Journal of Official Statistics, Vol. 24, No. 4, 2008, pp. 507–512

Discussion

Nathaniel Schenker[1] *and Trivellore E. Raghunathan*[2]

It is an honor for us to contribute to the collection of articles and discussions based on the Morris Hansen Lectures, commemorating one of the heroes of our field. It is also a pleasure for us to discuss this thoughtful article by Joe Sedransk, who has contributed to our profession in a number of ways, such as via the development and application of Bayesian methods for inference about finite population quantities.

In Section 1 of this discussion, we highlight one of the potential benefits of Bayesian inference. In Sections 2 and 3, we place Sedransk's article in the context of some Hansen-related literature, in particular the famous article by Hansen, Madow, and Tepping (1983), and we focus on the need to use models that are attentive to features of the sample design. We are flattered that Sedransk chose to include a discussion of Raghunathan et al. (2007) as part of his article, and in Section 4, we respond briefly to some of his comments. Finally, in Section 5, we mention a few areas for further research to supplement those discussed by Sedransk.

1. Highlighting a Potential Benefit of Bayesian Inference

Sedransk mentions several potential benefits of Bayesian inference, such as avoiding the need for large-sample approximations, increasing efficiency, facilitating the use of complicated models, facilitating the incorporation of prior data, and foundational arguments. A related benefit, which is one of our favorites, is the relative ease with which answers can be obtained in complicated situations by using Bayesian methods. This benefit has been enhanced greatly by the development of computational techniques, such as Markov chain Monte Carlo methods, for generating draws from multivariate posterior distributions.

The problems of small-area estimation and combining information that were discussed by Sedransk in Sections 2.2 and 2.4 are nice examples of complicated situations in which Bayesian simulation methods facilitated making inferences. Moreover, simulation methods are extremely useful for obtaining answers in missing-data problems. For example, Faucett, Schenker, and Elashoff (1998) used Gibbs sampling to obtain Bayesian

[1] Senior Scientist for Research and Methodology, National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, MD 20782, U.S.A. Email: nschenker@cdc.gov. Schenker was a formal discussant of Sedransk's presentation of the 2007 Morris Hansen Lecture.
[2] Professor, Department of Biostatistics, School of Public Health, and Research Professor, Institute for Social Research, University of Michigan, Ann Arbor, MI 48109, USA. Email: teraghu@umich.edu
Acknowledgments: The authors thank William W. Davis, Jennifer H. Madans, and Van L. Parsons for helpful comments. The findings and conclusions in this discussion are those of the authors and do not necessarily represent the views of the National Center for Health Statistics, Centers for Disease Control and Prevention.

inferences for the cumulative effect of post-operative smoking on lung cancer patients when information about smoking behavior was obtained only intermittently. As another example, Raghunathan et al. (2001) used ideas similar to those underlying Gibbs sampling to develop a multivariate technique for multiple imputation with variables of different types (e.g., continuous, categorical, and count).

2. Placing Sedransk's Article in the Context of Some Hansen-Related Literature

In their well-known and important article, Hansen, Madow, and Tepping (1983) compared design-based and model-based modes of inference for sample surveys. They concluded basically that for descriptive inference from reasonably large, well-designed sample surveys, design-based inference is to be preferred, because it avoids errors due to model misspecification that are possible with model-based inference, and because it loses little efficiency relative to model-based inference. Hansen, Madow, and Tepping (1983) acknowledged, however, that model-based methods for sample surveys can be useful and important in the contexts of sample design, inference for small samples, inference in the presence of nonsampling errors, and situations in which inferences under a model are of intrinsic interest. The discussion and rejoinder of Hansen, Madow, and Tepping (1983) also highlighted the importance of incorporating sample design features in Bayesian model-based inference. We discuss this issue further in Section 3.

Previous Hansen Lectures comparing design-based and model-based modes of inference for sample surveys have agreed substantially with the conclusions of Hansen, Madow, and Tepping (1983). See the articles based on Hansen Lectures by Smith (1994), who gave a philosophical discussion of the comparison, and Kalton (2002), who discussed various uses of models in the context of sample surveys.

In many ways, the current article by Sedransk is consistent with the findings of Hansen, Madow, and Tepping (1983) as well. This is not meant to imply that Sedransk has expressed agreement with those findings. Indeed, he clearly has an affinity for Bayesian inference. Rather, with one possible exception to be noted later, the examples used by Sedransk to illustrate the value of Bayesian methods fall into categories in which Hansen, Madow, and Tepping (1983) concluded that model-based methods can be useful.

One of Sedransk's examples is survey design (Section 2.3), which was mentioned earlier as an area where Hansen, Madow, and Tepping (1983) acknowledged that model-based methods can be useful. An interesting question that arises in this context is the following: "After Bayesian methods have been used at the design stage to formulate an optimal design, does Bayesian inference add a great deal at the analysis stage?" We suspect that the answer might be "no," as long as the models underlying the design are good, the design-based estimators are nearly optimal for the chosen design, and the data do not suffer from deficiencies such as small sample sizes or nonsampling errors. For example, if the survey design uses prior knowledge to determine inclusion probabilities that are roughly proportional to a survey outcome, it will be hard to beat the traditional Horvitz-Thompson estimator of the population mean for that outcome. If the inclusion probabilities are not roughly proportional to the outcome, however, a model-based approach may be able to provide a better estimator. Such considerations could be one reason why Hansen, Madow, and Tepping (1983) concluded, as mentioned earlier, that design-based inference loses little efficiency relative to model-based inference for reasonably large, well-designed sample surveys. Designing a sample well involves incorporating prior knowledge about the population into the design, which in a sense makes it a model-based endeavor.

Two of Sedransk's other examples that are obviously in areas where Hansen, Madow, and Tepping (1983) considered model-based inference to be appropriate are inferences for small subpopulations (Section 2.2) and pooling data from several sources (Section 2.4). These examples describe problems in which the objective is to deal with small sample sizes as well as certain nonsampling errors.

A final example from Sedransk's article that is consistent with the findings of Hansen, Madow, and Tepping (1983) is the use of Bayesian methods for provider profiling (Section 4). In a sense, this is a problem in which inferences under a model are of intrinsic interest, since provider profiling seeks to identify hospitals that are outliers under a specified model. Another reason why this example does not contradict the findings of Hansen, Madow, and Tepping (1983) is that provider profiling often is not a problem of sample surveys. Rather, the entire population is being analyzed. Thus, although the example still has a place in a discussion of inference about finite-population quantities (part of the title of Sedransk's article), it is an example in which design-based inference might not even be applicable.

The one example in Sedransk's article that might be considered counter to the findings of Hansen, Madow, and Tepping (1983) is that of inference for establishment surveys (Section 2.1). In fact, the main example used by Hansen, Madow, and Tepping (1983) to illustrate pitfalls of model-based inference was one in which they found that data on establishment surveys appeared consistent with a model but the model-based analysis was actually inaccurate. As discussed by Sedransk, in his example and many other situations with establishment surveys, the models are plausible and relatively simple and can be evaluated using vast amounts of census data. Thus, they may very well be accurate and appropriate for inference. Another positive characteristic of Sedransk's models for establishment surveys is that they incorporate features of the sample design. The use of models that are attentive to such features will now be discussed.

3. Using Models That Are Attentive to Features of the Sample Design

In his discussion of Hansen, Madow, and Tepping (1983), Rubin (1983) pointed out the importance of conditioning on the probabilities of selection in models for Bayesian inference from sample surveys, noting that this point is often misunderstood. In their rejoinder, Hansen, Madow, and Tepping (1983) agreed with Rubin's point. Other articles have made similar points. For example, Little (2004) wrote that "models need to properly reflect features of the sample design . . . or else inferences are likely to be distorted." See also Little (1982) and Reiter, Raghunathan, and Kinney (2006).

As noted in the previous section, the model for Bayesian inference from establishment surveys presented by Sedransk in Section 2.1 does incorporate features of the sample design. Specifically, the parameters of the model are allowed to vary by sampling stratum. This was not the case in the example given by Hansen, Madow, and Tepping (1983), although they argued that the data were consistent with their model.

The models in the other two examples presented by Sedransk concerning Bayesian inference for sample surveys incorporate the sample designs as well, at least to some extent. Specifically, for the problem of inference for small subpopulations discussed in Section 2.2, Malec et al. (1997) used a model that included random effects for counties as well as some covariates that were related to the sample design. Moreover, they examined whether other covariates, such as the sampling weights, added to the model, and they concluded that the additional information was negligible. For the problem of pooling data from several sources discussed by Sedransk in Section 2.4, Raghunathan et al. (2007) incorporated features of the sample design into their model by using design-based point estimates and variance estimates as a starting point, as discussed further in the next section.

4. Brief "Rejoinder" to Comments on Raghunathan et al. (2007)

As we mentioned in the introduction, we are flattered, as discussants of Sedransk's article, to be in the unusual position of providing a brief "rejoinder" to his comments on our own article (Raghunathan et al. 2007). This article is actually just one product of a joint project involving the University of Michigan, the University of Pennsylvania, the National Cancer Institute, the National Center for Health Statistics, the National Center for Chronic Disease Prevention and Health Promotion, and Information Management Services, with the goal being to develop small-area estimates of the prevalence of cancer risk factors and screening based on combined data from the Behavioral Risk Factor Surveillance System (BRFSS) and the National Health Interview Survey (NHIS).

In the first stage of the modeling effort, described in Raghunathan et al. (2007), for combining information from the BRFSS and the NHIS, we specified a model for the county-level, design-based prevalence estimates. In the distributions specified for these estimates, we incorporated design-based estimates of variance. Sedransk has suggested the possibility of modeling the person-level microdata at the first stage rather than beginning with county-level, design-based estimates.

An advantage, related to the discussion in Section 3, of starting with county-level estimates was that it provided a simple way to account for the within-county sample designs (e.g., clustering within secondary sampling units) as well as differences between the sample designs of the surveys involved. Such differences occurred not only because we were using two surveys, the BRFSS and the NHIS, but also because the BRFSS is actually a conglomeration of surveys conducted by the individual states, and the state-level surveys can have differing designs and post-survey adjustments as well. We feel that incorporating features of the sample designs and accounting for differences in the various designs would have been much more difficult had we been modeling the person-level microdata in the first stage.

A second point is that, in the application described in Raghunathan et al. (2007), information on the small-area population distribution of person-level covariates would have been needed in order for gains in efficiency from modeling the person-level microdata at the first stage to be achieved. The situation is analogous to that of regression estimation of the mean from a sample survey, for which the precision of the regression estimator depends on the availability, for both the sampled and nonsampled cases, of covariate data that are predictive of the outcome. Indeed, Malec et al. (1997), for the application described in Sedransk's Section 2.2, restricted the covariates in their model to age, race, and sex categories so that reliable estimated distributions would be available at the county level. For the application of Raghunathan et al. (2007), however, because estimation was being carried out separately for households with and without telephones, the covariate distributions would have been needed at the county level by household telephone status. Thus, it would have been very difficult to obtain reliable estimated distributions even just for age, race, and sex.

5. A Few Areas for Further Research

We second Sedransk's call for more efforts at modeling microdata for population-based surveys. As he suggests, the methods developed and lessons learned from such efforts would be quite valuable.

In addition, we would like to see more research on the intersection between design-based and model-based inference. Such research would be useful for situations in which model-based methods are used to compensate for deficiencies in data when the ultimate analysis of interest is design-based. An example of such a situation is the use of model-based multiple imputation (Rubin 1987), with design features incorporated into the imputation model, to handle missing data, followed by a design-based analysis of the data completed by imputation. Research on the intersection between design-based and model-based inference would also be useful for situations in which the ultimate analysis of interest is model-based but the data being analyzed are from a complex sample survey. An interesting example, mentioned in Section 2.2 of Sedransk's article, is a situation in which the model-based analysis of interest involves clusters that are different from those used in the sample design. In some such cases, methods for accounting for the two sets of clusters in the analysis would be needed.

Another general area for research that is related to the one mentioned in the previous paragraph and also the discussion in Section 3 is the development of methods for incorporating features of complex sample designs into model-based analyses. Although some such methods are already in use, such as specifying random effects for clusters, including survey weights as covariates, etc., further work in this area would be helpful.

Finally, as pointed out by Rubin (1984), a model-based inference should be well calibrated to real-world events. Thus, further research on evaluating models from a design-based point of view, especially in the context of complex sample designs, would be useful.

6. References

Faucett, C.L., Schenker, N., and Elashoff, R.M. (1998). Analysis of Censored Survival Data with Intermittently Observed Time-Dependent Binary Covariates. Journal of the American Statistical Association, 93, 427–437.

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

Kalton, G. (2002). Models in the Practice of Survey Sampling (Revisited) (with Discussion). Journal of Official Statistics, 18, 129–161.

Little, R.J.A. (1982). Models for Nonresponse in Sample Surveys. Journal of the American Statistical Association, 77, 237–250.

Little, R.J.A. (2004). To Model or Not To Model? Competing Modes of Inference for Finite Population Sampling. Journal of the American Statistical Association, 99, 546–556.

Malec, D., Sedransk, J., Moriarity, C.L., and LeClere, F.B. (1997). Small Area Inference for Binary Variables in the National Health Interview Survey. Journal of the American Statistical Association, 92, 815–826.

Raghunathan, T.E., Lepkowski, J.M., Van Hoewyk, J., and Solenberger, P. (2001). A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models. Survey Methodology, 27, 85–95.

Raghunathan, T.E., Xie, D., Schenker, N., Parsons, V.L., Davis, W.W., Dodd, K.W., and Feuer, E.J. (2007). Combining Information From Two Surveys to Estimate County-Level Prevalence Rates of Cancer Risk Factors and Screening. Journal of the American Statistical Association, 102, 474–486.

Reiter, J.P., Raghunathan, T.E., and Kinney, S. (2006). The Importance of Modeling the Sampling Design in Multiple Imputation for Missing Data. Survey Methodology, 32, 143–150.

Rubin, D.B. (1983). Probabilities of Selection and Their Role for Bayesian Modeling in Sample Surveys. Comment on Hansen, Madow, and Tepping (1983). Journal of the American Statistical Association, 78, 803–805.

Rubin, D.B. (1984). Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician. The Annals of Statistics, 12, 1151–1172.

Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: Wiley.

Smith, T.M.F. (1994). Sample Surveys 1975–1990: An Age of Reconciliation? (with Discussion and Rejoinder). International Statistical Review, 62, 5–34.

Received April 2008