Journal of Official Statistics, Vol. 14, No. 2, 1998, pp. 119–135

The Hansen Era: Statistical Research and its Implementation at the U.S. Census Bureau, 1940–1970

Joseph Waksberg[1]

I feel deeply honored to have been asked to deliver the 1996 Hansen Memorial Lecture, and to thus be thrust into the company of the distinguished group of statisticians who spoke in previous years: Fred Smith, Wayne Fuller, Norbert Schwartz, Leslie Kish, and Ivan Fellegi. I will try to keep to the standards they set, but I hope no one is grading the lectures. It is also, of course, a privilege to contribute to the memory of Morris Hansen and to pay tribute to his seminal role in advancing both theory and practice in survey methods. As many of you know, I was Morris's disciple, colleague, and personal friend, and during this talk I will reflect all three roles. However, in any discussion of history of sampling and related survey methods, Morris's name is bound to occur frequently.

1. Background

The committee appointed by the Washington Statistical Society to select topics and speakers for these lectures, as well as to handle the administrative arrangements, asked me to reminisce on the research carried out by the staff of the U.S. Census Bureau during the period when they were developing so many of the survey methods still in common use, and laying the seeds for further development. The years covered were mostly from 1940 to 1970, but I will occasionally stray from this time period in both directions. I will mainly concentrate on work carried out by the Census staff, but research is not carried out in isolation and other names will come up on occasion.

The organization of research on statistical methodology at the Bureau of the Census and its implementation to Bureau surveys was basically established in the latter half of the 1940s. I believe it is still in place. Mathematical statistical units were included in all of the major subject matter divisions, and they reported administratively to the division chiefs but technically to the Associate Director responsible for research and statistical methodology. There was, in addition, a small research staff reporting directly to the Associate Director for statistical methods. Research on issues that were broad enough to have multiple applications or that advanced theory rather than practice was supposed to be carried out by the research staff. Research required for the conduct or improvement of specific surveys or censuses was the responsibility of the statisticians in the subject matter divisions. However, no one worried about keeping strict lines between short term and long term research. Staff was mostly assigned to what had to be done.

For a long time I assumed that management analysts with strong sense of lines of

[1] Westat, Inc., 1650 Research Boulevard, Rockville, MD 20850, U.S.A.

authority would have viewed the organization as a monstrosity. However, Jim Bonnen has recently informed me that it is a fairly common, quite rational response to the management problem of providing a complex and critical input or service necessary to the performance of multiple parts of an agency. In large organizations such as the Census Bureau which produce complex outputs that must be frequently adjusted or redesigned, it is important to keep the interaction of research and development with the enterprise as informal and flexible as possible. In any case, I am not aware of anyone trying to change the organization. Anyone attempting to change it would have to overcome two obstacles. First, the mathematical statisticians who initially comprised most of the research staff were extremely productive both in the development of theory and monitoring its application, and it would have been difficult to justify changes for the sake of efficiency. Secondly, it would have involved arguing with Morris Hansen, who would have been a formidable opponent. In any case, I and most of the other staff members were happy with the flexibility of the organization. It permitted us to do things we liked and found stimulating, gave us opportunity to interact with other Bureau statisticians whose judgment we respected, and provided a forum for learning about developments in parts of the Bureau outside our immediate involvement.

I would like to give credit to the principal participants in those early years. The Associate Director responsible for research was at the top of the pyramid. Morris Hansen who created the organizational structure of the research activities held that position until 1968 when he retired from the Government. He was replaced by Joe Daly. I took over in 1971 and held that post until mid-1973. During almost all of the time that Morris was Associate Director, Bill Hurwitz was his close assistant and alter ego. Bill provided technical direction and acted as both an intellectual stimulus and someone who prodded us to get the work done. He was also a warm personal friend who took an interest in everyone on his staff.

The mathematical statisticians in these early years included Bill Madow, Ben Tepping, Max Bershad, Joe Daly, Marge Gurney, Blanche Skalak (later on Blanche Hurwitz), Hal Nisselson, Alice Rhodes, Joe Steinberg, Jack Ogus, Ralph Woodruff, Tom Jabine, Bob Hanson, Lee Gilford, and Irene Hess, among others. Ed Deming and Lester Frankel were on the staff for a few of the early years. In 1949, Joe Steinberg was building a staff to handle the statistical problems of the 1950 population census and asked me to join him. I was thus somewhat of a latecomer to the group. In the late 1940s, the staff expanded in another direction. Morris and Bill were moving into consideration of nonsampling errors and they added several psychologists and demographers to the staff, including Eli Marks, Leon Pritzker, and Parker Mauldin. Still later, Barbara Bailar, Walter Perkins, and Charlie Jones were added as part of a larger buildup for the 1960 Census. This rather small and intimate group transformed survey methods.

Morris and Bill established a small advisory committee on survey methods. It included Bill Cochran, H. O. Hartley, Bill Madow, Fred Stephan, and for a time Nathan Keyfitz. When Keyfitz left Statistics Canada, he was replaced by Ivan Fellegi. The Census Bureau thus had, either on its staff or as advisors, most of the statisticians in the United States and Canada making important contributions to sampling methods.

It is an oversimplification to state that all of the developments in sampling were being carried out at the Census Bureau. The Department of Agriculture was deeply involved in research on sampling and established a statistical laboratory at Iowa State University. Before Cochran and H. O. Hartley came to Iowa State, the research was rather narrowly focused on agricultural problems. In addition, there was slow progress in incorporating the research results into survey operations. This was in contrast to the Census Bureau, where there was constant pressure to introduce sampling and statistical thinking into as many fields as possible. Other organizations in the U.S. that were active in the field were the University of Michigan where Leslie Kish did much innovative work and BLS, principally led by Jerry Cornfield. However, none of these organizations had the resources of the Census Bureau, nor the strong motivation to combine theory and practice. Outside of the U.S., important developments were going on in India, Great Britain, and Canada. In the late 1940s, statisticians in India, mainly led by P. C. Mahalanobis and P. V. Sukhatme, were doing work that in many ways paralleled developments at the Census Bureau, in particular, in attempting to take nonsampling errors into account in survey design. The work in England tended to be more theoretic, although Neyman's article (1934) was fundamental to much of the future development. It probably had a more immediate effect in the United States than in England.

It is useful to review the historical setting in which the developments in theory and practice were made. Bershad and Tepping (1969) have pointed out that the status of sampling practice before Neyman's groundbreaking article was published was primitive by present standards. Completely uncontrolled selection from available lists was quite common. Quota sampling was perhaps the most prominent technique to control biases of selection. Even when random samples were used, there was no attempt to estimate sampling errors, and when it was done, simple random sampling models were used, whether it was appropriate or not. The state of the practice can be observed by comments made by Bowley, one of the most distinguished social statisticians in the early part of the 20th century, during the discussion following Neyman's article (Neyman 1934). ''If, in fact, one has made a stratified selection and writes down the precision as if it had been purely random, then one is on the safe side. I have myself generally been content to let it go at that for two reasons: (1) that it is very difficult to measure the additional precision due to stratification . . . at any rate it is a very lengthy business; and (2) . . . I have had to explain and try to justify the methods of sampling to nontechnical readers and therefore I have been obliged to leave out a great deal that Dr. Neyman would have put in.'' The remarkable progress made in the next several decades is obvious. Furthermore, the progress was not only in theory, it was also in educating the public to accept complex sampling methods, and also in some cases to understand them.

Morris, Bill Hurwitz, and the rest of the staff came on the scene at an opportune time. First, there was a sudden growth of interest in social and economics statistics, initially from the serious economic problems of the depression of the 1930s, followed by the readjustments of the postwar years. The demand for statistics blossomed, and most of this required sampling. Secondly, the Bureau of the Census was fortunate in having the top management sympathetic to funding research and willing to introduce research results into survey operations. Phil Hauser was deputy director and for a time acting director of the Bureau in the early postwar years and he was very supportive of new methodological approaches. I assume that the director during the late 1940s, J. C. Capt approved of these developments although I had very little personal involvement with him. Others at the

Bureau at the management level during the 1940s and 1950s were Con Taeuber, Howard Grieves, and Ross Eckler, and they were also receptive to new ideas and innovative approaches.

I have occasionally wondered whether survey methodology would have developed very differently if the Bureau had not supported the research and applications in those early years. I suspect that the developments would have been delayed somewhat, but not have been fundamentally changed. Morris and Bill, with their restless minds and creative energy would have moved to another location. The rest of the statistical staff would have missed the intellectual stimulation that kept them loyal to the Census Bureau for so many years, and would have scattered. Other organizations – universities, private companies, nonprofits – would have become the focus of the new developments, and at some point in time the Bureau of the Census would have had to both accept the new ideas, and join in the drive for further research. However, luckily this is just speculation. In practice, for the statisticians engaged in methodological research, the Bureau was an exciting and stimulating place to work.

2. Early Years of Sampling Research

Morris Hansen has described the start of the Census Bureau's involvement in sampling methods in 1937 (Olkin 1987). At the present time, when hardly a day goes by without newspapers and the broadcast media reporting new economic statistics, it is a jolt to realize the lack of information on the nation's most critical problem in the 1930s. The nation was in the depths of the great depression and estimates of unemployment in the newspapers varied from 3 to 15 million, depending on the estimation methods. A voluntary unemployment registration was carried out in 1937, handled by the Post Office. A small group of sociologists, including Cal Dedrick at the Census Bureau and several others loosely associated with the Bureau, thought that such a registration would yield numbers no one could interpret. They suggested a check census in a sample of areas. Morris Hansen was brought in to help design it. The report, issued in 1938, provided estimates of unemployment and of their standard errors. They were generally believed and accepted as valid measures.

It was a watershed experience for both the Bureau and Morris and the others involved in the statistical research. First, it was a demonstration of what area and probability sampling could do. (It is a pity that 60 years later, some members of Congress still have not recognized this!) Before that, the Census Bureau had the idea it could not do sampling because that would discredit results. For the research staff, it was also a demonstration of the philosophy of randomization as distinguished from modeling using purposive selection which was commonly done in those days. It was also the first major application of ratio estimation in which the independent variables were correlated with, but not identical to the variables measured. Both of these developments were strongly influenced by Neyman's 1934 article. Finally, the report projected the sample estimates to smaller areas through regression relationships; this is still the primary basis of small area estimates.

The unemployment survey was the precursor of what became the Current Population Survey (CPS). Some of the research staff at the Works Progress Administration (the Government agency responsible for administering depression-related programs) had participated in the 1937 survey and decided they could design a national survey to keep track of unemployment. Although the sample was not completely a probability design, it had many features that were innovative at the time, including a first stage sample of counties followed by a sample of blocks within urban areas. Shortly after the beginning of World War II, the WPA was abolished and the survey was transferred to the Census Bureau. Initially called the Labor Force Survey it later became the Current Population Survey (CPS).

Bill Hurwitz joined the Census staff in 1940, and shortly thereafter Hansen and Hurwitz started working as a team. The labor force survey provided a wonderful laboratory for their zeal for the development of efficient and relatively unbiased sampling methods. In 1942, shortly after the Bureau took over the labor force survey, they published an article on the role and impact of the intraclass correlation in cluster sampling (Hansen and Hurwitz 1942). The following year they produced an article that introduced the use of multistage sampling with probability proportionate to size (Hansen and Hurwitz 1943). In this article they developed the variance for multistage sampling with varying probabilities, with or without subsampling. The theory and methods in these articles are essentially still used. In the first few months of the survey, the estimates tended to bounce around; they developed the theory for poststratification to stabilize the estimates.

The subjects of these articles and their timing indicated an important feature of the research. They were driven by real needs of particular surveys. The research started off by examining the problems in existing and planned surveys and attempting to solve them. Frequently, this called for the development of new theory, or novel ways of implementing existing theory, which produced survey methods that were generally applicable. Most of the work at the Bureau at that time followed this pattern. I think the outstanding creativity of early Census research and development was in a large part due to the stimulation of responding to real statistical problems.

Other statisticians joined the statistical research staff during the 1940s and began to contribute to theory and methods. Bill Madow was very strong in statistical theory, and he, Morris, and Bill were responsible for the first book on sampling in which a modern approach was taken (Hansen, Hurwitz, and Madow 1953). There were contributions to the book by other staff members who developed some of the proofs of theorems, and who calculated some of the parameters. The research staff also started examining other sampling issues, including alternative rotation plans, estimation methods, and variance estimation. Ben Tepping and Max Bershad had particularly imaginative approaches to problems. Ralph Woodruff developed a procedure for establishing confidence bounds around medians which is still applicable. Most of the staff, however, was engaged in making sure that the new theories were operational.

Area sampling was the method of choice for the household surveys in the 1940s and 1950s, and considerable effort was devoted to the production of maps, and to the calculation of between PSU and segment components of variances. Up to the late 1950s the calculations were carried out with hand calculators, and they were tedious, and limited in scope. Nevertheless, it was recognized that in order to apply theory efficiently, estimates of variance components were necessary.

List sampling for household surveys was introduced in the CPS and other household surveys in the early 1960s, and I was responsible for that. As in most other developments, it was designed to overcome specific problems in area sampling. One of the problems was the quality of the maps. The second problem was the speed with which measures of size of segments based on the Census deteriorated. The 1950s and 1960s were periods of vast suburbanization of America, and the geography was changing rapidly. I remember that at one time, one of the segments in the CPS had grown so rapidly that if unbiased weighting had been used, it would have accounted for one-third of the African-American population of the United States. Obviously these effects were reduced, both in sampling and weighting, but with uncertain effect on statistics. In 1959, I worked on the development of statistics on housing starts which rely on the building permits issued in most localities. The building permits seemed to me to be the clue to solving the problems of area samples. I proposed that the second stage of area samples – the segments – be replaced by a list sample selected from the most recent census supplemented by a sample of building permits. I remember proposing it at one of Bill Hurwitz's staff meetings, and that a vigorous discussion followed. I do not know to whom Bill spoke or whether he had to overcome objections from other Bureau staff, but shortly afterwards, he told us to go ahead. When I think of the experience now, I am aghast at the speed with which the decision was made and implemented. Maybe it was the arrogance of youth, but we had a lot of self confidence and felt we were charting new paths. Of course, the Bureau was smaller and less complex than it is now, and it is probably easier to move more quickly under such circumstances.

Let me say a few words about the atmosphere at that time. Although most of the statistical staff was actually located in the various subject matter divisions, there was considerable interaction and interchange of ideas. There were weekly staff meetings of the senior research staff with Bill Hurwitz, although if particularly important issues were on the agenda, Morris would attend. I use the term senior and agenda rather loosely. Most of us were in our thirties, and there was rarely an agenda or organization to the meetings. Staff members brought up whatever problems were bothering them, and there was usually a spirited discussion. The decibel level of the arguments sometimes exceeded the amount of reasoning. However, for me at least, it was a valuable learning experience, both in sampling methods and in how to get things done. There was complete democracy, and we all felt free to argue and disagree with Bill and Morris. Of course, when we disagreed with them, they usually convinced us that we were wrong!

3. Extension of Research to Other Survey Methodological Issues

The 1950 Census had a number of striking innovations. Sampling was introduced for the first time as an inherent part of the data collection. The 1940 Census did include a sample component, but it was restricted to newer and less critical items, whereas in 1950 some of the basic items were collected from a sample. Secondly, a sample of enumeration districts was selected for early processing; they served as the basis of a set of preliminary reports of the Census. (The plan resembled the system currently used by the television networks to project election results before the counting of the votes is completed.) Thirdly, a large-scale Post Enumeration Survey (PES) was attached to the Census. A fourth innovation consisted of the randomization of a sample of interviewer assignments in order to measure the between-interviewer variance. Finally, an experiment was carried out – a mail census with self-enumeration.

At the same time, work began on the development of a model that would describe and permit more detailed analysis of nonsampling errors. This model partitioned nonsampling errors into bias, simple response variance, and correlated response variance. It had a particularly important effect on methods used for population and housing censuses.

Research on bias included examination of coverage error, and a little later on rotation group bias, nonresponse effects, and a detailed examination of reporting on labor force. Studies of correlations within interviewers' work were mostly restricted to the decennial censuses, although one experimental study was done in the National Health Interview Survey. However, the sample sizes in sample surveys were mostly too small to achieve stable estimates of intra-interviewer correlations.

Most of us in the research staff looked on these studies, and the sampling research, as a way of unlocking the secrets of effective survey design. We were trying to unravel the DNA of survey structure. We were, of course, not completely successful but I think we outlined a path that is still being followed. The emphasis on total error also led to attempts to control errors better in both censuses and sample surveys. Formal programs were instituted in both the censuses and sample surveys to control and to measure the quality of the data collection.

4. Changes in Census Method

In a paper prepared for a memorial meeting to honor Bill Hurwitz, Leon Pritzker and I described changes that took place in census methods, essentially between 1940 and 1970 (Waksberg and Pritzker 1969). I will briefly summarize the content of that paper.

The period from 1940 to 1970 was characterized by major revisions in methods of taking the census which in many ways paralleled the research being carried out. Starting in 1940, and with increasing emphasis in subsequent censuses, the entire range of population census activities was examined to determine what were the sources of errors or other inaccuracies in the census, and how, for a given budget, the total effect of all errors combined could be minimized. Two aspects of the effect of errors came under scrutiny. The first was the question of the required accuracy and the extent its improved accuracy is worth additional cost. Attempts to evaluate the cost of achieving different levels of accuracy and their effect on analyses of data were stimulated by the fact that this type of thinking was fundamental to sampling practice. The second was the increased knowledge of the relationship of the various sources of errors, and that the appropriate methodology was to minimize the total error. Simultaneously, the new developments in sampling theory and methods produced an awareness that practical methods of research existed for investigating methodological problems in census taking. A large number of research studies to measure the extent of problems in censuses and to see what methods of overcoming them were possible.

Let me list the major changes in demographic censuses during this period.

1. The first changes were an extension in the scope of sampling. The 1940 Census utilized sampling for the collection of some of the data and the preparation of some of the detailed tabulations. The use of sampling was significantly extended in 1950, but there were still some restrictions on the determination of the items that were collected on a sample basis. In 1960, the restrictions were almost

completely removed, and sampling was used for all items except those virtually required to define the population, and housing items needed for block statistics. This pattern was repeated in 1970, 1980, and 1990. For the year 2000 Census, the Bureau is attempting to rely even more on sampling.

Four factors brought about general acceptance of this extension of sampling. (a) The developments in theory produced an awareness of the scientific bases for sample estimates. This was bolstered by such experiences as the 1937 Enumerative Check Census. (b) The 1950 evaluation program provided evidence that for most census statistics, the introduction of sampling would have only a minor effect on the total mean squared error for a census tract or larger area (Bureau of the Census 1960). (c) If the cost savings from sampling were used for more intensive training of the field staff and more quality control, it was possible to attain equivalent or more accurate data through the use of sampling. Furthermore, the utility of data would be considerably increased by the reduction in the time interval between data collection and publication of results. This was an important consideration before computer processing was common, or the use of CATI and CAPI. (d) The sampling operations were carried out in the field operations reasonably well. This is not to say that some biases in sample selection did not turn up, but they were not large enough to have a serious effect on most uses of the data.

2. Probably the most drastic change in census methodology has been the introduction of self-enumeration. Through 1950, with minor exceptions, the information was collected by an enumerator reading the census questions to one respondent in each household and recording the replies. In the training of the field staff much of the emphasis was to read the questions exactly as worded. Even a cursory observation of enumerators showed serious lapses in following these instructions. The 1950 interviewer variance study showed that this resulted in very large response variances for small areas, such as tracts and small cities and counties. Self-enumeration's main purpose was to reduce this component of variance. Self-enumeration was first introduced in 1960, and comparisons of the quality of data in 1960 and 1950 showed that the expected improvement did occur. Furthermore, it was accompanied by reductions in bias and random response variance, probably because self-enumeration made it possible for more household members to participate in filling out the census forms.
3. A mailout-mailback census was introduced in about two-thirds of the United States in 1970, and covered almost the total country in 1980 and 1990. The Bureau plans to retain a mail census in the year 2000. After self-enumeration was accepted in the 1960 Census, there no longer seemed any need for interviewers to deliver or receive the questionnaires. Probably a more important incentive for a mail census was that the preparation of a mailing list provided more time and opportunity to check the coverage of households in the census. By 1970, the deficiencies in census coverage revealed by demographic analyses and the Post Enumeration Surveys had become well known, and checking and double checking mailing lists seemed a sensible way to attack part of the problem. Thirdly, the increasing entrance of women into the labor force was beginning to present difficulties in the recruitment of interviewers and a mail census reduced the number of enumerators needed.

4. Procedures for improving coverage were introduced in censuses. They involved such methods as: enlisting the support of city officials in checking mailing lists, attempting to gain the confidence of local communities by working with community groups, and much more intensive public relations efforts.

5. Sample Surveys

Since the work in sampling theory has already been discussed, I will just briefly mention other principal activities in sample survey research. Considerable analysis was carried out on alternative rotation group patterns and associated estimation methods. The relative advantages of list versus area samples were examined. Computers and FOSDIC were introduced into sample surveys in the mid-1950s. In addition to their drastically reducing the data processing time and the reduction of data entry errors, computers made possible the use of more sophisticated estimation methods such as composite estimation and calculation of variances in a fairly rigorous manner. Starting in the late 1950s, a half-sample replication method (nonbalanced) was used to estimate variances. Balancing was studied and proposed by Philip McCarthy a little later. The program for computing total variances was quite quickly adapted to estimate variance components and to seasonally adjusted estimates. Computers also brought about the possibility of better methods of seasonal adjustments and Julius Shiskin, who was on Hansen's staff in the late 1950s and early 1960s developed the X-11 program.

Most of the sampling methods currently used were developed by the early 1960s. There were later refinements, and they are still going on, but no major breakthroughs.

6. Some Personal Experiences

I have been asked to describe some unusual or memorable personal experiences, and I have chosen ones that illustrate different aspects of what I consider the appropriate role of the statistical methods staff in large organizations.

6.1. Do not be afraid to get involved in all parts of survey operations

The 1970 Census was carried out partly by mail and partly by having an enumerator drop off the questionnaire with later pick-up. The latter was referred to as the ''conventional census,'' as distinct from the ''mail census.'' By 1970, most of the data items were on a sample basis and there were separate questionnaires for sample and nonsample households. In the nonmail areas, covering about one-third of the United States, each enumerator entered all addresses in the assignment area on a listing sheet which was premarked to show the households in the sample. An examination of an early pretest showed a very large excess of vacant units in the sample and a smaller than expected number of occupied units. When my staff examined the completed listing sheets, they found evidence of enumerators deliberately arranging the order of their visits to get vacant units into the sample. This phenomenon had not occurred in the 1960 census that used a similar listing sheet. In exploring possible motivations to manipulate the order in which households were listed, we discovered that the payroll procedures were different in the two censuses. In both censuses, payment was by piecework. In 1960, the payment was based on the number of persons, and the number of households, separately for sample and nonsample cases. The payrates for each of these components corresponded generally to the amount of time the enumerator spent. In preparation for the 1970 census, the Bureau's field organization had decided the 1960 system was too cumbersome, and changed it to a flat rate per household. Since the proportions of sample households were almost identical among the enumeration districts, and since average household size was fairly stable, the new system did not appear to introduce inequities in payments. However, the field organization did not count on the law of unexpected consequences. It did not take long for interviewers to learn that since vacant units were easier to enumerate, they could make the same amount of money by arranging to put vacant housing units in the sample. Of course, not all enumerators operated this way and the enumerators who did, could conveniently manipulate only some of the vacant units, but the net effect was clearly visible. Enumerators had difficulty learning many of the census concepts and rules, but it did not take very long for them to learn what was to their advantage.

I went to see the Chief of the field operations division and told him what had happened and that it was necessary to change the method of paying enumerators. He immediately pointed out that management was his job, and in effect, to mind my own business. We argued for a while, but when I said I was going to the Director and tell him the integrity of the Census was threatened, he gave in. The method of payment was apparently the key to the problem because the Census did not have any excess of vacant units in the sample.

6.2. *Consider the implications of what you know about survey methodology on data analysis*

In 1960 or 1961, the Department of Housing and Urban Development (HUD) asked the Bureau to develop plans for what is now the American Housing Survey. HUD indicated that it wanted annual national data and statistics for a selected number of metropolitan areas. I believe the budget permitted about 14 metro areas. The analysts at the Housing Division picked out 14 areas which were large and had diverse characteristics and started to work on procedures. I was called in, presumably to work on the sampling. Before doing that, I considered the kinds of data that would be produced and came to the conclusion that it was unlikely that useful data on year to year movements for the metropolitan areas would be produced. First of all, housing characteristics change fairly slowly, and in most cities the sample sizes were not large enough to detect trends for periods as short as one year. This appeared to be the case even though the sample design and measurement methods were specifically focused on measuring change. Secondly, the data were subject to significant measurement problems, even including what should be considered a housing unit, and this would make the statistics on annual change even more uncertain.

I proposed expanding the set of metro areas by a factor of three or four, and rotating the areas in and out of the sample on a three- or four-year cycle. The survey could thus produce data for a much larger set of areas, and at the same time more interesting data for each area would be available. As in any proposal to change existing thinking, objections were raised. However, the proposal was accepted and the rotating system is still in place 30 years later.

Similar issues keep on recurring. Last year, in responding to a Request for Proposal for the household survey on drug abuse, I found that the same policy was being followed – a small number of metro areas was being surveyed annually. Measurement errors in data on drug abuse must be much larger than on housing characteristics so the same logic applied. In Westat's proposal, I made exactly the same suggestion for this survey that I had made 30 years earlier. Unfortunately, Westat did not win the contract, but I was told my proposal was accepted and put in place.

6.3. It is sometimes necessary to be brash

In 1973, I testified before a Congressional Committee relating to the re-authorization of the legislation that distributed funds to school districts based on statistics on the number of school children below poverty in each district. The committee chair had been informed that minorities were seriously undercounted in the censuses and he wanted to know the effect of the undercount on the allocation of funds. I told the Committee what we knew about undercoverage and speculated briefly about its possible effect on fund allocation. Then, I stated that coverage probably had a minor effect, but that there was a much more serious problem with the data. I informed the Committee that the allocations for 1973 were still based on projections of data from the 1960 Census; the 1970 Census small area tabulations were still not available. By the time the 1970 Census results were issued, they would be at least four years old, and, in turn, they would form the basis of uncertain projections for the next eight to ten years.

The chair asked me what could be done to improve the data, and I suggested a sample survey designed to provide state data. Within a short time, funds were appropriated for the Survey of Income and Education (SIE). This experience may have been in the back of my mind when I heard about the Bureau's plans for continuous measurement which seems to me a better way of anticipating and alleviating problems of using outdated census information.

My comments at the Congressional hearings on the need for a better basis for the annual estimates were not planned in advance, nor cleared with the Director, but they occurred to me during the discussion at the hearings. However, Vince Barabba who was the Director and who also was at the hearings approved of new ideas and was just as much interested in my suggestion as the Congressional Committee. I have a feeling that things are much more structured nowadays, and that there would be inhibitions about bringing up subjects that had not been agreed upon in advance.

6.4. At the same time, know when to be discreet

By the mid-1960s, the Census Bureau had firmed up plans for a mail census in most of the U.S. in 1970. We had done considerable examination of the pros and cons and had concluded that a mail census would improve the quality of the data. However, we did not expect any reduction in cost.

The plans were presented at a Congressional hearing around 1965 or 1966. The Director and Morris Hansen testified. I went along to see what happened. The chair of the committee was very sympathetic and indicated he would support the Bureau. He noted that although a mail census would probably not produce quite as good data, the cost savings would make up for that. This, of course, was exactly the reverse of what we expected. I watched the Director and Morris to see their reaction. Both kept quiet, and the mail census was implemented.

6.5. Focus on the important objectives: Establish priorities and keep them in mind

This is probably the most common issue in survey design, and I suspect it is frequently ignored. I will give two instances that illustrate the range of issues in which consideration of priorities is important.

By the mid-1960s, the coverage problems in the Census had become clear and the research staff was busily engaged in examining possible ways of improving coverage. The mail census with self-enumeration and more attention on coverage in the questionnaires were expected to improve the counts. In addition, plans were developed for more intensive relations with community groups, particularly in minority areas in which coverage problems were most severe. Fifteen million dollars was set aside for this program. In 1968, the Bureau faced a budget cut. After cuts were made in other Bureau activities, a $15,000,000 shortage still existed. The Bureau's executive group did the obvious – eliminated the new program. By this time I was convinced that coverage was becoming the most serious problem in the Census and that the decision was a mistake. I had a long meeting with Con Taeuber, the Associate Director responsible for the decennial census, and proposed the restoration of the coverage improvement program, to be paid for by a reduction in the sample from 25 to 20 per cent. I apparently convinced him because that was how the 1970 census was carried out. In subsequent censuses, there were increases in funding for community participation.

A second example that comes to mind is also associated with the 1970 census. Some early checks on quality of the data, while the data collection was still going on, indicated that there were a fair number of cases in which occupied units were reported as vacant units. A crash program was developed to get more information, and a national sample survey of the vacant units referred to as the Housing Vacancy Check, was carried out in the summer of 1970. We estimated that about 10 per cent of the reported vacants were actually occupied. We considered whether we should adjust the census data. Note that ugly word 'adjust.' At that time there were no adequate methods of analyzing the effect of such adjustments on the population counts for small areas – counties, cities, tracts, etc. I rather quickly developed some theory to study the mean squared error of such statistics (Waksberg 1970, 1971), theory that I a little later generalized to apply to a much broader variety of small area estimates (Gonzales and Waksberg 1973). Applying the theory to data from the Housing Vacancy Check indicated that adjustments would improve data in moderately large areas, generally areas with 50,000 or more persons, but reduce accuracy in a fair number of smaller areas.

I thought that the improvements in state data and in the large cities and counties were worth the losses in smaller areas. I recommended adjusting the census, and that was done.

Perhaps this experience influenced my view on adjusting the census, when the drive began to do this for counts of the total population. It seems to me exactly the same issues as for the vacancy check are involved. Most statisticians believe that adjustments will improve the population counts for the total U.S., states, and the larger cities and counties. Adjustments, however, could adversely affect the counts for very small areas. Let me give a simple example. Assume you have a city in which 90 per cent of the blocks are counted perfectly, with the remaining 10 per cent all undercounted by 20 per cent resulting in an overall 2 per cent undercount. The sample Post Enumeration Survey correctly estimates the 2 per cent undercoverage but cannot tell which blocks were undercounted, so the adjustment procedure adds 2 per cent to each block. You have what is almost a Lewis Carroll type of paradox – 90 per cent of the blocks will have poorer counts and the remaining 10 per cent will be only slightly improved, but the count for the city as a whole will become accurate. As in the case of the Housing Vacancy Check, I think that the small deterioration in the block data is a sensible sacrifice for the improvement in the city counts and that this favors adjustment. The crucial uses of census population counts – apportionment, redistricting, allocation of funds – are mostly at the large area level. In many of the arguments against adjustment that I have heard from competent and knowledgeable statisticians, I thought their priorities differed from mine.

6.6. Do not expect to win all arguments

Many of the attempts to introduce new survey and census methods that grew out of consideration of the implications of statistical research were bitterly opposed by some staff members of the Bureau. There was opposition to the extension of sampling in the census, to the use of self-enumeration and to the mail census. In CPS, there were arguments about the replacement of the area sample in the U.S. by a list sample. I remember another dispute which, looked at in retrospect, indicates how hard it is to change established modes of thought. The CPS started off as strictly a labor force survey, and for the first five years or so, employment, unemployment and related information constituted the sole content. In the late 1940s pressure mounted on the Bureau to provide additional social and economic data. CPS seemed to be the natural vehicle for this. Some of the economic analysis at the Bureau felt that this would be betrayal of the purpose of the CPS. The suggestion that income be included in CPS was particularly opposed. I remember a memo from a high level Census staff member who predicted that adding income would kill the CPS. It is difficult to be a prophet in the twentieth century. I should note that the arguments about extending the content of the CPS were mostly between different segments of the analytic and operations staff, rather than between the research staff and other parts of the Bureau. The fact that most of the innovations were accepted and implemented was, to a large extent, because of the extraordinary skill and energy of Morris Hansen and Bill Hurwitz. However, we did not win all arguments.

When a mail census was decided on for 1970, the field organization insisted on restricting it to the metropolitan areas, containing approximately two thirds of the U.S. population. The field staff was concerned that the nature of mailing addresses in rural areas would make it difficult for enumerators to locate nonrespondents. The research staff felt that the Bureau had enough experience to overcome these problems, and that a mixture of two procedures would complicate the public relation effort to get people to respond by mail. Also, the pretests had shown that mail response rates were much higher in rural than in urban areas.

I remember accompanying Morris Hansen to a meeting with the Director, Deputy Director, and the Associate Director and Division Chief responsible for field operations where the issue was hotly debated. We lost the argument, and the 1970 census used two procedures. Of course, 1980 and 1990 used mail procedures throughout virtually the entire United States, and mail seems to be firmly established for the future, so we ultimately won out. In retrospect, I am somewhat sympathetic to our opponents' point of view. It is probably sensible to phase in revolutionary changes a little more slowly than we were driving for.

One argument we did not win related to data collection in the CPS. About 1961, following the suggestion of an advisory committee examining labor force data, we proposed making CPS a continuous weekly survey to provide labor force data for the average of the month instead of one week a month. I was not involved in studying or pushing that proposal; I believe Hal Nisselson and Max Bershad were the main research staff members developing the case. We never could get BLS (Bureau of Labor Statistics) to agree to this. Many years later, in preparing material for a talk in 1989, I had occasion to look over the published reports on employment and unemployment for the preceding two or three years and was struck by the number of times the report indicated erratic behavior of the seasonal factors that created uncertainty in the implication of the data. Generally, this was due to quirks in our calendar, so that the shifting holidays and school closings had unpredictable effects on data that described activities during the CPS week. The statistics for June, September, November, and December are particularly susceptible. Other problems are caused by occasional but dramatic events such as blizzards and floods that affect construction, agriculture and other outdoor activities. I suggested that BLS reconsider the possibility of a weekly survey (Waksberg 1989). Once again, BLS did not agree.

6.7. *Do not treat statistical procedures as mechanical operations; be prepared for the unexpected*

One interesting example of the law of unexpected consequences relates to statistical programs for early estimates on election night. The event took place a little before I became involved in these predictions. In 1966, the CBS network decided to speed up the reporting of elections by using a sample of election precincts. Observers were stationed at these precincts and they called in the results to a central office which immediately processed the results. The entire program was contracted out to a private market research company. The company followed its standard practice in sampling and data processing with disastrous results in the Governor's race in Maryland.

The Maryland Democratic party had gone through a bitter primary that year in which there were two liberal Democrats and one very right wing candidate. The civil rights struggle was the most pressing domestic issue of the time, and the third candidate mainly campaigned against open housing legislation with the slogan, ''Your home is your castle.'' The other two candidates split the more liberal vote, and the right wing candidate won the primary. The Republican candidate kept a fairly low profile during the campaign.

Black and liberal white voters abandoned the Democratic party and the Republican, Spiro Agnew, won by a comfortable majority. The company handling election predictions in New York did not seem to be aware of this situation. The computer program used to process the data included a search for outliers, which were deleted from the data file. A precinct was considered an outlier if its vote differed drastically from earlier elections.

Black and very liberal precincts which were traditionally very Democratic went heavily Republican in 1966, and were consequently tagged as outliers and deleted. CBS predicted a strong Democratic majority in Maryland, exactly the reverse of what actually happened.

This experience was an important factor in CBS's decision to set up their own unit for early predictions. They hired Warren Mitofsky and he asked me to act as a consultant. From the beginning, we have been conscious of the importance of avoiding mechanical rules. We did have a program to detect outliers. When one was found, we first tried to determine whether it was a transcription error or some other clearly correctable problem. Probably more important, two versions of all estimates were made – one including the outliers and the other excluding them. If they pointed in different directions or even if one indicated the results were only marginally statistically significant, we almost always delayed the prediction until more data became available.

Incidentally, by now Mitofsky and I have been involved in election predictions for 30 years. We have made several thousand predictions. I think there have been less than 10 wrong calls, probably closer to 5. We seem to have been somewhat too conservative in applying tests of statistical significance. When Morris Hansen started work on the Enumerative Check Census of Unemployment, it was felt necessary to demonstrate that a sample could adequately reflect the true figures. Election predictions provide thousands of empirical examples that this does occur. Waiting for the final figures as they become available several hours after election predictions have been made is an interesting, although somewhat anxious, experience.

6.8. There have been steady improvements in survey practice

Comparing the struggles for acceptance of the implications of the research findings with current survey practice clearly shows the substantial progress in the field. Certainly among the most important improvements is the integration of computers in survey methods. Even leaving aside uses for data collection such as CATI and CAPI, editing, tabular presentation, and microdata files, it is hard to remember that before computers, estimation methods were severely limited and that calculation of variances for complex designs was practically never done.

I remember developing the first plans for carrying out variance computations on a computer, Univac I. A half-sample replication procedure for CPA was used. I essentially designed the system and Bob Hanson, one of the members of the research staff, programmed it. He was considered an experienced programmer because he was a little less of a novice than I was. Even so, it took him three or four months of intensive work. Univac I had very little storage capacity, and intermediate results were consequently written out on magnetic tape to be used as input in the next step.

This was considered one of the longest and most complex programs for which the computer was being used so it was normally carried out at night. It generally took all night to calculate variances for a handful of items for one month's CPS. I remember baby-sitting with the computer during the first few months of variance computations. Part of the reason it took all night was that the system operated on hundreds or thousands of vacuum tubes, and each night a few of them would blow. A technician had to be called to locate and replace them. Meanwhile, information was destroyed and some of the steps had to be repeated. Other unusual events occurred to delay the process. I remember at one time putting one of the magnetic reels with results of some intermediate processing on a table, and little later a telephone was put on top of it. When the phone rang, the magnetic resonance scrambled the data on the tape, and we had to go back to earlier stages of processing.

One of the most curious things about Univac I was the method of mounting tapes. Here was one of the marvels of the age, and small clips, basically staples, had to be bent to fasten the reels to the machine. About half the time I did this, I pricked my fingers, and my blood was over much of Univac I. My fingers still hurt when I think of it. When I think of Westat's programs for calculating variances for a wide variety of variables on PCs, I am once again impressed with the progress that has been made.

7. Conclusion

A lecture like this should probably end with some deep philosophical insights and advice for the younger generation. I have tried to weave whatever insights and advice I have into the body of my lecture and there is no need to repeat them. Instead, I will conclude with a personal statement. I have been incredibly lucky to have been thrust, almost by accident, into a job that has been intellectually stimulating and in which I felt that I was making an important contribution to society. The contribution was partly in helping develop theory that improved statistical methods, and partly because the statistical methods were being applied to programs that shed light on most of the important social and economic issues of the day, many of which we still face. The challenges in extending statistical methods to new areas and attempting to get them accepted were exciting. Many of my present friends are ones I developed 30 or 40 years ago at the Bureau. I hope all of you have as much fun from your work as I did.

8. References

Bershad, M.A. and Tepping, B.J. (1969). The Development of Household Sample Surveys. Journal of the American Statistical Association, 64, 1134–1140.

Gonzalez, M. and Waksberg, J. (1973). Estimation of the Error of Synthetic Estimates. Proceedings of the International Association of Survey Statisticians, Vienna, Austria, August.

Hansen, M.H. and Hurwitz, W.N. (1942). Relative Efficiencies of Various Sampling Units in Populations Inquiries. Journal of the American Statistical Association, 37, 89–94.

Hansen, M.H. and Hurwitz, W.N. (1943). On the Theory of Sampling from Finite Population. Annals of Mathematical Statistics, 14, 333–362.

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

Neyman, A.J. (1934). On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. Journal of the Royal Statistical Society, 97, 558–625.

Olkin, I. (1987). A Conversation with Morris Hansen. Statistical Science, 2, 162–179.

Stephan, F.F. (1948). History of the Uses of Modern Sampling Procedures. Journal of the American Statistical Association, 43, 12–39.

U.S. Bureau of the Census (1960). The Accuracy of Census Statistics With and Without Sampling. Technical Paper No. 2.

Waksberg, J. (1970). Analysis of Synthetic Estimates. Internal U.S. Census Bureau Memorandum, J. Waksberg to T. Jabine, December 10.

Waksberg, J. (1971). Mean Square Error of Revisions in Population Count from Vacancy Recheck. Internal U.S. Census Bureau Memorandum, J. Waksberg to J. Daly, January 26.

Waksberg, J. (1989). New Directions for Some Household Surveys and Associated Research Needs: Discussion. Proceedings of the Fifth Annual Research Conference, U.S. Bureau of the Census, 48–53.

Waksberg, J. and Pritzker, L. (1969). Changes in Census Methods. Journal of the American Statistical Association, 64, 1141–1149.

Received July 1997