# DISCUSSION

Richard Valliant, U.S. Bureau of Labor Statistics
2 Massachusetts Ave. NE, Room 4915, Washington DC 20212

It is an honor to be a discussant in this memorial session for Morris Hansen. I worked with Morris, Joe Waksberg, and Ben Tepping for about five years at Westat learning the theory and practice of design-based survey sampling. Morris was a kind and generous man who always got the most from his staff by giving us challenging assignments and making it clear that he felt we could do the job regardless of its difficulty. At the end of his career, Morris was actively involved in the controversies about inference from finite population samples that remain some of the most interesting and fundamental issues in sampling. Fred Smith has also been an active participant in these controversies and is a most appropriate speaker for this session.

This paper appears to be a radical departure from previous positions taken by T.M.F. Smith. However, based on the variety of interpretations made by different readers of this paper, the author's current position is ambiguous. The conclusion of the paper contains the statements: "I now find the case for hard-line randomization inference based on the unconditional (randomization) distribution to be acceptable" and "The conclusions drawn from procedural inferences ... now seem to me to be particularly appropriate to official statistics where the objective is purely descriptive." These assertions may be interpreted as supporting a brand of randomization sampling theory in which models are eschewed entirely, but this interpretation appears to be more extreme than Smith intends.

Based on the oral presentation of the paper, the author seems to be granting a more narrow role to randomization inference related to the distinction he draws between "scientific" or "predictive" inference and "procedural" inference. Procedural inference would apply to the descriptive statistics published from most sample surveys. Predictive inference can be entirely model-based, but procedural inference, according to Smith, can be acceptably done using the unconditional randomization distribution. An example of where this thinking would lead in a problem that is fairly well understood, such as ratio estimation, might clarify his position. Is the model-assisted approach advocated in Särndal, Swensson, and Wretman (1992) in concert with his current thinking? As Smith notes, the approach there is to use models in the design and derivation of estimators, but to use the randomization distribution to compute statistical properties. Any system, including the popular model-assisted approach, that relies only on the randomization distribution for the inference step, and allows no role for some type of conditioning, will inherit many, if not all, of the problems associated with a purely design-based approach.

The inferential principles supporting the use of models are well-known and do not need to be repeated here. If a model were exactly correct, then the model should be used for inference, whether it be predictive or procedural. Since models are never exactly correct, the practical question is: how useful are they? Considerable evidence exists that ignoring an approximately correct model at the inference stage is a major error. This discussion contains a few comments on some non-technical objections to randomization theory, on problems in making conditional inferences, and finally, on the implications of using only the distribution associated with the sample design at the inference stage.

The origin of this new acceptance of randomization inference is, in part, the feeling that good models in the social science may be nonexistent or, at best, too weak to be used as the basis for inference. If we have no model, we have no likelihood (other than the uninformative one associated with randomization), and the appeal to general principles of inference to justify model-based theory loses its force. Certainly, in social science, models do not exist in the same sense as in the natural sciences. The extreme view, as summed up by Ziman (1978, p.171), is that

> "the behavioural sciences are cluttered with innumerable half-articulated speculative models ... that have never been subjected to critical validation. ... Many of the 'pictures' (models) in the minds of research workers and practitioners are sheer fantasy, contradictory in themselves and having no basis in reality."

Ziman, who is a physicist, had in mind models of human behavior rather than of social science survey data, but that position seems close to ones taken by critics of model-based inference in their less charitable moments.

Maintaining that models in the social sciences are too weak to be used for inference calls into question the logic of doing analytic work with survey data. This may go beyond what the author intends, but I expect that some readers will draw this conclusion. A great deal of effort has been expended recently on determining how to perform regressions, contingency table analyses, and related procedures on complex survey data. Surely, this effort has not all been feckless. If so, I will have to remove the fine book by Skinner, Holt, and Smith (1989) from my shelf.

One objection to a purely randomization-based approach is that, in my experience, samplers rarely select purely random samples. The vigorous defenses mounted for design-based theory have always seemed to me a bit disingenuous for that reason. Selection methods invariably involve systematic sampling from a sorted list or some other technique that deliberately restricts the configuration of possible samples and that is not amenable to randomization analysis. In fact, the traditional methods for analyzing the properties of systematic sampling are model-based. One might argue that units are listed in a random order or in an order unrelated to the variable of interest, but this is seldom really the case. During my years at Westat, I do not recall Morris Hansen ever selecting a simple random sample, even within strata. When selecting samples with unequal probabilities, we would also use systematic sampling. Joe Waksberg and I once actually did select a stratified simple random sample of hospitals, but I think if the client had not resisted, Joe would have used a systematic sample. The reasons for restricting the randomization are good ones, and are well known to practitioners, but the only hope of analyzing estimators under those conditions is through the use of models. Notable is the fact that simulation studies in the literature testing properties of estimators usually employ random selection methods to which randomization analysis does apply and avoid the systematic methods more common in practice.

One of my main, non-technical objections to purely design-based inference is that it has little pedagogical interest. This may seem a niggling complaint, but it has an effect on who specializes in sampling and, therefore, on survey practice. Sampling is full of interesting problems, but the randomization approach is not interesting to students. As both Cochran (1978) and Godambe (1978) noted, in discussing an early version of the HMT paper, graduate courses in sampling are not offered by many statistics departments, and, where they are offered, are not popular among students. Speaking as someone who attempts to hire students with some education in sampling, the situation has not changed much since 1978. Problems in traditional sampling texts are considered in an *ad hoc* or piecemeal approach with little in the way of unifying principles. Population structure is used in vague ways that are more artful than mathematical. In the hands of someone with the advanced intuition of Morris Hansen this is fine, but most of the rest of us need more specific guidance. The traditional approach leads, for example, to the ratio and regression estimators being presented in some texts without reference to the fact that they are motivated by models. Explicit appeal to models often simplifies problem formulation and certainly makes the subject more comprehensible to anyone who is a novice at sampling but has some acquaintance with the rest of statistics.

Strict use of the design-based approach leads to ignoring obvious cases where conditioning is appropriate. Artificial examples can be given where unconditional inference is obviously poor, but more subtle cases exist. One such example is the study of the ratio estimator by Royall and Cumberland (1981). Design-based analyses had given no indication of the conditional problems with the ratio estimator and estimators of its variance. The HMT study was meant to answer such criticisms of the design-based approach. In their simulation study, HMT generated a finite population from a model of the form $E(y|x) = \alpha + \beta x$. Two of the estimators of totals they included were the combined regression and ratio estimators. As shown empirically in Valliant (1987), the combined ratio estimator is conditionally biased in such a case, even in stratified sampling. HMT, on the other hand, felt that "stratification has satisfactorily controlled the biases of the regression and ratio estimators" (HMT 1983, p.783). They found no conditional biases because they only performed unconditional analyses. Related to this is the fact that, even when randomization confidence intervals give correct coverage probabilities over all samples, the coverage probabilities in identifiable subgroups may be quite poor (Holt and Smith 1979, Royall and Cumberland 1985).

The asymptotic approach described by Robinson (1987) does provide a way of doing design-based conditioning in some cases. In the post-stratification problem, for example, suppose that $\hat{\mathbf{T}}$ is the vector of estimated post-stratum totals and that $\hat{\mathbf{N}}$ is the vector of estimated numbers of units in the post-strata. In large samples, if $(\hat{\mathbf{T}}, \hat{\mathbf{N}})$ is approximately multivariate normal with respect to the design, then the conditional distribution of $\hat{\mathbf{T}}|\hat{\mathbf{N}}$ can be studied. The conditional mean of $\hat{\mathbf{T}}|\hat{\mathbf{N}}$ suggests a type of regression estimator which is discussed by Casady and Valliant (1992) in a paper being presented at this conference. Rao (1985) also discusses design-based conditioning but restricts himself to cases in which the sample size is random. Even in those cases, a finite sample theory for design-based conditioning is difficult if not impossible to implement in practice.

In the post-stratification problem conditioning on the vector of post-stratum sample sizes in simple random sampling or on the vector of estimated post-stratum sizes in more complex designs seems reasonable. In other situations, though, how to condition in a non-Bayesian way is far from a settled question. This is true not only in sampling but also in the rest of statistics. Kiefer (1977) presents methods for assigning conditional confidence coefficients to Neyman-Pearson-Wald (NPW) decision procedures by partitioning a sample space into a family of subsets. The confidence coefficient reported for a procedure varies depending on the subset into which the data fall. Kiefer's investigation was motivated by a variety of examples in which NPW procedures lead to disturbing, counter-intuitive results. The difficulties with standard procedures were, in part, remedied by conditional analyses, but, in a given problem, many partitions are possible that could be used for conditioning. Kiefer was unable to formulate procedures for uniquely determining partitions, and he and one of the paper's discussants, J. Wolfowitz, felt that, in general, this would be impossible. Similar problems carry over to model-based conditioning in finite population inference. Analogies to the examples presented by Kiefer can probably be concocted for finite population inference in which unconditional results are disturbing but how to condition using models is ambiguous.

A compromise position between the design- and model-based sampling theories is the model-assisted approach, which, to a large extent, is becoming accepted practice. This compromise approach is likely to be much better received for descriptive statistics published by governments than is a purely model-based approach. But, to quote Lovie (1990) in a review of a book on a unified theory of sampling: "if backstabbing, well-poisoning, tunnel vision and the like are as prevalent as is intimated in the book, the chance that such a compromise will survive does not look too good."

Aside from the personality conflicts of participants on each side of the issue, the compromise, model-assisted approach has technical difficulties. The last step of the approach, which computes statistical properties using the unconditional randomization distribution, ignores the possibility of conditioning, as noted earlier. Model-based analyses, which explicitly compute such quantities as variances with respect to a model, have elucidated situations where certain procedures do or do not work well. Even if a model is not used for inference, a model-based analysis can contribute to the improvement of design-based practice. The variance estimators derived by Royall and Cumberland (1978) for the ratio and regression estimators, for example, are robust within a certain class of models, but are also design-consistent under simple random sampling. There are many other examples where model-based thinking has clarified problems in survey sampling, including distribution functions, generalized variances, price indexes, small area estimation, estimation in the presence of outliers, and systematic sampling. In a number of these cases, existing estimators have been improved or new ones derived which have both good design- and model-based properties.

My own opinion is that randomization theory, unaided by the structure imposed by good models, is inadequate. One of the achievements of model-based theory is that it has pulled sampling under the same umbrella as the rest of statistics. How to incorporate the powerful techniques of mainstream statistical theory into the solution of our problems is much clearer with the model-based approach. Not only will this incorporation lead to better solutions, but it will also encourage more talented students to specialize in sampling theory.

## Additional References


Casady, R.J. and Valliant, R. (1992), "Conditional Properties of Post-stratified Estimators Under Normal Theory," *Proc. of the Sec. on Surv. Meth. Res.*, American Statistical Association, to appear.

Cochran, W.G.; Godambe, V.P. (1978), Discussion of "On Inference and Estimation from Sample Surveys," by M.H. Hansen, W.G. Madow, and B.J. Tepping, *Proc. of the Sec. on Surv. Meth. Res.* American Statistical Association, 95 and 92-93.

Kiefer, J. (1977), "Conditional Confidence Statements and Confidence Estimators," (with discussion) *Journal of the American Statistical Association*, **72**, 789-827.

Lovie, P. (1990), Review of *Unified Theory and Strategies of Survey Sampling* by A. Chaudhuri and J.W.E. Vos, Amsterdam: North Holland in *Journal of the Royal Statistical Society A*, **153**, 104-105.

Royall, R.M. and Cumberland, W.G. (1978), "Variance Estimation in Finite Population Sampling," *Journal of the American Statistical Association*, **73**, 351-358.

—— (1981), "An Empirical Study of the Ratio Estimator and Estimators of its Variance," *Journal of the American Statistical Association*, **76**, 66-77.

—— (1985), "Conditional Coverage Properties of Finite Population Confidence Intervals," *Journal of the American Statistical Association*, **80**, 355-359.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992), *Model Assisted Survey Sampling*, New York: Springer-Verlag.

Valliant, R. (1987), "Conditional Properties of Some Estimators in Stratified Sampling," *Journal of the American Statistical Association*, **82**, 509-519.

Ziman, J. (1978), *Reliable Knowledge: An Exploration of the Grounds for Belief in Science*, London: Cambridge University Press.