Statistical Science
1987, Vol. 2, No. 2, 180–190

Some History and Reminiscences on Survey Sampling

Morris H. Hansen

Abstract. This is a personal account and reminiscences concerning some of the highlights of the history and important events in the development of survey sampling as perceived by the author. It gives special emphasis to the developments at the Bureau of the Census in the late 1930s through the 1960s, reviews early acquisition of Univac and related equipment at the Bureau of the Census and briefly discusses the more recent issue that has received attention concerning the role of probability sampling versus model-dependent sampling.

Key words and phrases: Bureau of Census history, Film Optical Sensing Device for Input to Computers, inference from samples, model-dependent sampling, probability sampling, Univac.

INTRODUCTION

When I joined the Bureau of the Census in 1935 the Bureau was in the beginning stages of a transition from a staid, old-line organization to an innovative organization that was to stimulate the development of and make significant contributions to various aspects of the theory and practice of designing and taking sample surveys and censuses. The Bureau had already done pioneering work in the development, construction and application of punched-card data processing equipment, but was relatively lethargic in other areas.

I was fortunate to join the Census Bureau staff at a time when opportunities were open, and as a young member of an evolving team that was prepared to create and take advantage of the opportunities. The New Deal, and shortly thereafter World War II, created major needs and opportunities for statistical development.

I will make a few brief remarks on the history of sample surveys prior to this time, and continue by summarizing a few of the highlights of early contributions to survey sampling and census methods and theory generally, including especially, those at the Bureau of the Census.

It is important to distinguish two distinct types of inferences based on sample survey and census results. The first is descriptive, that is, the aim is to describe the characteristics of a specified finite population. For this case a complete census is sometimes available or can be taken. If the census covers the desired subjects and is complete and accurate, it would be sufficient. Ordinarily, a complete census is not feasible, especially for providing current information on many studies and topics, and a sample survey is used to provide estimates of what would be obtained by a complete census. The second type of inference is concerned with the causes that produced certain characteristics of the population. Such problems may be particularly important but inference may be more difficult. The distinction between the two types of inference is sometimes not clearly recognized, especially since both may use the same data and the same or similar methodology. My discussion, except where specifically indicated otherwise, relates to the first of these two types of inference, that is, estimating the characteristics of a well-defined finite population.

SOME EARLY SAMPLE SURVEY HISTORY

There are many illustrations of early applications of sample survey methodology, both in Europe and in the U. S., during the nineteenth century, and even earlier (see Stephan, 1948; Kruskal and Mosteller, 1979; Hansen, Dalenius and Tepping, 1985). Many of the methods used today were conceived as desirable

---

Morris H. Hansen is Chairman of the Board, Westat, Inc., 1650 Research Boulevard, Rockville, Maryland 20850. *Much of the content of this paper was originally presented in the videotaped Pfizer Colloquium lecture given at the University of Connecticut on October 30, 1985. That videotape is in the American Statistical Association collection of lectures by distinguished statisticians and reproduction here from that lecture is with permission. The material assembled and published in earlier papers (cited in the references) by the author together with Tore Dalenius, William G. Madow and Benjamin J. Tepping provided background information and was drawn on freely for this paper.*

or useful on intuitive grounds and applied in some of these early surveys, but with little or no support and guidance from theory. The methods applied included stratification, cluster and multistage sampling, post-stratification and others. Kiaer, in an 1897 paper (see Kiaer, 1976, for a translation), gave a relatively sophisticated discussion of the principles and uses of such methods, including recognition of some issues and limitations, and presented recommendations to the International Statistical Institute concerning the use of sampling to acquire information.

Bowley supported Kiaer's recommendation, and in 1903 the International Statistical Institute (ISI) passed a resolution supporting the use of the "representative method." The ISI appointed a commission in 1924 that made a number of important recommendations in a 1926 report (Jensen, 1926). Bowley's 1926 memorandum in the commission's report emphasized the need for random selection of the sample. He also emphasized the need for some kind of a comprehensive list covering the population, either of elements or of clusters, from which to draw the sample, so that random or systematic procedures (treated as equivalent to random) could be used for sample selection. He gave the theory for proportionate stratified sampling, and emphasized that nonresponse may be a problem that needs attention (Bowley, 1926).

In the U. S., in the early and mid-1930s, the Great Depression and the New Deal programs, including especially the Federal Emergency Relief Administration and the Work Projects Administration (WPA), needed various types of information. Such needs stimulated and provided support for a number of sample surveys. A wide range of surveys was taken, often using systematic and random sampling methods sometimes within purposefully selected samples of large areas. Stephan (1948) reviews many of these surveys.

A NEW ERA: ROTHAMSTED EXPERIMENTAL STATION

A new era was introduced in the 1920s with the remarkable developments in statistical theory and practice by R. A. Fisher at the Rothamsted Experimental Station (Fisher, 1925). Among other things, Fisher stressed randomization, replication and local control in experimental designs. These principles led to important contributions by Yates and others at Rothamsted to the theory and practice of survey sampling in the mid-1930s (Yates and Zacapony, 1935; Cochran, 1939, 1940). They used randomization and replication which make the sample design self-contained; i.e. enabled estimation of the precision of the sample estimates solely on the basis of the information in the sample. Local control (stratification) was also stressed to reduce the magnitude of the sampling errors. Multistage sampling was introduced in the early agricultural sampling at Rothamsted, although in these early applications the theory assumed primary units that were of the same size. The analysis of variance provided variance estimates and estimates of the components of error. Criteria were also suggested for jointly considering and balancing cost and accuracy.

NEYMAN'S CLASSIC 1934 PAPER

In 1934, Neyman presented his now classic paper on "the two different aspects of the representative method" to a meeting of the Royal Statistical Society in London (Neyman, 1934). It was a remarkable paper in which he presented the first well-rounded discussion of inferences from samples of a finite population on the basis of randomization introduced by the sample selection procedures, that is, of what is now known as probability sampling. The paper also contains a comparative evaluation of purposive selection and random sampling. He critically assessed the basic assumptions which must be met if purposive selection is to give satisfactory results, and indicated that the regression of the variable to be estimated on the control variables must be linear (or nearly so) and concluded: "I think it is rather dangerous to assume any definitive hypotheses concerning the shape of the regression line."

The concept of confidence intervals was defined for the first time in this paper. Neyman discussed best linear unbiased and consistent estimators, and noted that the method of random sampling allows a consistent estimate of the average of a variable, X, whatever the properties of the population, and makes possible an estimate of the precision of the results obtained in the form of confidence intervals. He also discussed the use of ratio estimators and presented needed theory. He indicated that with large enough samples the distribution of the estimate is practically normal and thus there are no difficulties in calculating a confidence interval.

He emphasized that the important consideration is that confidence intervals be computed appropriate to the particular sample selection plan and estimation procedures used, without requiring that the estimation procedures be "best" in some sense or that the confidence intervals be the shortest possible. He noted that short confidence intervals are preferable to long ones, but that in practical terms it isn't necessary to seek the shortest.

Neyman provided the theory for optimum allocation of sampling units to strata in his 1934 paper, which he developed independently of Tschuprow, whose earlier result (in 1923) appears to have been overlooked at the time by statisticians generally.

In 1937, W. E. Deming arranged for Neyman to give a series of lectures at the U. S. Department of

Agriculture in Washington, D. C. One of these lectures dealt with sampling of human populations. In the course of this lecture, Milton Friedman and Sidney Wilcox presented a problem concerning what has become known as "double sampling" or "two-phase sampling." Neyman later considered the problem in some detail and developed and presented the theory of double sampling (Neyman, 1938). In that paper Neyman introduced explicit cost functions, in what was one of the forerunners of the joint use of both cost and variance functions to optimize sample design within a specified class of designs.

In the late 1930s and early 1940s, a few major centers emerged for research in and application of sampling methodology. One was the Statistical Laboratory at Ames, Iowa, another was the Bureau of the Census. At the same time major advances were occurring in India and at a few other locations. Before discussing work in the U. S., I will mention some developments in India.

INDIA

P. C. Mahalanobis organized the Indian Statistical Institute in the 1930s. The Institute contributed to the development and application of survey sampling to provide support for social and economic planning in India. Mahalanobis advocated the view that it was imperative not merely to imitate procedures in advanced countries, but to adjust methods to the prevailing conditions in India. This is the appropriate approach in sampling; that is, the methods used, in addition to focusing on the particular goals to be served, should identify and make use of relevant available resources. He developed a philosophy of "statistical engineering," and carried out designs in experimental stages, over several years, developing and basing designs on cost and variance estimates. Among other things, he introduced a program of interpenetrating samples to help control and evaluate contributions to total survey error (Mahalanobis, 1946).

Sukhatme, in the Indian Council of Agricultural Research, also did extensive work on agricultural sampling. He made various theoretical and practical contributions, including some on stratified sampling. He also made important contributions concerning the control of nonsampling errors by showing empirically that small plots, as used as sampling units by Mahalanobis for measuring crops, were vulnerable to boundary biases resulting from the tendency to include in a plot plants on the boundary that, in fact, should not be included. These biases led to overestimates that varied inversely with the size of the plot, and the biases in yield estimates could be substantial if small plots were used (Sukhatme, 1946).

Sukhatme and Mahalanobis disagreed strongly on some of the practical statistical activities, and they carried on a more or less continuing controversy. Later, Sukhatme went to the Food and Agriculture Organization in Rome where he continued to extend methods and theory on sampling and measurement errors in surveys.

Mahalanobis visited us at the Bureau of the Census shortly after World War II and we interchanged ideas and exchanged experiences. Mahalanobis made the observation which was surprising to me initially but with which I fully agreed after learning more of his work, that the "statistical engineering" activities, as he referred to them, at the U. S. Bureau of the Census and those in his organization in India, were the only two really similar large scale sample survey programs in the world. Both were concerned with total survey design, based on empirical studies, and covering not only statistical but operating problems and procedures. Nevertheless, both based designs on a strong theoretical foundation, and the need to develop theory to meet new problems that were encountered. Mahalanobis had great influence in India and was a close adviser to Nehru. I first visited India in 1949 at which time he arranged for us to have lunch with Nehru.

I had the pleasure of making a second visit to India at Mahalanobis' invitation in 1956 to serve on a review committee for the National Sample Survey. R. A. Fisher was chairman, and the committee included Frank Yates, Tosio Kitagawa and others. It was a remarkable experience. I believe that Frank Yates and I were the only ones willing to challenge Fisher on some positions, which he seemed to take to support Mahalanobis on possibly political rather than technical considerations. At these meetings we also had the opportunity to meet Chou-en Lai, then Premier of China, whom Mahalanobis had invited to visit the Indian Statistical Institute while he was visiting India. We both spoke to the same meeting.

The Indian Statistical Institute, which had a strong statistical staff, was an academic and training center as well as an operating statistical organization, taking the national sample survey and doing other statistical projects.

Professor Mahalanobis retained his vigor and charm in communicating his ideas until his death in 1971.

THE DEPARTMENT OF AGRICULTURE AND THE STATISTICAL LABORATORY AT IOWA STATE UNIVERSITY

Surveys of crop acreages and yields were being taken by the Department of Agriculture by the middle of the nineteenth century, using relatively crude sampling and estimation methods. Advances using regression were introduced later, and extended by Charles Sarle, Mordecai Ezekiel, Louis Bean and others in the 1920s

(Statistical Reporting Service, 1969). These methods, with continuing improvements, are still in use, but supplemented and extended in important ways by probability samples (Statistical Reporting Service, 1964).

The Department of Agriculture was one of the early organizations in the U. S. to initiate research and development work on probability sampling, and they established a cooperative research program with the Statistical Laboratory at Iowa State University in 1938, through the initiative of C. F. Sarle. It was a fortunate development that resulted in a group doing research on sampling, led initially by Arnold King and Ray Jessen, that has made important and continuing contributions to theory and practice. Iowa State is also one of the few universities that has featured training in sample survey theory and methods as an important part of statistical training.

Bill Cochran came to the Laboratory in 1939 from the Rothamsted Experimental Station, where he had been working with Yates. I mention only a few of the contributions from the Laboratory.

Jessen published a classic paper in 1942 in which he investigated the problem of approximating the optimum sizes of sampling units for agricultural studies. These studies, along with earlier work by H. F. Smith, guided the development—in cooperation with the U. S. Bureau of the Census—of the Master Sample of Agriculture, which had numerous and important early applications and stimulated developments in sampling (King and Jessen, 1945). The Census participation put the master sample on a large scale for use in the 1945 Census of Agriculture. This also enhanced its uses in other surveys. The joint work contributed not only to the Master Sample design but to the principles, methods and theory of sample surveys as a result of stimulating activities and discussions between the staffs at Ames and at the Bureau of the Census.

Another development was concerned with the design of surveys to estimate population characteristics on two successive occasions. In the 1942 paper, Jessen presented the theory for optimum allocation of the matched and unmatched portions of a sample on each of two occasions. This initial step stimulated the later development of designs and theory for rotating samples for surveys taken on successive occasions for time series estimation.

In 1942, Cochran made a particularly important contribution on the use of regression estimation in sample surveys, and in 1946 he extended work done by the Madows (Madow and Madow, 1944) on systematic sampling (Cochran, 1942, 1946).

U. S. BUREAU OF THE CENSUS

After I joined the Bureau of Census in 1933, I worked in the Personnel Division for a year, doing work in personnel classification and getting broadly acquainted with the Bureau. Calvert L. Dedrick then asked me to join the small staff in the Statistical Research Division. A principal initial assignment was to learn about early work on sampling.

The 1937 Enumerative Check Census

In 1936 the Bureau of the Census embarked on a limited experimental and research program on sampling and its potential applications. A stimulant occurred on August 30, 1937 when the Congress, with the strong support of President Roosevelt, authorized a national voluntary registration of the unemployed and partially unemployed. The nation was in crisis, and estimates on the magnitude of unemployment were wide-ranging and highly controversial. Roosevelt's signature was on the questionnaire to be delivered to every household requesting cooperation. The field work on the voluntary registration was carried through by the Post Office, and they did a remarkably fine job.

Fortunately, a staff consisting of S. A. Stouffer and Frederick F. Stephan, as consultants, and Calvert L. Dedrick and others foresaw the lack of validity of the results of this tremendous voluntary registration undertaking, and persuaded the administration to conduct an enumerative check census in a sample of areas. This check census involved interviewing all households within a probability sample of postal delivery routes; the interviewing was done by the mail carriers. They also, by sorting the voluntary mail returns as though they were going to deliver them back to the senders, identified the voluntary mail returns for the sampled postal routes. This made possible the use of ratio estimation based on the voluntary registration returns.

With Cal Dedrick, I had the responsibility for the analytical interpretation of the results (Dedrick and Hansen, 1938). It was a tremendous learning experience, and it also accomplished the goals that were intended to but could not be satisfactorily served by the voluntary registration. A prime result was the convincing demonstration of the usefulness of sampling on an exceedingly important and visible national study. Such a role for sampling had been doubted and resisted by many, in the Congress and elsewhere. Some of the top staff in the Bureau of the Census thought that sampling rather than complete coverage could destroy the reputation the Bureau had achieved for accuracy of its statistics, and that sampling was for others. They did not recognize some of the serious deficiencies that could be demonstrated to exist in some census results.

The voluntary registration was to take place by November 20, 1937, and the household canvass was done by postal carriers during the week ending December 4, 1937. Remarkably, with the overriding support it received, this massive undertaking, along with a preliminary evaluation from the sample, progressed at a sufficient pace that some preliminary results became available on New Year's Eve (end of 1937), with many of us working all that night to make it possible (I called my wife at midnight to wish her a Happy New Year).

The 1937 Enumerative Check Census not only showed what could be done with sampling in measuring unemployment, it contributed advances in methods for measuring employment, labor force participation and various characteristics of these classes and of nonlabor force participants. The effectiveness of the survey depended on ratio estimates of the enumerative check results to the voluntary registration, and we learned about ratio estimation the hard way, guided by Neyman's paper and Neyman's 1937 visit. It was a stimulating and exceedingly fascinating experience that contributed much to the welfare of the U. S. and to the future acceptance of sampling, at least in the Bureau of the Census, and I believe in much wider circles.

The Enumerative Check census occurred shortly after the well-known debacle of the *Literary Digest* poll in predicting the 1936 presidential elections, along with the successful performance of the Gallup Poll which was founded on sounder sampling methods, although still not using probability sampling. The Gallup Poll success undoubtedly had the greatest impact on public acceptance of sampling. However, the Enumerative Check Census, based on probability sampling methods, was supported by theory as well as intuitive considerations and set a strong precedent for future use of sampling in official U. S. statistics.

Sampling in the 1940 Census

The Bureau of the Census was already giving consideration to the first use of sampling as a part of a decennial census to collect supplemental information that could not be obtained in the complete census at reasonable cost. The success of the Enumerative Check Census helped assure acceptance of this approach. Phil Hauser (Philip M.), who had been brought in as Assistant Chief of the Population Division, was a stalwart, along with Dedrick, Stephan and others in getting such an approach adopted. He brought Ed Deming (W. Edwards Deming) in to the Population Division and we worked jointly, along with Fred Stephan, in developing the sample in the 1940 Population Census. It was the beginning of an exceedingly rewarding personal as well as professional relationship among all of us.

The census was taken by listing successively on a page one family after another, and continuing to the next page, one line to a person. The sample information was collected for persons listed on several patterns of predesignated line numbers, simultaneously with collection of the census information. There were challenging problems of avoiding or reducing biases because of the impact of the order of listing households and persons within households. Also, this procedure could not be fully controlled when the enumerator could to some extent determine the order of enumeration. Nevertheless, despite such biases, the success of the undertaking again laid the groundwork for future extensions of sampling in census-taking as well as in current sample surveys (Stephan, Deming and Hansen, 1940). That sampling would have an important role in census work was now assured.

Some notable additions to the Census Bureau staff occurred during the 1940 Census, including some highly qualified people recruited to work as clerks (under severe depression conditions) but soon identified and given more appropriate assignments. Some others were recruited as professionals with mathematical, statistical or related backgrounds. One recruited as a statistician was Bill Hurwitz (William N. Hurwitz). He and I began to work as an effective team that was far more productive than the two of us working separately. It was an exceedingly rewarding working relationship.

The Labor Force Survey

A third major development in the Census Bureau that was to follow would strengthen the role of sampling in basic continuing surveys of national importance. During the Great Depression of the 1930s the focus was on unemployment. With U. S. entry into World War II near the end of 1941, the statistical needs changed and instead of unemployment the focus was on employment.

Steve Stock (J. S. Stock), Les Frankel (L. R. Frankel), John Webb and others had initiated a pioneering national multistage sample in the WPA to measure unemployment, following the 1937 Enumerative Check Census (Frankel and Stock, 1942). After U. S. entrance into the war, the WPA was abolished and in 1942 this survey was transferred to the Bureau of the Census, with the principal focus now on the measurement of employment and labor force participation. Such information was needed, among other reasons, to guide an effective draft policy that would not damage the performance of important industries.

We reviewed and undertook a revision of the survey design. The outcome was to place the survey for the first time on a full probability sampling basis, along with the introduction of some new sampling methods and theory that enhanced the validity of the survey results, increased efficiency and provided a design framework that has guided many survey designs throughout the world.

A number of new sample design features were introduced along with supporting theory. I will mention only one of them—the introduction of sampling with probabilities proportionate to measures of size at successive stages of sampling, but such that the overall probability of selection for the sample could be uniform, if desired, for each member of the population (Hansen and Hurwitz, 1943). This procedure had the effect of equalizing interviewer workloads and thus facilitating more effective administration. At the same time, it reduced variances as compared to alternative procedures.

These probability proportionate to size (PPS) selection procedures were combined with the procedures for estimation that made use of supplementary information at the primary sampling unit level, along with poststratification to independent figures on the total population of the U. S. by age, sex and race. The basic theory was provided for the complex selection and estimation procedures.

Some interesting lessons were learned in the Labor Force Survey during and after the war.

The initial design of the Labor Force Survey at WPA was a probability design at the first two stages of selection. At the final stage, in rural areas, random points were selected with equal probability. The households selected were those closest to the selected points. While this provided a probability sample of points, the probabilities of selection of households were unknown. When the measurements from the revised survey became available in 1943, they showed striking differences in agricultural employment, differences that were explainable (after they were known) by the nonprobability-sampling mechanism. The results substantially altered wartime manpower policy when the differences became known. The design was now such that it would reflect changes as they took place without depending on the validity of an assumed relationship or model.

Immediately after the war, in October 1945, I participated in taking a sample survey of the Japanese population, and was in Japan for a few weeks. While there a friend came and asked me if I had heard of the demise of the Labor Force Survey. He indicated that with the discharging of millions of soliders from the Army everyone knew that unemployment had surged upward, and that the survey was not showing this. My response was that the survey was designed to measure effectively under such changing circumstances, and I had full confidence in what it was showing within the range of the estimates of sampling variability. Subsequent developments showed this to be well-placed confidence, the discharged soldiers had not gone into the labor market immediately. It demonstrated that the survey could meet the challenges of effective measurement under sharply changing economic circumstances.

The 1954 CPS Revision

In 1954 the Census Bureau undertook a revision of the sample design of the Labor Force Survey (Hansen, Hurwitz, Nisselson and Steinberg, 1955), now known as the Current Population Survey, by spreading the same size sample of about 25,000 households over a much larger number of primary sampling units (from the original 68 to 230). The transition took place in 1954 and there was great concern when the estimates of unemployment differed between the new and old samples by more than could be explained by sampling variability. After intensive study, including appointment of a review committee, it was concluded that differences arose because of the reduced supervision and training of interviewers in the original sample of 68 areas, while extensive attention was being given to training the new interviewers in the new and much larger number of primary sampling units. The control of measurement had always been given serious attention, but this emphasized the great sensitivity of quality of such sensitive measurements to the conditions under which they were taken. New and increased controls based on process control procedures were introduced. These made use of sample inspection of the work of enumerators at periodic intervals. The inspection took the form of both observation and reinterview, in addition to an edit of all completed questionnaires, with feedback of errors found for taking corrective action (Hansen and Steinberg, 1956).

Another innovation was introduced at this time, applicable to recurring surveys. This was a system of sample rotation and an approximately optimum system of composite estimation based on the sample rotation to improve estimates of both coverage and of changes over time (Hansen, Hurwitz, Nisselson and Steinberg, 1955). These procedures had been designed with supporting theory and introduced earlier in the Census Bureau's Retail Trade Survey (Woodruff, 1959).

Independently, Patterson (1950) gave a general theory for optimum sample rotation and estimation, although this theory did not meet the needs of practical sample design because it called for revision of all prior estimates at each new period. Margaret Gurney and Joseph Daly of the census staff extended the theory (Gurney and Daly, 1965).

Extension of Sampling to Other Subject Areas

With the advances and success of the Labor Force Survey, the next step was the extension of sampling to other subject areas. The first efforts sometimes encountered the attitude that while sampling would work with populations that were relatively homogeneous, like households and people, they would not work with business populations, in which the distributions are exceedingly skewed. We were able to show how to take advantage of the skewness, with approximately optimum allocation of the sample, and how to provide effective samples through the joint use of area and list sampling, even in the absence of up to date lists. Sampling was successfully extended, with new principles, methods and theory developed as needed, to manufacturing, retail trade, wholesale trade, agriculture, governmental units and other subjects areas. See for example Bershad (pages 516–558) and Ogus (pages 582–588) in Volume 1 of Hansen, Hurwitz and Madow (1953). Progress on these and later developments included many contributions from the able statistical research staff that was recruited and trained at the time of the 1940 Census and later. Leaders of this staff included Max Bershad, Joe Daly, Leon Gilford, Harold Nisselson, Eli Marks, Jack Ogus, Leon Pritzker, Joe Steinberg, Ben Tepping, Joe Waksberg, Ralph Woodruff and others.

The Redesign of the Major Censuses

The reasonable control of measurement errors is a challenge in both censuses and sample surveys. A great deal of attention was devoted to experimental studies for the control of measurement errors in the decennial censuses. A number of experimental studies were conducted before and after the 1950 Census, and in addition, extensive experimental studies were incorporated into the ongoing censuses, beginning with the 1950 Census. Included were matching studies to other records, to the prior census and to births. Also, during the census there were separate intensive re-enumerations in a sample of small areas, one focused on evaluating coverage, and another one on content errors in the census. Also included were randomization studies in which two enumeration districts were randomly assigned to each enumerator within the experimental areas. Such randomization made it possible to estimate the variance of response errors between and within enumerators, and thus the correlation of errors within an enumerator's work. Also included were self-enumeration studies in which respondents were requested to fill out their own questionnaires, and the enumerator's role was greatly reduced.

The designs of the response-error studies were guided by the development of response-error models with supporting theory (Hansen, Hurwitz, Marks and Mauldin, 1951; Hansen, Hurwitz and Bershad, 1961). Empirical evaluation of model parameters through the experimental studies just mentioned indicated especially the substantial impact of correlated response errors within the work of interviewers.

On the basis of these studies it was concluded that correlated response errors within the work of enumerators in the decennial censuses constituted a serious problem, especially for small area statistics, and that self-enumeration substantially reduced these, and also improved accuracy of response on most items. The work of the editors and coders could be reasonably controlled by quality process control methods, as could the work of interviewers in continuing ongoing surveys, and for these the contributions of correlated response errors were less serious problems. These results, with others that I will now discuss, led to a revolution in census taking in subsequent censuses, with self-enumeration as the principal means of data collection.

Univac and Related Electronic Advances

Shortly after the end of World War II, I was approached, as the person in the Census Bureau with responsibility for research and development, by J. P. Eckert and John Mauchly (the designers of the Univac), with remarkable proposals and asking the Census Bureau's support in building a large-scale general-purpose electronic computer. The Census Bureau had always been in the vanguard in the development and application of punched-card equipment. With the participation of the National Bureau of Standards, we contracted to support the design and construction of the first Univac. We participated in some design decisions. Delivery was accepted on March 31, 1951. Of course we had been preparing for it, had programs already written and tested, and immediately put it to use, 24 hours a day, 7 days a week, on important parts of the data processing of the 1950 Census.

After the 1950 Census we placed the Current Population Survey and other sample surveys on the computer, and began planning for use of the computer on the forthcoming Censuses of Business and of Manufactures as well as the next decennial Census. We also began developing new methods for machine editing, for imputation of missing values, and sample estimation that took advantage of the computer's capabilities. Of course, by today's standards Univac I was massive in physical size, slow and limited in memory, but, nevertheless, it was a sensational advance. There were, of course, additional remarkable computer developments to come.

The 1960 Census

In planning the 1960 Census (which began with the 1950 Census experimental and evaluation studies), it was immediately clear that things were unbalanced if we had electronic computers for data processing, but we still required hundreds of operators to manually key the information into punched cards. We had already discussed with IBM the possibility of developing mark-reading equipment in time for use in the 1950 Census, along with the anticipated electronic computer for data processing. After some developmental work, IBM concluded that they could not prepare the equipment for massive paper handling in time for the census, and the 1950 Census data were again recorded by manual keying.

We immediately discussed the problem with the National Bureau of Standards, and they undertook with us the development of mark-reading equipment that could handle the massive data-conversion job for the 1960 Census. The result was FOSDIC—Film Optical Sensing Device for Input to Computers. The questionnaires were microfilmed and then read by FOSDIC. Filming was proposed as a more reliable system than reflective reading of pencil marks. The equipment was developed, tested and constructed in time for its highly successful use in the 1960 Census.

To reduce costs and to greatly speed up the processing of the 1960 Census only a few key questions were covered in the complete census. Most topics were covered in relatively large samples of a magnitude of 5, 20 and 25%. The 100% questionnaires were designed so as not to include items that involved manual editing or coding. After extensive testing we concluded that for the sample items respondents would be asked to prepare their own responses on FOSDIC-readable questionnaires. Special paper handling equipment was designed, tested and built to facilitate microfilming the approximately 40 million returned questionnaires. The microfilm for the complete census items was processed through FOSDIC and the computer while the sample questionnaires were being edited and coded and subsequently processed through FOSDIC and the computer. Public cooperation was relatively high. Interviewers followed up as necessary. When these procedures, along with self-enumeration, and others were applied in the 1960 Census the result was far more timely information as well as greater accuracy and reduced cost (U. S. Bureau of the Census, 1966).

Process control procedures were successfully instituted on the various phases of printing the questionnaires, microfilming, FOSDIC reading and editing and coding of the sample returns. In many respects it was a reasonably close approach to a well-balanced total design.

The Panel of Statistical Consultants

The Census Bureau greatly benefited 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 members included Fred Stephan (Frederick F. Stephan) and Bill Madow (William G. Madow) for the full time period, Ivan Fellegi from Statistics Canada, H. O. Hartley and others. All were exceedingly able, but 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.

Cochran took the initial role in suggesting the creation of the Panel of Statistical Consultants, and it proved a highly beneficial arrangement. We had close and continuing personal as well as official relationships with the panel members. On evenings between the usual two-day meetings of the panel we often had the panel in our home and enjoyed a drink or two as well as good conversation—solving the problems of the world as well as of statistics. Cochran was an exceedingly effective chairman. Few, if any, have contributed more to statistics as consultant, teacher, researcher and author. I once took him sailing on my Sailfish, and due to my lack of skill we turned it over and came into the house wet and bedraggled. He asked for a cigarette, and suggested maybe a fresh pack would be reasonable compensation for what had happened to him.

Fred Stephan was a highly effective statistical consultant over the years, to the Census Bureau and others, from the beginning of my participation in the Census Bureau.

Ivan Fellegi's participation in the panel was one aspect of the cooperation between the Census Bureau and Statistics Canada. He came to Canada as a young Hungarian refugee. He has contributed importantly to the advancement of statistics, and to the panel, as well as to Statistics Canada. He was just recently made the head of Statistics Canada.

Bill Madow was in the Census Bureau earlier, made contributions to sampling theory, notably to systematic sampling theory, was a joint author of what is now often referred to in Westat as "the Bible," and continued as an adviser over the years.

SOME ADDITIONAL REMARKS

I have given only a brief review of a few highlights of statistical activities in the Census Bureau and elsewhere. The Census Bureau developments involved cooperation between top management, operating staff and statistical methods staff. I became part of top management as well as promoter and leader of the statistical methods staff, and I believe this joint role facilitated statistical progress. Progress also depended on constant questioning why things were done as they were, or as they were proposed to be done. We regarded it as important not to be afraid to be right, even in the face of unanimous opposition, and to be willing to ask foolish questions—they often turned out not to be foolish. We learned through experimental studies and experiments that things that were "known" to be true on the basis of long experience often were not. Generally decisions were made after examination of alternatives as well as extensive advance development of tentative plans, testing, evaluation and revision, with these applying to questionnaire design, sampling procedures, enumeration procedures and all the many other aspects of taking a census or major sample survey.

My participation in the Census Bureau ended near the end of 1968, when I retired from the government and joined Westat, Inc. where the work has continued along similar lines in many respects, but often in different subject areas. I note with sadness that Bill Hurwitz died in March 1969. Hurwitz's leadership, insistence on high standards and contributions cannot be sufficiently emphasized. I owe him a great debt.

SOME OTHER EARLY CONTRIBUTORS

I should add that while Ed Deming never served on the panel, we were in constant communication over the years, wrote joint papers and interchanged ideas. He left the Census Bureau after the 1940 Census and joined the Budget Bureau, where he influenced national and international statistical developments. He then left the government and became a consultant, primarily to private industry, on sample surveys, statistical quality control and other statistical problems. His contributions to quality control in Japan, and now throughout the world, are being recognized by the public as well as statisticians and other professionals, and he has become world famous. I have had the pleasure of working as a personal consultant to him, and have spent many Saturday or Sunday mornings over the years discussing his problems. He recognizes that a private consultant working alone needs someone to discuss problems and argue with, and it has been my great pleasure to serve this role.

Ed was always willing to help students from around the world. He arranged many social evenings at his home, inviting both budding and well-known statisticians and others. He enjoyed good banter and fun as well as serious discussion. One evening he and Fred Stephan prepared a beer malted milk for me. It tasted like it sounds.

I will mention only three others that made important early contributions to sample surveys. Hartley was at the Statistical Laboratory at Ames after Cochran, and then moved to Texas A. and M. University, and both places trained students and contributed to advances in sample survey methodology. Les Kish was at the University of Michigan Survey Research Center, where he took the lead on survey methodology, contributed to theory and practice and trained many from all over the world in sample survey theory and methods. Tore Dalenius, in Sweden and at Brown University in the U. S., has trained many and made a number of contributions, including a widely used method for approximating strata boundaries with optimum stratification (Dalenius, 1957).

I will close with a brief discussion of a topic that has received considerable attention in recent years.

THE ROLE OF MODELS IN SURVEY SAMPLING

The preceding discussion has been concerned primarily with the use of probability sampling. Probability samples make joint use of a sample selection plan in which each element of the population has a known positive probability of being selected and estimators such that confidence intervals can be computed which, for large enough samples, will include the parameters being estimated with probability closely approximating the specified level. Survey practitioners commonly depend on probability samples for large-scale surveys or for important surveys in which public acceptance is desired. Of course, for probability samples to be effective the survey design must be operationally feasible, and carried out so that performance conforms reasonably closely to specifications.

Another school of thought holds that the sample selection plan and estimators should be chosen so as to have attractive characteristics (e.g., minimum variance) under an assumed model. We have referred to this as model-dependent sampling. The model may be based on prior information from earlier studies or on information from the observed sample or both. With this approach the inferences from the sample depend upon the model that is assumed. Also, with a model-dependent approach, explicit randomization may or may not be used in selecting the sample, although there is coming to be considerable agreement that randomization is desirable because of its appearance of objectivity, whether or not it is otherwise regarded as necessary by the model-dependent sampler.

In making inferences to a finite population the model-dependent approach is ordinarily either equivalent or superior to the probability sampling approach *if* the assumed model, in fact, accurately describes the population being sampled. But if the model is not fully realistic, the model-dependent approach may result in misleading inferences. From the viewpoint of the practicing survey statistician, it is important to realize that a deviation between the model and the population too small to detect on the basis of statistical tests for the available observations may, in fact, cause problems in inference. Even if the selection is done with known probabilities, but an estimator is used that is not unbiased or consistent under probability sampling, the outcome may be seriously biased estimates of the characteristics of the finite population. The sizes of the bias will be unknown, and the nominal confidence intervals may make the results appear more accurate than they are. These points are illustrated by an example in Hansen, Madow and Tepping (1983). In that illustration samples of 100 are sufficiently large for the probability sampling approach to have a mean square error about half as large as the mean square error with model-dependent sampling. That example illustrates how probability sampling avoids the biases inherent in an assumed model that is not fully descriptive of the population but that appears to hold well and that fits acceptably on the basis of statistical tests.

On the other hand, in making inferences about the causal system that generated a particular finite population—or to make predictions about future realizations of that causal system—there is, however, no alternative to the use of a model-dependent approach and consequent risks.

Probability samplers make extensive use of superpopulation models to great advantage in the context of probability sampling for a finite population. The sample designs may use models in inference, only to the extent that the validity of the sample estimates does not depend on the validity of a model. However, models, whether or not fully valid, can provide an effective guide in choosing probability sampling designs. A model of the population may, for example, suggest how to assign approximately optimum allocation or probabilities of selection, and how to choose an estimator among available alternative consistent estimators. The resulting variances will be smaller if the model used is indeed a reasonably accurate description of the population than if it is not. But even if the model is not particularly good, with probability sampling the estimators are unbiased or consistent, as are the estimators of their variances.

The above discussion is a brief summary of some of the principal points and results presented in Hansen, Madow and Tepping (1983). The reader is referred to that paper for fuller details and for comments by discussants representing various points of view. It seems to me that there may be an approaching consensus on the use of probability sampling in large-scale surveys, although there remain some able and dissident voices (see, for example, Royall and Herson, 1973a, 1973b; Smith, 1976).

REFERENCES

Many additional references appear in the references cited below.

Bowley, A. L. (1926). Measurement of the precision attained in sampling. *Proc. Internat. Statist. Inst.* **12** 6–62.

Cochran, W. G. (1939). The use of the analysis of variance in enumeration by sampling. *J. Amer. Statist. Assoc.* **34** 492–510.

Cochran, W. G. (1940). The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *J. Agr. Sci.* **30** 262–275.

Cochran, W. G. (1942). Sampling theory when the sampling-units are of unequal sizes. *J. Amer. Statist. Assoc.* **37** 199–212.

Cochran, W. G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. *Ann. Math. Statist.* **17** 164–177.

Dalenius, T. (1957). *Sampling in Sweden. Contributions to the Methods and Theories of Sample Survey Practice.* Almqvist and Wiksell, Stockholm.

Dedrick, C. L. and Hansen, M. H. (1938). *Census of Partial Employment, Unemployment and Occupations: 1937.* **4.** *The Enumerative Check Census.* U. S. Government Printing Office, Washington.

Fisher, R. A. (1925). *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh.

Frankel, L. R. and Stock, J. S. (1942). On the sample survey of unemployment. *J. Amer. Statist. Assoc.* **37** 77–80.

Gurney, M. and Daly, J. F. (1965). A multivariate approach to estimation in periodic sample surveys. *Proc. Social Statist. Sect. Amer. Statist. Assoc.* 247–257.

Hansen, M. H., Dalenius, T. and Tepping, B. J. (1984). Some recollections and expectations in survey sampling. In *Statistics: An Appraisal, Proceedings 50th Anniversary Conference, Iowa State Statistical Laboratory* (H. E. David and H. T. David, eds.). The Iowa State Univ. Press, Ames, Iowa.

Hansen, M. H., Dalenius, T. and Tepping, B. J. (1985). The development of sample surveys of finite populations. In *A Celebration of Statistics: The ISI Centenary Volume* (A. C. Atkinson and S. E. Fienberg, eds.). Springer, New York.

Hansen, M. H. and Hurwitz, W. N. (1943). On the theory of sampling from finite populations. *Ann. Math. Statist.* **14** 333–362.

Hansen, M. H., Hurwitz, W. N. and Bershad, M. A. (1961). Measurement errors in censuses and surveys. *Proc. Internat. Statist. Inst.* **38** 359–374.

Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (1953). *Sample Survey Methods and Theory* **1, 2.** Wiley, New York.

Hansen, M. H., Hurwitz, W. N., Marks, E. S. and Mauldin, W. P. (1951). Response errors in surveys. *J. Amer. Statist. Assoc.* **46** 147–190.

Hansen, M. H., Hurwitz, W. N., Nisselson, H. and Steinberg, J. (1955). The redesign of the census current population survey. *J. Amer. Statist. Assoc.* **50** 701–719.

Hansen, M. H. and Madow, W. G. (1976). Some important events in the historical development of sample surveys. In *On the History of Statistics and Probability* (D. B. Owen, ed.). Dekker, New York.

Hansen, M. H., Madow, W. G. and Tepping, B. J. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *J. Amer. Statist. Assoc.* **78** 776–793 (Comments and rejoinder, 794–807).

Hansen, M. H. and Steinberg, J. (1956). Control of errors in surveys. *Biometrics* **12** 462–474.

Jensen, A. (1926). Report on the representative method in statistics and the representative method in practice. *Proc. Internat. Statist. Inst.* **12** 357–439.

Jessen, R. J. (1942). *Statistical Investigation of a Sample Survey for Obtaining Farm Facts.* Agricultural Experiment Station, Iowa State College of Agriculture and Mechanic Arts, 304, Ames, Iowa.

Kiaer, A. N. (1976). The representative method of statistical surveys (original in Norwegian; translation for new edition). *Kristiania Videnskabsselskabets Skrifter, Historisk-filosofiske Klasse,* 1897 (4), 37–56. Statistisk Sentralbyra, Oslo.

King, A. J. and Jessen, R. J. (1945). The master sample of agriculture. *J. Amer. Statist. Assoc.* **40** 38–56.

Kruskal, W. H. and Mosteller, F. (1979). Representative sampling, III: The current statistical literature. *Internat. Statist. Rev.* **47** 245–265.

Madow, W. G. and Madow, L. H. (1944). On the theory of systematic sampling, I. *Ann. Math. Statist.* **15** 1–24.

Mahalanobis, P. C. (1946). On large-scale sample surveys. *Philos. Trans. Roy. Soc. London Ser. B* **231** 329–451.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection (with discussion 607–625). *J. Roy. Statist. Soc.* **97** 558–606.

NEYMAN, J. (1938). Contribution to the theory of sampling human populations. *J. Amer. Statist. Assoc.* **33** 101–116.

PATTERSON, H. D. (1950). Sampling on successive occasions with partial replacement of units. *J. Roy. Statist. Soc. Ser. B* **12** 241–255.

ROYALL, R. M. and HERSON, J. H. (1973a). Robust estimation in finite populations I. *J. Amer. Statist. Assoc.* **68** 880–889.

ROYALL, R. M. and HERSON, J. H. (1973b). Robust estimation in finite populations II: Stratification on a size variable. *J. Amer. Statist. Assoc.* **68** 890–893.

SMITH, H. F. (1938). An empirical law describing heterogeneity in the yields of agricultural crops. *J. Agric. Sci.* **28** 1–23.

SMITH, T. M. F. (1976). The foundations of survey sampling: A review (with discussion 196–204). *J. Roy. Statist. Soc. Ser. A* **139** 183–195.

STATISTICAL REPORTING SERVICE. (1964). Miscellaneous Publication No. 967. U. S. Dept. Agriculture, Washington.

STATISTICAL REPORTING SERVICE. (1969). *The Story of U. S. Agricultural Estimates*. Miscellaneous Publication No. 1088. U. S. Dept. Agriculture, Washington.

STEPHAN, F. F. (1948). History of the uses of modern sampling procedures. *J. Amer. Statist. Assoc.* **43** 12.

STEPHAN, F. F., DEMING, W. E. and HANSEN, M. H. (1940). The sampling procedure of the 1940 Population Census. *J. Amer. Statist. Assoc.* **35** 615–630.

SUKHATME, P. V. (1946). Bias in the use of small-size plots in sample surveys for yield. *Nature* **157** 630.

TSCHUPROW, A. A. (1923). On the mathematical expectation of the moments of frequency distributions in the case of correlated observation. *Metron* **2** 461–493, 646–680.

U. S. BUREAU OF THE CENSUS. (1966). *1960 Census of Population and Housing: Procedural History*. U. S. Government Printing Office, Washington.

WOODRUFF, R. S. (1959). The use of rotating samples in the Census Bureau's monthly surveys. *Proc. Soc. Statist. Sect. Amer. Statist. Assoc.* 130–138.

YATES, F. and ZACAPONY, I. (1935). The estimation of the efficiency of sampling with special reference to sampling for yield in cereal experiments. *J. Agr. Sci.* **25** 543–577.