Survey Methodology, June 1990
Vol. 16, No. 1, pp. 81–86
Statistics Canada

Comments on Articles in the Special Section

MORRIS H. HANSEN[1]

These are excellent papers that I enjoyed reading. Three of these papers focus primarily on historical and current developments and to some extent looking to the future. The paper by Kish is focused on and is an effort to influence some important future developments. I will attempt to add a little clarification from my own personal history and point of view on the historical summaries, and a little perspective, again from my personal point of view, on Kish's proposal for rolling censuses to replace the more traditional censuses.

Rao and Bellhouse have given a compact but useful survey of sampling development. Their summary begins, after a few preliminaries, at about the time that I first began to participate in censuses and sample surveys, and their improvement.

Their survey is done about as well as can be accomplished in such a compact summary, without elaborating on details. However, I would like to provide a slightly different view than they present on the development of sampling with probabilities proportionate to size or to measures of size (PPS). They accurately indicate that we (Hansen and Hurwitz) developed the theory for PPS sampling with replacement as an approximation. We were unsuccessful in solving the problem of variance estimation with varying probabilities when sampling without replacement that was soon solved by Horvitz-Thompson and others. However, with possibly rare exceptions, we never proposed the use of or used sampling with replacement. In practice, we did PPS sampling without replacement, usually either by choosing two or more units from a stratum by a systematic sampling procedure with the units arranged in a random or systematic sequence, or by choosing one unit per stratum. Units that would have had high probabilities of selection were selected with certainty. We prepared estimates of aggregates and functions of these by weighting by the reciprocals of the probabilities, exactly as in what has come to be referred to as the Horvitz-Thompson estimator. The variance estimators resulted in moderate overestimates because they assumed sampling with replacement as a simplification. Ordinarily, we have not regarded moderate overestimates of variance as a serious concern. The ultimate cluster variance estimator was often used. This is a very simple approximate variance estimator that involves weighting (if subsampling has been used) within the first stage units up to the first stage unit level, and then computing the variance between such first-stage unit estimates (see Hansen, Hurwitz, and Madow, p. 257). Horvitz and Thompson provided the initial breakthrough in variance estimation when sampling more than one unit per stratum with varying probabilities.

Sampling with PPS had the advantages that Rao and Bellhouse briefly describe. In addition, its use was a great convenience in multistage sampling, with probabilities proportionate to measures of size at each stage up to the final. The probabilities at the final stage were often set to achieve uniform overall probabilities of selection of the elementary units.

I add one other comment on their paper with respect to jackknife variance estimation. They indicate that the jackknife variance estimators are known to be inconsistent for nonsmooth functions like quantiles, even in the case of simple random sampling. They might have said, especially in the case of simple random sampling of the elements that are the units of analysis. We have recently demonstrated empirically that variances of medians and (in this case)

[1] Morris H. Hansen, Westat, 1650 Research Boulevard, Rockville, MD, 20850, U.S.A.

of 10th and 90th percentiles can be well estimated with the usual ultimate cluster jackknife variance estimation procedure with multistage sampling in which two or more first-stage units or combinations of them are identified in a stratum (one dropped and the other doubled, to form a replicate). We hypothesize that jackknife worked well in these applications because each ultimate cluster associated with a first-stage unit contains a substantial number of elementary units in the sample. We anticipate that it would work equally well, although we have not demonstrated it, when the jackknife replicates are formed by another procedure often followed, in which a simple random (or stratified random) sample is divided into m simple random subsamples (or stratified random subsamples utilizing the same strata to the extent feasible), and dropping one subsample at a time.

Fienberg and Tanur have presented an interesting perspective on the influence of the institutional setting in which survey research has developed. I agree with their view that an improved understanding of the development of survey methods is achieved by an understanding of the institutions through which survey research and surveys are done. At least those survey developments in which I have participated have arisen largely out of the institutional setting, and the need and opportunity to solve problems that occurred in accomplishing programs of the institution. Again, I have comments on some of the details in the developments in which I was a participant.

Fienberg and Tanur properly indicate that the design of what is now known as the Current Population Survey or CPS (earlier known as the Labour Force Survey) had a key role in the evolution of sampling theory and its application that has influenced other developments. However, they incorrectly suggest that its principal origins were in the experimental Trial Census of Unemployment carried out in late 1933 and early 1934 as a Civil Works Administration (CWA) project in three cities. There is some confusion in their paper of the 1933-34 CWA trial census with the 1937 "Enumerative Check Census" that accompanied the 1937 "Unemployment Census". It was the latter that, as they mention, Dedrick, Hansen, Stouffer, and Stephan jointly worked on, and that was the progenitor of the CPS. The 1937 Unemployment Census was a national registration done through the Post Office. The Enumerative Check Census was taken by mail carriers in a national probability sample of postal routes - they took a complete census of each postal route in the sample. New concepts for measuring labour force and unemployment were developed and applied in it based on behavior in a prior week. It was also a first application of nationwide area probability sampling. Its purpose was to evaluate the 1937 national registration of the unemployed (as discussed in the accompanying paper by Barbara Bailar). That sample survey taught us much, and was the seed for the monthly Labor Force Survey, later to become the Current Population Survey. Again, I was an active participant. Bailar desbribes it well. Stock, Frankel, and Webb and others at the Work Projects Administration (WPA) also had a role in the design of the national registration and of the Enumerative Check Census. Those were the days of dire unemployment, and the need for a continuing measure was obvious and urgent.

With this experience Stock, Frankel, and Webb, along with their colleagues at WPA perceived the opportunity and need for a continuing survey. They initiated a monthly unemployment and labor force survey, introducing some imaginative concepts in survey design (but also some problems that needed later correction). The monthly survey was just getting well established when Pearl Harbor and U.S. entry into World War II occurred, and the needs for information were radically changed. Labor shortage rather than high unemployment became the problem. The WPA was no longer needed and was abolished, and the survey was transferred to the Bureau of the Census to become a labor force survey to measure especially war-time implications of labor force participation and employment. When the survey was transferred to the Bureau of the Census we perceived some problems in the original design and developed

solutions to them, which led to the introduction, among other things, of PPS sampling and other design innovations. These developments for the labor force survey (now the CPS with a much broader role) have had a substantial impact on sample methodology, and more important, on meeting the needs of the nation for up-to-date information, not only on labor force but on many other subjects - demographic, social, and economic.

Feinberg and Tanur might also have emphasized the remarkable consequences of bringing together census-taking and sampling, along with computerization and automated reading of position marks on census questionnaires. In modern censuses in the United States, beginning with the 1960 Census, the questionnaires used for collecting information from all households are relatively brief in content. The principal content of the censuses is now obtained through samples taken simultaneously with and as part of the census, and, of course, on an exceedingly large scale in order to produce useful data for perhaps 40,000 small areas. A related development was the introduction in the 1960 Census of self-enumeration methods. The decision to introduce self-enumeration was guided by the application of the response error model to which Feinberg and Tanur refer, and by associated research and experiments on response errors, and especially on the correlated response errors associated with the work of enumerators. These innovations were guided by large-scale experiments that were done prior to and as part of the 1950 Census and in later censuses as well as in separate experiments. Another contribution was FOSDIC (Film Optical Sensing Device for Input to Computers), a device for reading position marks designed by the Bureau of Standards at the Census Bureau's request, in response to Census Bureau needs to replace the massive key-punching effort in a census. A consequence of the innovations that were introduced was more timely results and generally more accurate censuses, as well as lower costs. The opportunities for progress arose in view of the problems of large-scale census taking, and how they might be solved with the application of sampling and self-enumeration, along with the remarkable advances made possible by the development and application of electronic computers and FOSDIC, in which the Census Bureau was a pioneer.

In the late 1930's, some of the top Census Bureau staff, as well as members of Congress, were reluctant to see sampling introduced into the work of the Census Bureau. Complete enumeration had been the tradition. The use of probability sampling in the 1937 enumerative check census associated with the national unemployment registration was an important factor in achieving the acceptance of sampling as a methodology appropriate to the Bureau of the Census, again as more fully told in the accompanying paper by Bailar. The 1940 population census was a pioneering effort in the application of sampling in the collection of supplemental items of information in a census. In this effort Deming and I worked as colleagues. I was working with Calvert Dedrick, and Deming with Philip Hauser, with effective consultation and advice from Fred Stephan, and we all worked as a team in developing this important milestone in the application of sampling.

I have little in the way of comments to add to the paper by Barbara Bailar. As the paper indicates, I was an active participant along with Bill Hurwitz and our colleagues, in the developments she describes so well. I do have a minor correction. Feinberg and Tanur correctly identify the 1951 paper on response error models by Hansen, Hurwitz, Marks, and Mauldin as the original publication on the model, which Bailar credits to a later (1960) paper by Hansen, Hurwitz, and Bershad. The later paper elaborated those results, and included empirical data from the application of the model in large-scale randomization experiments involving the random assignment of enumerators in the 1950 Census. Analysis of these results as summarized in the 1960 paper showed the substantial and striking impact on small area census statistics of correlated errors within the work of interviewers. Earlier memoranda containing the results reported in that paper, and associated studies, were the principal vehicles that led

to the use of self-enumeration as the procedure for collecting the principal content items in the 1960 Census. They also led to transferring the collection of much of the information to a large sample instead of a complete census, with substantial cost reduction implications, improved timing, and generally improved quality. Bailar's paper provides an excellent summary description.

I should note, in this connection, the remarkable contribution to these developments that came from Bill (William N.) Hurwitz. He and I worked as a team that was far more effective than the sum of our individual contributions. In addition, I cannot give enough credit to our colleagues that we recruited and helped to stimulate and to some extent train, and who became the backbone of developments in the Census Bureau in the application of sampling, quality control, and operational research methods to the successful design and conduct of samples and censuses in wide ranging subject areas. Leaders among these colleagues included Max Bershad, Joseph Daly, Leon Gilford, William Madow, Eli Marks, Harold Nisselson, Jack Ogus, Leon Pritzker, Joseph Steinberg, Benjamin Tepping, Joe Waksberg, Ralph Woodruff, and others. I often get much of the credit, but without Bill Hurwitz, especially, and our colleagues, it could not have occurred.

I should mention that we benefited greatly, also, from the participation and advice from a panel of statistical consultants, with Bill Cochran (William G. Cochran) as chairman, over the years from 1955 until I left the Bureau in 1968. Other principal members included Fred Stephan (Frederick F. Stephan) and Bill Madow (William G. Madow) for the full time period, and Ivan Fellegi from Statistics Canada, H.O. Hartley, and others for part of the time. All were exceedingly able. However, we did not look to them as experts whose advice would simply be sought and generally followed. Instead, we operated on an interactive basis. We discussed specific issues or problems as well as all phases of total survey design for a particular survey, experiment or census. We received much useful advice; they also learned from us.

The paper by Leslie Kish moves the emphasis from historical background and recent and current advances to proposals for taking censuses of the future, through the introduction of what he calls rolling censuses. He also describes rolling samples in various forms.

Each of the kinds of rolling samples that he discusses, with and without overlapping panels are, as he indicates, in use for various purposes at the present time, and his discussion of these does not propose anything new. I suppose he introduces them for generality and as a means of suggesting their potential relationship to a rolling census.

The particular rolling census he describes is a weekly sample, with the total population of housing units at each point of time subdivided into 520 subsamples, one to be covered each week over a 10-year period. Thus, the entire population of housing units would be covered in a decade except for new additions of housing units in samples that had already been covered earlier in the decade. If the procedure were continued over time, then at any point in time the aggregate of the 520 samples for the prior ten years would provide average census results, representing the average situation over the prior 10-year period. It is an interesting and imaginative proposal. However, there are also problems.

He suggests a rolling census without any overlap in the coverage in successive weeks or other periods, except after the full decade when it starts all over again. Such an approach would provide a large national cross section sample each week, as well as average or aggregate results for each month, each year, and for other periods. However, without any overlap in the samples, it will be a relatively crude instrument for measuring changes occurring in small areas from week to week, from month to month, or even from year to year. Overlapping samples might be introduced, as he indicates, but would add greatly to costs. Of course, changes can be measured with the proposed rolling samples, but without partially overlapping samples the result would be large sampling errors of estimates of change for small areas. Providing data

for small areas is a primary purpose of the Decennial Census. I believe that reliably measuring such changes may be as important as providing aggregate measures for points in time. While Kish recognizes this, he seems to dismiss it.

Undercoverage of the population would likely be a particularly serious problem with a rolling census. Because of the general recognition by the public of the need for censuses, along with the intense publicity that is feasible for a census, the completeness of coverage of the censuses has traditionally been much greater than that in even the best sample surveys (although coverage still remains a problem in the censuses). The problem of net undercoverage in sample surveys is quite general – even including the Current Population Survey in the U.S. which is often taken as a model. Public interest with continuing weekly publicity for a rolling census could not conceivably be maintained.

Another issue in my judgment is the likely high cost of such a system. Kish recognizes this, also, and then seems to dismiss it. While I have not seen any cost estimates, I would not be surprised that over a decade the rolling census would cost substantially more than the cost of taking complete censuses quinquennially, plus the cost of relatively large-scale monthly samples to provide measures of change and information on various subjects for states and large areas within most states. Moreover, I anticipate that quinquennial censuses would be easier to interpret and more useful by providing measures for small areas at points in time, or for short intervals of time, rather than providing average measures over periods up to ten years.

The Census Bureau, influenced, in part, by Kish's earlier recommendations for such a rolling census, and the desire to spread the workloads has come up with some proposed alternatives for consideration for taking a brief decennial census along with rotating censuses. They consider some alternative approaches to rotating censuses of whole states over a decade. It is an innovative proposal intended to spread the workload while avoiding the high cost of a rolling census such as described by Kish.

I am one who believes that a quinquennial census, along with ongoing large-scale current surveys, are well worth a substantial cost. However, I believe that if a rolling census were adopted, as proposed by Kish, overlapping samples should be used. A rolling census, even without overlapping samples, may cost considerably more than the cost of the current census program extended to include a quinquennial census. I question if it is worth the added cost, or that it has advantages over a quinquennial census plus substantial intercensal samples. I anticipate that the rolling census approach would yield less useful information than quinquennial censuses for most purposes because it would provide complete census counts only for averages over a 10-year period. Quinquennial censuses, along with sufficiently large current samples to provide relatively up-to-date information for large areas, along with other procedures for providing data for state, county, and perhaps also small area population estimates, seem to have advantages from a cost-benefit point of view.

Kish is to be commended for his efforts to solve some of the census problems by a radical new approach. However, to me, the rolling census does not appear to be the answer. Perhaps more effective utilization of administrative records can provide results that hold more promise, again along with current samples and a decennial, or, hopefully, quinquennial censuses. Perhaps the remarkable new computerized mapping and coding system (known as TIGER) developed by the Census Bureau for the 1990 Census holds much promise for improving census-taking, and for current sample surveys. In addition, incorporating the TIGER geographic coding into the major administration records systems might make them more accessible for population estimates and for other uses. Up-to-date maintenance of TIGER, along with a currently maintained address register, are hopefully to be included in the Census Bureau's future plans.

REFERENCES

HANSEN, M., HURWITZ, W., and MADOW, W. (1953). *Sample Survey Methods and Theory*. New York: John Wiley & Sons.

HANSEN, M. (1987). Some History and Reminiscences on Survey Sampling. *Statistical Science*, 2, 180-190.

OLKIN, I. (1987). A Conversation with Morris Hansen. *Statistical Science*, 2, 162-179.

DUNCAN, J.W., AND SHELTON, W.C. (1978). *Revolution in United States Government Statistics, 1926-1976*. U.S. Government Printing Office, Washington, D.C.

Survey Methodology, June 1990
Vol. 16, No. 1, pp. 87–88
Statistics Canada

Reply

J.N.K. RAO and D.R. BELLHOUSE

We thank the discussants, Hansen and Smith, for their useful comments.

Hansen provided important observations on the development of PPS sampling. He is correct in saying that Hansen and Hurwitz (1943) did not propose the use of sampling with replacement and that only for variance estimation they assumed sampling with replacement. Incidentally, Murthy (1967. p. 184) notes that Mahalanobis (1938) has referred to PPS sampling and the associated unbiased estimator of a total in the context of sampling plots for a crop survey.

Hansen also made some interesting observations on the use of delete-1 cluster jackknife variance estimator for nonsmooth functions like quantiles. It is now well-known that the delete-1 jackknife variance estimator of a quantile is inconsistent under simple random sampling. Empirical results in Kovar, Rao and Wu (1988) indicate that it is also inconsistent under stratified simple random sampling. It is also likely inconsistent under stratified cluster sampling if the subsamples from the clusters are small or if the intra-cluster correlations are significant. In Hansen's application the subsamples from the clusters are quite large and the intra-cluster correlations very small. In this case, the delete-1 cluster jackknife variance estimator may be well-behaved in view of Shao and Wu's (1989) result that the delete-d jackknife variance estimator, under simple random sampling, is consistent, provided $n^{1/2}/d \to 0$ and $n-d \to \infty$ as the sample size $n \to \infty$.

The method of dividing a simple random sample into m subsamples, each of size d say, and dropping one subsample at a time, as suggested by Hansen, is similar to Shao and Wu's delete-d jackknife except that they consider all $\binom{n}{d}$ subsamples in constructing the variance estimator. However, the delete-d jackknife variance estimator is likely to be more stable. Shao and Wu also consider balanced subsampling requiring only b subsets of size $n-d$, where b ($\geq n$) is the number of blocks in a balanced incomplete block design.

Smith provided some important observations on the foundational aspects of sample survey theory, in particular, on the importance of Ericson's (1969) work on Bayesian estimation of a total under exchangeable priors. In this connection, we note that equivalent results for the posterior mean and the posterior variance, under simple random sampling, were also obtained by Hartley and Rao (1968). A. Scott pointed out the similarity of the two approaches in his discussion of Ericson's paper. However, an advantage of the Hartley-Rao approach is that the inferences depend on the sample design, unlike Ericson's approach. Their approach also yields useful classical inferences. Rao and Ghangurde (1972) extended the Hartley-Rao results to stratified random samping, double samping with unknown strata sizes, the Hansen-Hurwitz method for handling nonresponse, and two-stage random sampling.

The GUT approach for inference, proposed by Smith looks very promising. We agree with Smith that the point estimators using the different approaches rarely differ very much in practice, and that the issue essentially reduces to the choice of a measure of uncertainty, as noted in our paper.

We also agree with Smith on the importance of measuring total survey error from ongoing surveys.

ADDITIONAL REFERENCES

MURTHY, M.N. (1967). *Sampling Theory and Methods*. Calcutta: Statistical Publishing Society, Indian Statistical Institute.

MAHALANOBIS, P.C. (1938). *Statistical report on the experimental crop census*, 1937. Indian Central Jute Committee.

RAO, J.N.K., and GHANGURDE, P.D. (1972). Bayesian optimization in sampling finite populations. *Journal of the American Statistical Association*, 67, 439-443.

SHAO, J., and WU, C.F.J. (1989). A general theory for jackknife variance estimations. *Annals of Statistics*, 17, 1176-1197.

Survey Methodology, June 1990
Vol. 16, No. 1, pp. 89–90
Statistics Canada

Reply

STEPHEN E. FIENBERG and JUDITH M. TANUR

We are grateful to Bob Groves and Morris Hansen for their insightful comments and to the editor of *Survey Methodology* for the opportunity to update our thinking in 1990 rather than waiting for 2040. Groves and Hansen make several important points; we shall attempt to react to them in turn.

We very much like Groves' summary to the effect that governments emphasizing service for the welfare of the populace demand more information about their services than do those pursuing other goals. Consistent with this thesis is the fact that the most substantial new national survey launched in the United States during the 1980s, a decade not noted for an emphasis by the federal government on expanding welfare services, was the Survey of Income and Program Participation, one of whose primary purposes has been to monitor the impact of government welfare programs on income and assets. Moreover, as the countries of Eastern Europe democratize and turn to the West for assistance in upgrading their statistical systems, including the development of infrastructures for the conduct of large scale surveys, we see additional support for such a thesis. Thus it seems to us that Groves shares our belief that the institutional bases for survey research shape the content and direction of such surveys. Whether they provided homes or incubators for the best and the brightest seems to us akin to the nature/nurture debate – more a framework for discussion than an either-or choice. Indeed, we agree with Groves that the purposes of the various sectors shaped their choice of tasks, at least in part. In line with his urging of a cross national perspective, however, we note that institutional roles differ across countries. For example, there has been a widely-held view in the United States that the Federal government should not be in the business of collecting survey data on subjective phenomena (*e.g.*, see Turner and Martin 1984, 31-39) – a quite different stance has been taken by the British government, especially in connection with its annual report, *Social Trends* (Turner and Martin 1984, p.4).

Groves suggests that the membranes between sectors (academic, commercial, and governmental) are less permeable than we suggest. Neither we nor he have collected systematic empirical evidence on this question, but we point again to our concept of bridging institutions which bring together representatives of the various sectors, for the interchange of ideas if not personnel. And we hasten to point out that Groves' own recent appointment to the position of Associate Director of the U.S. Bureau of the Census, as well as Hansen's movement from that position into the commercial domain back in 1968, indicate the value, if not the ease, of membrane crossing.

Groves indirectly speculates that we choose to focus on technological advances, longitudinal surveys, and cognitive aspects of surveys because these are our areas of interest and experience, and he suggests several other developments that are worthy of consideration. Of course he is correct in suggesting that we have focussed on the developments that fit with our interests, but surely technological advances as a topic subsumes Groves' first two additional areas of importance: (1) development of generalized statistical software packages and (2) existence of survey data archives. We wonder, however, if the technological advances we both note, coupled with the ubiquity of surveys that we also both note, do not have negative as well as positive consequences. For example, the complex analyses of survey data by undergraduates (or indeed

any beginners) using statistical software packages often show neither an understanding of the data being analyzed nor the appropriateness of the packaged statistical methods used.

The ubiquity of surveys is a consequence not only of the demand for information but also of the relative ease with which surveys can be carried out and the data analyzed given current technology. (And we believe that the availability of survey data for reanalysis will only increase with the advent and adoption of new storage technologies such as CD-ROM and optical disks). Such ease is a mixed blessing. As Groves notes, the 1980s have seen a growing problem of nonresponse in the United States, a pattern that manifested itself earlier and (so far) more seriously in Europe. We do not need to postulate a growing trend toward demands for privacy to explain this decline in response rates, though such a trend may well exist. We need only look at the major nonresponse problems currently being encountered in the conduct of the U.S. 1990 decennial census, in both the mail-out-mail-back and in the door-to-door phases, to see evidence to support the contention that respondents are merely getting tired of being surveyed so frequently.

Further, as Groves points out, survey research has not been central to the self-image of academe, because survey research has not fully evolved into a separate identifiable discipline, with specified standards and training criteria. Since there are no departments of survey research on university campuses, almost anyone who cares can mount a survey or carry out analyses of survey data. While some people do these tasks well, others do them poorly thereby giving the whole survey enterprise a bad name. Thus, if we are to present the optimistic report on the state of the survey enterprise in 2040 that Groves envisages, it seems to us that the innovations in education and training that neither he nor we are currently able to chronicle will have to become institutionalized.

We are especially pleased to have Hansen's embellishment on our brief account of the development of the survey enterprise in the U.S. government in the 1930s and 1940s. His comments supply some of the human drama that Groves says is lacking in our institutional focus.

Hansen also expands on our account of the link between censuses and sampling and the introduction of self-enumeration into U.S, censustaking, that was guided by the study of response errors. The major decline in completion rates for self-enumeration in the 1990 decennial census suggests the need to reexamine the implications of the various components in the Hansen-Hurwitz-Marks-Mauldin model for non-sampling errors. In addition we note that as part of the 1990 census, the Bureau of the Census will mount a new Post-Enumeration Survey (PES) of 150,000 households whose results will be used to evaluate census coverage. The technological advances in computerized data management and in computer-based matching of files between the PES and the census were essential ingredients to the launching of this major new government survey and its planned use to measure both under-and over-coverage of the household-based population.

ADDITIONAL REFERENCES

TURNER, C.F., and MARTIN, E. (eds.) (1984). *Surveying Subjective Phenomena. Vol. 1.* New York: Russell Sage Foundation.

Survey Methodology, June 1990
Vol. 16, No. 1, p. 91
Statistics Canada

Reply

BARBARA BAILAR

The comments from the discussants describe even more contributions of the Federal Government to the world of statistics. I am very grateful to Gordon Brackstone and Morris Hansen for mentioning these additional topics. The topic I omitted that may have had the biggest impact on statistics as well as other quantitative fields was the development of the computer for data processing and data analysis purposes. Again, the team of Hansen and Hurwitz were the prime movers, urging and funding the development of UNIVAC I and then bringing it into the Census.

Morris Hansen describes the remarkable team at Census who worked with him and Bill Hurwitz on so many topics. I feel very fortunate that I began my career at the Census Bureau when these people were there and that I was able to work with most of them for many years. It is rare that one gets that kind of apprenticeship.

Gordon Brackstone questions whether the statistical methodology developed by the Census Bureau had a benefit to the wider world of statistics. Certainly, given the amount of interaction among government statistical offices, the Bureau of the Census has influenced government statistical operations in other countries. Brackstone finds the impact of the Census Bureau development on university statistics departments rather mixed. He may be correct as far as course offerings are concerned, but I believe the ASA-NSF-Census Fellowship program and the Agriculture Fellowship program have had a big impact. More university professors and graduate students are aware of and working on non-sampling error, disclosure avoidance, and time series problems. The recent addition of Fellowship programs at the Bureau of Labour Statistics and the National Center for Education Statistics have also highlighted these research areas. The NSF now receives many proposals based on research started at one of the government agencies.

The main problem now is to make sure that research results are used. Many government programs are slow to accept new methodology because change is disruptive. Yet, to make sure that methods are improving, change is necessary.

Survey Methodology, June 1990
Vol. 16, No. 1, pp. 93–94
Statistics Canada

Reply

LESLIE KISH

In his fine discussion Fritz Scheuren complements our comparisons of alternative census methods by advocating administrative registers for the USA. I support his expert plea to study what these methods could offer as additions, as complements to the decennial censuses. They are coming to many countries and we would like to know where, when, and how? It is even likely that they will not only complement, but even replace decennial censuses soon in some places. When in the USA? I don't know; we were comparatively slow and late in adopting a successful registry of births and deaths. And even now their reporting is rather slow.

Rolling samples could be designed for quick reporting, and timeliness is only one of the advantages of rolling samples. Thus it is biased to compare rolling censuses with traditional censuses, both as regards costs and benefits, only on the basis of the single output for which decennial censuses are designed. It would take detailed, technical investigations to compare the factors of costs, coverage, timeliness, content, *etc.* of rolling versus decennial censuses in the USA. But 10 to 15 million dollars monthly can go far. The issue of adequate censuses is most salient in 1990 in the USA and elsewhere, but the other uses of samples should not be forgotten, as we plan for the last decade of our twentieth century.

My contribution aims mainly to advance the *diverse* advantages of cumulations from periodic samples, which have been neglected in favor of the other benefits that can be obtained from the growing numbers of periodic surveys. Rolling censuses may become someday one of those benefits, and rolling samples have been used already – though not often enough, I believe. Asymmetrical cumulations may exist rarely and obscurely, and the split-panel designs that I propose, not at all.

Furthermore my scope is not merely national (the USA), nor even continental (North America): it is intercontinental and international. For example, registers have come to the Nordic countries and they may come to Canada before the USA. Rolling censuses pose a much smaller expansion of the Labour Force survey in Canada because it is one-tenth the size of the USA, as Fritz and I both show. But some other country may well use them before either.

Not only international, rolling samples and cumulations are also aimed to be inter-disciplinary, not only for making population counts. Good many of the other needs of statistical offices – and *of other institutions for data collections*! – would be better served by a trained "permanent" staff than by a hurriedly hired huge army whose training time roughly equals their brief employment.

Scheuren is most complimentary when he calls rolling censuses a new paradigm. It is true that, as all new paradigms, they meet three big mental blocks when I present cumulations and rolling samples: a) averaging of variable data instead of an arbitrary date like April 1, of the decennial year; b) accepting some of the mobility of human populations instead of fixing them to unique sites; c) rolling samples to replace fixed primary sampling areas. So it may seem paradoxical when Morris Hansen notes that my "discussion of these does not propose anything new." Hansen may have encountered all of these proposals, and perhaps dismissed some of them. Personally I have described rolling samples since at least 1961 and proposed rolling censuses since 1965. But I also found that for many people they come as new ideas, and often as strange new ideas.

Finally let me only add two important origins in the '40's for sampling, although for me personalities and priorities are only minor aspects of the history of any science. Iowa State at Ames should be mentioned, where, under George Snedecor and Henry Wallace, Bill Cochran started in the spring of 1939 the first course of sampling and turned out pioneer MA's, then PhD's in sampling. Then Henry Wallace (again) in the US Dept. of Agriculture started the Division of Program Surveys, hired me in 1941 and Steve Stock in 1942 for the first national samples in Washington in 1942, followed by the 1943 sample at the USBC. Stock, Frankel and Webb (from the WPA samples) began the second sampling course in fall 1939 at the USDA graduate School, which became famous and productive under Hansen, Hurwitz and their Census staff. Among influential courses there I shall testify especially to those of Deming, the major figure at the school. The teaching and learning of samples in the forties was done mostly at Ames and in the USDA, as well as at the USBC.