear averages whereas the small area estimates are generally not. Even if the small area estimates align with what is actually measured in the ACS, multiyear averages are conceptually distinct from single-year estimates, so using the ACS for validation is not perfect.

A number of alternative approaches to validation have been used, beginning with careful examination of model diagnostics. A popular approach to validating final estimates involves first constructing an artificial population, drawing subsamples, creating small area estimates from the subsample data, and comparing the results to the "truth," as reflected in the simulated population. Several federal programs have used this approach. SAMHSA staff used a large survey—the BRFSS—with related measures for comparison and mapped the small area estimates with the estimates from this survey in order to determine if the two sources showed consistent patterns, even though the estimates at the national level were not identical. CDC staff assessed the extent to which small area estimates preserved known correlations among the characteristics for which small area estimates were produced—for example, diabetes and obesity. Cross-validation has also been used. This involves removing a subset of areas, re-estimating the model, and then applying the new model to the excluded areas to determine how closely the two sets of estimates agree.

Mapping can also be exceedingly valuable for validation. Mapping the estimates will show whether the geographic variation reflects plausible patterns or excessive noise. This can be helpful for validation—particularly if the error estimates imply less noise than the geographic variation suggests.

6. Interpreting and Communicating Results

Interpretation and communication of results requires careful attention—especially with a new program. Getting users to understand the limitations of direct estimates from small samples has been a continuing challenge. No less than with direct estimates, the producers of small area estimates need to consider how best to communicate the variability of their estimates. This is particularly true when estimates are for smaller areas than an agency has released previously. (This will characterize most new estimates.)

Users interested only in point estimates pose a particularly difficult challenge, but for all users prominent displays of error statistics are important. Some producers have developed informative graphics to assist users in making comparisons across areas. Such displays work well with state estimates, where the number of areas is relatively small. They present more of a challenge for substate areas, however, given their sheer numbers.

Maps, which we have noted are useful for validation, can also be important aids to interpreting and communicating results. Mapped estimates can reveal geographic patterns that may convey important information about sources of variation. Comparing maps of different small area characteristics may also provide evidence of relationships that are not evident in summary statistics.

When there are explicit stakeholders for a set of estimates—for example, state program directors—it may be advisable to meet with the stakeholders to discuss the estimates, how they were created, and what is known about their quality. Giving stakeholders an opportunity to raise questions and express concerns is the best way to make sure that the explanations offered by the producers are being understood. When methods are new to the stakeholders, effective communication may require multiple meetings, but the benefits in terms of support for the program and suitable use of the estimates can be invaluable. Such stakeholder education was particularly important when FNS began its current program of state estimates of infants and children eligible for WIC, which were to be used in funding allocation for the program.

B. Sources of Auxiliary Data and Issues in Using Them

The principal sources of auxiliary data for small area models historically have been administrative records and decennial census data. But the replacement of the census long form by the ACS is shifting usage away from the census. In addition to these sources, we also discuss HRSA's Area Health Resources Files and some prospective new sources below.

1. Administrative Records

Administrative data are widely used as auxiliary variables because they typically have no variance. When incorporated into models, they can help to generate very precise estimates. Some administrative data are available as individual records (that is, microdata), and in that case they can be aggregated into whatever geographic units are being estimated, provided that sufficient locational information is included. Zip codes provide great flexibility, but full addresses are even better because they can be geo coded in many cases. Another advantage of individual records is that they can be used to construct new variables that cannot be created from aggregate data. In its SAIPE estimates, for example, the Census Bureau constructs a "tax return poverty rate" by calculating the number of personal exemptions on returns with adjusted gross income below the official poverty threshold for a family of the size corresponding to the number of exemptions, and dividing the number by an estimate of the population for the area. No other agency producing small area estimates has access to federal tax data, however, so the use of these data has been limited to applications by the Census Bureau.

FNS compiles counts of various nutrition program participants by county from data submitted by the states. These include monthly counts of SNAP participants by state and county and October counts of students certified for free and reduced-price school lunches. The latter have become an important indicator of the prevalence of low-income students in schools. Similarly, the Social Security Administration's monthly counts of Supplemental Security Income recipients aged 65 and older by state are an indicator of the incidence of low income among this population.

The BEA's state and county estimates of per capita income by source, discussed in Chapter II, are indirect estimates constructed by disaggregating a combination of administrative and survey data. As one of the only readily accessible annual sources of small area income data, they have been used as auxiliary variables in a number of SAE programs. For example, the Census Bureau uses the proportion of total personal income derived from government transfers and growth in personal income since the last census as predictors in its SAIPE estimates of median household income.

There are a number of issues with the use of administrative records as a source of auxiliary variables. Timeliness is a factor for some administrative data, as the release of such data often lags that of annual survey data by many months. Quality is a greater concern. A number of the major sources of administrative data (SNAP, Medicaid, UI) are collected by the states, and although there may be efforts at the federal level to ensure that the data are collected in the same way, the quality of the data may not be uniform. Within a state, when data are collected locally, data quality may vary from area to area. In addition, although there is a tendency among users to view administrative data as having higher general quality than survey data, it is important to recognize how the data are generated and used in administering the program. Data on benefits paid by the agency reflect actual administrative transactions and, therefore, are among the most accurate. Data on the characteristics of participants, on the other hand, may come from a variety of sources and may not be uniformly verified. Items of income or expenses that fall below a threshold value may not even be recorded—showing up instead as zeroes in a field that has nonzero values for some records. Some data, such as race or ethnicity, may be based on caseworker observation. Residential location—critical for small area estimates—can be especially problematic. For example, the addresses reported on tax returns are not required to be residential, much less the taxpayer's principal residence. Unknown numbers of taxpayers report their business addresses or those of their tax preparers, and the residential addresses do not necessarily reflect where the taxpayers were living at the time they filed. As with variables on surveys, variables collected in administrative records can change definition over time or can be removed from collection. Some administrative variables may be affected by one-time events—such as a natural disaster or a policy change. Changes to administrative variables may be less well documented than changes to survey variables, underscoring the need for users of administrative data to consult periodically with the program staff that collects or compiles the administrative data.

2. Decennial Census Data

Another source of data with little or no variance is the decennial census. Before the census long form was eliminated (beginning with the 2010 census), long form data were collected from 20 percent of households in most areas and a higher fraction in small areas. The Census Bureau published a wide range of estimates based on the full, long form sample. These were a primary source of auxiliary variables for SAE. (Public use microdata included only samples of the long form data, so they were much less useful for applications below the state level.) The challenge to modelers was that the data represented just one year in 10, so they reflected area variation at one point in time and did not capture subsequent change. Models that used census variables often included additional variables representing ratios of more current variables to census values. Such variables might be based on administrative data or on survey variables measured at the state or national level. The elimination of the census long form removes all but basic demographic information from the census, although it does provide the ACS as an alternative (see below).

3. The ACS

With the replacement of the census long form by the ACS, the long form data that were previously used as auxiliary variables in small area models and in validation of small area methods are no longer available. However, the conversion of the decennial long form into a continuous survey offers several advantages. First, the ACS provides precise, annual estimates for many characteristics and areas that would have required SAE previously. Second, the ACS provides a wide range of potential auxiliary variables for small area estimates of characteristics that the ACS does not measure. Third, related to this, the ACS expands the opportunities for generating small area estimates of a wide range of characteristics by pairing the ACS with other surveys. Fourth, drawing covariates from the ACS rather than the decennial census eliminates the need to deal with a variable time lag and predictive ability that declines over time. Fifth, the ACS includes design features that strengthen the quality of its data relative to what was collected in the census long form.

At the same time, the ACS shows the limits of even large-scale surveys in addressing the need for estimates of substate areas. For the majority of counties and nearly all subcounty areas, ACS data must be pooled over three or five years to provide estimates that the Census Bureau is willing to release. These pooled estimates are still based on samples smaller than comparable estimates from the census long form, which could not support very precise estimates for many counties, much less sub-county areas. Small area methods can provide more timely estimates with an annual reference period. In many cases these estimates will be more precise than what can be obtained from the ACS.

4. The Area Health Resources Files

The Area Resource File, a database of county-level variables maintained by HRSA, has been a principal source of data for small area analysis and SAE since its creation. Renamed the Area Health Resources Files and substantially expanded, this database is a compilation of more than 6,000 variables gathered from more than 50 sources, including the decennial census (short form and long form), postcensal population estimates, BEA economic data, BLS employment and unemployment statistics, and an extensive array of health-related sources. The latter include data from government sources such as HRSA, CDC, NCHS, and the Department of Veterans Affairs, and private sources such as the American Hospital Association, the American Medical Association, the American Dental Association, and the American Osteopathic Association. Health-related data include health care professionals by specialty, health professions training, health facilities, hospital utilization, and hospital expenditures. Part of the addition that led to the name change was the inclusion of state and national files that include more extensive demographic characteristics plus work force employment and training data for 50 different health care professions.

5. New Sources

The Affordable Care Act, through its reporting requirements, may be responsible for generating new data sources that can be used in small area models. For example, the IRS will be collecting data on health insurance coverage in order to enforce the individual mandate and to implement the premium tax credits for low-income individuals and families. At this point, however, we can only speculate about the data that may become available.

Through a partnership between the Census Bureau and the states, the bureau's Longitudinal Employer Household Dynamics program is generating a number of data elements that combine employer and employee data in ways that link employment characteristics to small domains. Using data from the UI system and the Quarterly Census of Employment and Wages, the bureau has been able to construct a unique set of measures of employment activity for detailed levels of geography and industry. Although these data cannot be released to other agencies or researchers for use in small area models (or other purposes), the bureau has created synthetic versions of selected variables that can be released. [8] These data represent a potentially valuable addition to the variables available for generating small area estimates of economic indicators.

C. New Developments in Methodology

Rao's 2003 text, *Small Area Estimation*, has been a standard reference for researchers building small area models, but the book is now 10 years old. Pfeffermann (2013) has provided an update through his review of new developments in the field (see Chapter II). On the modeling side, ways of applying causal modeling to large datasets have emerged from research in artificial intelligence by Judea Pearl at UCLA. In addition, a third generation of R software for person-level models has been released. The next frontier in SAE, panelists agreed, is multivariate estimation, where multiple, correlated variables are estimated jointly so that their small area estimates will incorporate their true correlations. [9]

D. Improving Interagency Collaboration on Methods and Data

Collaboration between agencies can be very helpful in terms of providing access to broader staff expertise and permitting access to data that would otherwise not be available. One panelist, for example, noted that collaboration with the Census Bureau was a potential means to working with data the bureau could not release. Through agreements developed and maintained over the years, the bureau has been able to obtain tax data that are critical to the estimation of internal migration and a primary source of auxiliary variables for its small area income estimates in multiple programs. In this case the data sharing has been in one direction, and IRS staff have had comparatively little joint work with Census Bureau staff. Another model of collaboration is the joint work between NCHS and NCI, which has also included partners at two universities, another CDC agency, and a private organization. The methodological developments produced by this work have led directly to a program of small area estimates that NCI expects to continue.

Establishing collaborative agreements can be challenging—particularly when efforts cross departments whose lawyers may interpret regulations differently and be guided by different traditions with respect to interagency research. It may take years to establish the agreements needed to move forward, which may outlast the interest of one or the other party. Staff changes may also play a role. One participant observed that common interest in a successful collaboration is critical.

[8] Synthetic variables are imputed versions of administrative or survey variables. They are created as a substitute for the original variables when the release of actual values would create a risk of disclosure. Synthetic variables are unrelated to the SAE methodology of synthetic estimation, discussed earlier in this report.

[9] An explanation of how this has been done by Mathematica in producing state estimates of SNAP participation rates among all eligible persons and the working poor is included in Appendix D.

One way in which collaboration can be fostered is through regular meetings of those who share research interests. An offer from the Statistical and Science Policy Office in OMB to create a small area working group under the FCSM so that those involved in producing small area estimates can communicate on a regular basis was greeted enthusiastically. Since the panel meeting, steps have been taken to convene the first such meeting, and the prospects appear good that a working group on small area estimates will be established.

This page has been left blank for double-sided copying.

V. CONCLUSIONS AND RECOMMENDATIONS

Surveys remain critical sources of information, but sample sizes limit their effectiveness in describing small domains, which can be states and substate areas, population subgroups, or combinations of these. The Census Bureau was able to transform the decennial census long form into a continuous survey—the ACS—by spreading its sample over the decade, in effect. Sample size increases are becoming increasingly more costly, and the ACS shows that even a survey that obtains data from two million households a year cannot address all of the small area data needs for the type of information that it collects. More specialized data collections that support critical policy needs in health and human services are often challenged to produce subnational estimates below the largest states. Small area estimation is a powerful tool for extending survey estimates to small domains.

A. Concluding Observations

Based on our review of federal government programs producing small area estimates coupled with the discussion at the expert panel meeting, we report the following observations:

- Within HHS, the use of SAE methods to produce small area estimates on a regular basis is limited, but these applications show a high level of sophistication and cutting-edge methods.
- The development of a program of small area estimates requires significant staff expertise, time, and resources, but these costs are small in comparison to those of new or expanded data collection.
- Software limitations are an especially challenging part of the development and implementation of small area methods.
- The identification and use of suitable auxiliary variables present a number of challenges.
- Comparative evaluations of alternative approaches are not as common as prospective users might wish.
- Validation of small area estimates remains a significant challenge; and varied approaches have been used.
- Interpretation and communication of small area estimates requires careful attention, but better communication of the limitations of direct estimates is needed as well.
- The ACS changes the landscape for small area estimates in a number of ways.
- Collaboration presents both opportunities and challenges, yet the sharing of technical resources across agency divisions and even across agencies may be necessary to fully exploit the potential of small area methods.

Below, we develop each of these points.

1. Applications within HHS

We identified several production applications of SAE within HHS, including the following uses:

- State estimates of age-group-specific prevalence of 25 substance-use behaviors measured in NSDUH (SAMHSA)
- County estimates of diabetes prevalence, incidence, and risk factors (CDC)
- State and county estimates of cancer risk factors and screening (NCI)
- State estimates of household substitution of wireless telephones for landlines (NCHS)
- Model-based estimates of vital statistics for geographic and demographic small domains (NCHS)

Most of these applications are relatively new. The CDC and NCI estimates and the NCHS estimates of wireless substitution were all initiated within the past decade. By contrast, the SAMHSA estimates began in the early 1990s. Reflecting the skills of those involved as well as advances in computing, these newer approaches and the SAMHSA estimates all employ empirical or hierarchical Bayes methods.

In addition to these production applications, small area estimates have been developed as part of methodological research at NCHS, and both CDC and NCI are working to extend their methods to other measures while AHRQ is developing experimental county-level estimates for its quality indicators program and exploring the production of additional state estimates from MEPS.

2. Requirements for Establishing an SAE Program

The development of a program of small area estimates requires significant expertise, time, and resources. Requisite skills include statistics or econometrics with an emphasis on modeling, high-level programming, and strong communication ability. Meeting these needs may require hiring new staff or obtaining contractor support. Development and implementation can take two years or longer. Nevertheless, these costs are dwarfed by the expense of new or expanded data collection to produce direct estimates of satisfactory precision for the areas for which estimates are needed.

3. Software Issues

Most of the formal SAE methods used by federal agencies and their contractors are computationally intensive. As such, they require specialized software or specialized routines within larger software packages. Computational speed is a particular issue for users, as is the extent of the diagnostic information that the software can provide. Some of the most sophisticated users of SAE methods have written their own software to address these limitations. This may not be an option for many agencies, however.

4. Auxiliary Variables

Administrative records at the individual level—persons or program units—are particularly useful as auxiliary variables in small area models because they allow users to create variables (such as the tax return poverty indicators used in SAIPE) and to construct estimates for non-standard geographic units (such as school districts and congressional districts). Access to microdata is often restricted, however. When individual records are not available, aggregate data may suffice, but such data may not cover all of the geographic areas for which estimates are needed.

The quality of administrative records can exceed that of survey data, which is one of the characteristics that make them so valuable for SAE. The quality may vary widely across items or areas, however. Users need to become familiar with the information on data quality that the producer has generated over the years but also be aware that such information may be insufficient for their purposes. Apart from general issues with the quality of specific variables, anomalies can occur as the result of administrative actions, natural disasters, or other factors. This underscores the importance of consulting with agency staff when using an agency's data.

Other considerations are relevant in choosing auxiliary variables. Experienced modelers caution that too much time may be spent searching for potential covariates among administrative sources when common variables from widely used sources may perform just as well. Even if a variable has been established as a good covariate, there may be reason to avoid using it in a small area model. A variable used in a small area model becomes incorporated into the estimates. Using a variable in this way may limit its value in future analyses of area variation.

5. Comparative Evaluations

Our literature search and review of federal programs identified few examples of comparative evaluations of rigorous alternative small area methods. The primary reason, we infer, is that the effort required to implement a method fully enough to conduct an empirical evaluation is substantial. Because of this, agency staff may research the alternatives through the literature or by talking to other users, but the choice of a method, typically, is made before significant development occurs. In refining the application of the chosen method, agency researchers may conduct comparative evaluations of selected options within the overall methodology, but even this type of methodological research sees limited use. Consequently, the federal statistical community has learned little about how different SAE approaches compare when applied to the same estimation problem. Thus, agencies that are developing new applications and seeking to choose the most suitable method cannot draw on external research to determine what methods are likely to work best for their particular applications. What they can learn is what methods have been applied successfully to similar problems and, if they choose one of these methods, what issues they may encounter and how to address them.

6. Anticipating Uses

The desired properties of an estimator should affect design choices at the outset and determine what criteria are given the most weight in evaluating the estimates. For instance, if the measurement of change at the area level is an important goal, this would place a premium on selecting auxiliary variables that reflect change at that level. Absent such variables, the small area estimates may be deficient in how well they represent actual change over time.

Small area estimates produced by federal agencies are sometimes used for funding allocation, and in those circumstances the quality of the estimates has direct implications for how well federal funds are targeted and how efficiently they are distributed. If the estimates for individual areas vary too much from year to year, program administrators or legislators may respond by establishing hold-harmless provisions or other administrative remedies that undermine the use of the estimates. Statistical solutions would be preferable but may be difficult to implement once the estimates are in use. In addition, by affecting the final estimates the choice of auxiliary variables has a bearing on how funds are allocated. Potential inequities resulting from the selection of particular variables need to be considered.

7. Validation

Evaluating small area estimates is perhaps the most challenging aspect of their production. That small area methods had to be used in the first place implies that direct estimates of sufficient precision are not available. When the census long form was still in place, small area estimates of census-like variables could be evaluated by applying the methodology to a census year and comparing the results to the census estimates. But this is no longer an option—except for the basic demographic characteristics measured on the decennial census short form.

Given the critical role of modeling in the development of small area estimates, careful examination of model diagnostics—that is, internal validation—may be the most common form of validation. Evaluations of the end results apply a variety of methods. Simulations using an artificial population are becoming more popular. Other approaches include comparing the small area estimates to the most precise direct estimates (for example, the largest states), assessing how well the small area estimates preserve known correlations, and using cross-validation.

8. Interpreting and Communicating Results

No less than with direct estimates, producers of small area estimates need to consider how best to communicate the uncertainty associated with their estimates. They may also have to explain why their estimates are better than direct estimates. Meeting with stakeholders can be exceedingly useful in explaining the notion of borrowing strength and how it is reflected in the small area estimates as well as in learning and addressing the concerns and questions that stakeholders may have.

Some producers of small area estimates have developed informative graphics to assist users in making comparisons across areas. State estimates, for example, lend themselves to visual displays showing which state estimates are significantly different from others. In addition, maps can be useful not only in validation, but also as an aid to interpretation and communication of findings.

9. Impact of the ACS

The replacement of the census long form by the ACS affects SAE in a number of ways. First, it eliminates long form variables previously used in models and validation—although in some cases the equivalent ACS variables can be substituted. Second, it provides direct estimates for many characteristics, geographic areas, and subpopulations that would have required SAE previously. Third, it provides a wide range of auxiliary variables that are contemporaneous with the direct estimates used in model development. Fourth, it eliminates the need to deal with the variable time lag that characterizes the use of auxiliary variables from the last decennial census. Fifth, it expands opportunities for SAE through pairing the ACS with other surveys that capture the relevant content but lack the sample size. At the same time, the ACS also shows the limits of even large-scale surveys. For most substate areas, only three-year or five-year averages are published, and they can be very imprecise. For such areas, SAE can provide timelier, more precise estimates.

10. Interagency Collaboration

Two important benefits of collaboration are providing access to broader staff expertise and providing access to data. The challenges to collaboration can be significant, however. Common interest in a successful collaboration is critical to establishing inter-agency agreements that will foster a successful partnership. Models of successful collaboration—such as that among NCI, NCHS, and two universities—can be studied for ideas on how to make such ventures effective.

In addition to formal collaboration, interagency working groups can be helpful for sharing ideas and keeping members up to date on activities that are ongoing or under consideration. Panelists responded enthusiastically to an OMB offer to create a group under the FCSM that would allow those who are working on small area projects or those who are interested in the topic to communicate regularly.

B. Recommendations

One of the goals of this project was to help ASPE communicate to others in the department what small area methods can and cannot do to address data needs involving states, substate areas, and other small domains. The successful SAE programs at CDC, NCHS, NCI, and SAMHSA, as well as the ongoing work at AHRQ to develop county estimates for the agency's quality indicators program provide examples that illustrate the potential of these methods to fill such data gaps, the amount of effort that this may entail, and the types of data that must be assembled. We recommend that ASPE draw on these examples to convey a realistic appreciation of what SAE can do and what its application entails.

Other agencies in HHS that routinely use small area statistics in their work recognize the value of such data but may not have the requisite staff or resources to develop their own small area estimates. Although we did not uncover pressing demands for particular small area estimates that are not produced currently, we suspect that each of these agencies may be able to identify needs for expanded statistics in the future. If that should occur, we recommend that each agency explore potential collaboration with NCHS, which has the deepest staff with significant experience with SAE techniques. We also offer agency-specific recommendations below.

Much of the data collected and used by CMS is administrative and covers the full population, so the need for SAE is not apparent. Although these data could provide a potential source of new auxiliary variables, it is very unlikely that CMS would find reason to develop such variables absent requests from outside the agency. We recommend that ASPE monitor the operating divisions with active small areas estimates programs to see if there are needs for auxiliary variables that CMS could address. In addition, the MCBS is too small to support direct estimates at the state level, but if there is interest in state estimates, the combination of survey and administrative data collected by the MCBS would provide a good start for an application of SAE. We recommend that ASPE explore interest inside and outside of CMS in state estimates from the MCBS and, if such interest is identified, work to put CMS in contact with NCHS for a possible collaboration.

The Division of Program Statistics in the Office of Public Health Support in the IHS produces two major reports: (1) Trends in Indian Health and (2) Regional Differences in Indian Health. The latter (last published about a decade ago) relied heavily on decennial census data as the source of statistics on educational attainment, unemployment, median household income, and poverty of the American Indian and Alaska Native populations by area office. There are 12 area offices, and they vary widely in the size of the population they serve, with the largest being more than 10 times the smallest. With the elimination of the census long form, the Division of Program Statistics will have to turn to the ACS to provide these same statistics on American Indians and Alaska Natives. This opens up the possibility of producing estimates more than once a decade to better document trends, but sample size then becomes an issue. Although five-year roll-ups of ACS data might provide adequate precision—albeit less than the census long form—five-year moving averages are not satisfactory for monitoring trends. Nonoverlapping five-year averages would provide more interpretable trend information. Although this would mean longer intervals, this would still be an improvement over the decennial census. To produce trend data for the American Indian and Alaska Native populations from one-year or two-year ACS samples might require SAE methods. We recommend that IHS assess the merits of using SAE methods to produce more frequent statistics on American Indians and Alaska Natives.

In addition to producing the Area Health Resources Files, HRSA also produces statistics on the health care labor force and has developed criteria for defining health professional shortage areas. Although SAE methods are not needed to identify shortage areas as they have been defined traditionally, by using actual counts of medical professionals in combination with Census Bureau population estimates, the application of small area methods could enable HRSA to broaden the way it defines shortage areas. We recommend that HRSA explore whether there might be value in an expanded approach.

Finally, some mechanism for sharing experiences with SAE would serve a broad array of current users and prospective users within the federal government. We encourage OMB to pursue the establishment of a working group on this topic, and we encourage ASPE to participate in this effort and inform the operating divisions of HHS of developments that are communicated through this group.

REFERENCES

Barker, Lawrence E., Theodore J. Thompson, Karen A. Kirtland, James P. Boyle, Linda S. Geiss, Mary M. McCauley, and Ann L. Albright. "Bayesian Small Area Estimates of Diabetes Incidence by United States County, 2009." *Journal of Data Science*, vol. 11 (2013): 249-267.

Baskin, Robert M., and John P. Sommers. "Comparison of Direct, Mixed Model, and Bayesian Metropolitan Statistical Area Estimates for the Insurance Component of the Medical Expenditure Panel Survey (MEPS)." *Proceedings of the Joint Statistical Meetings.* Alexandria, VA: American Statistical Association, 2008.

Blumberg, Stephen J., Nadarajasundaram Ganesh, Julian Luke, and Gilbert Gonzales. "Wireless Substitution: State-level Estimates from the National Health Interview Survey, 2012." *National Health Statistics Reports*, no. 70 (December 18, 2013), National Center for Health Statistics.

Cadwell, Betsy L., Theodore J. Thompson and Lawrence E. Barker. "Bayesian Small Area Estimates of Diabetes Prevalence by U.S. County, 2005." *Journal of Data Science*, vol. 8 (2010): 173-188.

Cunnyngham, Karen. "State Estimates of Children Ages 0 to 4 in Low-Income Families in 2010 Derived from CPS Data Using an Empirical Bayes Shrinkage Estimator." Memorandum to Grant Lovellette and Ed Harper, Food and Nutrition Service, August 9, 2012.

Cunnyngham, Karen, Laura Castner, and Amang Sukasih. "Empirical Bayes Shrinkage Estimates of State Supplemental Nutrition Assistance Program Participation Rates in 2008-2010 for All Eligible People and the Working Poor." Washington, DC: Mathematica Policy Research, February 2013.

Czajka, John L. and Thomas B. Jabine. "Using Survey Data to Allocate Federal Funds for the State Children's Health Insurance Program (SCHIP)." *Journal of Official Statistics,* vol. 18, no. 3, September 2002, pp. 409-428.

Federal Committee on Statistical Methodology. Statistical Policy Working Paper 21: Indirect Estimators in Federal Programs. Washington, DC: Office of Management and Budget, 1993.

Folsom, Ralph, Babu Shah, and Akhil Vaish. "Substance Abuse in States: A Methodological Report on Model Based Estimates from the 1994-1996 National Household Surveys on Drug Abuse." *Proceedings of the Joint Statistical Meetings.* Alexandria, VA: American Statistical Association, 1999.

Graven, Peter and Joanna Turner. "Modeling Health Insurance Coverage Estimates for Minnesota Counties." *Proceedings of the Joint Statistical Meetings.* Alexandria, VA: American Statistical Association, 2011.

Ingram, Deborah D., Jennifer D. Parker, Nathaniel Schenker, James A. Weed, Brady Hamilton, Elizabeth Arias, and Jennifer H. Madans. "United States Census 2000 Population with Bridged Race Categories." *Vital and Health Statistics*, series 2, no. 135, September 2003.

Mohadjer, Leyla, Graham Kalton, Tom Krenzke, Benmei Liu, Wendy Van de Kerckhove, Lin Li, Dan Sherman, Jennifer Dillman, Jon Rao, and Sheida White. *National Assessment of Adult Literacy: Indirect County and State Estimates of the Percentage of Adults at the Lowest Level of Literacy for 1992 and 2003* (NCES 2009-482). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC, 2009

National Research Council. *Statistical Issues in Allocating Funds by Formula.* Panel on Formula Allocations. Thomas A. Louis, Thomas B. Jabine, and Marisa A. Gerstein, editors. Committee on National Statistics. Washington, DC: National Academies Press, 2003.

National Research Council. *Small-Area Income and Poverty Estimates: Priorities for 2000 and Beyond.* Panel on Estimates of Poverty for Small Geographic Areas. Constance F. Citro and Graham Kalton, editors. Committee on National Statistics. Washington, DC: National Academies Press, 2000.

Pfeffermann, D. "New Important Developments in Small Area Estimation." *Statistical Science*, vol. 28, no. 1, 2013, pp. 40–68.

Raghanathan, Trivellore E., Dawei Xie, Nathaniel Schenker, Van L. Parsons, William W. Davis, Kevin W. Dodd, and Eric J. Feuer. "Combining Information from Two Surveys to Estimate County-Level Prevalence Rates of Cancer Risk Factors and Screening." *Journal of the American Statistical Association*, vol. 102, no. 478, June 2007, pp. 474-486.

Rao, J.N.K. *Small Area Estimation*. Hoboken, NJ: John Wiley & Sons, 2003.

Schirm, Allen and Alan Zaslavsky. "Model-Based Microsimulation Estimates for States When State Programs Vary." *Proceedings of the Joint Statistical Meetings.* Alexandria, VA: American Statistical Association, 1998.

Schirm, Allen, Alan Zaslavsky, and John Czajka. "Large Numbers of Estimates for Small Areas." *Proceedings of the Federal Committee on Statistical Methodology Research Conference*, Washington, DC: Federal Committee on Statistical Methodology, 1999.

Sommers, John P. "Producing State Estimates with the Medical Expenditure Panel Survey, household Component." Methodology Report No. 16. Rockville, MD: Agency for Healthcare Research and Quality, December 2005.

Substance Abuse and Mental Health Services Administration. "2010-2011 National Survey on Drug Use and Health: Guide to State Tables and Summary of Small Area Estimation Methodology." Rockville, MD: Substance Abuse and Mental Health Services Administration, 2012.

Wei, Rong, Robert N. Anderson, Lester R. Curtin, and Elizabeth Arias. "U.S. Decennial Life Tables for 1999-2001: State Life Tables." *National Vital Statistics Reports*, vol. 60, no. 9, September 14, 2012.

Zaslavsky, Alan M. and Allen L. Schirm. "Interactions between Survey Estimates and Federal Funding Formulas." *Journal of Official Statistics*, vol. 18, no. 3, September 2002, pp. 371-391.

APPENDIX A

AGENDA AND ATTENDEES

This page has been left blank for double-sided copying.

U.S. Department of Health and Human Services
Small Area Estimation: New Developments and Directions for HHS

Tuesday, June 4, 2013
9:00 – 4:00
Hubert H. Humphrey Building, Room 705
200 Independence Ave., SW, Washington, DC

Agenda

9:00 **Welcome and Introductions**

9:10 **Purpose of the Meeting**

9:15 **Overview of Applications Represented by the Panel**

Department of Health and Human Services
Other Federal Agencies
State Government

11:00 **Break**

11:15 **Topics in Using Small Area Estimation (SAE)**

What Types of Data Needs Can SAE Address?

Implementation Issues (technical staff, resources, development time)

Choosing a Method

Identifying Suitable Auxiliary Data

12:30 **Lunch (Humphrey Building Cafeteria)**

1:15 **Topics in Using SAE, continued**

Evaluation

Interpreting and Communicating Results (describing uncertainty)

2:00 **Sources of Auxiliary Data and Issues in Using Them**

Administrative Records
Other Surveys
Impact of the American Community Survey
New Sources

2:45 New Developments in Methodology

3:15 Improving Cross-Agency Collaboration on Methods and Data

3:45 Summary and Closing Remarks

LIST OF ATTENDEES

Name	Agency/Organization
Panelists	
Robert (Bob) Hornyak	Administration for Community Living
Robert (Bob) Baskin (by phone)	Agency for Healthcare Research and Quality
Carol Gotway Crawford	Centers for Disease Control and Prevention
Don Malec	National Center for Health Statistics
Vladislav (Vlad) Beresovsky	National Center for Health Statistics
Pavlina Rumcheva	National Center for Health Statistics
Geoff Gerhardt	Centers for Medicare & Medicaid Services (CMS)
Benmei Liu	National Cancer Institute (NIH)
Art Hughes	Substance Abuse and Mental Health Services Administration
Jennifer Nooney	Health Resources and Services Administration
David Powers	Census Bureau
Robin Fisher	Department of the Treasury
Dan Sherman	American Institutes for Research (contractor for NCES)
Joanna Turner	SHADAC (Univ. of Minnesota)
Kathleen Call	SHADAC (Univ. of Minnesota)
Partha Lahiri	University of Maryland
Akhil Vaish	Research Triangle Institute (contractor for SAMHSA)
Ralph Folsom (by phone)	Research Triangle Institute (contractor for SAMHSA)
Alan Zaslavsky (by phone)	Harvard Medical School
ASPE/MPR Staff at Panel Table	
Susan Queen	ASPE
Michael Millman	ASPE
Joan Turek	ASPE
John Czajka	Mathematica
Allen Schirm	Mathematica
Amang Sukasih	Mathematica
Alyssa Maccarone	Mathematica
Invited Guests	
Elena Fazio	Administration for Community Living, CDAP
Kristen Robinson	Administration for Community Living, CDAP
Connie Citro	Committee on National Statistics (National Academy of Sciences)
Ed Spar	Committee on National Statistics (National Academy of Sciences)
David Bott	CMS
Brian Harris-Kojetin	Office of Statistical and Science Policy, OMB
Steve Cohen	National Science Foundation
Jessica Banthin	Congressional Budget Office
Charlie Pineles-Mark	Congressional Budget Office
Xiaotong Niu	Congressional Budget Office

This page has been left blank for double-sided copying.

APPENDIX B

SUMMARIES OF PRESENTATIONS

This page has been left blank for double-sided copying.

The panel meeting, "Small Area Estimation: New Developments and Directions for HHS," was convened on Tuesday, June 4, 2013, from 9:00 AM to 4:00 PM at the Hubert H. Humphrey Building in Washington, DC. A list of attendees is provided in Appendix A. Each of the panelists representing an agency or organization engaged in small area estimation (SAE) was asked to prepare a 5 to 10 minute presentation summarizing one or more applications. Panelists representing departmental agencies that were not engaged in SAE were invited to provide a brief overview of their agency mission and whether SAE was under consideration as a way to meet specific data needs. Most of the presenters produced PowerPoint slides, which will be posted to the ASPE website. This appendix provides summaries of the presentations. Appendix C presents summaries of the discussion sessions.

SESSION 1
OVERVIEW OF APPLICATIONS REPRESENTED BY THE PANEL

A. Department of Health and Human Services

A.1. Administration for Community Living (ACL)

Presented by Robert Hornyak, Administration for Community Living

The Administration for Community Living (ACL) was created about a year ago with the goal to help all Americans—particularly older adults and people with disabilities—live at home with the supports they need, participating in communities that value their contributions. The units under ACL include the Administration on Aging (AoA), Administration on Intellectual and Developmental Disabilities (AIDD), Center for Disability and Aging Policy (CDAP), the Office of the Administrator and Center for Management and Budget (CMB).

ACL collects state aggregated information (secondary data). It receives special data from the decennial census and other national surveys. The data used include the Health and Retirement Study (HRS), the National Health Interview Survey (NHIS), and other state and substate (county) data. Currently, ACL still does not perform any small area estimation.

A.2. Centers for Medicare & Medicaid Services (CMS)

Presented by Geoff Gerhardt, Office of Information Products and Data Analytics, CMS

CMS data collection and statistical analysis have been primarily focused on utilization of Medicaid and Medicare services. CMS collects data on medical claims and aggregates Medicare demographic, spending, utilization, and quality information at the Hospital Referral Region (HRR)[10] level. These data are published on CMS.gov and the Institute of Medicine website. However, there is still no work on small area estimation based on data modeling within CMS.

[10] Hospital Referral Regions (HRRs) represent regional health care markets for tertiary medical care that generally requires the services of a major referral center. The regions were defined by determining where patients were referred for major cardiovascular surgical procedures and for neurosurgery. Each hospital service area (HSA) was examined to determine where most of its residents went for these services. The result was the aggregation of the 3,436 hospital service areas into 306 HRRs. Each HRR has at least one city where both major cardiovascular surgical procedures and neurosurgery are performed. (Source: the Dartmouth Atlas of Health Care, accessed from http://www.dartmouthatlas.org/data/region/).

CMS administers and produces the Medicare Current Beneficiary Survey (MCBS) that provides stable national estimates and information on the Medicare population, such as expenditures and sources of payment for services used by beneficiaries, changes in health status, satisfaction with care, and usual source of care. Again, there is no work on small area estimation with this survey.

CMS is now focusing on quality and outcomes; there has been a greater demand to measure and analyze outcomes.

A.3. Agency for Healthcare Research and Quality (AHRQ)

Presented by Robert Baskin; Division of Statistical Methods and Research; Center for Financing, Access and Cost Trends

SAE work within AHRQ includes:

- State level estimates of medical expenditures by type of expenditure, as part of the Medical Expenditure Panel Survey - Household Component (MEPS-HC)
- Metropolitan Statistical Area (MSA) level estimates of employer contributions to health insurance and premiums, as part of the Medical Expenditure Panel Survey - Insurance Component (MEPS-IC)
- Experimental county-level estimates of conditions such as diabetes and chronic obstructive pulmonary disease (COPD), as part of the Prevention Quality Indicators (PQI) measures, a component of the AHRQ Quality IndicatorsTM (QI)
- Hospital discharge rates and per capita costs for 832 Core Based Statistical Areas, as part of the Healthcare Cost and Utilization Project (HCUP)

The work on PQIs is intended to provide a county-level set of measures that can be used with hospital inpatient discharge data to identify quality of care for "ambulatory care sensitive conditions." These are conditions for which good outpatient care can potentially prevent the need for hospitalization or for which early intervention can prevent complications or more severe disease. It focuses on hospital admissions for diabetes, asthma, and COPD per 100,000 persons 18+ years. The work started with the diabetes population and is still in progress. AHRQ and the Battelle Memorial Institute are in the process of deriving estimates of diabetes at the county level for the entire county population (based on patient's county of residence). The primary sources of data include the CDC's BRFSS data and socioeconomic variables from the U.S. Census Bureau's American Community Survey (ACS). Bayesian area level models are being fit using WINBUGS software (Lunn et al. 2000), as well as Integrated Nested Laplace Approximations or INLA (Rue, Martino, and Chopin 2009).

The challenges in this work have included obtaining, assembling, and managing the data; model fitting; and accounting for the sampling design of the data (modeling design variances). AHRQ is also facing remaining issues that include judging the accuracy of estimates (by comparison with CDC county level estimates of diabetes), extending the model to more than diabetes, and ultimately using the estimates.

A.4. Centers for Disease Control and Prevention (CDC)

Presented by Carol Crawford, Division of Behavioral Surveillance

CDC administers the Behavioral Risk Factor Surveillance System (BRFSS), a telephone health survey system that collects data from U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services. BRFSS collects data in all 50 states as well as the District of Columbia and three U.S. territories from more than 400,000 interviews with adults each year.

Currently, BRFSS provides state-level estimates and their confidence intervals; Selected Metropolitan/Micropolitan Area Risk Trends (SMART) direct estimates and confidence intervals for cities and counties where sample size is sufficient (over 200 micropolitan statistical areas/MMSAs in 2012); and county-level indicators (7-year aggregation) that include Community Health Rankings, Community Health Status Indicators, Health Indicators Warehouse, MedMap, and Environmental Public Health Tracking Network.

Using data from the state of Florida for 2010, CDC conducted a study that compared methods[11] with a goal to improve over the current 7-year aggregation. In this study, the accuracy of estimates, their uncertainty, and ability to preserve relationships, and other criteria were assessed. These criteria included non-statistical criteria such as number of years of data used, computational time, simplicity and ease of communication, transparency and reproducibility, and resources and staffing needed for large-scale production.

CDC concluded that the biggest challenge in using a new method is resources and staffing. For BRFSS, the estimation must move to production quickly because of the substantial and increasing demand for estimates, limiting the time available for development and testing of a new method. Other challenges include: computationally time consuming estimation algorithms, the inability to reproduce results reported in published papers, adapting existing programming code to run on BRFSS data, and difficulties in obtaining and compiling auxiliary data.

A.5. Health Resources and Services Administration (HRSA)

Presented by Jennifer Nooney, Health Resources and Services Administration

HRSA has carried out many analyses for small areas that do not require SAE techniques—for example, using census data.

HRSA hired a contractor to develop health areas for analysis of Medicare claims data and also look at Medicaid databases. HRSA contracted with Dartmouth to provide maps.

HRSA developed shortage designation criteria and uses them to decide whether a geographic area, population group or facility is a Health Professional Shortage Area (HPSA) or a

[11] The current 7-year aggregation was compared with methods such as unweighted logistic random effects (Srebotnjak, Mokdad, and Murray 2010), scaled weighted logistic random effects (g-EBLUP, similar to Jiang and Lahiri 2006), multi-level models (for HHS by Zhang based on Congdon 2009), linear weighted and benchmarked methods (You and Rao 2002), Bayesian methods (Cadwell et al. 2002, Malec et al. 1997), and benchmarked and constrained to SMART created for BRFSS (Battaglia and Frankel 2011).

Medically Underserved Area/Population (MUA/P). HPSAs and MUA/Ps were created from collections of counties and sub-county geographic units. HHS has been thinking about redefining health shortage areas. Currently, the criteria are based on a physician to population ratio. For example, for a primary care HPSA the ratio is 1:3,500; when there are 3,500 or more people per primary care physician, an area is eligible to be designated as a primary care HPSA. New criteria are still being reviewed, and it may be a number of years before their release. The new criteria will include positions like Nurse Practitioners and Physician Assistants.

HRSA also maintains the Area Health Resources Files (AHRF), a county level database containing more than 6,000 variables for each of the nation's counties. The AHRF contains information on health facilities, health professions, resource scarcity, health status, economic activity, health training programs, and socioeconomic and environmental characteristics. In addition, the basic file contains geographic codes and descriptors that enable it to be linked to many other files and to allow aggregation of counties into various geographic groupings. The AHRF integrates data from numerous data sources including: the American Hospital Association, the American Medical Association, the American Dental Association, the American Osteopathic Association, the Bureau of the Census, the Centers for Medicare & Medicaid Services, the Bureau of Labor Statistics, the National Center for Health Statistics, and the Veterans Administration.

A.6. National Cancer Institute (NCI)

Presented by Benmei Liu, Division of Cancer Control and Population Science

The NCI has been utilizing small area estimation techniques for several projects including:

- Cancer risk factors and screening behaviors at the state, health service area, and county level by combining data from BRFSS and NHIS (joint work with NCHS and the Universities of Michigan and Pennsylvania)
- Tobacco related small area estimation using data from the Tobacco Use Supplement to the Current Population Survey (TUS-CPS)
- State level estimates for cancer related knowledge variables using data from the Health Information National Trend Survey (HINTS)

The outcomes of interest in the cancer risk factors and screening behaviors estimation are:

- Current and ever smoking prevalence (for age 18+ overall and by gender)
- Mammography Prevalence within 2 years (female age 40+)
- Pap Smear test prevalence within 3 years (female age 18+)

Estimates have been produced for 1997-1999, 2000-2003, 2004-2007, and 2008-2010.

For county level estimation (a similar approach is used for state level estimation), NCI implemented hierarchical Bayes three-level models. For each outcome of interest, in the first level (the sampling model), the distribution of direct estimates of county prevalence rates for telephone and non-telephone households for both the NHIS and the BRFSS was modeled conditional on unknown county parameters. In the second level of the model (linking model), the parameters of the first level model were entered in a regression model as dependent variables, and these unknown county parameters were predicted by 26 economic, demographic, and educational attainment measures obtained for all counties from the 2000 Census and other sources. In the third level, the parameters in the second level regression (such as the regression coefficients and the variance components) were modeled and estimated using a diffuse proper prior distribution.

The challenges faced by NCI in producing these small area estimates included model non-convergence and data accessibility for variables used in the models. In the future, NCI plans to add more outcome variables such as colon cancer screening and Body Mass Index (BMI).

The estimates for states, counties, and health service areas can be obtained from the NCI SAE website: http://sae.cancer.gov/. The state cancer profiles are available from http://statecancerprofiles.cancer.gov/. The county attribute data such as median income and incidence and mortality rates can be obtained from the NCI Surveillance, Epidemiology and End Results (SEER) website, located at: http://seer.cancer.gov/seerstat/variables/countyattribs/.

A.7. National Center for Health Statistics (NCHS)

Presented by Donald Malec, Vladislav Beresovsky, and Pavlina Rumcheva, National Center for Health Statistics

NCHS has been one of the most active federal agencies in producing small area estimates within its programs. Some examples of SAE work are presented below.

A.7.a. County estimates of smoking and cancer screening rates

In joint work with the University of Michigan, the University of Pennsylvania, and NCI (see Section A.6), NCHS combined the county estimates by telephone status from the NHIS with the estimates from BRFSS. The Bayesian small area technique strengthens county estimate using associations with socio-demographic variables. The estimates for 1997-1999 and 2000-2003 are available online from the NCI website: http://sae.cancer.gov. New estimates that modify the method to account for the cell-phone only population are under development.

A.7.b. State and substate estimates of people who use only wireless phones

Joint work with NORC and the University of Minnesota combined estimates of wireless rates from the NHIS with rates measured at other times (“borrowing strength” across time). To produce model-based estimates, SAE modeling techniques were used to combine direct survey estimates from the NHIS, direct survey estimates from the ACS, and auxiliary data that are representative of those geographies. The estimates of the proportion of people who lived in households that were wireless-only, wireless-mostly, dual-use, landline mostly, and landline-only for seven 6-month periods were produced. The empirical best linear unbiased prediction (EBLUP) technique was used to derive the model-based estimates. The estimates are available from:

- for 2010: http://www.cdc.gov/nchs/data/nhsr/nhsr039.pdf
- for 2011: http://www.cdc.gov/nchs/data/nhsr/nhsr061.pdf

The latest estimates are used to benchmark mixed-frame telephone surveys, including the National Immunization Survey (NIS).

A.7.c. Small area estimates from the Division of Vital Statistics

The Division of Vital Statistics periodically produces state-specific life tables for the 50 states and DC by race (white and black) and sex. For states with moderately small numbers of deaths (that is, for states with population subgroups having between 300 and 700 deaths), a mixed probability model was used to calculate the probability of dying (qx) between ages x and x+1. This model uses observed data for previous years and a “zero-inflated” model (Voulgaraki, Wei, and Kedem 2008) to estimate qx values for ages for which there are no deaths or insufficient deaths to calculate a reliable estimate. A more detailed description of this method is available at: http://www.cdc.gov/nchs/data/nvsr/nvsr60/nvsr60_09.pdf.

The same Division has also produced annual estimates of county age-adjusted mortality rates due to drug poisoning deaths. The estimates use the “Hurdle Model” with covariates, county random effects, and county by year random effects. The covariates are drawn from the Area Resource File, the Federal Bureau of Investigation Uniform Crime Reporting Program, the SAMHSA substate estimates of drug use, and decennial census data. Evaluation of the estimates utilized residual analysis and expert review.

A.7.d. Small domain estimates from the NHIS using a Modified Kalman Filter model

A model-based approach was developed for estimating the prevalence of diabetes for 11 domain groups including Native Hawaiians and Other Pacific Islanders. A Modified Kalman Filter model developed by RAND was used to fit the data (direct point and variance estimates from NHIS). This model assumes that true health status in each racial/ethnic group evolved according to a group-specific linear trend and autoregressive deviations around that trend. For evaluation, simulations were conducted to assess the fixed design variance assumption and model robustness. The RAND SAS macro and user’s guide accompanied by a detailed description of the methodology is at: http://www.rand.org/content/dam/rand/pubs/technical_reports/2011/RAND_TR997.pdf.

A.7.e. Fast screening for outcomes that vary by small area

The work on fast screening for outcomes that vary by small area is based on the premise that it is easier to estimate the variability across small areas than it is to derive estimates for each individual small area. The goal is to develop a procedure to determine outcomes with large variability across small areas. Resources can then be allocated for the development of small area statistical models for the outcomes with the greatest variability across small areas.

The method used has a simple model with no covariates (similar to the “inconsistency measure” in meta analysis). So far the work that is under evaluation discriminates among NHIS health insurance outcomes at the state level. Early findings have revealed that the method used is sensitive to assumptions about prior distributions if little data are available, and current research is being done to develop better priors.

In another study, using 2010 NHIS binary outcomes, NCHS implemented relative variability calculations (Coffey, Feingold, and Bromberg 1988) based on small-area proportions to identify outcomes with large small-area variability, accounting for demographic factors such as race, gender, and age.

A.7.f. Small area estimates from the NHIS utilizing block-linked American Community Survey data

In joint work with the U.S. Census Bureau, NCHS is creating an NHIS-ACS file at the block-level for developing small area estimates. From the ACS data, detailed estimates of health insurance, overall health, and socio-economic variables are available. The goal is to use ACS estimates as covariates to create "NHIS like" estimates for small areas, with targets to estimate health insurance and access to care outcomes for states and for counties bordering Mexico (based on SAHIE experience).

Research is still in progress. NCHS spent over one year to arrange access to Title 13 data and to ensure that there is no duplication of effort. Work that has been accomplished was fixed-effect modeling to determine covariates and important design variables. Future work may include estimation research on: (1) improving design-based state and local estimates by calibrating to ACS estimates, (2) being as "non-parametric as possible" using multinomial distributions to associate block characteristics (household types, outcomes types between NHIS and ACS) and using the relationships to "fill-in" empty NHIS blocks, and (3) investigating an alternative that combines the best parts of SAHIE and SAIPE experiences (without IRS data) with the NCI's NHIS-BRFSS modeling.

A.7.g. Small area estimation from the Electronic Medical Records supplement to the National Ambulatory Medical Care Survey

The National Ambulatory Medical Care Survey (NAMCS) is an annual national survey of office-based physicians and providers practicing in community health centers (CHC). It collects information about physicians and practice characteristics. Data collection is done by field representatives dispatched to sampled facilities. The Electronic Medical Records (EMR) supplement to NAMCS is a mail survey of office-based physicians designed to collect data about the adoption of electronic medical and health records (EMR/EHR) systems. A combined American Medical and Osteopathic Association (AMA and AOA) database serves as the frame for both surveys. NAMCS also samples CHCs from a separate administrative database. The question of how national growth of EMR adoption is reflected locally in states provides a motivation to produce small area estimates of the adoption of EMR/EHR systems by office-based physicians. To facilitate state-level SAE, the EMR supplement to NAMCS is stratified by states.

NCHS implemented area-level Fay-Herriot models (Fay and Herriot 1979) applied to the log of in-state proportions of any and basic components of EMR systems. County-level covariates from the Area Resource File (ARF) such as demographic, economic and healthcare information from various surveys and administrative databases were aggregated to create state-level covariates.

In this work model selection significantly affects estimated proportions. The issues faced include: (1) determining how much fit should be achieved to attain an optimal balance of direct and regression parts of the composite estimator and (2) ensuring that the proportions adopting EMR systems on any and basic levels are correlated. Where adoptions of both types of systems are growing annually in similar ways, modeling the joint distribution of these variables may be reasonable. In addition, accounting for obvious correlations between these two variables may reduce errors of estimation.

A.8. Substance Abuse and Mental Health Services Administration (SAMHSA)

Presented by Art Hughes, Substance Abuse and Mental Health Services Administration

SAMHSA sponsors the National Survey on Drug Use and Health (NSDUH), which collects data on the prevalence, patterns, and consequences of alcohol, tobacco, and illegal drug use and abuse in the general U.S. civilian non-institutionalized population, age 12 and older. The survey has been conducted since 1971, and annually administered since 1996. The outcomes of interest are binary variables indicating drug use, marijuana use, and health issues such as depression and anxiety. NSDUH oversamples youth ages 12-17 because it is interested in prevention, risk factors, and behavior in schools.

In the early 1990s there was discussion about the need to produce state estimates. However, the survey was designed as a telephone survey, and the response rate was low. This was due in part to the topic. Congress passed legislation mandating state estimates. The state level estimates started in 1999. It was originally thought that direct estimates would be used for the largest states and SAE for the smaller ones; but it was decided that estimates for all states would be derived using SAE methods.

The small area estimates are developed using logistic models with mixed effects at the state and sub-state level. The focus is on binary variables, although there are some continuous variables. The models are hierarchal linear models and incorporate weights where the fixed effects are weighted (survey-weighted hierarchical Bayes). For cross-tabulated variables age, gender, and race, the estimates add up to the state level, and the state SAEs add up to the national.

Auxiliary data for the models include crime reports, treatment rates for alcohol and drug use, overall and drug-related deaths, and unemployment information. The models also use census long-form data from 1990 and 2000 and now will start to include 5-year tract and block level as well as county level ACS data.

Model validation was done by taking large state estimates as the truth, and taking subsamples (replicates) from large states and performing SAE. SAMHSA has been pleased with the results.

To detect change over two-year periods for a state, SAMHSA used to oversample selected states, but the sample still did not have enough power to detect plausible levels of change. As a result, SAMHSA switched to producing two-year moving averages, and this is what is done currently.

One of the challenges in this work has been to convince states that the estimates from SAE are statistically valid. Often the direct estimates and estimates from SAE are very close except in the tails. Another challenge is to meet state requests to produce sub-state estimates. Sub-state areas are defined by the states, but the state defined areas do not always match with sampling units in the multi-stage sample design.

B. Other Federal Agencies

B.1. Food and Nutrition Service (FNS)

Presented by Allen Schirm, Mathematica Policy Research (contractor for FNS)

For nearly two decades Mathematica Policy Research has been conducting small area estimation to produce state level estimates for the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) and the Supplemental Nutrition Assistance Program (SNAP, formerly known as the Food Stamp Program) for all states and DC. For WIC, the estimated numbers of infants and children 1-4 in families with incomes at or below 185 percent of the federal poverty level have been used since 1994 to allocate WIC funds to the states. For SNAP, the estimated numbers of eligible people measure the need for the program, and the estimated SNAP participation rates measure the program's performance in reaching its target population.

For both programs, estimates are derived using empirical Bayes shrinkage estimation methods and data from the CPS, the ACS, and administrative records (currently, mainly income tax data). The estimators combine direct sample estimates and regression predictions and are multivariate, using data for several years to borrow strength not only across areas but also over time. SNAP estimates for all eligible people and the working poor are derived jointly, borrowing strength across these two groups.

The primary means of disseminating the SNAP estimates to a wide audience of policy makers, program administrators, researchers, and others is an accessible and attractive color pamphlet that is available in both printed form and as a PDF. The pamphlet uses "caterpillar" graphics to illustrate the estimates and their uncertainty and to facilitate appropriate comparisons across the states. These estimates can be found at: http://www.fns.usda.gov/ops/supplemental-nutrition-assistance-program-snap-research.

In other work for FNS, Mathematica has carried out microsimulation of SNAP for about 40 years, simulating thousands of proposed changes to the program's eligibility and benefit rules. In the last couple of decades, with welfare reform and other program changes, there has been a much greater need for state estimates. Yet the state samples in the primary survey and administrative databases are small. To improve the precision of the state estimates, Mathematica has reweighted these databases to borrow strength across states. The reweighting method fits a Poisson regression model to obtain an estimated prevalence in each state of every household in the database. This model is specified to control important aggregates at the state level, and the prevalences are expressed as a matrix of weights, with each household having a weight for every state. Estimates for a state are obtained by passing through the microsimulation model all households in the database, not just the households actually in that state. By applying the appropriate weight for each household, the database is weighted to look like the state, rather than the whole country (Schirm and Zaslavsky 1998). An important feature of this method is that the reweighting is performed just once. Thus, modeling does not need to be repeated every time a change to SNAP rules is proposed. An evaluation of the reweighting method found substantial gains in accuracy relative to direct estimation, with increases in precision far exceeding the errors from any biases introduced.

B.2. U.S. Census Bureau

Presented by Wes Basel, Small Area Estimates Branch, U.S. Census Bureau

The U.S. Census Bureau runs one of the oldest small area estimation programs, the Small Area Income and Poverty Estimates (SAIPE) program that provides annual estimates of income and poverty statistics for all school districts, counties, and states. The main objective of this program is to provide estimates of income and poverty for the administration of federal programs and the allocation of federal funds to local jurisdictions. In addition to these federal programs, state and local programs use the income and poverty estimates for distributing funds and managing programs. SAIPE was initiated in the early 1990s when there was a need for median income statistics by state, and the CPS—the primary source of such data at the national level—could not provide such statistics at the state level. SAIPE was developed to address this need and expanded to provide estimates of poverty.

The SAIPE program currently produces the following county and state estimates: total number of people in poverty, number of children under age 5 in poverty (for states only), number of related children ages 5 to 17 in families in poverty, number of children under age 18 in poverty, and median household income. In addition, in order to implement provisions of the No Child Left Behind Act of 2001, SAIPE produces the following estimates for school districts: total population, number of children ages 5 to 17, and number of related children ages 5 to 17 in families in poverty. The estimates are annual and do not borrow strength across years.

SAIPE uses a Bayesian shrinkage estimator that combines direct survey-based estimates and indirect estimates from regression model predictions. In earlier years the direct estimates used data from the CPS, but now data from the ACS are used. For the prediction model SAIPE implemented a Fay-Herriot two-level model (Fay and Herriot 1979) with predictors drawn from auxiliary data such as IRS tax returns, SNAP enrollment, and Census Bureau population estimates. For school district level estimates, due to a lack of predictors at this level, the estimation does not use a regression model but allocates the updated county estimates among school districts in the same proportions that poor school-age children were distributed across the districts in the decennial census data.

In addition to SAIPE, the U.S. Census Bureau also runs a similar SAE program for estimating health insurance coverage for states and all counties: the Small Area Health Insurance Estimates (SAHIE) program, which focuses on demographic details. Similar to SAIPE, SAIHE was created in response to a CDC need for more information on cancer screening for low income, uninsured, and underserved females. SAHIE builds on the work of the SAIPE program.

The biggest challenges faced in the SAIPE and SAIHE programs are data preparation (such as for IRS tax return data), collecting county-level SNAP data, understanding tax return data at the county level, and identifying sub-county boundaries (every year that school district boundaries are updated). Among the goals with regard to auxiliary data, the Census Bureau needs to do a better job of associating tax returns with areas. In addition, since there is no more census long form data (which was used as the gold standard), the evaluation of sub county estimates presents a challenge, which is being addressed with ACS data—the replacement for the census long form. Other issues faced include developing greater transparency, disseminating the estimates, understanding synthetic estimates, and creating better documentation. Plans for future development of these programs include extending and improving the input data; for example, using tract-level data to improve the ACS estimates and using the ACS jointly with other surveys such as the Survey of Income and Program Participation (SIPP).

B.3. National Center for Education Statistics (NCES)

Presented by Dan Sherman, Education Program, American Institutes for Research (contractor for NCES)

NCES administers assessments to measure different dimensions of achievement and literacy, including national and international assessments that measure skills and achievement in reading and math at the national level. NCES is interested in state and county-level data to identify high-need areas to help direct funding such as federal grants to states for adult education and allocation of state funds to counties. These data can also help in targeted social marketing; that is, how best to create materials that will be understood by the population in an area.

NCES assessment data include the National Assessment of Educational Progress (NAEP), an ongoing annual assessment that measures 4th and 8th grade reading and math achievement; the 1992 National Adult Literacy Survey (NALS), which had state-level samples (with emphasis on subgroups); and the 2002 National Assessment of Adult Literacy (NAAL), which measured adult literacy in English, and health literacy at the national and state levels. In addition, the U.S. has participated in international surveys including the Trends in International Math and Science Study – TIMSS), which has been conducted since 1999 and covers 4th and 8th graders, and the Program for International Assessment of Adult Competencies (PIAAC), which was conducted in 2011-2012.

The primary interest of NCES in small area estimates has been in geographic areas rather than demographic subgroups. Data from national surveys are often very thin for state level analysis. For example, the 2002 NAAL covered only 11 percent of counties, many with less than 20 individuals. The 2007 TIMSS covered 250 schools; only 5 states had 10 or more schools.

For the 2009 NAAL/NALS, NCES conducted a study with the goal to produce state/county-level estimates of percent “below basic” literacy. The study used a complex Hierarchical Bayes (HB) estimation model to make predictions and put “credible intervals” around the point estimates. The model used the 1990 and 2000 census data for predictor variables at the county level. Computations were performed using the Markov Chain Monte Carlo (MCMC) method (Gelman et al. 2004) and implemented using the WinBUGS software (Lunn et al. 2000). Among the challenges faced was that the model was that the indirect model (using mostly demographic predictors) did not fit very well given the small amounts of direct information in most sample counties, so the county estimates used in the predictive model were very noisy.

Given current sample sizes, NCES will likely have very limited direct information for SAE. Since there is little strength from which to borrow, the key for indirect estimation is to have a good model. In practice, it is preferable to work with models that can be easily programmed and quickly run. For communicating the results to clients/data users, it is very useful to present the point estimates and confidence intervals using snake (or caterpillar) charts. This type of chart allows ready comparison of unit estimates and highlights the ability to distinguish areas from one another.

The state or county level estimates can be obtained from the NCES website at: http://nces.ed.gov/naal/estimates/StateEstimates.aspx.

B.4. Department of the Treasury

Presented by Robin Fisher, Office of Tax Analysis, Department of the Treasury

The Office of Tax Analysis (OTA) in the Department of the Treasury is responsible for the development, analysis, and implementation of tax policies and programs, such as analyses of the effects of the existing tax law and alternative tax programs. OTA develops and operates several major microsimulation models and maintains large statistical databases to analyze the economic, distributional, and revenue effects of alternative tax proposals and tax systems. Many of the large microdata files used in OTA's models are developed from samples of tax returns prepared by the Internal Revenue Service (IRS) Statistics of Income (SOI) Division with direction from OTA.

OTA has conducted microsimulation that produce state level estimates. For these microsimulations, OTA used the Individual Income Tax Model (ITM) sample, which is a stratified random sample of the population of Form 1040 returns selected from the IRS individual master file and edited and imputed so that data no longer have out-of-range values, inconsistencies, or missing values. The ITM samples are weighted to represent the entire population of taxpayers. To produce state estimates, a regression model that predicts the probability that a sample return belongs to each state, conditional on the covariates, is estimated for a given year. For any return in the sample the vector of estimated probabilities is multiplied by the sample weight, yielding 51 state weights.

This approach was evaluated by comparing the ITM estimates to the population file. This produced satisfactory results for most of the estimates. However, some estimates clearly showed significant differences which were attributed to estimation or editing issues. The variance estimation was only partly implemented, but OTA wanted to know if certain of the results were an artifact of sampling or modeling. The modeling effort includes using extensive covariates to control modeling error. OTA is also looking for ways to control sampling error.

C. State Government

C.1. State Health Access Data Assistance Center (SHADAC), University of Minnesota

Presented by Kathleen Call, State Health Access Data Assistance Center, University of Minnesota

SHADAC is a program funded by the Robert Wood Johnson Foundation and a part of the Health Policy and Management Division of the School of Public Health at the University of Minnesota. It helps states collect and analyze data to inform state health policy decisions relating to health insurance coverage and access to care. Its goal is to help states bridge the gap between health data and the policy-making process.

For Minnesota, SHADAC utilized data from the Minnesota Health Access Survey to estimate county-level uninsured rates for the year 2009, along with estimates of uncertainty. In choosing the SAE method, SHADAC used accessible methods that can be applied to other states. To produce the small area estimates of uninsured rates, SHADAC used direct estimates from the

Minnesota Health Access Survey, 5-year estimates of demographic variables from the ACS, and the 2007 SAHIE. These data sources were combined via a simultaneous equation model.

SHADAC has expanded the work to the state of Oklahoma.

This page has been left blank for double-sided copying.

REFERENCES

Battaglia, M.P., M.R. Frankel, M. Town, and L. Balluz. "A virtual popluation approach to small area estimation." *Proceedings of the Section on Survey Research Methods.* Alexandria, VA: American Statistical Association, 2011.

Cadwell, B.L., T.J. Thompson, J.P. Boyle, L.E. Barker. "Bayesian small area estimates of diabetes prevalence by US county, 2005." *Journal of Data Sciences*, vol. 8, 2010: pp. 173-188.

Coffey, M.P., M. Feingold, and J. Bromberg. "A normed measures of variability among proportions." *Computational Statistics & Data Analysis,* vol. 7, no. 2, 1988, pp. 127-141.

Fay, R., and R. Herriot. "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data." *Journal of the American Statistical Association,* vol. 74, no. 366, 1979, pp. 269–277.

Gelman, A., J. Carlin, H. Stern, and D. Rubin. *Bayesian Data Analysis.* 2nd edition. Texts in statistical science. Chapman and Hall/CRC, 2004.

Jiang, J., and P. Lahiri. "Estimation of finite population domain means: A model-assisted empirical best prediction approach." *Journal of the American Statistical Association,* vol. 101, no. 473, 2006, pp. 301-401.

Lunn, D.J., A. Thomas, N. Best, and D. Spiegelhalter. "WinBUGS—A Bayesian modeling framework: Concepts, structure, and extensibility." *Statistics and Computing,* 10(4), 2000, pp. 325-337.

Malec, D., J. Sedransk, C. Moriarity, and F. Leclere. "Small-Area Inference for Binary Variables in the National Health Interview Survey." *Journal of the American Statistical Association,* vol. 92, no. 439, 1997, pp. 815–826.

Rue, H., S. Martino, and N. Chopin. "Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations." *Journal of the Royal Statistical Society,* Series B, vol. 71, part 2, 2009, pp. 319-392.

Schirm, A. and A. Zaslavsky. "Model-Based Microsimulation Estimates for States when State Programs Vary." *Proceedings of the Section on Survey Research Methods.* Alexandria, VA: American Statistical Association, 1998, pp. 339-344.

Srebotnjak T, Mokdad A, Murray C. A novel framework for validating and applying standardized small area measurement strategies. *Population Health Metrics,* vol. 8, no. 6, 2010, pp. 1-13.

Voulgaraki, A., R. Wei, and B. Kedem. "An Approach to Handle the Problem of Zero Deaths in Estimating Mortality Rates." *Proceedings of the Joint Statistical Meetings, Section on Statistics in Epidemiology.* Alexandria, VA: American Statistical Association, 2008.

You, Y., and J.N.K. Rao. "A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights." *The Canadian Journal of Statistics,* vol. 30, no. 3, 2002, pp. 431-439.

APPENDIX C

DISCUSSION SESSIONS

This page has been left blank for double-sided copying.

Prior to the meeting, participants were issued a set of discussion questions for each of the topics and subtopics in the five discussion sessions. These discussion questions are listed below. Following the questions we present summaries of the discussion in each of the sessions.

Topics in Using Small Area Estimation (SAE)

What Types of Data Needs Can SAE Address?

- What motivated your agency's development of the small area estimation program that you described earlier? Were you seeking to improve existing estimates, or were you seeking to develop new estimates?
- If your methodology combines direct and indirect estimation, do you prepare estimates for areas for which direct estimates are not possible?
- Did you find that SAE was not suitable for particular estimates that your agency wanted to generate?
- If you do not currently apply SAE methods, did you find that your needs could be met without such methods, or were there other considerations?
- Would you like to expand your use of SAE methods?

Implementation Issues (technical staff, resources, development time)

- Does your organization have personnel who can develop and implement SAE methods?
- Has your organization used external resources to develop or implement your methods? Resources from other government agencies? Expert panels? Consultants or contractors?
- Have staffing issues (technical expertise or staff size) limited the development or application of SAE methods?
- How long did it take to develop the methods that you now use?
- Are there other issues you faced in implementing SAE?

Choosing a Method

- To what extent did you explore alternative methodologies in developing your current approach to producing small area/domain estimates?
- Did you develop and compare estimates from more than one methodology? Variants on one methodology?
- If yes to either, what criteria did you use to select a method?
- Do you have any recommendations for other agencies that are seeking to apply SAE?

Identifying Suitable Auxiliary Data

- Did you conduct an extensive search for auxiliary data, or were you already aware of suitable data when you undertook the development of your small area estimates?

- What factors were important in selecting suitable auxiliary data?
- Are you seeking to improve your auxiliary data?
- If yes, are there particular properties that are lacking or need to be enhanced?

Evaluation

- How did you evaluate the quality of your small area estimates?
- Did you have an independent estimate of "truth," or did you use another approach?

Interpreting and Communicating Results (describing uncertainty)

- Can and do you generate measures of precision for your small area estimates?
- Have you documented the improved precision gained from your small area estimates compared to direct estimates?
- Do you present estimates of uncertainty with your estimates?
- If not, how do you characterize the uncertainty for your users?
- Have you sought to determine a best way to describe uncertainty?
- How do you communicate results and uncertainty in a way that is meaningful to policy makers?
- Have the estimates of uncertainty generated any issues with users or with senior managers?

Sources of Auxiliary Data and Issues in Using Them

Administrative Records

- Are there issues of timeliness or quality with the administrative data you use?
- What other limitations have you encountered?
- Are there administrative data that you would like to use but cannot obtain access?

Other Surveys

- Do you use more than one survey to produce your estimates?
- In what ways are the additional survey data used?
- Are there issues of timeliness or quality with these data?

Impact of the American Community Survey (ACS)

- For those of you who used decennial census long form data, how does its replacement by the ACS affect your methodology/estimates?
- Will the ACS supplant any of your estimates?
- Will it potentially enhance any of your estimates? How?

New Sources

- What new data sources are in development or about to become available?
- Will the Affordable Care Act generate new data?

New Developments in Methodology

- Are you currently exploring any new methodologies?
- Are you aware of new methods that you would like to consider?
- Having heard the several presentations, do you have any suggestions from your own experience? Or from your familiarity with the methodological literature?

Improving Cross-Agency Collaboration on Methods and Data

- Have you collaborated with other agencies in developing or implementing your methods?
- If not, would you find such collaboration useful? In what ways?
- If yes, how have you found this collaboration useful? In what ways (for help with methodology or data or both)? Would you expand such collaboration if you could? In what ways?

SESSION 2

TOPICS IN USING SMALL AREA ESTIMATION (SAE)

2a. What Types of Data Needs Can SAE Address?

State estimates are a goal of many surveys, but there is a growing need for substate estimates. For example, communities are required to produce needs assessments in certain areas related to health care, and small area statistics provide a potential data source. BRFSS with its state-based design and large state samples is better positioned to address such needs than most federal surveys. But survey coverage of small areas is a challenge. The ACS provides such coverage, but the data must be aggregated over 3 or 5 years to provide estimates for most substate areas. Timeliness becomes an issue when data are pooled over time, and data aggregated over time do not lend themselves to measuring change as well as single-year estimates. SAE can fill the need that these circumstances create, but some small areas may model better than others. For example, county boundaries, which are political in origin, may not align with the phenomena being modeled. Geocoding of areas that do not align with political boundaries may be necessary. Ideally, this should be done by the entities requesting the estimates. In addition model development takes time. If the estimates take too long to produce, their value diminishes.

2b. Implementation Issues (technical staff, resources, development time)

The application of SAE requires technically skilled staff, but production work may not attract Ph.D. statisticians. For a federal agency, contracting provides an alternative to hiring staff, but there might not be sufficient funds readily available for contracting to provide a long-term solution. Collaboration between agencies and partnering with universities can be an effective way for agencies to gain access to experienced and skilled staff. The recent collaboration among NCI, NCHS/CDC and the University of Pennsylvania in a project involving BRFSS and NHIS data provides a good example.

Panelists had a lot to say about software issues. Software packages reduce the need for high-level programmers, but they may not be able to handle more complex applications or large datasets. WINBUGS is a popular software package for SAE. Several of the panelists have made use of it, and a number continue to do so. But several users reported problems with convergence, slow speed, and its inability to detect model over-specification—a problem noted with R and SAS as well. The Census Bureau and SAMHSA have programmed their estimation entirely in C. The CDC has used R but calls C routines while NCHS has used FORTRAN. The up-front development costs in moving away from packaged software are high, requiring a commitment to devote sufficient resources. Furthermore, the need for a strong programmer in such languages continues beyond the initial development, as the software requires modifications over time. But the payoff for customized software can be substantial. At the same time, there are drawbacks. States may want to be able to reproduce the estimates that are provided to them. Software choices may have to be made with such considerations in mind. Transparency is important.

The issue of benchmarking was raised—that is, adjusting state estimates so that they sum to national totals, or adjusting substate estimates to sum to state totals. One panelist noted that benchmarking adds a great deal of time to the production and complicates the overall model and wondered if the gain was worth the effort. Another issue is what to do with areas that have strong direct estimates; should they be included in the benchmarking, which may weaken their estimates, or should they be allowed to retain their direct estimates, and the benchmarking confined to the balance of areas? Other panelists discussed ways in which the benchmarking could be incorporated into the modeling process itself. Variance estimation is also made more problematic by benchmarking, although two panelists suggested ways in which the variance added by benchmarking could be incorporated into an overall variance estimate without complicating the modeling process.

The session closed with an observation that states that receive funding to collect survey data may be opposed to model-based estimation because it could threaten future funding to collect more data.

2c. Choosing a Method

With regard to the choice among alternative methods, no one reported conducting a thorough, empirical comparative evaluation of different methods. Typically, the choice of method was made up front, based largely on a review of the literature. Implementing a complex method requires a lot of time and resources (as discussed above). If alternative methods were considered, a decision was reached without an empirical evaluation. Once a basic method was selected, an empirical evaluation may have been conducted to demonstrate the performance of the method against a direct estimator and one or more simple SAE alternatives.

The Census Bureau's SAIPE program was the subject of a National Academy of Sciences (NAS) panel study. Implementation of SAE methods for SAIPE was mandated by law—to produce more up-to-date estimates for Title 1 funding. The NAS panel reviewed the methodology in light of the needs. The panel endorsed the proposed methodology but thought the specific regression model being used was not good enough, so pending model improvements, they recommended that the initial estimates be averaged, 50-50, with estimates from the decennial census. Ultimately, SAIPE worked out very well. The users—states and school districts—have become accustomed to these estimates, and they are rarely challenged. The availability of data from the ACS has not reduced the need for SAIPE, which is more timely at the level of geography required, where the ACS estimates need three or five years of data. A substantially larger ACS sample size would be required to change this, and such an increase is very unlikely. The next step for SAIPE, which produces independent estimates of multiple characteristics, is to explore modeling multiple dependent variables or find another way to link the estimates of different characteristics. What is needed is a method that can account for correlations among the outcomes.

Panelists had some advice for agencies just starting SAE programs. One panelist underscored the need for modeling expertise, as this is a critical component of nearly all SAE methods. Attempting to apply methods "off the shelf" is dangerous without the modeling expertise. In addition to the modeling expertise, those who apply SAE methods need to understand the design of the surveys they are using and the properties of auxiliary data sources.

There were several suggestions with regard to good literature on SAE methods. Rao's texbook is a widely used reference—particularly for Bayesian methods. The book is addressed to a technical audience, however. There was a suggestion to look for presentation slides on SAE—perhaps from a professor. Pfeffermann published an article in 2013 on new developments in SAE, although it, too, is quite technical. The final report of the NAS SAIPE panel, from 2000, contains both technical parts and parts that speak to the broader user community. What is particularly helpful is that it provides a very deep discussion of the application. The 2006/2007 SAIPE report from the Census Bureau was recommended as being very good for people who are new to SAE. The Bureau includes equation-by-equation detail; standard errors; and graphs as well.

Several panelists addressed the question of when one should opt for a small area methodology over direct estimation. On a note of caution, it was pointed out that variance estimation is much more difficult with SAE than with direct estimation—although the latter is often not that easy, given the frequent use of complex sample designs, and may require design information that is not typically included on public use files. With SAE, good variance estimation is not only more difficult; it may be essential to the method working correctly. One panelist mentioned seeing papers in which high CVs prevented the SAE methods from working, but the true problem may have been that the estimates of the CVs were flawed. Something that can help in evaluating variance estimates for SAE is that when the sample sizes for individual areas grow large, the SAE variance estimates should converge to the variance estimates from the survey estimates.

Often the decision to use SAE methods is dictated by the need for an estimate that cannot be produced reliably from a survey. As was noted earlier, the need for particular estimates and even the requirement to use SAE methods may derive from a law, which leaves no option. When the requirement is not a legal one, the need must be for estimates from many areas, as SAE is not feasible for a small number of areas.

A challenge in evaluating estimates from SAE is that the variances from, say, a Bayesian model and the variances from direct estimates may not be strictly comparable. The concept of a CV derives from direct estimation. Moreover, the variances of model-based estimates depend on the correctness of the model. If the model is good, the variance estimates are probably good. If the model is weak, the variance estimates may be too high or too low. Measuring the gain in precision from SAE by comparing CVs is not sufficient, as the results may be misleading.

This latter point was reiterated in a discussion of model over-specification. One panelist asked how one could tell if a model was over-specified. Overly small standard errors may be an indication. An example was cited in which many of the models that were tested had comparable standard errors but one model had very low standard errors. Another example involved a Bayesian model with small standard errors but negative estimates for the variable being estimated, which should take only positive values. With Bayesian methods, a problem with the model may be reflected in the posterior distribution: a posterior mean may not exist. Another panelist noted that state estimates present a particular challenge because there are a small number of degrees of freedom relative to the great many auxiliary variables available, potentially. Looking at the residuals can be very informative. The improvement from adding a covariate may be confined to one state, indicating over-fitting.

Another modeling issue applicable to estimates that are repeated each year is what to do if a variable that was used in the base year model is no longer significant. More generally, changing the model may shift the entire series of estimates. When funding allocation depends on the estimates, changing the model may be appropriate statistically, but it could mean that a state receives less money. From a policy perspective, this outcome may not be good. The possibility that this might occur is informative to both the policy maker and the statistician. It was noted that the issue of re-specifying models or even changing methods has an analog in the survey field, where small changes in question wording can have a big effect on the responses. In such cases, the change must be introduced gradually.

A panelist noted that model-based estimates can have better properties than direct estimates. The final report of the NAS panel on formula allocation, which followed the SAIPE panel, has a good discussion of how the properties of allocation systems interact with properties of the estimates.

2d. Identifying Suitable Auxiliary Data

The choice of auxiliary variables is obviously important to the application of SAE methods, but to what extent do users evaluate potential auxiliary variables versus selecting those that are most readily available? In response to this question, panelists identified a number of issues with respect to the selection of auxiliary variables to include in SAE models.

The quality of auxiliary variables is clearly important to the quality of small area estimates that use those variables. Evaluations of auxiliary variables conducted by the agency that produced the variables may not be sufficient for applications outside of those for which the variables were collected. Given the opportunity, it may be advisable for the agency using the auxiliary variables for SAE to reassess their quality.

Another panelist noted that a lot of effort may be expended on finding potential covariates when models using gender, age, and race may work just as well in certain applications.

Continued availability of variables over time is an issue, but so is the fact that variables used in the modeling are taken off the table for future analyses, when they might be of interest, for example, as factors predicting between-area variation in an outcome. In other words, including particular variables in the SAE model for an outcome may limit the usefulness of future research on why that outcome varies among areas.

Another panelist pointed out that if one of the goals in preparing small area estimates is to measure change over time, this requires that the auxiliary variables include indicators that will reflect change in the outcome being measured. Decennial census variables may be good predictors of an outcome cross-sectionally, but because they are fixed in time they will not predict change. In modeling health insurance coverage, Medicare and Medicaid administrative data provide good indicators of change in public coverage, but there is no counterpart for private insurance coverage, which accounts for most health insurance.

It was noted that at one time, the Area Resource File (ARF) seemed to be the only source of administrative data that was available. One could talk to people who worked with the ARF to find variables that were correlated with the outcome. It is also important that auxiliary variables be uniform in quality across small areas; otherwise the differences in quality will contribute to model error. How many datasets are uniformly collected across areas? There may be issues with respect to the uniformity of definitions on particular variables, and each agency has its own quality control.

IRS data provide a relevant example. Tax returns contain addresses, which are used to assign returns to small areas, but taxpayers are not required to report a home address. Some filers report a work address or an accountant's address. It is not known what fraction of addresses in tax data are home addresses or how this varies by state or area. The IRS publishes tax data by ZIP Code, and a researcher outside the agency prepared a paper on the income of residents of buildings with their own ZIP Codes. It turned out, however, that none of the taxpayers filing returns with these ZIP Codes actually lived in those buildings.

Sometimes auxiliary variables do not measure what they appear to measure. This may be due to how an agency defines those variables, or to systematic measurement error, as in the IRS data. If the data come from the states, each state may have its own quality control. Uniformly poor data may be better than a mix of good and bad data if one cannot determine which is which. On this point it was noted that Medicaid claims data are provided by the states, and states may have different programs for different areas within their boundaries as well as different collection and validation methods. This may be okay for estimates informing state policy, but it becomes problematic for estimates used to inform national policy or conduct comparisons across states.

One panelist asked how many agency representatives were using the ACS as a source of auxiliary data and wondered if it was viewed as the best thing out there. SAMHSA is using it to replace the variables that were formerly collected on the census long form, but they are not sure what to make of changes in question wording. SAIPE and SAIHE are using ACS data as well.

One panelist noted that auxiliary variables often come from a sample survey, which poses new technical challenges. Another cited the imputation of missing data in the ACS as a specific example of such a challenge.

Multivariate modeling was raised again in response to a comment that the Census Bureau could combine health and poverty information at the county level. Each individual outcome is modeled separately with a different set of covariates.

In response to a question about other auxiliary variables, the Census Bureau panelist mentioned SAIPE's use of tax data. The Census Bureau receives microdata from the IRS. The tax data provide variables that can approximate poverty at the county level. If the tax data were not available, other variables could make important contributions, but they do not add much when tax data are included. The other datasets miss the same things that the tax data do.

2e. Evaluation

For evaluation, the SAIPE program had the decennial census long form data, which provided the best available estimates of small area poverty every ten years. With the ACS replacing the long form, the ACS now provides the best small area poverty data for evaluating the SAIPE estimates. The ACS 5-year sample is being used as an artificial population. A sample is pulled and used to compute a set of SAIPE estimates that are evaluated against this population. An alternative approach is to split the ACS sample in half and use one half to evaluate predictions made using data from the other half.

Mathematica's work for FNS also used simulations. After creating an artificial population, many samples were drawn, and direct and indirect estimates were derived for each sample and compared with the true values from the constructed population. Statistics such as mean absolute error and mean squared error, which reflected both sampling error and bias, were calculated to assess the relative accuracy of alternative estimators.

SAMHSA uses a similar approach to evaluate small area estimates from NSDUH based on subsamples taken from large states. In addition, BRFSS provides data for an external evaluation, since it also captures data on alcohol abuse. Maps are produced to show the extent of correspondence between the estimates from the two surveys.

A question was posed as to how one designs a simulation study of a Bayesian model. It was noted that a paper by Lahiri, Liu and Kalton used simulation to look at the design-based properties of SAE. Some models that are good from a Bayesian perspective may not have good design-based properties. For example, a model may have 95% Bayesian coverage of the true value but not necessarily 95% coverage from a repeated sampling perspective. NCI has compared direct and model-based estimates. The model-based estimates were compared to cancer registry data and matched well. NCI is evaluating three models of tobacco use, based on data from the CPS.

In its own evaluations, the CDC assessed how well SAE preserved known correlations—for example, between obesity and diabetes. Most of the methods that they tried preserved these relationships.

Early NCHS modeling used cross-validation. For example, 20 percent of counties were removed, the models were re-estimated, and then applied to estimate the counties that were removed. Posterior predictive checks were used as well. Such checks can be used to evaluate the model by checking how well the model predicts design-based estimates. More recently, NCHS has used simulations. They assume an observed proportion is the true proportion and calculate confidence intervals for the predictions, repeating this many times to see how often the confidence interval contains the assumed true value.

2f. Interpreting and Communicating Results (Describing Uncertainty)

Two panelists spoke to the value of maps for both presenting results to non-statisticians and helping researchers identify patterns. For example, in a map of doctor visits, the area along the Appalachian Mountains stood out, which also supported the face validity of the estimates. In addition, NCI showed focus groups their website, which contained documentation explaining the methods, but users did not have time to read the details. They thought that a map would have been helpful.

An issue that NCI has confronted is how to advise users that particular estimates are model-based versus direct estimates and that standard errors alone may not be sufficient to characterize uncertainty. The confidence intervals for small area estimates can be smaller than those of the direct estimates, but one has to interpret the results from the two methods differently. To assist users, NCI developed locally specific information about the use of local versus borrowed data.

There was discussion of how NCI used BRFSS data, prompted by users wanting to know how much information was obtained by adding additional datasets. NCI observed that, for BRFSS, the answer depends on the outcome. Perhaps because it is a random digit dial survey, BRFSS reports higher prevalence for some outcomes than housing unit sampling. The modeling that NCI did assumed that NHIS was the gold standard, but NHIS does not have data for counties. Direct estimates for counties are taken from BRFSS, with a bias correction factor included in the modeling stage. Three separate direct estimates are produced—from NHIS households with telephones, NHIS households without telephones, and BRFSS. These are combined to create a joint sampling distribution. In the absence of a gold standard for county-level estimates, BRFSS is used to borrow strength.

With regard to sitting down with policymakers when they request more estimates, SHADAC representatives noted that originally their customers wanted only direct estimates, which encouraged them to modify the design of their state survey to provide better information for counties. SHADAC has started creating more small area estimates, but policymakers are still unsure about them, so they are not being used. Now we have the ACS, which provides estimates that were produced previously by SAE. The combination of 5-year ACS estimates and SAIHE estimates is hard to top in the eyes of customers.

From ASPE’s experience, both a nontechnical explanation of the methodology and an explanation of uncertainty were desirable. Congress, on the other hand, wants point estimates. The issue there is how to communicate that the level of uncertainty may be different than for sample estimates.

For clients who are more technically oriented, a re-simulation of the survey data, comparing errors from direct estimation versus SAE, may help to boost confidence in the estimates.

Another panelist added that a lot of times faith in direct estimates is misplaced. One principle of effective communication is to provide a balanced presentation. Often we focus on the properties and problems of SAE without doing the same for direct estimates. Direct estimates and their standard errors are often provided without a thorough discussion of the uncertainty in the estinates and the implications of that uncertainty for important uses of the estimates. Another principle of communication is to not debate the merits of alternative SAE methods in the context of a particular set of estimates that identifies potential winners and losers. Instead, bring together the relevant stakeholders and discuss the evidence pertaining to the strengths and weaknesses of alternative methods, including the method (if any) that has been used to date and new methods under consideration. Based on that discussion, reach agreement on how to proceed. Another panelist added that it may be helpful to focus on the purpose of the estimates. That is, find out exactly what the user wants to compare. The answers to these questions establish the pros and cons for direct estimates versus modeling. For example, if the user wants to compare estimates over time, then 5-year ACS estimates will not meet that need.

SESSION 3

SOURCES OF AUXILIARY DATA AND ISSUES IN USING THEM

There are issues in obtaining different types of administrative records. In the case of the Census Bureau, the agency receives tax data under an inter-agency agreement, and there is a system in place for getting projects approved to use these data within the Bureau.

For those seeking access to internal Census Bureau microdata for SAE, direct collaboration with the Bureau may be the only option. An example involving validation of Medicaid Social Security numbers underscored this point.

Some types of data sharing may be possible only with changes in the law. But changing the law will not ensure that a data series valuable for SAE continues to be produced. Several years ago Canada changed its tax laws and removed the number of exemptions from the tax return. This greatly reduced the value of Canadian tax data for population estimation. When SAE relies on an external source of administrative records, there is always a risk of having to work with less effective auxiliary variables because the administrative data have been changed or even eliminated.

NCHS used administrative records from AHRQ on hospital emergency departments to model components of electronic medical and health records that were measured in an NCHS survey. The survey was not well-suited for SAE, so the model predictions of the survey estimates were poor. Nevertheless, the model-based estimates were shown to be superior to the direct estimates. This was a good exercise for convincing clients of the value of modeling.

A panelist noted that administrative records have their unique strengths and weaknesses. It can take a while to find someone who knows the data well enough to explain these strengths and weaknesses. Another panelist mentioned that with medical claims data, there are situations where two different types of providers may perform the same service but code them differently.

Another panelist cautioned that because users lack control over the administrative data in their models, they need to be alert to potential inconsistencies and errors or to changes over time. For example, differences among regions may have explanations that can be found in news events, and changes over time may reflect changes in policy incentives that affect responses. Building on this point, another panelist underscored the importance of involving program administrators and subject matter experts, who can explain data anomalies, such as the impact of major storms on counts of program beneficiaries. Other panelists noted that states may send updates to their earlier data submissions; users need to be aware that this may occur. In general, information on the limitations of administrative databases is not widespread among users. One reason is the lack of a good mechanism for disseminating such information. Journals rarely publish articles reporting poor data quality, although one panelist noted an example in which a particularly bad dataset was the topic of an article.

Besides the ACS, other new sources of auxiliary variables are few in number. Implementation of the Affordable Care Act will generate new data and perhaps improve the quality of address information in tax data for at least part of the population. The Longitudinal Employer and Household Dynamics (LEHD) program at the Census Bureau has created a synthetic cross tabulation of residence by place of work—at the block level. The program has produced a nice tool for displaying such data, called On the Map. Users can download a dataset. Another panelist observed that while the data are synthetic (imputed from the actual data, which cannot be released), they are still useful. Three states do not participate, but this panelist imputed them using ACS data. The greater errors that characterize synthetic data are a concern. HRSA is enhancing the ARF by developing a "minimum" dataset for each state. It is not yet clear where the data will ultimately reside or how it will be made available.

SESSION 4

NEW DEVELOPMENTS IN METHODOLOGY

Panelists were familiar with only two examples of new methodology. One involved causal modeling using large datasets, which came out of Artificial Intelligence research by [Judea] Pearl at UCLA. It provides an alternative to correlational modeling. Another panelist reported that there is now a third generation of R software for person-level models. The user can incorporate weights using a pseudo-likelihood, although it is not as easy to incorporate design components (stratification and clustering) into the model, although he and his colleagues are trying an approach to doing this. They have been able to produce MSE estimates that incorporate all of the design features, optimizing the model parameters and variance components to minimize design-based MSEs.

SESSION 5

IMPROVING CROSS-AGENCY COLLABORATION ON METHODS AND DATA

The principal comment on cross-agency collaboration was an observation that it is difficult enough to cross divisions within an organization, let alone go across agencies. Issues include changes in personnel that set back progress and unevenness in the distribution of costs and benefits. Common interest in a successful collaboration is very important.

SESSION 6

SUMMARY AND CLOSING REMARKS

The representative of the Office of Statistical Policy within OMB asked if there was interest in creating a group under the FCSM that could communicate on a more regular basis—similar to an existing group on data confidentiality, which is thriving. Several participants endorsed the idea and felt that such a group would make a useful contribution.

This page has been left blank for double-sided copying.

APPENDIX D

SMALL AREA ESTIMATION IN THE FEDERAL GOVERNMENT: A REVIEW OF LITERATURE

This page has been left blank for double-sided copying.

I. INTRODUCTION

Sample surveys are an important source of data on a wide range of topics, but surveys are limited by their size, which is constrained by cost, burden on respondents, and other considerations. Users who value the estimates that particular surveys produce often want to obtain estimates for subpopulations, or domains, that are smaller than the surveys were designed to support. Frequently such domains are geographic, but they may be socio-demographic as well (for example, age groups or racial/ethnic groups).[12] Censuses have a different type of limitation. Sample size may be no issue in producing estimates from a full population census, and depending on the size of the domain it may present only a minor issue for an associated sample survey conducted with a substantial fraction of census households (such as the census long form),[13] but in the U.S. a census is conducted only once every 10 years and provides no data for the 9 years that precede the next census.

When direct estimates for individual domains—that is, estimates based on the sample observations for those domains alone—lack the statistical precision or the timeliness that is needed for specific applications, users may elect to use the estimates anyway (perhaps unaware of their limitations), but better options are available. Small area estimation (SAE) encompasses a broad range of statistical techniques that were developed to address the situation where direct sample estimates are inadequate or simply not available for all the domains of interest. SAE methods make use of indirect estimation approaches (Federal Committee on Statistical Methodology 1993). Techniques for SAE—and indirect estimation generally—date back many decades, but they have experienced considerable growth in both their variety and use in the past twenty years.

Central to SAE is the notion of "borrowing strength," which involves using data from related areas or time periods and other data sources to increase the effective sample size when direct estimates are not sufficiently precise (Rao 2003). Typically, but not always, these data are brought into the estimation through a model (which may be explicit or implicit) that links the area of interest to the related areas or time periods. This linkage is accomplished through supplementary information—auxiliary variables—associated with the variable of interest.

Related areas or time periods generally do not provide unbiased estimates of the variable of interest in the area and time period for which an estimate is desired, and the bias does not diminish with effective sample size, but if the bias from the model is small enough, the reduction in variance may be sufficient to reduce the mean squared error (MSE). That is the goal of SAE—to reduce the variance of a small area estimate by more than enough to compensate for the bias that is introduced by borrowing strength and thus produce an estimate with a smaller MSE than is obtained with direct estimation. SAE can also be applied in situations where a direct estimate

[12] There is a literature on small area analysis, and much of it lies in the field of public health. In this literature the small area unit is not a given but rather a subject of analysis. For example, a research question may be posed in these terms: what is the right definition of local area for studying access to health care? For a different problem the question might be posed: how do we define neighborhood for the purpose of studying neighborhood influences on behavior? Siordia (2013) addresses the first question, and Cutchin et al. (2011) address the second. In SAE, small area definitions are fixed at the outset.

[13] For very small areas, including many counties and school districts, even the long form samples were small and could not provide precise estimates of very many characteristics.

of a variable for the area of interest may not exist but information that can be used to generate an indirect estimate is available.

This review focuses on applications that illustrate current and recent uses of SAE—primarily in the federal government. Chapter II provides an overview of methods used to produce small area estimates, including illustrative examples. Chapter III details a number of production applications of SAE, and Chapter IV discusses methodological research and programs in development.

II. METHODS USED TO PRODUCE SMALL AREA ESTIMATES

Rao's 2003 book is the most current text on SAE, and it has become a standard reference and handbook on SAE methods. In Rao's presentation, borrowing strength is accomplished through models that link the small area to the other areas or time periods that contribute to the small area estimates. The models can be implicit or explicit. Both types of models make use of supplemental data, but explicit models take account of the variation between areas that is not explained by this additional information. Most of the examples of SAE in the federal government discussed in the later chapters involve the use of explicit models, but methods using implicit models remain in use today—sometimes in combination with methods involving explicit models.

Borrowing strength can also be accomplished without an implicit or explicit model. Examples of techniques for producing small area estimates by using external information to adjust the survey weights or combining estimates across areas or over time are discussed in Section A. Section B discusses SAE methods based on implicit models, and Section C discusses methods using explicit models. Section D summarizes a number of comparative evaluations that encompass all of these techniques.

A. Small Area Estimates without Models

Three approaches to developing small area estimates without models are considered here: (1) adjustment of survey weights using external information to enhance precision, (2) smoothing the estimates for individual areas using the estimates from surrounding areas, and (3) pooling estimates over time.

1. Adjustment of Survey Weights

The precision of direct estimators can be improved through adjustments to the weights or to the estimates themselves. In fact, calibration of survey weights has attracted considerable interest in recent years. Adjustment of the weights or the estimates themselves is also a way to improve the quality of small area estimates from surveys that were not designed to produce such estimates with satisfactory precision. By reducing variability or bias—or both—the adjustments can reduce the MSE of direct estimates, making them more useful than the unadjusted estimates.

The National Survey of Children with Special Health Care Needs is sponsored by the Maternal and Child Health Bureau and conducted by the National Center for Health Statistics (NCHS). The survey is a module of the State and Local Area Integrated Telephone Survey and was designed to support national and state estimates but not substate estimates. Nevertheless, the state samples are comparatively large. Nationally, interviews were conducted with nearly 200,000 households, and almost twice that many children were screened. Children with special health care needs (CSHCN) are defined as children with physical, developmental, behavioral, or emotional conditions that require health or related services beyond what is needed by children typically. Bramlett and Blumberg (2008) developed estimates of the prevalence of CSHCN for metropolitan and micropolitan statistical areas and individual counties. By applying ratio adjustments to the survey weights in counties with sufficient sample to generate prevalence estimates for specific subpopulations, they were able to produce satisfactory estimates for 74 metropolitan areas and 45 individual counties. The sample size threshold was determined from a power analysis of the minimum detectable difference between two areas. For each small area, estimates of the prevalence of children with special health care needs were produced for subpopulations of children by age (0–4, 5–9, 10–14, 15–17), sex, and ethnicity (Hispanic/non-Hispanic).

Both the Behavioral Risk Factor Surveillance System (BRFSS) and the National Health Interview Survey (NHIS) collect data on cancer risk factors. The BRFSS is a state-based telephone survey with a sample that can produce reliable estimates for states and large substate areas, but it excludes non-telephone households and has a much lower response rate than the NHIS. Elliott and Davis (2005) proposed an extension of dual-frame estimation that allows each survey's strengths to compensate for the other survey's weaknesses. Specifically, they combined the two datasets, and using variables common to the two datasets they estimated the propensity that a record in the combined dataset was in the NHIS.[14] Good predictors for this purpose are those that are related to the outcomes of interest (specific cancer risk factors) and whose estimates differ between the two surveys. They then adjusted the BRFSS weights so that the distribution of propensity scores among the BRFSS cases approximated the distribution among the NHIS cases. With the adjusted weights they estimated male smoking prevalence and female mammogram usage for all counties with at least 50 BRFSS respondents in the relevant subpopulation.

2. Spatial Data Smoothing

Health researchers studying the distribution of health practitioners, health conditions or outcomes, and the use of medical services often make use of spatial smoothing techniques to reduce the variability of estimates. For very small geographic areas such techniques may be applied even when the data source is administrative records (such as vital statistics) rather than a survey, as estimates with no sampling error may nevertheless have other random variation that can obscure patterns. Smoothing algorithms borrow strength across areas in a very obvious way. Numerous types of spatial smoothing algorithms have been applied—some more formal than others. One example that is illustrative of the general approach is discussed here.

Pickle and Su (2002) implemented a spatial data smoothing technique whereby direct survey estimates of county-level indicators of health insurance coverage and health risk factors were smoothed using an iterative algorithm called "weighted head-banging" (Mungiole et al. 1999). In this application, each county had up to 27 neighboring counties. For each county, first, the median estimated prevalence rate was calculated for neighboring counties. Second, these neighboring counties were then grouped according to whether their estimated prevalence rates fell above the median or below the median. At this step, each county had two groups of surrounding counties: "high counties" and "low counties." Third, a weighted median within each group was calculated, where the weight was based on county population. That is, at the end of this step each county had a median from the high group (called the "high screen") and a median from the low group (called the "low screen"). Fourth, the original prevalence rate for the county was then compared to the high and low screens, and the smoothed prevalence rate was calculated as: (1) the low screen, if the original prevalence rate was below the low screen, (2) the high screen, if the original prevalence rate was above the high screen (3) the original prevalence rate, if the original prevalence rate was between the low and high screens (including equal to one or

[14] While this method of combining the surveys uses a model, the model does not involve small areas either implicitly or explicitly.

the other screen). In addition, if a county's sample was above a specified size, the county retained its original estimate. This process was performed for all counties and then repeated ten times using the results from the prior iteration each time.

3. Pooling Direct Estimates over Time

For surveys that are conducted annually, pooling estimates over time is a very common technique for increasing effective sample size. Samples from consecutive years are not ordinarily independent. Some surveys, like the Current Population Survey (CPS) include panel components in order to reduce estimates of change over time. In the CPS, 75 percent of the addresses sampled in consecutive months and 50 percent of the addresses sampled in the same month one year apart are identical by design. Estimates that combine consecutive years of CPS data must take account of the substantial sample overlap. Even if a survey lacks a longitudinal component, the units sampled in consecutive years are often drawn from the same local areas—a strategy used to reduce the costs associated with drawing the sample. The effective sample size when two consecutive annual samples are combined is something less than the sum of the two samples, although the reduction is not nearly as substantial as it is with the CPS. The ACS, on the other hand, is designed for estimates to be combined over time. Because of its scale, the ACS can draw samples from all counties and other substate areas in every year. The Census Bureau releases estimates that combine sample data in three- and five-year rolling averages and releases three- and five-year public use samples with weights that combine the survey years. With all of these approaches it is important to recognize that a multi-year average obscures variation between years and that, depending on how the data are weighted, the difference between two multi-year estimates is determined largely if not entirely by the nonoverlapping years.

B. Small Area Estimates Based on Implicit Models

As examples of SAE with implicit linking models, Rao includes traditional demographic techniques used to estimate area population size as well as methods of indirect estimation exemplified by synthetic estimation and composite estimation.

1. Demographic Methods

Demographers have relied on a variety of methods to develop population estimates for small (and large) areas for the years following decennial censuses. These include the use of demographic accounting, where population change in an area is equal to the base (census) population plus births minus deaths plus net migration (in-migration minus out-migration). Vital statistics provide estimates of births and deaths at the national, state, and local levels, but in the U.S. there are no administrative data that provide equally complete coverage of migration. Enrollment records for Medicare, which covers over 90 percent of the population 65 and older, are an important source of information on where part of the population lives. Estimates of both gross and net migration have been developed from the addresses reported on tax returns, which cover a very substantial fraction of the population but systematically exclude those with low income. Demographers have developed procedures—including regression models—to estimate net migration from "symptomatic indicators," which include changes in housing stock and school enrollment, for example. These approaches differ from most of the other small area methods discussed in this report in that they generally do not involve sample survey data.

2. Synthetic Estimation

Synthetic estimation encompasses a class of estimators that use relationships measured reliably for a large area to estimate the characteristics of small areas within the large area. In Rao's terminology, synthetic estimation borrows strength through an implicit model linking the small area to the large area. In its simplest form, with no auxiliary variables, synthetic estimation uses the estimated mean or proportion for a large area to estimate the mean or proportion for a small area within that larger area. The implicit model is that the small area proportion approximately equals the overall (or larger area) proportion. In this case, the estimate of the small area proportion is calculated using the estimate of overall proportion.

If information on the composition of the area population is available, that information can be incorporated into the synthetic estimator. Using the unemployment rate as an example, reliable information on demographic composition may be available at the county level while estimates of the unemployment rate for demographic subgroups may be available at the state level. The state unemployment rates for these subgroups (defined, for example, by age, sex, race, or ethnicity) can be applied to the composition of the population (the auxiliary information) at the county level. The county (small area) estimate of the unemployment rate is thus the weighted sum of the subgroup unemployment estimates, where the weights reflect each county's demographic composition. The implicit model applied here is that specific relationships observed at the large area level—such as subgroup unemployment rates—are true for all small areas within the larger area.

By way of illustration, Gonzalez, Placek, and Scott (1993) showed how they applied synthetic estimation to estimate the percentage of jaundiced births, by state, in 1980. For each state they classified live hospital births in that year by 25 combinations of the mother's race, age, and live birth order. For each class they applied a national estimate of the percentage of live births that were jaundiced as reported in the National Maternal and Infant Health Survey. Using the 25 class counts in each state as weights, they calculated a weighted sum of the percentage of births that were jaundiced in each state, obtaining a state total, which they then expressed as a percentage of the total hospital births in the state in that year. This same methodology was applied to develop state estimates of a number of characteristics reported in the survey.

Synthetic estimation was used extensively in a number of federal agencies in the 1970s and 1980s, and according to one set of observers writing in 2004 was still the most commonly used SAE technique in the field of public health (Jia, Muennig, and Borawski 2004). Overall, however, its use has diminished with the development of better estimators—and the computational power to apply them.

3. Composite Estimation

As a general class, composite estimators, which combine two estimates, may be the most widely used in the federal government today. Typically, the composite estimator of a small area characteristic is a weighted average of two alternative estimates, where the weights are ωi and $(1 - \omega i)$ with ωi bounded by 0 and 1. The general idea is to select weights that minimize the error of the composite estimate, given the error associated with each of the components, but this can be approached in different ways. The simplest scheme assigns a uniform value to ωi for all areas, but this may not be particularly useful when the error associated with the two estimators varies differentially across areas such that one estimator is better is some areas and the other estimator is better in other areas.

The most popular form of composite estimator for SAE is one that combines a direct estimator and an indirect estimator. The James-Stein estimator can be expressed as a composite estimator of this type and is also known as a shrinkage estimator, as it shrinks the direct estimator toward the indirect estimator (Rao 2003). The direct estimator is assumed to be unbiased while the indirect estimator is biased but has smaller variance. The indirect estimator can come from many different methods, such as a synthetic estimator (Schaible 1978), an estimate from a larger area that includes the area of interest (Sommers 2005), or an estimator derived from an explicit model, such as a regression estimator (Jia, Muennig and Borawski 2004).

In estimating six types of health care usage and expenditures for the 30 largest states, Sommers (2005) applied a composite estimator that combined a direct estimate for each state and an indirect estimate consisting of the direct estimate for the census division in which the state was located. Thus the indirect estimate for a particular state borrowed strength from the other states within the same division. The division estimate was chosen as the indirect estimate because there is a correlation between the estimates for states within the same division. For the composite estimate for each state, the state and division estimates were weighted to minimize MSE. Despite being biased, the composite estimator reduced relative MSE compared to two unbiased estimators: direct state estimates and state estimates post-stratified to state population totals. The improvement made it possible to produce estimates with satisfactory precision for even the smallest of the 30 states, which was not true of either of the alternative estimators.

C. Small Area Estimation with Explicit Models

When composite (or shrinkage) estimators are used for applications in the federal government today, they typically rely on an explicit rather than implicit model to generate one of the component estimates. While implicit models use auxiliary variables to provide a link between the area being estimated and related areas, explicit models take account of variation between areas. For example, an implicit model in a synthetic estimate of the small area unemployment rate would incorporate differences in the unemployment rate among subpopulations whereas an explicit model would seek to predict the unemployment rate from a set of covariates that contribute to explaining inter-area differences in unemployment rates.

Rao (2003) lists four advantages of explicit models:

1. Model diagnostics can be applied to identify models that fit the data well
2. An area-specific measure of precision can be associated with each area estimate (in contrast to the global measures of precision derived from methods with implicit models)
3. A wide variety of types of models and data structures can be utilized
4. Recent methodological developments with respect to random effects models can be incorporated to improve the accuracy of small area inferences

These benefits of modeling are reflected in the applications discussed below.

1. Types of Models

Predictive models borrow strength by combining data over areas or over time (or both) to estimate the model parameters. For example, a state-level model uses data from all the states to make predictions for each individual state, based on the values of the covariates for that state. Predictive models also borrow strength by drawing their predictors from additional data sources.

Models can be at the area level or the unit level. An area-level model uses characteristics measured at the area level (from other data sources) to predict the area level characteristic that is being estimated. For example, a state or county-level model is considered an area-level model. A unit-level model uses characteristics of disaggregated units—say, persons—to predict the characteristic of interest at the unit level. The unit-level predictions can then be aggregated to the area level. Area-level models are much more common for applications of SAE than unit-level models. Nearly all of the applications discussed in later chapters involve area-level models or multilevel models, which represent fixed effects at different levels of area (for example, county and state) or at both the unit and area level. Mixed models, which combine fixed and random effects (the latter at the area level), are popular as well.

When linear models involve combinations of fixed and random effects, they can be estimated using best linear unbiased prediction (BLUP) estimators, which minimize MSE among a class of unbiased estimators. The BLUP estimator can be expressed as a weighted sum of a direct estimate and a regression synthetic estimate, which gives it the form of a composite estimator. BLUP estimation depends on variance parameters that are typically unknown, but empirical best linear unbiased prediction (EBLUP) provides a way to estimate a BLUP model. Fay and Herriot (1979) used an EBLUP estimator in their seminal work on estimating income for small areas, discussed below.

The EBLUP estimator can be used with linear mixed models, which encompass a wide range of SAE applications, but such models are inappropriate for binary data or count data (Rao 2003). Moreover, accurate estimates of MSE for EBLUP estimators generally require normality of the errors. Empirical Bayes and hierarchical Bayes methods can handle binary data and count data and do not depend on normality. With a normal linear mixed model, empirical Bayes and EBLUP are identical. Hierarchical Bayes is a fully Bayesian method whereas empirical Bayes is not. A Bayesian approach begins with a prior distribution of the parameters to be estimated. This prior distribution may be based on earlier research (an “informative” prior) or not (a “diffuse” or noninformative prior). The final, “posterior” distribution is derived from the prior distribution, given the data. Hierarchical Bayes methods offer greater flexibility than non-Bayesian methods, but their estimation can be more complex computationally because Bayesian estimators typically capture more sources of error, which contributes to their appeal. Different approaches have been developed to estimate such models. Markov Chain Monte Carlo (MCMC) methods are popular currently and are found in the most widely used software.

2. Non-Bayesian Examples

During the 1970s, federal funds were distributed to approximately 39,000 local governmental units under the General Revenue Sharing Program. Within states the funding allocation was based on population, taxes, and per capita income. The estimates of per capita income were based on decennial census data, updated each year to reflect growth in per capita income at higher levels of aggregation. More than a third of the 39,000 areas had populations smaller than 500 persons, which meant that their decennial census estimates of per capita income, which were derived from the long form's 20 percent sample, were very imprecise, so a decision was made to substitute county averages for these areas. Fay and Herriot (1979) developed an alternative, EBLUP estimator that could be used for all 39,000 areas and provide more precise estimates for the smallest areas. Producing the alternative estimates involved first fitting a regression equation to the census sample estimates, using as predictors the county averages, tax return data for 1969 (the income reference year for the census estimates), and census housing data. The goodness of fit between the regression predictions and the underlying true values was then estimated and used to calculate a weighted average of the regression and direct sample estimates for each area based on the average lack of fit of the regression estimates and the variance of the sample estimate. The weighted average was constrained to fall within one standard error of the sample estimate in each case. An evaluation of the results demonstrated the superiority of the model-based estimates to both the county averages and the direct sample estimates. Variations on the Fay-Herriot model have been used widely although often with empirical Bayes methods.

To estimate the county-level prevalence of severe work disability, Jia, Muenning, and Borawski (2004) implemented a multilevel mixed logistic regression model that included county as a random effect. That is, in addition to including specific characteristics of counties as fixed effects, the model included a term for each county. In this model the fixed effect covariates included both individual and area-level predictors: the BRFSS survey respondents' demographic characteristics (age, gender, race/ethnicity), and county-specific variables reduced through a factor analysis to four factors. The four factors were: (1) socioeconomic characteristics (unemployment rate, poverty rate, per capita income, median household income, and the proportion of adults with fewer than 12 years of education), mortality (age-adjusted rates by gender and race), demographic characteristics (proportion of persons living in rural areas and proportion of adults married), and health care service resources and utilization (hospitals, hospital beds, and hospital admission rates). Estimated prevalence rates by demographic group and county were calculated from the model and then aggregated to the county level.

Battaglia et al. (2011) developed a model-based estimator involving two datasets: (1) the survey data of interest and (2) an external larger survey data or census that was treated as a "virtual population." The approach included steps designed to ensure that the small area estimates generated by this approach would be consistent with direct estimates that the survey had sufficient sample to support. For this application, the BRFSS was the survey of interest, and a five-year ACS file was the virtual population. Within each state in the BRFSS a unit-level fixed-effect logistic regression model was developed for each of several health outcomes measured as binary variables. The predictors included person-level variables available in both the BRFSS and ACS and a set of county-level auxiliary variables from external sources. Using the estimated regression coefficients, a predicted probability of each health outcome was calculated for each person in the virtual population. These predicted probabilities were then raked iteratively (within each state) by BRFSS key domains so that the estimated prevalence in the virtual population matched that in the BRFSS by state and domain. The weighted mean of these adjusted predicted probabilities within each county became the estimated prevalence for the county.

Zhang et al. (2011) used a unit-level mixed model to estimate obesity prevalence in Mississippi counties from the state's BRFSS data for 2007-2009 and a set of auxiliary variables. First they estimated a model predicting the logit of obsesity status at the individual level from a combination of individual and county-level variables represented as both fixed effects and random effects. The fixed-effect regression coefficients (the random effects were not significant) were then used to predict the logit of obesity prevalence at the county level using external county-level auxiliary data drawn from 2000 census data and indicators compiled by the Economic Research Service of the U.S. Department of Agriculture (USDA). In the final step the predicted logits were converted to proportions.

When developing national estimates for small racial/ethnic groups, borrowing strength from other areas is not an option, and borrowing strength from other racial/ethnic groups may result in excessive bias, yielding estimates that understate the differences between the small groups of interest and the other, dominant groups. This can diminish the policy value of the estimates. Borrowing strength over time by pooling survey estimates across years can improve the precision of the estimates without the need to combine subgroups for which separate estimates are desired.

Srebotnjak, Mokdad, and Murray (2010) produced estimates of Type 2 diabetes prevalence for all U.S. counties by combining three approaches: (1) pooling data from multiple survey years, (2) incorporating spatial correlation via a spatial component, and (3) utilizing the relationship between the outcome variable and domain-specific covariates. The survey estimates were drawn from BRFSS. The authors compared 16 different models varying whether the following components were included: (1) a coefficient on the survey year (versus a trend), (2) a spatial covariate consisting of the county random intercept estimated from a simpler model and averaged over neighboring counties, (3) county-level covariates, and (4) an individual race variable. All of the models included an age group covariate, and separate models were estimated for males and females. To produce county-level predictions using the person-level covariates, the coefficients were applied to the demographic composition of each county. The models were evaluated against benchmark estimates, which were direct estimates for large counties. The benchmarks were compared to model-based estimates and direct estimates computed for simulated small areas, which were simple random samples (with replacement) of varying numbers of cases selected within each large area. The samples were replicated multiple times in order to permit calculation of the statistics needed for evaluation. The estimates were evaluated based on the correlation between the model-based estimates and the benchmark estimates and the average squared difference between the model estimates and the benchmark estimates. The comparisons were made for three different sample sizes. Even the simplest model showed marked improvement over the direct sample estimates, with the advantage increasing as sample size declined. For example, the correlation between the direct estimate and the benchmark for males 30 and older fell to zero with the smallest sample size while it remained above 0.7 for the best models. The most complex models performed the best. The results demonstrated that model-based estimation could improve the estimates as much as a several-fold increase in sample size.

Ganesh et al. (2009) applied a simple small area model accounting only for trends and state effects to produce county-level estimates of vaccination coverage rates from multiple years of an annual survey. Estimates were produced for every county and two-year period with a sample size of at least 35 for the two-year period and at least 15 for each year. The model, which used no auxiliary variables, included three components: (1) a mean for the first two-year period for counties in states in which no other counties met the sample size requirements, (2) a state effect for counties in states with multiple county estimates, and (3) a period effect for each two-year period. The EBLUP model also included random county and county-by time effects. The model-based estimates demonstrated greater stability over time than the direct estimates and had confidence intervals that were more than 40 percent narrower. The authors note that their approach offers an improvement over ad hoc techniques such as pooling estimates over multiple years.

When a small area model borrows strength over time, concepts from time series analysis can be used. A good example of time series modeling applied to SAE in the federal government is the estimation of state and large substate area employment and unemployment, discussed in Chapter III. In addition, Elliott et al. (2009), reviewed in Chapter IV, developed an approach for pooling data across years that combines a time series prediction of the current year value with a direct estimate of the current year value and used this method to estimate the prevalence of health outcomes among rare racial groups.

3. Empirical Bayes Examples

Chattopadhyay et al. (1999) apply simple empirical Bayes estimators (without auxiliary variables) to produce county-level estimates of the prevalence of alcohol dependence, using survey data collected by the Gallup Organization. In developing their empirical Bayes approach to this small area problem, the authors contrast the approach with synthetic estimation and a composite estimator combining the direct estimate with a synthetic estimate. Unlike the synthetic estimator the empirical Bayes estimator that they apply allows some variation in subpopulation prevalence rates among counties within a region. The results show a marked reduction in MSE relative to direct estimation.

In actual applications, the distinction between empirical Bayes and fully Bayesian models may blur. As an example, Cunnyngham, Castner, and Sukasih (2012), discussed in Chapter III, provide an exceptionally detailed description of their application of empirical Bayes methods to the estimation of state-level program participation rates and program eligibles. Of particular note is their discussion of the parameter estimation and derivation of the final participation rates and counts of eligible persons.

4. Hierarchical Bayes Examples

To estimate cancer risk factors and screening at the county level, researchers at the University of Michigan, the University of Pennsylvania, NCHS, the National Cancer Institute (NCI), the National Center for Chronic Disease Prevention and Health Promotion, and Information Management Services collaborated in the development of a hierarchical Bayes approach that combined data from two surveys: BRFSS and the NHIS. This approach is being applied by NCI currently to produce estimates of these same outcomes, as discussed in Chapter III.

While BFRSS includes adequate sample size for state-level and some county-level estimates, it has potential nonresponse bias due to low response rates and potential coverage bias due to the exclusion of households without telephones (and, until recently, households with only cell phone service). The NHIS has a higher response rate and covers households regardless of telephone service, but the sample was not designed to provide estimates at the state level much less the county level. To capitalize on the relative strengths of the two surveys, Raghunathan et al. (2007) developed a hierarchical Bayesian model with three stages or levels.[15] This model was applied to produce county estimates of six outcomes: two measures of smoking estimated separately for men and women and two measures of cancer screening (mammography and pap smear prevalence) for women alone. For each outcome at the county level the first stage modeled the sampling distribution of direct estimates of the outcome for NHIS telephone households, NHIS non-telephone households, and BRFSS households, conditional on three population parameters: the prevalence of the outcome for households with telephones, the prevalence of the outcome for households without telephones, and the bias of the BRFSS estimate for telephone households. This formulation assumes that the NHIS estimates are unbiased, and including the bias term provides a way to use the combined surveys to estimate the prevalence of the outcome for telephone households. The second stage modeled the between-county variation in the three population parameters using 26 county-level covariates (auxiliary variables) obtained from the 2000 census and other sources. Because the outcomes of interest are known to vary by socio-economic status, the county-level variables include measures of per capita income, percent below poverty, median home value, and the percent of the population graduating from college. Strength is borrowed across areas and other data sources through this second-stage model. The third stage is the assumed prior distribution for the unknown parameters of the second-stage model—that is, the regression coefficients and the 3x3 covariance matrix of the three population parameters.

Application of a Bayesian methodology involves generating a posterior distribution for each of the three unknown parameters at the county level and then calculating the mean and variance from each of these distributions. This is accomplished through a simulation (which is simpler than performing a complex integration), which involves making thousands of draws from the joint posterior distribution of all of the parameters, which number in the tens of thousands (three for every combination of county and year and three for every regression coefficient plus an additional six values for the covariance matrix).[16] In the end, the estimated prevalence of the outcome for each county is calculated as a weighted sum of the estimated prevalence among persons in telephone and non-telephone households, where the weights are derived from census 2000 estimates of telephone coverage.

Malec (1997) developed a hierarchical Bayes model to estimate poverty and housing unit characteristics at the tract level with data from the test phase of the ACS. The model allowed for uncertainty with respect to within-tract variability. Within-area variability is often treated as fixed in other small area methods. The author proposed a unit-level model to account for within-area variability. Both individual and housing unit characteristics were modeled, with individuals being modeled within housing units. Specifically, housing unit composition and the poverty

[15] The term "levels" has also been used in reference to the unit of analysis (individual versus area) and the geography (county, state, or nation) for which estimates are prepared and effects are estimated. Rather than create our own terminology we use the terms applied in the literature but try to clarify the interpretation in each case.

[16] Software is an issue for Bayesian small area estimates because of the volume of calculations. For this application the computations were programmed in Gauss and run on a desktop computer.

status of individuals within the housing unit were modeled. All members of a family were assumed to have the same poverty status, consistent with the family-based definition of poverty, but unrelated individuals within the same household could have different, although potentially related, poverty status. The model parameters were estimated using the MCMC algorithm, and estimates of tract-level poverty and selected characteristics were produced along with estimates of their accuracy.

5. New Developments

Pfeffermann (2013) reviews developments in SAE in the years since the publication of Rao's book. Calibration of the survey base weights to known totals of auxiliary variables can improve the precision of design-based estimates of small areas, and there are different approaches to calibration. Model-dependent and model-based direct estimates may also be useful when sufficient sample data for direct estimation are available. Developments in model-based SAE include new ways of estimating the MSE of the prediction model, computation of prediction intervals, new approaches to benchmarking, ways of accounting for measurement error in the auxiliary variables, identification of outlying estimates, methods to enhance the robustness of models, and new approaches to the problem of predicting ordered means. Pfeffermann also discusses approaches to SAE when confronted with two related problems: informative sampling (that is, selection related to the variables of interest) and nonresponse that cannot be characterized as missing at random. Finally, model selection and validation are perennial issues with model-based SAE. Pfeffermann reviews advances in those areas as well.

D. Comparative Evaluations

Perhaps because the development of small area estimates using any method but direct estimation is laborious, there are very few comparative evaluations presented in the literature, and the exceptions tend to involve comparisons of less powerful methods by today's standards. For example, in a review of SAE at NCHS in the early 1990s, Malec (1993) summarized five comparative evaluations involving synthetic estimation and one or two relatively simple alternatives. These studies were conducted between 1968 and 1977, when applications of synthetic estimation were growing.

Even much later, in a study of county-level disability prevalence, Jia, Muenning, and Borawski (2004) applied the synthetic method, spatial data smoothing, and regression and compared their estimates of severe work disability—derived from BRFSS data—to estimates of a similar concept from Census 2000. The comparisons utilized correlations, MSE, mean absolute difference, and a rank statistic. The synthetic method produced substantial improvement over direct estimates, increasing the correlation with the Census 2000 estimates from .26 to .55 and reducing all of the other measures of error substantially. Spatial smoothing improved upon synthetic estimation by some but not all measures, capturing more local variation but producing greater overall error. Of the three indirect methods, the regression method, which used a multilevel logistic regression model described earlier, was clearly the most effective, raising the correlation to .80 and producing smaller average error by every measure.

Goodman (2010) also compared direct estimation, synthetic estimation, spatial data smoothing, and regression as methods for producing county-level prevalence estimates of asthma, diabetes, and hypertension by race. The estimates generated by these methods were evaluated against publicly-available data collected at the local level for a subset of U.S. counties.

The data source for the direct estimates was BRFSS, and the comparison was limited to counties that met minimum sample size criteria for racial subgroups. Of the four methods, regression performed best overall, although the differences among the methods varied across the three health conditions. Regression worked particularly well for asthma, producing the highest correlations (in the range of .77 to .83) with the gold standard data and the smallest error. The synthetic method was nearly as good whereas the direct estimates and spatially smoothed estimates produced correlations near zero and markedly larger error. Regression was clearly superior for diabetes as well, but the correlations were much lower (.13 to .23). Synthetic estimation was not nearly as sharply differentiated from spatial smoothing and direct estimation as for asthma. For hypertension the four alternatives were even less sharply differentiated. Regression was best on only four of the six measures rather than all six, for example, and one of the two correlations was near zero.

For estimates of poverty rates at the state level, Schirm (1994) compared the direct sample estimator with three alternative indirect estimators: (1) a pooled sample estimator that combines the nonoverlapping households in three March CPS samples, (2) a regression estimator obtained by regressing direct estimates of state poverty rates on as many as five symptomatic indicators, and (3) a shrinkage estimator that combines the direct sample and regression estimates in an optimal way for each state. The four estimators were compared by drawing multiple samples from the March 1990 CPS sample, which was treated as the population. The estimators were compared with respect to their RMSE, their accuracy in identifying the 10 states with the highest poverty rates, and the coverage of their 95 percent confidence intervals (that is, do the confidence intervals include the true state poverty rate 95 percent of the time?). Compared to the direct estimator, the shrinkage estimator reduced the RMSE 97 percent of the time. Compared to the pooled sample estimator the shrinkage estimator reduced the RMSE 90 percent of the time, and compared to the regression estimator the shrinkage estimator reduced the RMSE 100 percent of the time. The shrinkage estimator was also much more likely to correctly identify 9 of the 10 states with the highest poverty rates, followed by the regression estimator, the pooled sample estimator, and the direct estimator. Finally, both the shrinkage estimator and the direct estimator included the true poverty rate in their 95 percent confidence intervals over 93 percent of the time. This compared to 85 percent for the pooled sample estimator and just over 50 percent for the regression estimator.

In investigating indirect estimation techniques to produce county-level estimates of proportions from the NHIS, Malec et al. (1997) compared estimates from hierarchical Bayes and empirical Bayes procedures to synthetic and direct design-based estimates. The variable estimated was the proportion of persons with a doctor visit in the past 12 months. Estimates were prepared at the county-level for 72 combinations of age, race, and sex. Direct estimates were not even possible for most of the 3,000 counties, as they had no sample observations in the NCHS, and many of the counties that did have sample observations could not support estimates for all 72 demographic groups. Consistent with other results, synthetic estimates underestimated variability across domains—to the point of implausibility. Empirical Bayes estimates performed well when domain sample sizes were moderately large, which generally meant the state level, but did not fare well for subpopulations within states, much less counties. Hierarchical Bayes methods, on the other hand, yielded satisfactory estimates of parameters and measures of precision even for small subpopulations.

Baskin and Sommers (2008) explored the application of composite estimation, non-Bayesian mixed modeling, and Bayesian modeling techniques to the estimation of means at the

Metropolitan Statistical Area (MSA) level using data from the Insurance Component (IC) of the Medical Expenditure Panel Survey (MEPS). The IC is a survey of employers, and direct estimates are produced for the nation, the states, and the 20 largest MSAs. This research applied SAE methods to develop estimates of mean insurance premiums and employee contributions for smaller MSAs within 10 states. Direct estimates for these areas showed fluctuation across years, which was thought to be inaccurate because it was contrary to trends in the state and nation. The composite estimate combined the direct estimate for the small area and an estimate for the rest of the state but did not borrow strength over time whereas the estimates based on explicit modeling borrowed strength across time as well as across domains. The evaluation compared the alternative estimators with respect to the average Relative Mean Squared Error (RMSE) of their estimates. All of the estimators using small area methods had smaller average RMSE than the direct estimates. The composite estimator had the lowest RMSE in all six years (2000-2005) but did not account for variation over time whereas the mixed model and Bayesian model incorporated a linear time trend, which preliminary analysis showed was the most effective of the time models considered. The Bayesian estimates had slightly smaller error than the mixed model estimates and good theoretical properties but are more difficult to produce, requiring different software than all of the other approaches.

This page has been left blank for double-sided copying.

III. APPLICATIONS IN PRODUCTION BY FEDERAL AGENCIES

More than 20 years ago the Subcommittee on Small Area Estimation of the FCSM surveyed the federal government for examples of indirect estimation and prepared summaries of all eight programs that used indirect estimators to produce estimates for publication (FCSM 1993). Beginning with the earliest, these eight programs included:

- County estimates of crop acreage, production, and yields and livestock inventories; National Agricultural Statistics Service (NASS)
- State, metropolitan area, and county personal income estimation; Bureau of Economic Analysis (BEA)
- Post-censal population estimates for states, counties and places; U.S. Census Bureau
- State and local area estimates of employment and unemployment; Bureau of Labor Statistics (BLS)
- County estimation of crop acreage using satellite data; NASS
- Synthetic estimation in surveys of individuals sampled from vital records; NCHS
- Estimation of median income for four-person families by state; U.S. Census Bureau
- Exploratory model-based state estimates from the NHIS; NCHS

County agricultural estimates date back to the 1920s while the first state estimates of personal income were produced in 1939. Post-censal population estimates for states started in the 1940s, and state and federal agencies began to produce state and local employment and unemployment estimates in that decade as well. All four of these programs have evolved substantially in the decades since they were initiated, but only the BLS state estimates currently use methods that would be described as SAE today. Among the more recent programs, synthetic estimation at NCHS has been largely supplanted by explicit model-based methods, and the Census Bureau's estimates of median income have been replaced in part by an expanded small area income estimates program and in part by the ACS, which can provide reliable direct estimates.

In this chapter we review contemporary production applications of small area methods within the Department of Health and Human Services (HHS) and, on a selected basis, elsewhere in the federal government. First, however, we discuss key issues that arise in developing an SAE program of regular estimates within the federal government.

A. Issues in Developing an SAE Program in the Federal Government

Both the Substance Abuse and Mental Health Services Administration (SAMHSA) and BLS addressed growing needs for small area estimates by developing SAE programs. Issues that they faced in producing small area estimates of substance abuse and employment were discussed in a conference sponsored by FCSM and are summarized below. Kalton (2004) reviewed their assessments and provided more general observations focusing on the dependence of models on auxiliary variables.

1. Estimates of Substance Abuse

For SAMHSA, implementation of a regular program of state estimates of about 20 measures for three age groups (12-17, 18-25, and 26 and older) required a redesign of a survey sample to provide a minimum of 400 observations per state (Wright 2004). Prior research had demonstrated the feasibility of applying small area techniques to produce the needed estimates, but the research also indicated that the production of annual estimates would require a minimum sample size of about 400 persons per state to meet the goals of the estimates.

A notable aspect of the presentation of results involved ranking the states with respect to each of the measures. In doing so, SAMHSA placed more emphasis on comparing states to the national average than to each other, as the estimates for many states were not statistically different. The estimates were accompanied by a technical appendix that provided guidance on interpreting the results, explained potential sources of bias, described the model validation efforts, and reported prediction intervals, which were notably shorter than design-based confidence intervals. In addition, simultaneous with the data release, every governor's office was sent the results to allow time for preparing for reporters' questions.

Estimates of year-to-year change are of great interest to the states, but when the second year of estimates became available, it was quickly evident that hardly any state experienced a significant change. A review of the findings suggested that the estimates lacked sufficient precision to capture meaningful change, which tended to develop slowly. SAMHSA considered options for dealing with this problem, including ways to better incorporate change into the hierarchical Bayes models, which are described in the next section. Suitable predictors could not be identified, however. Such predictors have to be at the county or lower level to match the way the estimates are prepared. To address these shortcomings, SAMHSA opted to produce two-year rather than single-year estimates and has been happy with the increased precision. Ultimately, the value of the estimates will depend on the extent to which they are useful in documenting successful effects of prevention, treatment, and other efforts directed at reducing substance abuse.

2. Estimates of Employment

The Current Employment Statistics (CES) survey is a monthly survey of establishments conducted for BLS via a federal-state cooperative program. While the sample design and estimation approaches for the CES were focused on state and national estimates, users were interested in smaller domains, such as detailed industries for individual states or even metropolitan areas. Consequently, BLS invested in the development of model-based small domain estimation methods.

Eltinge (2004) discusses the practical constraints on the development, implementation, and perceived value of small domain programs. For some small domain programs, Eltinge suggests that the presence of multiple stakeholders and multiple constraints can affect the development and implementation of these programs. If SAE methods are viewed as a form of technology, the literature on the adoption and diffusion of technology can provide insights that may be of value in the development and implementation of such programs.

Eltinge contrasts the Small Area Income and Poverty Estimates (SAIPE) program, which the Census Bureau developed to address specific funding allocation needs of the Department of

Education, with BLS's CES program, which does not have a dominant user or primary use. The CES's many stakeholders had diverse views on the ways in which the CES estimates could best be used. The CES program is also subject to multiple constraints, including quick publication deadlines following data collection, less timely availability of potential covariates, and issues of compatibility with legacy production systems. While the mathematical statistics literature focuses on the optimization of an objective function, the multiple stakeholders and multiple constraints imply multiple objective functions.

If these multiple objective functions are viewed as an opportunity rather than an obstacle, they can provide a rich set of problems for mathematical statisticians to address and thus enhance their potential contribution to small area methods. Efficient management of the scarce resources that mathematical statisticians represent implies focusing these resources on a moderate number of high priority topics—those most likely to lead to substantial improvements in the agency's statistical products.

Eltinge suggests that some of the insights that might be gained from the adoption/diffusion literature include focusing efforts initially on stakeholder requests that are most likely to produce high levels of reward relative to risk and giving balanced consideration to quantitative and qualitative aspects of the development of the new estimates, where the qualitative aspects include, for example, trade-offs between non-publication and release of estimates that lead to erroneous inference.

3. Model Dependence on Auxiliary Variables

In his discussion of Wright and Eltinge, Kalton (2004) notes the dependence of small area estimates on models, which have grown in sophistication and power, but observes that "model estimates can be no better than the auxiliary data on which they are based." Auxiliary variables need to be strong predictors of the characteristic to be estimated, but they also need to be measured uniformly across all small areas or able to be adjusted to compensate for such differences as may exist. Frequently the raw auxiliary data must be converted into suitable indices. To illustrate these points, Kalton cites the Census Bureau's experience with an indicator of food stamp receipt used to predict the number of poor school-age children in states and counties. Initially, the Census Bureau tried the number of recipients in July but found that the average monthly number of recipients centered on January 1 was a better predictor. In addition, reducing counts that were inflated by specific natural disasters improved the performance of the predictors.

These issues underscore the importance of having an in-depth understanding of the data on which the small area models depend, as idiosyncratic features—some of them specifi to the time period being modeled—can have pronounced effects on model estimates. Even with such knowledge, modelers need to conduct a thorough evaluation of both the model and the small area estimates that the model produces. When multiple estimates are needed, such as the 20 measures cited by Wright, or estimates of both levels and change, the producers must give attention to all of these needs in developing their models.

Picking up on Eltinge's discussion of the diffusion of technology, Kalton observes that applications of SAE are likely to grow but cautions that those who seek to apply these methods need to be aware of their resource requirements—particularly staff expertise—and they need to understand and communicate how the model-based estimates differ from direct estimates.

B. Small Area Estimates Currently in Production

Several applications of small area methods to the production of state and substate estimates by federal agencies are discussed below, beginning with a series of applications within HHS.

1. Drugs, Alcohol, and Tobacco Use

SAMHSA sponsors the National Survey on Drug Use and Health (NSDUH), an annual survey that serves as the principal source of data on the use of illicit drugs, alcohol, and tobacco among persons 12 and older in the civilian, noninstitutional population of the U.S. In 1999 the NSDUH sample was expanded to support state-level estimates, and since that time the Center for Behavioral Health Statistics and Quality (CBHSQ) has produced such estimates. Despite the sample size increase, the samples for most states are too small to produce annual estimates with adequate precision. Working with a contractor, the CBHSQ combines adjacent survey years and applies small area estimation methods to produce state estimates of 25 measures currently (see Hughes et al. 2011). The estimates are derived from a survey-weighted hierarchical Bayes (SWHB) methodology described by Wright (2003) and Folsom, Shah, and Vaish (1999). Like the methods employed for the Census Bureau's SAIPE program, discussed below, the SWHB methodology combines a direct sample estimate and a regression model-based estimate using weights that reflect that relative precision of the survey and model-based estimates for each state. The predictors are drawn from eight separate data series including, for example, Uniform Crime Report arrest totals and BEA per capita income. In 2002 a benchmarking procedure was introduced so that the national sum of the state estimates matches the national design-based survey estimates. Recently, SAMHSA has begun to use two-year moving averages to produce the state estimates. Estimates from the 2011 and 2012 surveys can be retrieved from the following website, which also contains links to extensive documentation as well as estimates from earlier years: http://www.samhsa.gov/data/NSDUH/2k12State/NSDUHsae2012/Index.aspx.

2. Prevalence of Disease and Risk Factors

BRFSS is administered by the Centers for Disease Control and Prevention (CDC) to provide state-level estimates of various risk factors and health behaviors, preventive health practices, and health care access primarily related to chronic disease and injury. There is a growing interest in obtaining estimates of such indicators for sub-state levels—in particular, the county (Jia, Muennig, and Borawski 2004). The BRFSS sample size, however, does not support estimation for more than a handful of areas smaller than the state, and findings from the national or state level often may not generalize to the county level. SAE techniques provide a means to extend survey estimates down to substate levels.

CDC has developed an SAE methodology for producing estimates of the prevalence of self-reported diabetes, the incidence of diagnosed diabetes, and the prevalence of selected risk factors for states and all U.S. counties and has published estimates for the years 2004 through 2009.[17] All of these estimates use data from BRFSS, a telephone survey of noninstitutionalized adults that is designed to produce state estimates of health status and factors affecting health. The estimates are based on an expansion of a Bayesian model proposed by Malec et al. (1997) that is applied to BRFSS data and demographic data from the Census Bureau's population estimates program (Cadwell et al. 2010). Multi-level logistic regression models with random effects for age, race, and sex at the county and state levels were fit using the MCMC method. To evaluate the model-based estimates during development, the final estimates were compared with direct survey estimates for large counties. Diabetes incidence was estimated using a model that conditions predicted incidence on predicted prevalence, using a variation on this methodology (Barker et al. 2013). Because incidence is so much lower than prevalence, the estimates of incidence were evaluated by comparisons with direct estimates at the state level, as direct estimates for even the largest counties are too imprecise. This work has been extended to the estimation of selected risk factors—obesity and leisure-time physical inactivity—at the county level.

In a collaborative project, NCI partnered with the University of Michigan, the University of Pennsylvania, NCHS, the National Center for Chronic Disease Prevention and Health Promotion (also in CDC), and Information Management Services to produce estimates of the prevalence rates of cancer risk factors and screening behaviors at the state, health service area, and county levels. The methodology, described by Raghunathan et al. (2007), was discussed as an example of a hierarchical Bayesian approach, in Chapter II. Estimates for states, counties, and health service areas (combinations of counties) are reported on an NCI website along with documentation of the methodology: http://www.sae.cancer.gov . Currently, estimates for just 1997-1999 and 2000-2003 are available, but estimates for the periods 2004-2007 and 2008-2010 are in production.

3. Household Telephone Use

The NHIS is unique among national household surveys in its collection of data on the type of telephone service—landline, wireless, or both—used by households in its sample. Because NHIS interviews are conducted in person, the survey is also able to identify households with no telephone service. The ACS also identifies households without telephone service. Such households are rare nationally, although the frequency differs among states and by other geographic as well as personal characteristics. Households with only wireless phone service (those that have substituted wireless for landline service) have grown dramatically, however, to the point where telephone surveys that sample solely from landline phones exclude a substantial and still growing proportion of the population with a decidedly different demographic than households with landlines (including those with both wireless and landline service).

In the past four years NCHS has collaborated with the University of Minnesota and NORC at the University of Chicago to publish three sets of state-level estimates of telephone usage by

[17] Prevalence refers to the number of cases or proportion of the population with a specific characteristic, and incidence refers to the number or proportion acquiring or diagnosed with this characteristic over a specified period of time (often a year).

type. The estimates differ in the length of the reference period over which the estimates were calculated, the additional geography provided, the level of detail on household characteristics, and the type of SAE methodology used to produce the estimates. The most recent estimates cover the two years 2011-2012 for 93 nonoverlapping areas consisting of individual states and county groups (Blumberg et al. 2013). Collectively the 93 areas cover the entire U.S.

Estimates were prepared for the proportion of people living in households defined as: (1) wireless-only, (2) wireless-mostly, (3) dual-use, (4) landline-mostly, and (5) landline-only. The proportion living in households with no telephone service was not modeled as this proportion can be estimated precisely from the ACS. Final estimates of the modeled proportions were adjusted to agree with the ACS estimate of households with no telephone service. Separate estimates were developed for adults (ages 18 and over) and children (under 18) for each of 12 six-month periods: January-June and July-December for each year from 2007 through 2012.

The estimates were derived from an EBLUP estimator as the weighted sum of three estimates for each area for a six-month period: (1) a direct estimate from the NHIS, (2) a synthetic estimate based on a regression of the NHIS estimate on survey data from the ACS and auxiliary information on listed telephone lines per capita, and (3) an adjusted direct estimate from the NHIS for all 10 six-month periods. With the third component, the estimator borrows strength over time as well as across geographic areas. Models for the 12 six-month periods were estimated jointly. A more complete description of the estimation methodology is provided in Blumberg et al. (2011). The most recent published estimates add two 12-month periods to those reported previously: July 2011-June 2012 and January-December 2012.

It is noteworthy that estimates were not reported for three states and three substate areas. In these six areas the direct estimates and synthetic estimates differed by more than a factor of two. The authors interpreted this as evidence that the direct estimates may be biased, violating one of the key assumptions of model-based estimation. They did not speculate as to the source of the bias.

The principal users of these estimates are survey organizations—including federal agencies—that require such information to improve the weighting of their samples. They may also use the information to enhance their telephone sample designs.

4. Small Area Income and Poverty Estimates

The Census Bureau's SAIPE program applies SAE methods to estimate the number of persons in poverty, by age group, at the state, county, and school district levels and to estimate median household income at the state and county levels. Unlike the small area per capita income statistics produced by BEA, which are derived by dividing an area estimate of aggregate personal income (the sum of all the income received by residents of a geographic area) by a population estimate for that area, estimates of median income and persons in poverty require household or family-level data. Under a Congressional mandate, the U.S. Department of Education commissioned a panel of the National Academy of Sciences (NAS) Committee on National Statistics to review the Bureau's data and methods, consider alternative data sources and approaches, and make recommendations (see, for example, National Research Council 2000). Building on the earlier work of Fay and Herriot (1979) at the Census Bureau, the SAIPE program applies an empirical Bayes methodology that combines direct survey estimates and predictions of these direct estimates from a regression model using predictors drawn from administrative records, BEA estimates of personal income, and income and poverty statistics from the 2000 decennial census (the last to collect such information). Detailed information on the SAIPE methodology along with the annual estimates can be found at: http://www.census.gov/did/www/saipe/index.html.

With the advent of the ACS, which collects data from about two million households per year and replaced the decennial census long form in 2010, estimates of income and poverty for geographic areas with at least 65,000 in population are available annually, but estimates for the many counties below this size require three- or five-year averages, and the ACS does not produce estimates for school districts. For counties below 65,000 in population, the model-based SAIPE estimates are considered to be more reflective of current conditions than the multi-year ACS estimates. Beginning with the estimates for 2005, when the ACS reached full scale, the ACS replaced the CPS Annual Social and Economic Supplement (ASEC) as the source of direct survey estimates in the SAIPE methodology.

At the state level the SAIPE program produces annual estimates of:

- Total number of people in poverty
- Number of children under 5 in poverty
- Number of related children 5-17 in poverty
- Number of children under 18 in poverty
- Median household income

"Related" children are children related to the household head. Poverty estimates for related children are required by the Department of Education.[18]

Separate models of poverty ratios are estimated for four age groups: 0 to 4, 5-17, 18-64, and 65 and older. For the age group 5-17, two equations are estimated: one for related children 5-17 and one for all children 5-17. Estimates of the latter are needed to construct the estimates of the total number of people in poverty and the number of children under 18 in poverty. The five models use single-year direct estimates from the ACS as the dependent variables. The predictor variables, calculated at the state level, are:

- The tax return poverty rate for the age group, where the numerator is derived from the number of exemptions reported on returns with adjusted gross income below the official poverty threshold for a family of the size implied by the total number of exemptions claimed, and the denominator is a Census Bureau population estimate for the age group used in the numerator, except for ages 5-17, where the denominator is total child exemptions[19]

[18] This description of the state and county estimation methodology is based on "2010-2012 County-Level Estimation Details," presented at: www.census.gov/did/www/saipe/methods/statecounty/20102012.html, which was accessed on February 18, 2014.

[19] Age is not reported on tax returns, but there are separate exemptions for "children" and persons 65 and older. The exemptions for children are used for children 5-17, and the exemptions for persons 65 and older are used for persons in that age group. Total exemptions for persons under 65 are used for children under 5 and persons 18-64.

- The nonfiler rate for the population under 65 (used for the under 65 age groups) or 65 and older, calculated as the difference between the estimated population in that age group and the number of exemptions claimed for persons in that age group, expressed as a percentage of that population
- The Supplemental Nutrition Assistance Program (SNAP, formerly the food stamp program) participation rate, calculated as the average monthly number of SNAP participants of all ages during the 12 months ending in June of the model year expressed as a proportion of the total population (used in the under 65 models)
- The Supplemental Security Income (SSI) recipiency rate, calculated as the average monthly number of SSI recipients in the prior year divided by the state population in the model year (used in the 65 and older model)
- The Census 2000 poverty ratio for the population 65 and older (for the 65 and older model) or (for the younger age group models) the residuals from a regression of the Census 2000 poverty ratios for the other age groups on 1999 versions of the other predictors for those models

The Census 2000 poverty ratios were based on data collected on the long form, which was replaced by the ACS. It is likely that future models will continue to use the Census 2000 poverty ratios until they lose their effectiveness as predictors.

An empirical Bayes methodology is used to combine the regression predictions and the direct estimates to produce state estimates of the five poverty ratios. Each state estimate is a weighted sum of the direct estimate and model-based prediction, where the weights sum to 1.0. The poverty ratios derived in this manner are then multiplied by Census Bureau population estimates (not ACS estimates, although very close) for the corresponding age groups. The resulting state-level estimates of persons in poverty are adjusted so that they sum to direct national estimates from the ACS for the estimation year. The separate estimates for children under 5 and 5-17 are combined to produce the estimate of children under 18 in poverty, and the separate estimates for all four age groups are summed to produce the final estimates of the total number of people in poverty.

The estimation of median income is done somewhat differently. The dependent variable is a single-year direct state estimate of median household income. There are two predictors: (1) the prior year state median adjusted gross income (AGI), obtained from tax returns filed in the ACS survey year, and (2) residuals from a regression of the census 2000 state median household income on the 1999 state median AGI, obtained from returns filed in 2000. The regression predictions and the single-year direct ACS estimates are then combined using the same empirical Bayes methodology used to estimate the five poverty ratios.

At the county level the SAIPE program produces annual estimates of:

- Total number of people in poverty
- Number of related children 5-17 in poverty
- Number of children under 18 in poverty
- Median household income

The county-level estimates of the first three are designed to sum to the SAIPE state-level estimates.

Unlike the state models, the county-level models predict the number of people in poverty rather than a poverty rate or ratio. Separate models are estimated for the total number of people in poverty, the number of related children age 5-17 in families in poverty, and the number of people under 18 in poverty. The dependent variable in each case is the log of the ACS direct estimate. The predictors are logs of number as well, which makes each model multiplicative (expressing the log of a number as the sum of the logs of several other numbers is equivalent to expressing the number as a function of the product of several other numbers). Predictors are similar to those used to predict the state-level ratios, except that they represent logs of numbers. For example, the model of the total number of people in poverty includes as predictors the logs of the number of poor [?] tax return exemptions, the number of SNAP benefit recipients in July of the previous year, the total resident population as of July 1, the number of tax return exemptions, and the Census 2000 estimate of the total number of people in poverty. Median income is predicted with a model that includes logs of the Census 2000 estimate of county median household income, the median adjusted income from tax returns, the proportion of the BEA estimate of total personal income derived from government outlays, the growth of BEA total personal income from 1999 through the target year, and the nonfiler rate, calculated from a combination of tax return and ACS data. The empirical Bayes estimates of county numbers of poor by age are raked to the SAIPE state estimates.

At the school district level the SAIPE program produces annual estimates of:

- The total population
- The number of children ages 5-17
- The number of related children 5-17 in poverty

Under the direction of the No Child Left Behind Act of 2001, the estimates of children in poverty are used by the U.S. Department of Education to allocate federal funds to school districts. Both of the other estimates have a role in funding allocation as well. [20]

[20] This description is based on "2010-2012 Overview of School District Estimates," which was obtained from the website: www.census.gov/did/www/saipe/methods/schools/data/20102012.html, which was access on February 18, 2014..

Unlike the state and county estimates, the school district estimates are not produced with empirical Bayes techniques, but in the end they are adjusted to be consistent with the county-level estimates. In this sense the school district estimates can be viewed as a disaggregation of the county estimates. Using data from federal tax returns geocoded to school districts and ACS county-level poverty rates for related children 5-17, the Census Bureau constructs a child poverty rate for each district.[21] In computing this poverty rate the tax return data must be adjusted using ACS data to reflect the target 5-17 population, as the return data do not contain children's ages.[22] The district-level child poverty rate is multiplied by an ACS estimate of related children 5-17, and the resulting counts of children in poverty are raked to the SAIPE county-level estimates of related children 5-17 in poverty. For school districts that cross county lines, a separate piece is estimated for each county. After raking, the estimates of portions from separate counties are summed to produce the final school district estimates. The school district estimates of total population and children 5-17 are not raked to higher level totals, as there are no SAIPE counterparts at the state and national levels.

5. Small Area Health Insurance Estimates

The Census Bureau applied the SAIPE approach to the development of the Small Area Health Insurance Estimates (SAHIE) program, which produces estimates of the number and proportion of nonelderly persons with and without health insurance coverage, by demographic group and relative income, for states and counties (Bauder, Luery, and Szelepka 2012).[23] Such estimates present a challenge for modeling because the best covariates for predicting insurance coverage at the state and local levels are informative only about public coverage (specifically, Medicaid and CHIP enrollment). The SAHIE approach, which has evolved substantially since the earliest estimates, uses direct estimates from the ACS along with auxiliary variables from the decennial census; tax returns filed with the IRS; enrollment data from SNAP, Medicaid, and CHIP; and demographic variables from the Census Bureau's population estimates program. A noteworthy feature of the SAHIE approach is that some of the auxiliary variables are treated as possibly biased measures of the variables of interest and are themselves modeled, following an approach developed by Fisher and Gee (2004) for the SAIPE estimates and extended to health insurance estimates by Fisher, O'Hara, and Riesz (2006). Because the estimates of insurance coverage are produced by poverty level, the modeling approach generates estimates of the number of persons in each income category in addition to the number with and without insurance coverage. These estimates from the insurance and income portions of the overall model are combined to generate the final estimates of coverage by income class.

The SAHIE estimates for 2010 were released in August 2012. The estimates along with extensive documentation of the methodology can be obtained from the SAHIE web page: http://www.census.gov/did/www/sahie/index.html.

---

[21] The estimates for the year 2010 use tax return data from returns filed in 2010 (for the 2009 tax year) and ACS five-year estimates for the years 2006-2010.

[22] While the funding allocation is for public schools, the estimates of school-age children are intended to include all children who reside within each district's boundaries.

[23] The income categories used currently reflect the needs of CDC, CHIP, and the Affordable Care Act.

6. State Estimates of Eligibility for Nutrition Programs

For nearly two decades, small area methods have been used to develop state estimates of participation rates for persons eligible for SNAP and children eligible to receive benefits from the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC). These applications, which are performed under contract to the Food and Nutrition Service (FNS) by Mathematica Policy Research, are described below.

a. SNAP Participation Rates and Eligible Persons

State estimates of participation rates and persons eligible for SNAP are produced using an empirical Bayes shrinkage estimator that combines direct estimates with regression predictions (Cunnyngham, Castner, and Sukasih 2013). Four variables are estimated: (1) the SNAP participation rate among all eligible persons; (2) the number of eligible persons; (3) the SNAP participation rate among the working poor, defined as eligible persons in households in which a member received income from employment; and (4) the number of eligible working poor. Direct estimates of the two participation rates, which serve as dependent variables in the participation rate regressions, are constructed from a combination of SNAP data on participants, estimates of eligible persons derived from a microsimulation model that applies SNAP eligibility rules to sample households in the CPS ASEC, and Census Bureau estimates of the resident population. The predictors in the regression models, which are estimated at the state level, include the percentage of the state population receiving SNAP benefits, the percentage of school-age children certified to receive a free lunch, a child poverty rate calculated from tax data (obtained from the Census Bureau's SAIPE program), and four measures produced from a three-year roll-up of ACS data and published on the Census Bureau website. The four measures are median family income, the percentage of the population 25 and older with a bachelor's degree, the percentage of foreign-born individuals who arrived in 2000 or later, and the percentage of children under 18 with income below 50 percent of the poverty level. The predictors are a subset of a much larger set of potential predictors evaluated during model development.

In order to borrow strength over time as well as across states, the estimates are produced for three years at once. Separate equations for participation rates among all eligible persons and among the working poor are estimated for three consecutive years. The six equations are estimated jointly, which means that prediction errors are allowed to be correlated and the between-group and inter-year correlations among the direct sample estimates of the six dependent variables are taken into account. Because of this, the shrinkage estimates for either population in any one year are determined by the regression predictions and direct sample estimates for both groups in all three years. As a result, the shrinkage estimate for a given state and year is not constrained to fall between the regression prediction and direct sample estimate for that population and year—although it generally does.

Preliminary shrinkage estimates of participation rates generated by the empirical Bayes estimator are adjusted so that the counts of eligible persons implied by the combination of participation rates and counts of participants sum to the direct estimates from the simulation model. In addition, where the estimated number of eligible persons falls short of the number of participants, the number of eligibles is increased so as to reduce the estimated participation rate to 100 percent. The estimates of eligible persons in the remaining states are reduced proportionately to maintain consistency with the national estimates in each year. The adjusted estimates of eligible persons are then used to recalculate the participation rates. This process is repeated for the working poor. The end result is a set of state estimates of the four variables for each of the three years.

The shrinkage estimates of state participation rates can be found on the FNS website at: http://www.fns.usda.gov/ops/supplemental-nutrition-assistance-program-snap-research.

b. Estimates of WIC-Eligible Children for Funding Allocation

Also for FNS, Mathematica has estimated the state-level number and percentage of infants and children 1-4 in families with incomes at or below 185 percent of the federal poverty level and therefore income-eligible for WIC, which FNS administers. Program regulations specify that such estimates be used to allocate federal WIC funds to the states (Cunnyngham 2012). The estimates are derived using an empirical Bayes shrinkage estimator similar to the one used to estimate SNAP participation rates and eligible persons. Direct sample estimates from the CPS ASEC are combined with model-based estimates of eligible percentages obtained from a state-level regression of the direct estimates on predictors drawn from tax data and the ACS. The estimates for 2010 were the first to use regression predictors drawn from the ACS. The previous estimates used predictors from Census 2000.

Production of the estimates for 2010 involved five steps:

1. State-level direct estimates of the percentage of children ages 0 to 4 in families at or below 185 percent of poverty (using the HHS poverty guidelines rather than the poverty thresholds used by SAIPE) were produced from the 2008 through 2011 CPS ASEC, reflecting annual income in 2007 through 2010
2. State-level regression models predicting the direct estimates in each of the four years were estimated jointly
3. The direct estimates and regression predictions were averaged using empirical Bayes shrinkage methods to obtain preliminary state estimates for each of the four years
4. These preliminary percentages were multiplied by Census Bureau demographic estimates of the state populations of infants and children 1 to 4 in each year to obtain preliminary estimates of children in the two age groups eligible for WIC
5. The preliminary estimates of eligible children in each year were ratio-adjusted to the national CPS ASEC estimates of eligible children 0 to 4 to obtain the final estimates

Infants and children 1 to 4 were combined in the CPS direct estimates and later separated using demographic estimates because the way that the CPS ASEC is currently weighted underestimates infants and overestimates children 1 to 4.

c. Estimates from Microsimulation Modeling of SNAP

In policy analysis, the outcomes of interest may change from one day to the next. While estimates of impacts at the state level may be desired, it is not feasible to develop small area models for every outcome—or even selected outcomes. To address this problem, Schirm and Zaslavsky (1998) developed a methodology for estimating a vector of state weights that could be assigned to every individual or household in a policy microsimulation model (see also Schirm, Zaslavsky, and Czajka 1999). Each sample person receives a nonzero weight for every state, where the state weight is the product of the individual's sample weight and an estimated probability of residing in that state, given the individual's characteristics. Using a predictive model to assign the probabilities gives each observation a relatively greater weight for those states whose residents that individual more closely resembles. At the same time, using the entire sample to estimate outcomes for each state increases the precision of the state estimates. Mathematica has made use of this methodology in its Survey of Income and Program Participation-based microsimulation model for policy analysis of SNAP.

7. State and Substate Estimates of Employment and Unemployment

Each month BLS publishes estimates of employment and unemployment for states and approximately 7,300 substate areas.[24] The estimates are based primarily on two monthly surveys: the CES and the CPS. Both surveys produce estimates of employment, but the estimates differ conceptually. The CES is a survey of establishments (work sites) and collects data on jobs on nonfarm payrolls by place of work. The CPS is a survey of households and collects data on the labor force status of household members by place of residence. A given individual may hold more than one job during the survey week, and both would be represented in the CES. In the CPS, each sample member of working age is classified as employed, unemployed, or out of the labor force during the survey week. A sample person with multiple jobs is counted as one employed person.

Under a federal-state cooperative program, detailed industry-level estimates of jobs, hours worked, and earnings are produced from the CES for the 50 states and DC, 372 metropolitan areas, and 34 metropolitan divisions. Most of the estimates utilize the "link relative" method, in which a matched sample of establishments responding to the current and prior month surveys is used to estimate the change in employment (or average hours worked or mean earnings), which is then added to the prior month's estimate of employment (or hours worked or earnings) to produce estimates for the current month. The employment estimates are benchmarked to estimates from the Quarterly Census of Employment (QCEW), which collects employment and earnings data from state Unemployment Insurance (UI) tax records covering 97 percent of nonfarm, civilian jobs.

For small industries the monthly samples are not sufficiently large to produce estimates with satisfactory precision. To generate estimates for these industries, BLS applies a small area method in which the estimate for a given domain is a weighted sum of the direct estimate of the month-to-month change and a time series forecast of the relative change. The estimator is based on the Fay-Herriot model, where the weight assigned to each component is a function of the variance of the direct estimate and the predictive accuracy of the time series forecast.

Also under a federal-state cooperative program, estimates of the civilian labor force and the unemployed are produced from the CPS and other sources for the same geographic areas as the CES estimates plus hundreds of additional areas. Estimates for the states and a small subset of metropolitan areas and state balances are based on time-series models with benchmarking to

[24] This narrative is based on descriptions provided in the *BLS Handbook of Methods*, chapters 2 and 4, accessed at http://www.bls.gov/opub/hom/; "Local Area Unemployment Statistics Estimation Methodology," accessed at http://data.bls.gov/lau/laumehd.htm on February 6, 2013; and "Notice of Changes to the CES Small Domain Model," accessed at http://data.bls.gov/cgi-bin/print.pl/sae/saesdm.htm on May 22, 2013.

national totals. Estimates for the remaining areas are derived through a building block approach that dates back to the late 1950s, although it has been modified a number of times since.

The time series models used for states and large substate areas represent the actual CPS monthly estimates in the form of a signal plus noise, where each is modeled separately. The signal is estimated from a model that decomposes the time series into stochastic trend, seasonal, and irregular components and includes predictors from the UI and CES programs as well as historical relationships in the CPS itself, as reflected in the time series. The general approach was proposed by Scott and Smith (1974) and developed more fully by Bell and Hillmer (1990) and Pfeffermann (1991). The modeling also includes outlier identification and removal. The noise component is estimated from a model that reflects the autocorrelation of sampling error in the CPS, where 75 percent of the samples for consecutive months—and 50 percent of those 12 months apart—are drawn from the same addresses (Tiller 1992). [25]

Separate models of the employment and unemployment levels are estimated for each state plus New York City and the Los Angeles metropolitan division, and estimates of the labor force level and the unemployment rate are derived from the outputs of the two models (the labor force level as the sum of the employment and unemployment levels, and the unemployment rate as the unemployment level divided by the labor force level). The employment model includes nonagricultural payroll estimates from the CES as a covariate while the unemployment model includes UI claims. [26]

The state model estimates are benchmarked to CPS national estimates of employment and unemployment for that month (described as "real-time" benchmarking). The benchmarking accounts for factors that would show up only slowly in the model estimates themselves, which rely on historical data. It also renders the estimates comparable across levels of geography.

The methodology for producing state estimates has evolved significantly in the past five decades. Starting in 1978, monthly estimates for the 10 largest states, New York City, and the Los Angeles-Long Beach metropolitan area were derived as direct estimates from the CPS. Estimates for the remaining states and areas were developed using the building block method. The number of direct-use states was expanded to 11 in 1985, and in 1989 the time-series approach was introduced for the remaining 39 states and selected large areas. A second generation time-series approach was introduced in 1994, coinciding with a major redesign of the CPS (including the introduction of computer-assisted interviewing). With subsequent sample cuts, it became difficult to maintain the direct estimates with the required precision, so time-series modeling was extended to the direct-use states and areas. A third generation of time-series models that included real-time benchmarking was introduced in 2005, and a method of smoothing the seasonally adjusted estimates was added in 2010.

Below the state level and the largest metropolitan areas, UI claims provide counts of the unemployed for all labor market areas and some smaller areas, but many of the unemployed do not receive UI benefits. Those who do not include:

[25] A detailed description of the models and their estimation is provided by Tiller (2006).

[26] The models that are estimated for five additional large metropolitan areas and their respective state balances do not include the CES and UI covariates.

- Former UI recipients who have exhausted their benefits but continue to look for work (that is, continue to meet the definition of unemployed)
- Persons who worked in jobs that are not covered by the UI system
- Persons who are entering the labor force for the first time
- Persons who are re-entering the labor force after an absence

New entrants and re-entrants were not employed immediately prior to their current spell of unemployment and, therefore, do not qualify to receive UI benefits.

Under the building block method, UI claimants are estimated directly from their monthly UI claims.[27] Former claimants who have exhausted their benefits (exhaustees) are estimated from final UI payments received during the reference week and prior weeks, based on an empirical model of how many weeks those who have exhausted their benefits continue to look for work. Both entrants and re-entrants are estimated from monthly state CPS data using a five-year weighted average. These state estimates are allocated to labor market areas within the state based on population shares (of persons 16-19 for new entrants and 20 and older for re-entrants). Unemployment associated with jobs not covered by the UI system is estimated from ratios derived from noncovered employment.[28] Estimates of these components of unemployment are summed, and the resulting unemployment estimates are adjusted to the state totals derived from the time-series models (specifically, each area's share of the state building block total is multiplied by the state unemployment estimate).

Estimates of employment for most metropolitan areas and large labor market areas are taken from the CES but with an adjustment for place of residence. The adjustment is based on commuting patterns reported in the decennial census (and now the ACS). For small metropolitan areas and labor market areas the employment estimates are obtained instead from the QCEW.

State agencies also produce estimates for the remaining metropolitan areas, counties, labor market areas, cities with 25,000 or more residents, all cities and towns in New England regardless of size, and areas used for various federal assistance programs—more than 7,300 unique geographic areas in all. The labor market area estimates of the unemployed are disaggregated to component counties and cities using UI claims, if available, and Census Bureau population estimates. The covered unemployed (claimants plus exhaustees) are disaggregated based on UI claims while entrants and re-entrants are disaggregated based on population shares of teens and adults. The unemployed who worked in jobs not covered by the UI system are disaggregated separately. Labor market area estimates of employment are disaggregated using employment shares from the ACS (previously the decennial census).

[27] In the late 1970s, the UI claims data were standardized to be more consistent with the concept of unemployment underlying the state and local area estimates. The improvements focused on the claims data produced for the CPS reference week. These claims data are based on residence and exclude claimants with any earnings during the reference week.

[28] The estimates of noncovered employment were based on the decennial census long form prior to 2010 and will be replaced, presumably, by estimates based on the ACS.

This page has been left blank for double-sided copying.

IV. RESEARCH AND DEVELOPMENT

Small area estimates of a wide range of variables have been developed experimentally—some with the goal of putting them into production and others to demonstrate what small area methods can accomplish. For at least one agency, the production of small area estimates was explored as part of a broader redesign of a major survey. In this chapter we review several examples of SAE research and development conducted within the federal government or with direct application to federal surveys, beginning with the aforementioned redesign effort.

A. Survey Redesign to Support Small Area Estimates

The National Crime Victimization Survey (NCVS) was initiated by the Bureau of Justice Statistics (BJS) in 1972 to produce national estimates of the levels and characteristics of criminal victimization, including crime not reported to the police. For the past several years the BJS has been engaged in an effort to improve the capacity of the NCVS to produce subnational estimates. Because the NCVS was not designed to produce estimates below the national level, revisions to the sample design have been included among the options being considered.

As part of a redesign of the survey, the BJS contracted with Westat to explore the production of small area victimization estimates. As part of this work, preliminary state models and estimates have been developed. Westat reviewed three possible sources for predictor variables—the ACS, BJS administrative statistics, and the Uniform Crime Report Program (Cantor et al. 2010). Westat's report discussed model development related issues such as selecting a small set of predictor variables from a larger set of potential predictors, finding the correct functions of those predictors (linear, quadratic, a polycotomous categorization), defining important interactions between the selected predictors, deciding on a final model for generating the indirect estimates (empirical Bayes versus hierarchical Bayes), specifying a methodology for generating the estimates (for example, MCMC), exploring unit-level vs. area-level approaches, and considering an extension of the modeling to time series models that take advantage of the longitudinal aspect of the NCVS. In its assessment, Westat evaluated two alternatives to SAE: direct estimation with the present sample design and adding sample by alternative methods. In recommending next steps, Westat advised that none of the three approaches be ruled out as yet, and efforts have continued on parallel tracks.

Li, Diallo, and Fay (2012) discuss efforts to develop an approach to producing small area estimates using the core NCVS and auxiliary variables that are available currently. They present a candidate approach, which involves combining observations across several years and using the EBLUP form of the BLUP estimator to derive the estimates for small areas. They also describe their investigation of potential auxiliary variables. While one of the factors motivating the NCVS was the underreporting of crime to the police, the Federal Bureau of Investigation's Uniform Crime Reports (UCR) remains a key source of predictors. Based on state-level regressions, the best UCR predictors of several NCVS crime rates were identified. Continuing efforts are focusing on refining the small area model, producing performance and diagnostic indicators to assess model quality and reliability, and examining other output related to estimates on domestic violence, victim demographics, and other geography (cities, counties).

One of the survey design options under consideration is to field a separate, lower-cost crime survey in parallel with the NCVS, which would then be reduced to core questions. The separate survey might be administered by mail, for example, and targeted to the areas for which estimates are desired. Edwards, Brick, and Lohr (2012) discuss issues in the design of this companion survey. One approach is to use the companion survey to collect better auxiliary variables to support model-based estimates of subnational areas using the core survey as the source of direct estimates. Another strategy is to use the low cost of the companion survey to collect more extensive sample data on key estimands and then combine the estimates of these variables from the two surveys.

Fay and Li (2012) discuss a third broad option, which is to expand the core survey sample and revise the design, taking into account the goal of producing state victimization rates. Taking note of the design of the ACS, their analysis is based on combining three years of survey data to produce state estimates. With the 2011 sample size, the NCVS could support state estimates with CVs under 10 percent for just three states. With a relatively modest sample size increase, three-year CVs below 10 percent could be achieved for the 12 largest states. To achieve a 10 percent CV for all 50 states and the District of Columbia, however, the annual sample would have to be expanded to nearly three times its 2011 size—to almost 300,000 persons. If funding for such an expansion cannot be secured, a compromise design could be developed that would support direct state estimates for some but not all states while providing strong national estimates.

B. Research Applications

We review applications of small area methods to topics related to health services utilization, health status and risk factors, and tax analysis.

1. Health Services Utilization

Schneider et al. (2009) used small area estimation to generate information on county-level disparities in mammography use by race and age group. Their synthetic regression approach involved first estimating a model predicting the prevalence of mammography for each of nine race-age categories. Potential predictors included individual variables from the 2000 BRFSS and county-level variables drawn from the 2000 Decennial Census, the Health Resources and Services Administration's (HRSA's) Area Resource File, and the number of local mammography facilities certified by the Food and Drug Administration (FDA) per 10,000 women aged 40 or older. The predictive models were used to prepare a synthetic estimate of the proportion of women in each county who reported having had a mammogram in the previous two years (the BRFSS measure). The separate predictions by race and age were combined based on the demographic composition of each county as observed in the census.

The National Immunization Survey (NIS), conducted by the CDC, monitors the vaccination rates of children who are between 19 and 35 months old. It produces estimates for the national population and selected local areas. For surveillance reasons, health administrators are interested in vaccination rates at the sub-state level, which the design of the NIS does not accommodate directly. Ganesh et al. (2009), discussed in Chapter II, derived model-based estimates for counties not included in the NIS design. Specifically, using data from counties with direct estimates they developed a small area model that incorporates temporal trends. The model produced county-level vaccination coverage rates with narrower confidence intervals that those from direct estimates.

When an unexpected shortage of influenza vaccines developed during the 2004-2005 influenza season, survey questions about vaccination were added to the BRFSS, which is conducted throughout the year. This enabled real-time, monthly estimates of vaccination coverage from October 2004 through January 2005, but while the BRFSS provides direct estimates for selected counties and metropolitan areas, it is not designed to support estimates for all counties. Jia et al. (2006) applied a combination of model-based SAE procedures (specifically, a generalized linear mixed model) and nonparametric spatial smoothing (Pickle and Su 2002, discussed in Chapter II) to the BRFSS data to estimate monthly vaccination rates for all counties—including those with few or no survey respondents.

The National Ambulatory Medical Care Survey (NAMCS) presents estimates of visits to office-based physicians and selected community health centers. It is conducted by NCHS. In 2012, the NAMCS was redesigned with a substantial increase in sample size; however, within small areas, direct survey estimates may remain imprecise. Model misspecification in model-based estimates may induce unintended biases, depending on the choice to model fixed or random effects. Beresovsky and Malec (2012) demonstrate that, in the context of the NAMCS, the logistic-normal model is more robust to misspecification of fixed effects and more efficient than direct estimates. The related National Hospital Ambulatory Medical Care Survey (NHAMCS) produces national estimates of health service utilization at emergency, outpatient, and ambulatory surgery departments on noninstitutional, non-Federal general hospitals. The study of utilization in small areas, such as states and counties, motivates the development of model-based estimates, proposed by Beresovsky et al. (2011) and Beresovsky et al. (2010). They incorporated auxiliary data from the Area Resource File and the Verispan Hospital Database to create small area prediction models and compared their estimates to administrative, nationally representative data from the Healthcare Cost and Utilization Project databases of the Agency for Healthcare Research and Quality.

2. Health Status and Risk Factors

The BRFSS collects data on six individual-level behavioral health risk factors associated with the leading causes of premature mortality and morbidity among adults, namely: cigarette smoking, alcohol use, physical activity, diet, hypertension, and safety belt use. The BRFSS is a state-based health survey, but county-level estimates are often required by programs, for example, to develop estimates of disease prevalence for allocation of funds according to local need. To accomplish this, auxiliary information on the targeted small-area population can be used in a hierarchical modeling framework. For example, a study by Zhang et al. (2011), discussed as an example of non-Bayesian modeling in Chapter II, used data from the 2000 decennial census and the U.S. Department of Agriculture 2004 county typology codes, in addition to BRFSS, to study the prevalence of obesity in Mississippi. Related methodological work for small area estimates of body mass index was under-taken by Nandram and Choi (2010), who investigated implications of nonresponse and selection bias in small domains, and by Ybarra and Lohr (2008), who developed a small area estimator that accounts for measurement error in auxiliary information. These latter two methodological projects applied their methods to the National Health and Nutrition Examination Survey.

Congdon (2009) proposed a multilevel prevalence model for cardiovascular outcomes that incorporates both patient risk factors from BRFSS survey data and their variation by geographic location. The model improves explanation of variation in disease prevalence by geography—in this case ZIP Code tabulation areas—not otherwise accounted for in the distribution of patient risk factors.

Survey estimates for small domains that are not geographic, such as small racial/ethnic groups, can be improved by pooling estimates over multiple years—that is, by borrowing strength over time. However, the policy context may change over time so that pooled estimates may not accurately reflect current needs. Using NHIS data from 1997 to 2004, Elliott et al. (2009) developed an approach for combining data across years using the Kalman filter (Kalman 1960), a computational algorithm for separating signal from noise. They used prior year data to predict the current year value based on the strength of the autocorrelation in successive year-to-year values and then combined this prediction with the current year direct estimate using weights that depend on how much the current year mean differs from both the trend and the recent past.[29] More weight was given to the current year estimate when the difference, measured as a variance, was large. They applied this approach to improve the accuracy of current health prevalence estimates for rare racial/ethnic groups by aggregating cross-sectional estimates. The outcomes that were estimated included chronic and acute health conditions, risk factors, and health care utilization. The accuracy of the estimates was improved by pooling while the currency of the pooled estimates was enhanced by applying greater weights to more recent estimates and smaller weights to earlier estimates. For 18 of 19 outcomes, the error for separate estimates of American Indians/Alaskan Natives and Chinese Americans—which represent only ½ to 1 percent of the U.S. population—was reduced by 25 to 30 percent compared to single-year direct estimates.

3. State-level Tax Analysis

The Office of Tax Analysis (OTA) in the Treasury Department uses a microsimulation model to estimate the impact of hypothetical changes to individual tax law. The database for the model is a sample of individual tax returns drawn from the population of tax returns filed in a given year. The sample—which is edited, enhanced with modeled tax variables, and extrapolated to a future tax year—is large enough to support estimates for the nation and the largest states, but estimates for the remaining states lack sufficient precision. To address this need, OTA applied the methodology of Schirm and Zaslavsky (1998, 2002) and generated for each observation in the tax model database a vector of state probabilities. These probabilities were calculated from a set of tabulations produced from unedited data for the entire population of returns. Each element in the vector represents an estimate of the probability that the taxpayer filed in a given state. These probabilities sum to 1 for each record. When applied to an observation's sample weight, these probabilities yield a vector of state weights. Across observations, a given state weight sums to the total returns filed in that state. The state weights can be used to estimate state totals of a wide variety of tax variables. To evaluate the results, state estimates from the model were compared to published tabulations of tax items by state. A paper describing the approach to deriving and evaluating the state weights is forthcoming.

4. People in Poverty

The modeling approach used by the Census Bureau in its SAIPE program involves predicting the log of the number of poor children in an area. Counties with direct estimates of zero poor children must be excluded from the analysis because the log of zero is undefined.

[29] Setodji et al. (2011) provide documentation for a SAS macro that implements the MKF algorithm.

During four of the five years from 2005 through 2009, more than 5 percent of counties had to be excluded from the model estimation. Other limitations have prompted consideration of modeling poverty rates rather than levels. To address both issues, Wieczorek and Hawala (2011) proposed a “zero-one inflated beta” regression model, which combines the use of a beta distribution to model the distribution of poverty rates between 0 and 1, excluding the end points, and a multinomial model to handle rates of 0 and 1. Wieczorek, Nugent, and Hawala (2012) expanded and evaluated this model, using draws from a simulated population developed from ACS data. The new model was given a Bayesian formulation and fit with MCMC methods and compared to both direct and Fay-Herriot estimators of the poverty rate in 100 counties. The new model showed improvements upon the Fay-Herriot model with respect to bias, MSE, and coverage of confidence intervals and was only slightly more biased than the direct estimator.

This page has been left blank for double-sided copying.

REFERENCES

Barker, L.E., T.J. Thompson, K.A. Kirtland, J.P. Boyle, L.S. Geiss, M.M. McCauley, and A.L. Albright. "Bayesian Small Area Estimates of Diabetes Incidence by United States County, 2009", *Journal of Data Science*, vol. 11, 2013, pp. 249-267.

Baskin, R.M., and J.P. Sommers. "Comparison of direct, mixed model, and Bayesian metropolitan statistical area estimates for the insurance component of the Medical Expenditure Panel Survey." *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association. 2008, pp. 1405-1408

Battaglia, M.P., M.R. Frankel, M. Town, and L. Balluz. "A virtual popluation approach to small area estimation." *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, 2011.

Bauder, M., D. Luery, and S. Szelepka. "Small Area Estimation of Health Insurance Coverage in 2008 and 2009." Technical report. Washington, DC: U.S. Census Bureau, February 16, 2012.

Bell, W.R. and S.C. Hillmer. "The Time Series Approach to Estimation for Repeated Surveys." *Survey Methodology*, vol. 16, 1990, pp. 195-215.

Beresovsky, V., and D. Malec. "Effect of Model Misspecification on Small-Area Estimation of Proportions from the National Ambulatory Medical Care Survey Data." FCSM 2012 Research Conference Proceedings. Washington, DC: Federal Committee on Statistical Methodology, 2012.

Beresovsky, V., C. Burt, V. Parsons, and N. Schenker. "Application of Small-Area Estimation Methods to Emergency Department Data from the National Hospital Ambulatory Medical Care Survey." *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, 2010.

Beresovsky, V., C. Burt, V. Parsons, N. Schenker, and R. Mutter. "Application of Hierarchical Bayesian Models with Poststratification for Small-Area Estimation from Complex Survey Data." *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, 2011.

Blumberg, S.J., N. Ganesh, J.V. Luke, and G. Gonzales. "Wireless Substitution: State-Level Estimates from the National Health Interview Survey, 2012." *National Health Statistics Reports*, no. 70. Hyattsville, MD: National Center for Health Statistics, December 18, 2013.

Blumberg, S.J., J.V. Luke, N. Ganesh, M.E. Davern, M.H. Boudreaux, and K. Soderberg. "Wireless Substitution: State-level Estimates from the National Health Interview Survey, January 2007-June 2010." *National Health Statistics Reports*, no. 39. Hyattsville, MD: National Center for Health Statistics, April 20, 2011.

Bramlett, M., and S. Blumberg. "Prevalence of Children with Special Health Care Needs in Metropolitan and Micropolitan Statistical Areas in the United States." *Maternal and Child Health Journal*, vol. 12, no. 4, July 2008, pp. 488–498.

Cadwell, B.L., T.J. Thompson, J.P. Boyle, L.E. Barker. "Bayesian small area estimates of diabetes prevalence by US county, 2005." *Journal of Data Sciences*, vol. 8, 2010: pp. 173-188.

Cantor, D., T. Krentzke, D. Stukel, and L. Rizzo. "NCVS Task 4 Report: Summary of Options Relating to Local Area Estimation." Final report to the Bureau of Justice Statistics. Rockville, MD: Westat, May 19, 2010.

Chattopadhyay, M. P. Lahiri, M. Larsen, and J. Reimnitz. "Composite Estimation of Dru Prevalences for Sub-State Areas." *Survey Methodology*, vol. 25, no. 1 (June 1999), pp. 81-86.

Congdon, P.A. "A multilevel model for cardiovascular disease prevalence in the US and its application to micro area prevalence estimates." *International Journal of Health Geographics*, vol. 8, no. 6, 2009, pp. 1-14.

Cunnyngham, Karen. "State Estimates of Children Ages 0 to 4 in Low-Income Families in 2010 Derived from CPS Data Using an Empirical Bayes Shrinkage Estimator." Memorandum to Grant Lovellette and Ed Harper, Food and Nutrition Service, August 9, 2012.

Cunnyngham, K., L. Castner, and A. Sukasih. "Empirical Bayes Shrinkage Estimates of State Supplemental Nutrition Assistance Program Participation Rates in 2008-2010 for All Eligible People and the Working Poor." Report submitted to the USDA Food and Nutrition Service. Washington, DC: Mathematica Policy Research, February 2013.

Cutchin, M.P., K.Eschbach, C.R. Mair, H. Ju, and J.S. Goodwin. "The Socio-spatial Neighborhood Estimation Method, An Approach to Operationalizing the Neighborhood Concept." *Health Place*, vol. 17, no.. 5 (September 2011): pp. 1-23.

Edwards, W.S., J. M. Brick, and S. Lohr. "Designing a Low(er)-Cost Complement to the National Crime Victimization Survey." FCSM 2012 Research Conference Proceedings. Washington, DC: Federal Committee on Statistical Methodology, 2012.

Elliott, M., and W. Davis. "Obtaining Cancer Risk Factor Prevalence Estimates in Small Areas: Combining Data from Two Surveys." *Journal of the Royal Statistical Society, Series C (Applied Statistics),* vol. 54, no. 3, 2005, pp. 595–609.

Elliott, M., D. McCaffrey, B. Finch, D. Klein, N. Orr, M. Beckett, and N. Lurie. "Improving Disparity Estimates for Rare Racial/Ethnic Groups with Trend Estimation and Kalman Filtering: An Application to the National Health Interview Survey." *Health Services Research*,
vol. 44, no. 5, part I, 2009.

Eltinge, J.L. "Small Domain Estimation for the U.S. Current Employment Statistics Program: Management Implications of Multiple Stakeholders and Multiple Constraints." Statistical Policy Working Paper 35. Federal Committee on Statistical Methodology. October 2004.

Fay, R., and R. Herriot. "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data." *Journal of the American Statistical Association*, vol. 74, no. 366, 1979, pp. 269–277.

Fay, R.E. and J. Li. "Rethinking the NCVS: Subnational Goals through Direct Estimation." FCSM 2012 Research Conference Proceedings. Washington, DC: Federal Committee on Statistical Methodology, 2012.

Federal Committee on Statistical Methodology. Statistical Policy Working Paper 21: Indirect Estimators in Federal Programs. Washington, DC: Office of Management and Budget, 1993.

Fisher, R., and G. Gee. "Errors-in-Variables County Poverty and Income Models." In *Proceedings of the Joint Statistical Meetings, Sections on Government and Social Statistics*. Alexandria, VA: American Statistical Association, 2004.

Fisher, R., B. O'Hara, and S. Riesz. "Small-Area Estimation of Health Insurance Coverage: State-Level Estimates for Demographic Groups." *Proceedings of the Joint Statistical Meetings, Section on Government and Social Statistics*. Alexandria, VA: American Statistical Association, 2006.

Folsom, R., B. Shah, A. Vaish. "Substance Abuse In States: A Methodological Report On Model Based Estimates From The 1994-1996 National Household Surveys On Drug Abuse." *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association. 1999, pp. 371-375

Ganesh, N., P. Smith, R. Kennon, K. Copeland, and M. Wolter. "Modeling County-Level Vaccination Coverage Rates Across Time." *Proceedings of the Section on Survey Research Methods.* Alexandria, VA: American Statistical Association, 2009.

Gonzalez, J. F., P.J. Placek, and C. Scott. "Synthetic Estimation in Followback Surveys at the National Center for Health Statistics." In Federal Committee on Statistical Methodology. Statistical Policy Working Paper 21: Indirect Estimators in Federal Programs. Washington, DC: Office of Management and Budget, 1993.

Goodman, M. "Comparison of Small-Area Analysis Techniques for Estimating Prevalence by Race." *Preventing Chronic Disease*, vol. 7, no. 2, March 2010.

Hughes, A., P. Muhuri, N. Sathe, and K. Spagnola. "State Estimates of Substance Use and Mental Disorders from the 2008–2009 National Surveys on Drug Use and Health." Rockville, MD: Substance Abuse and Mental Health Services Administration, 2011.

Jia, H., M. Link, J. Holt, A.H. Mokdad, L. Li, and P.S. Levy. "Monitoring County-Level Vaccination Coverage During the 2004-2005 Influenza Season." *American Journal of Preventive Medicine*, vol. 31, no. 4, 2006, pp. 275-280.

Jia, H., P. Muenning, and E. Borawski. "Comparison of Small-Area Analysis Techniques for Estimating County-Level Outcomes." *American Journal of Preventive Medicine*, vol. 26, no. 5, 2004, pp. 453–460.

Kalton, G. "Providing Small Area Estimates Discussion." Statistical Policy Working Paper 35. Federal Committee on Statistical Methodology. October 2004.

Kalman, R. "A New Approach to Linear Filtering and Prediction Problems." *Transactions of the ASME—Journal of Basic Engineering*, vol. 82, 1960, pp. 35-45.

Li, J., M.S. Diallo, and R.E. Fay. "Rethinking the NCVS: Small Area Estimation Approaches to Estimating Crime." FCSM 2012 Research Conference Proceedings. Washington, DC: Federal Committee on Statistical Methodology, 2012.

Malec, D. "Model Based State Estimates for the National Health Interview Survey." In Federal Committee on Statistical Methodology. Statistical Policy Working Paper 21: Indirect Estimators in Federal Programs. Washington, DC: Office of Management and Budget, 1993.

Malec, D., J. Sedransk, C. Moriarity, and F. Leclere. "Small-Area Inference for Binary Variables in the National Health Interview Survey." *Journal of the American Statistical Association*, vol. 92, no. 439, 1997, pp. 815–826.

Mungiole, M., L.W. Pickle, and K.H. Simonson. "Application of a Weighted Head-banging Algorithm to Mortality Data Maps." *Statistics in Medicine*, vol. 18, 1999, pp. 3201-3209.

Nandram, B., and J. Choi. "A Bayesian Analysis of Body Mass Index Data From Small Domains Under Nonignorable Nonresponse and Selection." *Journal of the American Statistical Association*, vol. 105, no. 489, March 2010, pp. 120–135.

National Research Council. *Small-Area Income and Poverty Estimates: Priorities for 2000 and Beyond.* Panel on Estimates of Poverty for Small Geographic Areas, edited by C. Citro and G. Kalton, Committee on National Statistics. Washington, DC: National Academy Press, 2000.

Pfeffermann, D. "New Important Developments in Small Area Estimation." *Statistical Science*, vol. 28, no. 1, 2013, pp. 40–68.

Pfeffermann, D. "Estimation and Seasonal Adjustment of Population Means Using Data from Repeated Surveys." *Journal of Business and Economic Statistics*, vol. 9, no. 2, April 1991, pp. 163-175.

Pickle, L.W. and Y. Su. "Within-State Geographic Patterns of Health Insurance Coverage and Health Risk Factors in the United States." *American Journal of Preventive Medicine*, vol. 22, no. 2, February 2002, pp. 75-83.

Raghunathan, T., X. Dawei, N. Schenker, V. Parsons, W. Davis, K. Dod, and E. Feuer. "Combining Information from Two Surveys to Estimate County-Level Prevalence Rates of Cancer Risk Factors and Screening." *Journal of the American Statistical Association*, vol. 102, no. 478, June 2007, pp. 474-486.

Rao, J.N.K. *Small Area Estimation*. Hoboken, NJ: John Wiley & Sons, 2003

Schaible, W.L. "Choosing Weights for Composite Estimators or Small Area Statistics." Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods. Alexandria, VA: American Statistical Association, 1978: pp. 741-46.

Schirm, A. "The Relative Accuracy of Direct and Indirect Estimates of State Poverty Rates." Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods. Alexandria, VA: American Statistical Association, 1994: pp. 1071-1076.

Schirm, A., and A. Zaslavsky. "Reweighting a National Database to Improve the Accuracy of State Estimates." *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, 2002.

Schirm, A. and A. Zaslavsky. "Model-Based Microsimulation Estimates for States when State Programs Vary." *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, 1998, pp. 339-344.

Schirm, A., A. Zaslavsky, and J. Czajka. "Large Numbers of Estimates for Small Areas." FCSM 1999 Research Conference Proceedings. Washington, DC: Federal Committee on Statistical Methodology, 1999.

Schneider, K.L., K.L. Lapane, M.A. Clark, and W. Rakowski. "Using Small-Area Estimation to Describe County-Level Disparities in Mammography." *Preventing Chronic Disease*, vol. 6, no. 4, A125, October 2009.

Scott, A.J. and T.M.F. Smith. "Analysis of Repeated Surveys Using Time Series Methods." *Journal of the American Statistical Association*, vol. 69, 1974, pp. 674-678.

Setodji, C.M., J.R. Lockwood, D.F. McCaffrey, M.N. Elliott, and J.L. Adams. "The Modified Kalman Filter Macro: User's Guide." Santa Monica, CA: RAND Corporation, 2011.

Siordia, C. "Benefits of Small Area Measurements: A Spatial Clustering Analysis on Medicare Beneficiaries in the USA." *Human Geographies: Journal of Studies and Research in Human Geography* vol. 7.1 (2013): 53-59.

Sommers, J. "Methodology Report #16: Producing State Estimates with the Medical Expenditure Panel Survey, Household Component." December 2005. Available at [http://meps.ahrq.gov/mepsweb/data_files/publications/mr16/mr16.pdf]. Accessed February 8, 2013.

Srebotnjak T, Mokdad A, Murray C. A novel framework for validating and applying standardized small area measurement strategies. *Population Health Metrics*, vol. 8, no. 6, 2010, pp. 1-13.

Tiller, R. "Model-Based Labor Force Estimates for Sub-National Areas with Large Survey Errors." Washington, DC: Bureau of Labor Statistics, March 2006. Available at http://www.bls.gov/ore/pdf/st060010.pdf.

Tiller, R.B. "An Application of Time Series methods to Labor Force Estimation Using CPS Data." *Journal of Official Statistics*, vol. 8, 1992, pp. 149-166.

Wieczorek, J., and S. Hawala. "A Bayesian Zero-One Inflated Beta Model for Estimating Poverty in U.S. Counties." *Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods.* Alexandria, VA: American Statistical Association, 2011.

Wieczorek, J., C. Nugent, and S. Hawala. "A Bayesian Zero-One Inflated Beta Model for Small Area Shrinkage Estimation." *Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods.* Alexandria, VA: American Statistical Association, 2012, pp. 3896-3910.

Wright, D. "Policy Considerations in the Development of State Estimates of Substance Use Rates." Statistical Policy Working Paper 35. Federal Committee on Statistical Methodology. October 2004.

Wright, D. "State Estimates of Substance Use from the 2001 National Household Survey on Drug Abuse: Volume II. Individual State Tables and Technical Appendices." HHS Publication No. SMA 03-3826, NHSDA Series H-20. Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies, 2003.

Ybarra, L., and S. Lohr. "Small area estimation when auxiliary information is measured with error." *Biometrika*, vol. 95, no. 4, 2008, pp. 919–931.

Zhang, Z., L. Zhang, A. Penman, and W. May. "Using Small-Area Estimation Method to Calculate County-Level Prevalence of Obesity in Mississippi, 2007-2009." *Preventing Chronic Disease*, vol. 8, no. 4, July 2011.