Session 10

Providing Small Area Estimates

Small Domain Estimation for the U.S. Current Employment Statistics Program: Management Implications of Multiple Stakeholders and Multiple Constraints

John L. Eltinge[29]
Bureau of Labor Statistics

1. Introduction

The Current Employment Statistics (CES) survey is a large-scale monthly establishment survey conducted for the Bureau of Labor Statistics through a joint federal-state program. This survey currently is completing a transition to a probability sample design that is intended to produce relatively precise estimators of monthly total employment for each state in the U.S. For some general background on the CES and related technical and policy issues, see, e.g., American Statistical Association (1994), Butani, Harter, and Wolter (1997), Butani, Stamas and Brick (1997), West et al. (1997) and Werking (1997).

Although the CES sample design and estimation work focused primarily on production of estimates at the state and national levels, many stakeholders have strong interest in estimation for considerably smaller domains, e.g., for a specified major industrial division within a given metropolitan statistical area. Consequently, the Bureau of Labor Statistics is developing model-based small domain estimation methods for the CES. In general, small domain estimation involves very rich set of technical, management and policy issues. For some general background, see, e.g., Fay and Herriot (1979), Platek et al. (1987), Ghosh and Rao (1994), Schaible (1996), Shen and Louis (1998), Singh et al. (1998), Schirm et al. (1999), Marker (1999, 2001), National Research Council (2000 a, b, 2001), Rao (2002) and references cited therein.

In keeping with the theme of this year's FCSM Statistical Policy Conference, "Challenges to the Federal Statistical System in Fostering Access to Federal Statistics," I will focus this presentation on some policy and management issues that appear to have a substantial effect on the degree and nature of many stakeholders' use of small area estimates. Specifically, I will suggest that for some small domain estimation programs, the presence of multiple stakeholders and multiple constraints can have a substantial effect on the development and implementation of our small area estimation methods, and on the best strategies for communication of small domain information to various groups of stakeholders. In addition, I will suggest that it can be useful to view small domain estimation methodology as a form of technology; and that the development and implementation of small domain programs may benefit from previous literature on the adoption and diffusion of technology.

2. Multiple Stakeholders and Multiple Utility Functions Small Domain Programs

Some small domain estimation programs have been developed primarily for a single relatively well-identified purpose. A prominent example of this is the he Small Area Income and Poverty Estimates, or SAIPE, program of the U.S. Census Bureau and the U.S. Department of Education) which is focused primarily on funding allocation formulas, and for which other uses

[29] Office of Survey Methods Research, PSB 1950, Bureau of Labor Statistics, 2 Massachusetts Avenue NE, Washington, DC 20212. Eltinge_J@bls.gov

are somewhat secondary. On the other hand, other small domain estimation programs, including the one currently under development for the Current Employment Statistics program, have been motivated by a wide range of potential uses, without a single dominant user, and without a corresponding dominant consensus regarding priorities among estimands and the objective functions appropriate for evaluation of estimator performance.

Instead, extensive discussions with states and other stakeholders indicated a wide range of views regarding, e.g., the relative importance of estimates of total employment, one-month change, three-month change and twelve-month change; and an equally wide range on the relative importance assigned to relative bias, absolute bias, relative variance and absolute variance. In addition, some stakeholders use published small domain estimates as one of several sources of information regarding local economic conditions, and do not use these estimates in ways that would lend themselves to a solid characterization of, e.g., the relative benefits of reducing bias or reducing variance.

Finally, we have also encountered some data users who focus heavily on published CES small domain estimates, but do so in a form that may be described as informal simultaneous inference. For example, a metropolitan area analyst may examine time plots of employment estimates for the past three to eighteen months for a given metropolitan area-industry combination, and then attempt to make statements like, "employment is approximately constant," "employment has an upward trend" or "employment growth is very similar to what we are seeing in the rest of the state." This suggests that it may be worthwhile for the BLS and other statistical agencies to consider two areas for additional work. First, the small domain literature has tended to focus a substantial amount of attention on estimation of variances or mean squared errors. Given the abovementioned attempts at informal inference, it may be appropriate to develop related procedures for simultaneous confidence sets for linear functions of a vector of monthly estimands; and to develop related user-friendly graphical tools. Second, some of the abovementioned informal analyses appear to involve a large number of implicit hypothesis tests. Consequently, it would be worthwhile to explore the extent to which we can help the analysts incorporate appropriate measures of uncertainty into such tests; and to explore the extent to which one may use "false discovery rates" (e.g., Benjamini and Hochberg, 1995; DuMouchel, 1999; Genovese and Wasserman, 2002; and Storey, 2002) and related tools to provide appropriate quantitative measures for the risks incurred in this type of multiple testing.

3. Multiple Constraints in Small Domain Programs

The literature on statistical policy often notes – either explicitly or implicitly – that practical constraints can have a dominant role in the development, implementation and perceived value of a given statistical program. See, e.g., Bonnen (1988), Eckler (1972), Felligi (1996), Kaysen et al. (1969), Lehnen (1988), Moser (1976), O'Hare and Pollard (1998), Parke et al. (1976), Reynolds (1988), Rosenberg and Myers (1977), Shiskin (1970) and Weiner (1974). Similarly, the current development of the CES small domain estimation program has been heavily influenced by practical constraints on, e.g., the relatively brief time between data collection and publication deadlines, the timely availability of specific forms of microdata, a preference for compatibility with previously implemented methodology for national- and state-level estimation, and compatibility with legacy production systems. In many cases, a full-scale quantitative characterization of these constraints is not readily available, and work toward such a quantitative characterization can itself involve substantial costs and cognitive burden.

In contrast with this, the mathematical statistics literature in survey sampling tends to focus on optimization of an objective function (e.g., variance, mean squared error, or a pseudolikelihood expression) in the presence of a relatively complex stochastic structure induced by a complex sample design or a hierarchical model. In such work, constraints often are viewed as being relatively mild or otherwise of somewhat secondary interest.

Thus, in a qualitative sense, our optimization work follows the pattern displayed in Figure 1. (For some related discussion and partial exceptions to this, see, e.g., Ahsan and Khan, 1982; Cochran, 1977, Section 5A.3; Grzesiak and Johnson, 1989; Hansen et al., 1983; Harris, 1972; Kish, 1976; Neumann, 1999; and Renner, 1976.) On the other hand, the presence of substantial operational constraints, and the components of uncertainty associated with some of these constraints, could be characterized in a schematic form by Figure 2. Similarly, given the presence of many stakeholders with distinct utility functions, one could extend Figure 2 to include multiple objective-function curves.

4. Implications for Managers and Mathematical Statisticians

Comparison and contrast of the ideas in Sections 2 and 3 lead to some suggestions regarding efficient development of small domain estimation programs. First, for small domain work that is not dominated by a single objective (e.g., funding allocation formulas), some of the traditional optimization insights offered by mathematical statistics (as in Section 3) may be dominated by the presence of multiple utility functions, multiple constraints, and limitations of the quantification of these utility functions and constraints.

Second, this domination result is fundamentally an opportunity, rather than a barrier. From a pure research point of view, this provides a very rich set of mathematical statistics problems that, as we have seen, can be of very serious practical interest. In addition, this expanded set of problems appears to have substantial connections with some unresolved traditional problems in the analysis of survey data, e.g., issues raised by Hansen et al. (1983). This in turn potentially enhances the contribution that mathematical statistics can make to operational aspects of small domain programs. For example, comparison of Figures 1 and 2 suggests that one carry out a form of triage to identify specific constraints that may have the largest relative effect on the performance of the estimation program.

Third, this also leads to suggestions regarding efficient management of mathematical statistics research projects related to small domain estimation. Within agencies, mathematical statisticians constitute a relatively rare resource. Consequently, it is of interest to focus that resource on a moderate number of high-priority research areas that are most likely to lead to substantial improvements in agency production work. The ideas of Sections 2 and 3 suggest that for problems involving multiple utility functions and multiple constraints, one focus on cases involving sufficient common methodological structure and quantitative structure to lead to substantial improvements in perceived utility functions.

5. Access to Small Domain Estimates, and the Adoption and Diffusion of Related Innovations

Now let's return to the theme of this FCSM Policy Seminar, "Challenges to the Federal Statistical System in Fostering Access to Federal Statistics." For cases in which small domain estimates are used primarily for funding allocation formulas, it is possible that an agency may reasonably focus of development of appropriate methods, review of these methods by the National Academy of Sciences and other responsible outside scientific groups, and provision of public access to the estimates and associated measures of uncertainty. On the other hand, for a small domain program that does not have a single dominant purpose (e.g., the CES small domain program), it may not suffice to focus on relatively passive forms of access as such. In such a case, the practical value for multiple stakeholders may depend heavily on the extent (possibly limited) to which we can convey to these stakeholders a relatively refined sense of the information that is, and is not, conveyed by a given set of small domain estimates.

In thinking about programmatic ways in which we can address this need, it is useful to think about small domain estimation work as a technology, and to examine the extent to which we may obtain some management and statistical policy insights form previous studies of the ways in which multiple stakeholders explore and make decisions about a relatively new technology. These studies generally fall under the rubric of "adoption and diffusion of technology," and have arisen in several disciplines, including rural sociology, software engineering, military science, communications and marketing. A prominent early study by Ryan and Gross (1943) considered the processes by which groups of farmers in Iowa adopted hybrid seed corn. For detailed exposition and critique of the general literature on technology adoption and diffusion, see Rogers (1995) and references cited therein.

My initial reading of some of this literature led to several points of potential value for small domain work. Today, I'll emphasize five of these points, taken primarily from Rogers (1995). First, Rogers (1995, p. 11) defines an innovation broadly as, "an idea, practice or object that is perceived as new by an individual or other unit…." Thus, tables of small domain estimates, or new methods for the production and interpretation of these estimates would fit readily under this definition.

Second, Rogers (1995, p. 10) defines diffusion as a "process by which an innovation is communicated through certain channels over time among members of a social system." Note especially that this definition requires us to have a fairly concrete vision of the "members of the social system" who may use our estimates, and the channels through which we will communicate with them. This starts to push us beyond relatively passive notions of "access" and toward more active engagement with specific subsets of our multiple stakeholders.

Third, Rogers (1995, pp. 15-16) emphasizes five characteristics of innovations that he considers important in analysis of adoption and diffusion processes: relative advantage, compatibility, complexity, trialability and observability. Each of these clearly is applicable to small domain work, and may help us explain some of the responses of our stakeholders to new small domain estimation programs. For example, in keeping with the comments in Section 3, perceptions of relative advantage may be dominated initially by the ability to produce estimates for subpopulations that were not previously available. Subsequently, these perceptions may be influenced by observation of both Type I and Type II error rates encountered in either formal or informal inferences drawn from reported small domain estimates and associated measures of uncertainty.

Fourth, Rogers (1995, pp. 28-30) distinguishes among three types of innovation decisions: optional, collective and authority-based. For example, competition in a free market among several comparable technologies may result in optional innovation decisisons. On the other hand, an innovation decision driven by imposition of standards by a voluntary industry trade group would involve a mixture of collective and authority-based decisions. From the standpoint of states and other stakeholders, development and implementation of the CES small domain program may be viewed as largely authority-based, but extensive consultation with the Current Employment Statistics Policy Council added a component of collective decision-making to the process.

Fifth, Rogers (1995) and other authors often describe the process of adoption and diffusion with schematic diagrams like the one displayed in Figure 3. Along the horizontal axis, we have the elapsed time to adoption of a given innovation. The literature partitions the population of potential users into subpopulations according to their anticipated time to adoption. The resulting diagram uses a Gaussian density curve to indicate the approximate sizes of the subpopulations.

I would view the Gaussian approximation and specific subpopulation cutoffs with a considerable amount of caution, but I believe that in a qualitative sense, this schematic device can be useful. For example, on the left side of the time scale is a group labeled “innovators,” who are directly involved with research, development and very early use of a technology. For small domain estimation, this group would tend to include mathematical statisticians in academic institutions and in government statistical research and methods groups. Note especially that the values, technical sophistication and organizational dynamics within the “innovators” group tend to be quite different from those in the other groups. For instance, the next group, labeled “early adopters,” tend to work outside of the initial research and development environment, but potentially have an active interest in the development effort. For small domain work, this might include stakeholders directly involved with funding allocation formula work or with state Current Employment Statistics program offices.

The general suggestion from Figure 3 is that in the course of time, additional groups (labeled “early majority,” “late majority” and “very late or never”) may also begin to make use of a technology. Adoption decisions by these later groups often may be attributed to a combination of communication of this technology to members of these groups, and to the maturation of the technology as such. In that process, two important factors are the observable reward/risk ratio and the degree of standardization of the technology. The innovators and early adopters may be intensely interested in the technology because they anticipate that they be able to obtain a substantial (though perhaps rather uncertain) perceived reward in exchange for a given amount of investment risk. In contrast with this, the later groups may expect a higher degree of predictability in their observable reward/risk ratio, and would expect some degree of assurance that this ratio will exceed a reasonable threshold. In parallel with this, innovators and early adopters may tend to use highly customized technologies, while the later groups may expect a substantially higher degree of standardization.

As with any schematic description or generalization of human and organizational behavior, one should view the adoption and diffusion literature with a reasonable degree of caution. For example, some of this literature assumes a bit too readily that one should adopt the technology in question, and may not place enough emphasis on observable risk/reward ratios. Nonetheless, I believe that this literature does offer us some useful insights into management of small domain projects that do not involve a single dominant stakeholder.

First, as an agency and its stakeholders work with a large number of requests for small domain estimates, schematic diagrams like Figure 3 can help one set priorities and gain some perspective on the current state of development and adoption of a given set of small domain estimation methods. For example, if a program is in a relatively early stage of development, it may be best to focus efforts on requests from stakeholders with a realistic chance of obtaining fairly high levels of observable reward, relative to risk, and who have a relatively high tolerance for associated risks. Within this context, note that the literature on adoption and diffusion of technology often reports a sharp distinction, or "gap," between the "early adopter" group and subsequent groups. Consequently, many technologies that in principle could be used by a relatively large group do not, in practice, go beyond the "early adopter" stage.

Second, for small domain estimation, a serious assessment of relative rewards and risks tends to require balanced consideration of quantitative and qualitative components. Quantitative components include multiple components of uncertainty, e.g., sampling error, nonsampling error, model equation error and model misspecification effects. Somewhat more qualitative terms involve trade-offs among the costs of non-publication, relative costs of Type I and Type II errors, and the latter factors are often complicated by the presence of implicit multiple testing. Evaluation of these risks may require a relatively high level of customization, and thus, may require allocation of a relatively high level of resources by both the agency and by some data users. Within agencies, attempts to balance resource requirements against the abovementioned reward/risk calculations may be complicated by the fact that agencies generally do not have the same market mechanisms that are observed in some other areas of technology adoption and diffusion.

Acknowledgements

The author thanks the following persons for many helpful discussions of the Current Employment Statistics Program and small domain estimation: David Binder, Shail Butani, Moon Jung Cho, Cathryn Dippo, Wayne Fuller, Jack Galvin, Patricia Getz, Julie Gershunskaya, Rachel Harter, Steve Haslett, Larry Huff, Laurie Kujawa, Partha Lahiri, Kirk Muller, Lois Orr, Ken Shipp, Clyde Tucker, Kirk Wolter, Alan Zaslavsky and the members of the Current Employment Statistics Policy Council. In addition, the author thanks Fred Conrad, Norbert Schwarz and Roger Torangeau for comments on prospect theory; and Moon Jung Cho for preparing Figures 1 through 3. The opinions expressed in this paper are those of the author and do not necessarily represent the policies of the U.S. Bureau of Labor Statistics.

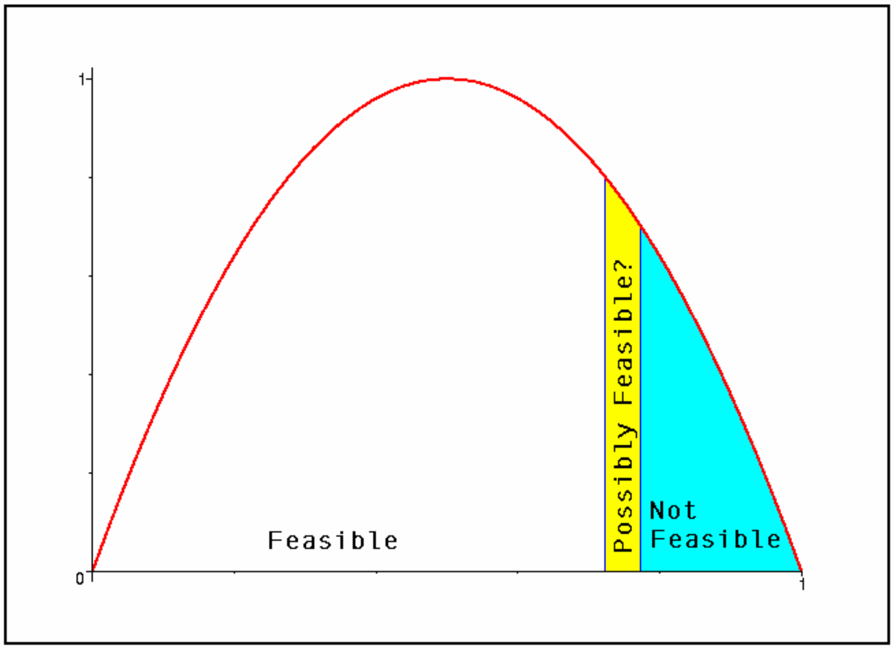

Figure 1: A Case in Which Constraints and Related Uncertainties Do Not Dominate Optimization

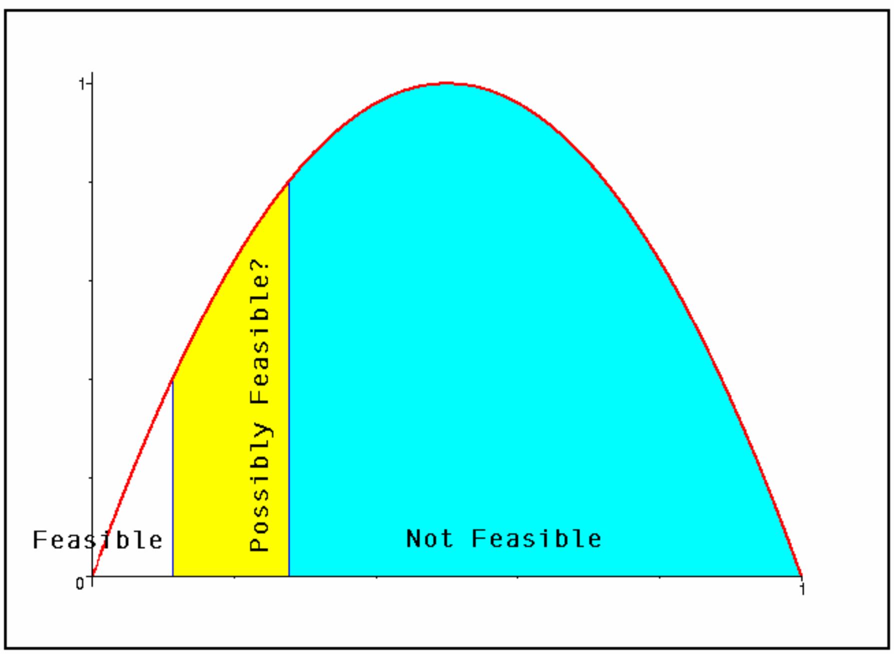

Figure 2: A Case in Which Constraints and Related Uncertainties Dominate Optimization

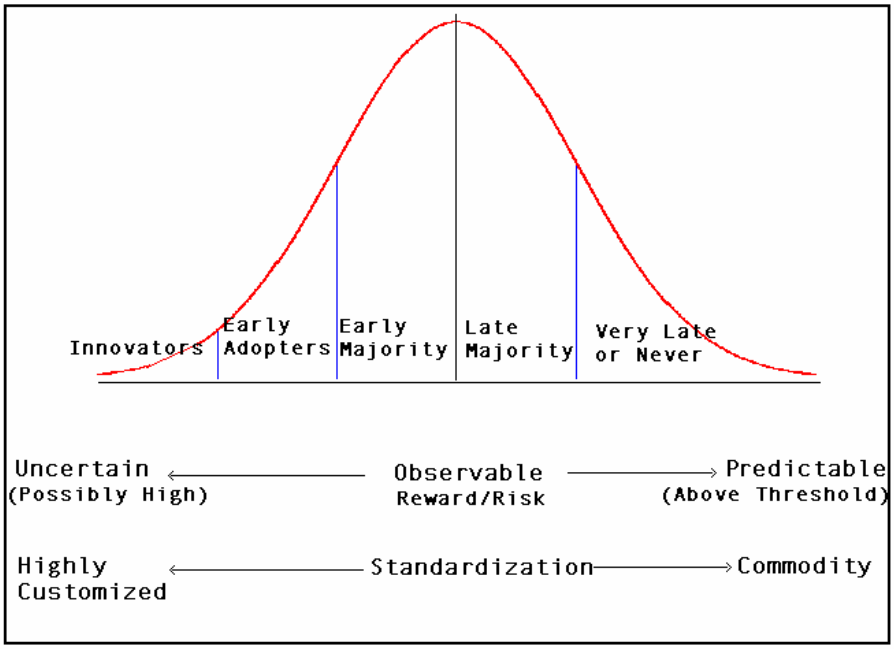

Figure 3: Customary Schematic Depiction
of the "Diffusion of Innovation"
(Adapted from Rogers, 1995 and Others)

References

Ahsan, J.J. and S.U. Khan (1982), Optimum Allocation in Multivariate Stratified Random Sampling with Overhead Cost. *Metrika* **29**, 71-78.

American Statistical Association (1994). A Research Agenda to Guide and Improve the Current Employment Statistics Survey. Report of the American Statistical Association Panel for the Bureau of Labor Statistics' Current Employment Statistics Survey.

Benjamini, Y. and Y. Hochberg (1995), Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Series B* **57**, 289-300.

Bonnen, J. (1988). Comment on, "Statistical Policy for State and Local Governments." *The American Statistician* **42**, 16-18.

Brackstone, G. (1999). Managing Data Quality in a Statistical Agency. *Survey Methodology* **25**, 139-149.

Butani, S., Harter, R. and Wolter, K. (1997). Estimation Procedures for the Bureau of Labor Statistics Current Employment Statistics Program. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 523-528.

Butani, S., Stamas, G. and Brick, M. (1997). Sample Redesign for the Current Employment Statistics Survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 517-522.

Cochran, W.G. (1977). *Sampling Techniques, Third Edition.* New York: Wiley.

Dillman, D.A. (1996). Why Innovation Is Difficult in Government Surveys (With Discussion). *Journal of Official Statistics* **12**, 113-197.

DuMouchel, W. (1999). Bayesian Data Mining in Large Frequency Tables, With an Application to the FDA Spontaneous Reporting System (with discussion). *The American Statistician* **53**, 177-202.

Duncan, J.W. (1988). Comment on, "Statistical Policy for State and Local Governments." *The American Statistician* **42**, 21-22.

Eckler, A.R. (1972), Some Comments on "Federal Statistics" (Report of the President's Commission on Federal Statistics – Volume I). *The American Statistician* **26**, 11-14.

Eltinge, J.L., J. Fields, R.C. Fisher, J. Gershunskaya, P. Getz, L. Huff, R. Tiller and D. Waddington (2001). Small Domain Estimation in the Current Employment Statistics Program. Paper presented to the Federal Economic Statistics Advisory Committee, June, 2001.

Fay, R.E. and R. Herriot (1979), Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association* **74**, 269-277.

Felligi, I.P. (1996). Characteristics of an Effective Statistical System (with discussion). *International Statistical Review* **64**, 165-197.

Fuller, W.A. (1999). Environmental Surveys Over Time. *Journal of Agricultural, Biological and Environmental Statistics* **4**, 331-345.

Genovese, C. and L. Wasserman (2002), Operating Characteristics and Extensions of the False Discovery Rate Procedure. *Journal of the Royal Statistical Society, Series B* 64, 499-517.

Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: an appraisal (with discussion). *Statistical Science* 9, 55-93.

Grzesiak, T.J. and C. Johnson (1989), Sample Allocation Meeting Multiple Estimate Reliability Requirements. *BLS Statistical Notes, Number 29.*

Hansen, M.H., W.G. Madow and B.J. Tepping (1983), An Evaluation of Model-Dependent and Probability-Sampling Inferences in Sample Surveys (With Discussion). *Journal of the American Statistical Association* **78**, 776-807.

Harris, L. (1972). A Note on the Decision Process of the Official Statistician. *The American Statistician* **26**, 23-24.

Harter, R.M., K.M. Wolter and M. Macaluso (1999), Small Domain Estimation of Employment Using CES and ES202 Data. *Statistical Policy Working Paper30: 1999 Federal Committee on Statistical Methodology Research Conference: Complete Proceedings*, 195-205.

Hartley, H.O. (1974), Multiple Frame Methodology and Selected Applications. *Sankhya, Series C* **36** 99-118.

Jenkins, R. and R. Chapman (1998). The Process of Adoption. *Proceedings of the 1998 Winter Simulation Conference*, 1547-1554.

Kaysen, C., C.C. Holt, R. Holton, G. Kozmetsky, H.R. Morrison and R. Ruggles (1969), Report of the Task Force on the Storage of and Access to Government Statistics. *The American Statistician* **23**, 11-19.

Kish, L. (1976). Optima and Proxima in Linear Sample Designs. *Journal of the Royal Statistical Society, Series B* **139**, 80-95.

Lehnen, R.G. (1988). Statistical Policy for State and Local Governments. *The American Statistician* **42**, 10-16.

Marker, D.A. (1999). Organization of Small Area Estimators Using a Generalized Linear Regression Framework. *Journal of Official Statistics* **15**, 1-24.

Marker, D.A. (2001). Producing Small Area Estimates From National Surveys: Methods for Minimizing Use of Indirect Estimators. *Survey Methodology* **27**, 183-188.

Moser, C. (1976). The Role of the Central Statistical Office in Assisting Public Policy Makers. *The American Statistician* **30**, 59-67.

National Research Council (2000a). *Small-Area Estimates of School-Age Children in Poverty: Evaluation of Current Methodology*. Panel on Estimates of Poverty for Small Geographic Areas, Constance F. Citro and Graham Kalton, editors. Committee on National Statistics. Washington, DC: National Academy Press.

National Research Council (2000b). *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 Academy Press.

National Research Council (2001). *Principles and Practices for a Federal Statistical Agency, Second Edition.* Committee on National Statistics. Margaret E. Martin, Miron L. Straf and Constance F. Citro, Editors. Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Neumann, K. (1999), Applied Systems Analysis in Official Statistics. *Journal of Official Statistics* **15**, 103-114.

O'Hare, W. and K.M. Pollard (1998). Assessing the Devolution Revolution: How Accurate Are State-Level Estimates from the Current Population Survey? *Population Research and Policy Review* **17**, 21-36.

Parke, R., C. Taeuber, J.H. Aiken, A.D. Biderman, B. Clyman, D. Greenwald, J.F. Kantner, R.E. Lewis, H.C. Passer, M.D. Wann, R. Loewenstein and F.C. Leone (1976), Improving the Federal Statistical System. *The American Statistician* **30**, 154-158.

Platek, R., J.N.K. Rao, C.-E. Sarndal, and M.P. Singh (1987). *Small Area Statistics.* New York: Wiley.

Rao, J.N.K. (2002). *Small Area Estimation.* New York: Wiley.

Renner, R.M. (1979), Estimation of Company Income Statistics. *The New Zealand Statistician* **14** (3), 19-30.

Reynolds, R.T. (1988), Comment on, "Statistical Policy for State and Local Governments." *The American Statistician* **42**, 18-20.

Rogers, E.M. (1995). *Diffusion of Innovations, Fourth Edition*. New York: The Free Press.

Rosenberg, H.M. and G.C. Myers (1977), State Demographic Centers: Their Current Status. *The American Statistician* **31**, 141-146.

Ryan, B. and N.C. Gross (1943), The Diffusion of Hybrid Seed Corn in Two Iowa Communities. *Rural Sociology* **8**, 15-24.

Savage, I.R. (1985), Hard-Soft Problems. *Journal of the American Statistical Association* **80**, 1-7.

Sackrowitz, H. and E. Samuel-Cahn (1999). *P* Values as Random Variables – Expected *P* Values. *The American Statistician* **53**, 326-331.

Schaible, W.L. (2000). *Indirect Estimators in U.S. Federal Programs.* Lecture Notes in Statistics, Volume 108. New York: Springer-Verlag.

Schirm, A.L, A.M. Zaslavsky and J.L. Czajka (1999), Large Numbers of Estimates for Small Areas. *Statistical Policy Working Paper 30: 1999 Federal Committee on Statistical Methodology Research Conference: Complete Proceedings*, 205-211.

Shen, W. and T.A. Louis (1998). Triple-Goal Estimates in Two-Stage Hierarchical Models. *Journal of the Royal Statistical Society, Series B* **60**, 455-471.

Shiskin, J. (1970), Strengthening Federal Statistics. *The American Statistician* **24**, 15-25.

Singh, A.C., D.M. Stukel and D. Pfeffermann (1998). Bayesian Versus Frequentist Measures of Error in Small Area Estimation. *Journal of the Royal Statistical Society, Series B* **60**, 377-396.

Smith, T.M.F. (1994), Sample Surveys 1975-1990: An Age of Reconciliation? (With Discussion). *International Statistical Review* **62**, 3-34.

Storey, J.D. (2002). A Direct Approach to False Discovery Rates. *Journal of the Royal Statistical Society, Series B* 64, 479-498.

Weiner, N.S. (1974). The Report of the President's Commission on Federal Statistics: A Summary of Recommendations, Views and Counterviews. *The American Statistician* 28, 42-46.

Werking, G. (1997). Overview of the CES Redesign. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 512-516.

West, S., Kratzke, D. and Grden, P. (1997). Estimators for Average Hourly Earnings and Average Weekly Hours for the Current Employment Statistics Survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 529-534.

Policy Considerations in the Development of State Estimates of Substance Use Rates
Doug Wright (SAMHSA)

The National Household Survey on Drug Abuse (NHSDA) is the primary source of information on the use of illicit drugs, alcohol, and tobacco by the civilian, noninstitutionalized population in the United States. The survey is sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA), and currently conducted by Research Triangle Institute under the direction of SAMHSA's Office of Applied Studies (OAS). Conducted since 1971, and annually since 1990, the survey collects data by administering questionnaires to a representative sample of the population through face-to-face interviews at their places of residence. Each year, the NHSDA interviews approximately 70,000 people age 12 years and older over a 12 month period. In 2000, SAMHSA published for the first time State estimates from the 1999 NHSDA using small area estimation techniques for the 50 states and DC. These were based on small area techniques that had been used on two different trial occasions, first using the 1991-1993 data combined, and later using the 1994-1996 data. That work demonstrated that the small area estimation methodology would work with a minimum sample size of approximately 400 persons per State. The 1999 NHSDA represented the first year in which the state sample sizes and allocation had been determined with the goal of making state-level estimates.

Prior to 1999, the NHSDA had a national design that utilized a first stage sample of approximately 120 counties (or groups of neighboring counties) and subsequent samples of block groups (either single blocks or groups of neighboring blocks), households, and persons. Sample sizes in the early 1990's ranged from about 18,000 to 28,000 respondents per year. In 1998, Congress requested SAMHSA to expand the NHSDA sample to provide State estimates. The purpose for the State estimates was to use them in conjunction with other available information to help determine those states having significant drug problems, the effectiveness of their programs, and the best allocation of block grant funds in order to reduce the national drug problem.[1]

The goals of the design for 1999 and later years were to provide state-level estimates of prevalence rates for approximately 20 measures for all persons 12 and older and separately for three age groups: 12-17, 18-25, and 26 and older. Approximately 900 cases were allocated evenly to the three age groups within each of the 42 states with smaller populations and the District of Columbia. In order to improve the estimates of precision at the national level, the remaining eight largest states were each allocated samples of approximately 3,600 cases. The total sample size was approximately 70,000 persons.

The conceptualization of the small area estimation is unique in many ways. Estimates are made at the block group level for every block group in the nation and summed up to the state level. The estimation method employs a full hierarchical Bayes approach that utilizes the sample weights, which are callibrated to approach the design-based estimate as the sample size increases. The model

---

[1]SAMHSA's Report to Congress: The Expansion of the NHSDA and SAMHSA's Plans to Improve Substance Abuse Services, Section III, page 3.

includes fixed and random effects which are estimated using Markov Chain Monte Carlo (MCMC) processes:

$$\lambda = \mathbf{X}\beta + \mathbf{Z}U$$

where $X\beta$ are the fixed effects and **ZU** are random effects.

The fixed effects include demographic variables from the Decennial Census at the block and tract level and county-level data related to substance use from a variety of other federal agencies. The random effects are estimated at the State and Field Interview (FI) region group levels. Each state has either 12 or 48 FI regions (geographic strata) depending on whether the State sample size is 900 or 3600. These regions are grouped into larger regions for estimating the random effects such that the small states (in terms of sample allocation) have 4 FI region groups and the large states have 16 such groups.

In order to obtain the best advice, an advisory panel of experts in small area estimation was constituted to advise the project by helping inform our review of the procedures and results, and the presentation and interpretation of those results.[2] This panel met once before the first release of the 1999 State estimates. Subsequently, it met a number of times to review and recommend how to publish the 2000 State estimates and what research to pursue to improve the precision of estimates of annual change at the State level.

The primary goal of state estimation was to rank the states from highest to lowest on a number of licit and illicit substance use measures and to measure the annual change at the state level. One of the crucial aspects of this project was to provide estimates of the precision of the small area estimates. Another was to validate the State estimates. Given the large sample sizes in the eight largest states, their small area estimates were heavily weighted toward the design-based estimates, and little weight was placed on the national model. However, for the other 42 states and the District of Columbia, in which the sample sizes were approximately 900, their precision rested more heavily on the fixed national model and validation was especially important for them.

The validation was based on combining large states so as to form four pairs of estimates. Thus, each of the four pairs of states had sample sizes of approximately 8,000 cases that could provide design-based estimates that were very accurate. The direct estimates for these quasi-states that were based on all 8,000 cases became the "true" values that would be compared to the small area estimates that were generated. To generate estimates that were comparable in design and process to our small area estimates for the smaller states, we exactly replicated the sampling process to produce "pseudo-

[2] The panel included William Bell of the U.S. Bureau of the Census, Partha Lahiri of the University of Nebraska, Balgobin Nandram of Worcester Polytechnic Institute and NCHS, Wesley Schaible formerly Associate Commissioner for Research and Evaluation at the Bureau of Labor Statistics, J.N.K. Rao of Carleton University, and Alan Zaslavsky of Harvard University.

states" with sample sizes of approximately 900 cases. Then the HB procedures were used to fit models and generate estimates for four substances for all persons age 12 and older, and for three age subgroups. We estimated the relative absolute bias for each measure. We also generated the 95% prediction intervals from the MCMC process and compared these to the 95% confidence intervals of the design-based estimates based on samples of the same size.

Results

Year 1. The results were released in two formats: tables of estimates alphabetically by state with their associated 95% prediction intervals and maps that reflected the ranking of states into quintiles. The states and DC were not ranked from 1 to 51 because there was significant bunching in the center of the distribution and prediction intervals were quite large (more so in 1999 when samples were based on a single year's data.). Quintiles were chosen to present the results. Since there were 51 estimates, 10 states were assigned to every quintile except the middle one which was assigned 11. On occasion, two or more States had identical estimates for a measure to the third decimal place or more. When this occurred at a boundary line between quintiles, those States were assigned to the lower quintile because it was desirable to "err" on the conservative side by assigning the states to a lower quintile.

The discussion in the report mainly focused on the highest and lowest quintiles: the States that had the highest or lowest prevalence rates. The maps of the states reflected the quintiles, assigning States in the highest quintile the color red and States in the lowest quintile the color white. Since States often had similar prevalence rates that were not statistically different, data users were encouraged to focus more on the prediction intervals for their estimates rather than the rankings themselves. Also, because the national average was more precise than any state estimate, the emphasis was placed on comparing States to the national average rather than to another state. An extensive technical appendix provided complete information on the validation, prediction intervals, interpretation of results, and sources of potential bias. Simultaneous with the data release, the governor's office in every State and the District of Columbia were sent the results in order to provide time to prepare for press inquiries.

In addition to discussing the states with the highest prevalence rates and the distribution of the top and bottom quintiles across the four Census regions, the discussion also covered the similarity of rankings for similar drugs, e.g. any illicit use and marijuana use, and the similarity across age groups. The 1999 HB estimates were quite good. Based on the validation, the relative absolute biases for the 12 and older age group ranged from a low of 1 percent for past month cigarette use to a high of 6 percent for past year cocaine use. In addition, the model-based prediction intervals were approximately 35% shorter than the corresponding design-based confidence intervals.

The NHSDA estimates have been useful in identifying states that have the highest and lowest prevalence rates, and raising questions about the possible reasons for those differences. For those states with low estimates, policy officials and researchers can ask what protective factors do those states possess and can their experience be transferred to the states with high prevalence rates. The rankings are also useful in identifying classes of drugs that have similar state rankings. The maps are useful in seeing the extent of regional similarity in the use of licit and illicit substances.

States are responsible for allocation of funds within the state. At this time, the only substate estimates that can be estimated from the HB process are those for the FI region groups where we estimate random effects. The FI region groups, however, don't necessarily match areas of interest to the States. At the state level, we have provided design-based estimates for a single year for those states with the large annual samples of approximately 3600 persons. For small states, we have combined two or more years to provide special design-based estimates.

Year 2. In the second year (2000), it was possible for the first time to estimate annual change at the state level using the HB methodology. However, evaluation of those estimates revealed that the changes were so small, and the prediction intervals so large, that there were almost no significant differences for any of the states for any of the measures. After meeting with our expert panel, it was agreed that we should not publish estimates of change based on just two years' data. Rather, we should combine two years' of data together in order to estimate a moving average and thus improve the precision of State estimates. Various options were considered for improving the estimates of change in the future, including estimating the difference of two consecutive moving averages, using retrospective estimates to improve precision of change, and simultaneously modeling two or more related variables. Another option discussed was to obtain better predictors of change in the HB model. Unfortunately, there do not appear to be any current national sources for county-level (or lower) data that reflect programmatic activity in the area of substance use prevention or treatment, or in other programs aimed at reducing substance use.

The 2-year estimates have been very precise with much less shrinkage toward the national model component, especially for the states with annual samples of about 900 persons. The rankings based on 2 years' data are very similar to those based on just the 1999 data. The relative absolute biases remained similar to those for 1999 for the 12 and older age group, ranging from about 1 percent for past month cigarette use and past month "binge" alcohol use, to about 8 percent for past year cocaine use. The prediction intervals based on 2 years' data were smaller than the corresponding prediction intervals based on a single year's data. For example, the prediction interval width for past month use of marijuana (persons age 12 and older) was 2.40 percent in 1999, but only 1.98 percent for 1999 and 2000 combined.

A public use file for the NHSDA has been developed; however, it does not include state identifiers for reasons of confidentiality - nor does it contain a linkage between sample respondents from the same household. The public use file is based on a complex disclosure method that utilizes subsampling and substitution subject to constraints that minimize the decrease in precision relative to the full file. At this time, we discussing the possibility of a license that would permit qualified users to conduct analysis with the full confidential file. We are also considering a data analysis system that might permit estimation of State-level crosstabulations, but provide no access to download the State identifiers.

Now in our third year of State estimation, we are obtaining our first estimates of change based on the difference of the 1999&2000 estimate from the 2000&2001 estimate. We believe that we may be starting to witness the emerging effects of prevention, treatment, and other programs aimed at reducing substance use, in that a few States that had high estimates in 1999 are beginning to show slight decreases in prevalence rates.

Response and nonresponse bias have always been a concern when collecting sensitive information. This is especially so given the varied response rates among states and the observation that there is a negative correlation across states between response rates and reported prevalence levels. Also, changes in the methodology, such as new field interviewer training to improve adherence to data collection standards and the use of monetary incentives to improve response rates, have apparently caused significant increases in prevalence rates, making it difficult or impossible to measure true year-to-year change net of any "field effects." We are currently studying these issues and the impact on our State small area estimation program.

Providing Small Area Estimates Discussion

Graham Kalton
Westat

Since the end of World War II the demand for survey data has experienced a continuous and ongoing expansion. In part, the expansion has been in the range of topics for which survey data are needed, and that has stimulated a number of methodological developments. Using newer data collection techniques, surveys are now used to collect data on topics—particularly sensitive topics—that in earlier years would have been considered beyond the realm of survey research. The National Household Survey of Drug Abuse (NHSDA)—the subject of Doug Wright's paper—is an example. In part, the expansion has also been in the sophistication of the demand for survey data. Whereas in the past policymakers would make do with often somewhat dated national estimates and estimates for a few large domains, their current demands are for timely data and for estimates for small domains. Some small domains are nongeographic subgroups, such as demographic domains (e.g., domains based on combinations of age, race, and sex) in population surveys and industrial division and size class in establishment surveys. For such domains, the production of small domain estimates of adequate precision may be achieved through increasing the survey's sample size and using methods to oversample the smaller domains. Another approach is to accumulate the sample over time, as is planned for the American Community Survey and is done in the NHSDA.

Other small domains are small geographic areas, such as states, metropolitan statistical areas, counties, and school districts. The expansion of sample required to produce reliable estimates even for states is often greater than resources can support. Moreover, the small area estimates of interest often relate to only a subdomain of the total population (such as the estimates for 12- to 17-year-olds in the NHSDA), in which case even larger sample sizes are needed. In such a situation the solution of accumulating sample over time may require too long a time period to satisfy the need for up-to-date estimates. Thus, alternative methods are needed.

The standard model for statistical inference in survey sampling is design-based inference. Design-based, or direct, estimates are not model-dependent, although they may be model-assisted. When a survey's sample size is inadequate to produce reliable direct estimates for small areas, it becomes necessary to employ indirect estimates. These indirect estimates are model-dependent, and there must therefore be concerns about model misspecification.

The essence of small area estimation is the use of auxiliary data available at the small-area level in a statistical model to predict the small area survey statistics of interest. The key requirement for this approach to be effective is the availability of good predictive auxiliary data. Such data can come from administrative records, a past census, or some other source. These data then need to be used in the careful development of a predictive model from which the small area estimates can be produced. An essential component of the approach is a thorough evaluation of the model and the estimates. Finally, valid measures of precision for the small area estimates need to be produced.

The idea of model-dependent estimation for producing small area estimates has a long history. An early example is to be found in the text by Hansen, Hurwitz, and Madow (1953, Volume I, pp. 483-486). Indirect estimates are now published on a regular basis by several Federal statistical agencies. A valuable review of eight Federal small area estimation programs is provided in the report of the Federal Committee on Statistical Methodology (FCSM) subcommittee on the subject (Schaible, 1993, 1996). Since that report was prepared, the U.S. Census Bureau has established its Small Area Income and Poverty Estimates (SAIPE) program and the Substance Abuse and Mental Health Services Administration (SAMHSA) is producing the state estimates of substance use described in Wright's paper. The Bureau of Labor Statistics is considering the introduction of small area estimates with its Current Employment Statistics (CES) program, and John Eltinge's paper discusses the issues involved.

The past decade has seen an explosion of theoretical research on small area estimation models and estimation methods, using in particular empirical best linear unbiased prediction, empirical Bayes, and hierarchical Bayes methods. A range of different models has been developed to cover dependent variables measured as categorical, continuous, or count variables and auxiliary variables measured at the area or unit levels. Also multivariate models have been developed to borrow strength for small area estimates for one subdomain from data for other subdomains and to borrow strength over time. Such methods are often computer intensive, using, for example, Markov Chain Monte Carlo methods. Current computing power and the availability of software, however, now make the application of these methods feasible (see Wright's paper for an example). Rao (2003) provides an excellent account of the current state of small area estimation methods.

While recognizing the importance of the recent theoretical developments, it remains the case that the model estimates can be no better than the auxiliary data on which they are based. Any small area estimation program should give a great deal of attention to finding auxiliary data sources, to checking the suitability of the data for use in models, and to constructing effective indices from the data for use in the models. The auxiliary data need to be measured uniformly across all small areas; alternatively, appropriate adjustments must to be made. The indices formed from those data need to be carefully constructed and thoroughly examined.

As an illustration, consider the estimation of poor school-age children for states and counties in the Census Bureau's SAIPE program. The total numbers of food stamp recipients, which are available monthly for states and annually for counties, are valuable auxiliary data, given that only poor households are eligible to receive food stamps. However, Alaska and Hawaii have higher income eligibility thresholds for food stamps than other states because of their higher costs of living. Thus adjustments need to be made to produce state and county numbers for Alaska and Hawaii that are comparable to the numbers of food stamp recipients in other states and counties in order to avoid distortion in the model-dependent small area estimates.

The indices constructed from the auxiliary data for use in a predictive small area model can significantly affect the quality of the small area estimates. In the initial formulation of the SAIPE state-level model, the food stamp index used in the model was based on the count of food stamp recipients in July of the reference year. Subsequent research led to a change to an index based on the monthly counts over a 12-month period centered on January 1 of the calendar year subsequent to the reference year. Also the counts were refined to remove persons who received food stamps due to specific natural disasters, and outliers were smoothed (Citro and Kalton, 2000, p. 28). As another example from SAIPE, evaluation of the initial county model estimates identified some distortion in the estimates in counties with high proportions of group quarters residents when an index based on the estimated number of persons aged under 21 was used in the model. This distortion was removed by changing the index to one based on the number of persons aged under 18, and this change also improved the model estimates in other respects (Citro and Kalton, 2000, p. 86).

This discussion draws attention to the importance of including a thorough evaluation of the model and the small area estimates in a small area estimation program. The SAIPE program provides a good illustration. In that program, the estimates of poor school age children have been evaluated by analyzing the regression residuals associated with some alternative models; by applying the models to the 1990 Census year and comparing the model estimates with Census estimates; by grouping counties on the basis of a variety of characteristics and comparing the SAIPE estimates with the direct estimates from the Current Population Survey for these groups; and by examining the stability of the models over time. In any such program imaginative ways to test the quality of the estimates should be sought. In planning a small area estimation program, considerable resources should be allocated to evaluation.

An important feature of small area estimation programs brought out in both Eltinge's and Wright's papers is that many different estimates may be needed for the small areas, not just the single estimate that is the focus of most theoretical work. Both papers point out the demand for estimates of both level and change, and that demand needs to be reflected in the modeling. To serve both these demands adds to complexity; a multivariate model is likely needed to produce valid estimates of change.

In the terms used by Rogers with regard to the diffusion of innovations, as discussed in Eltinge's paper, I would classify the present state of small area estimation in Federal statistics as being "early majority." The use of small area estimation methods is fairly well established in a number of areas, but I think that much greater use will likely be made of these methods in the future. Those who plan to develop new small area estimation programs need to appreciate the resources required and the properties of the resultant estimates. In particular, specialist skills are needed in the high-powered statistical methods that are used and expertise is needed in the auxiliary data sources and their properties. (The FCSM might play an important role in this area by facilitating the exchange of expertise in methods and data sources between Federal statistical agencies with small area estimation programs.) Careful model development and thorough testing are needed, and these are labor-intensive activities. Furthermore, the acquisition and checking of auxiliary data, model development, and testing are time-consuming activities that may seriously affect the ability to produce timely estimates.

The producers of small area indirect estimates need to make users aware of the model-dependent nature of the estimates and of the distinction between such estimates and the usual direct survey estimates. Small area estimation methods can be extremely valuable in addressing users' needs for such estimates. However, these methods should not be viewed as quick, easy, and inexpensive. Small area estimates need to be produced with great care and assessed with caution.

References

Citro, C.F. and Kalton, G. eds. (2000). *Small-Area Estimates of School-Age Children in Poverty. Evaluation of Current Methodology*. Panel on Estimates of Poverty for Small Geographic Areas, National Research Council. National Academy Press: Washington, D. C.

Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953). *Sample Survey Methods and Theory*. Wiley: New York.

Rao, J. N. K. (2003). *Small Area Estimation*. Wiley: New York.

Schaible, W. L. ed. (1993). *Indirect Estimators in U.S. Federal Programs.* Statistical Policy Working Paper 21. National Technical Information Service: Springfield, Virginia.

Schaible, W. L. ed. (1996). *Indirect Estimators in U.S. Federal Programs.* Springer-Verlag: New York.