Statistical Science
1992, Vol. 7, No. 3, 320–338

U.S. Government Contributions to Probability Sampling and Statistical Analysis

Joseph W. Duncan and William C. Shelton

Abstract. The Federal Government of the United States has collected and published an increasing volume of statistics from the founding of the republic, but its contributions to statistical theory and method did not really begin until 1933. Before then, the bulk of Federal statistics was done by tabulation and compilation, and methods were largely intuitive. The Roosevelt New Deal and the Committee on Government Statistics and Information Services (COGSIS) made probability sampling and statistical analysis a significant part of Government planning and operations. By early in World War II, Federal statisticians had become leaders rather than just followers in statistical theory and methods.

This article provides a summary of how this happened and especially of the subsequent development of survey sampling from finite populations. Attention is then turned to the development of statistical analysis in the Federal Government, a more diverse subject, which is both related to probability sampling in significant ways and very interesting because it is probably still in an early stage of development.

This paper also provides commentary on some recent developments in the Federal statistical system in general during the period 1977 to 1992.

Key words and phrases: Probability sampling in the Federal Government, survey sampling, sampling from finite populations, statistical analysis in the Federal Government.

INTRODUCTION

From a geographical point of view, world leadership in the theory and applications of probability sampling has passed through three periods:

1. From the early 1800s when Gauss, Laplace, Legendre and others set forth the mathematical theory of the "combination of observations" until about 1890, leadership was on the Continent of Europe.
2. From about 1890 when Karl Pearson (persuaded by Francis Galton), G. Udny Yule and others developed regression and correlation and later R. A. Fisher, Jerzy Neyman and others introduced small sample theory and sampling from finite populations, until about the beginning of World War II, leadership was in Great Britain.
3. Since about 1940 when U.S. Government statisticians, later led by Morris H. Hansen and William N. Hurwitz, developed sample survey methods and theory for human populations and academicians led by Harold Hotelling, George W. Snedecor, Abraham Wald, S. S. Wilks and others made broader but less intensive contributions, leadership has been provided by the United States.

In each of these periods, statisticians attacked quite different practical problems and formulated different theoretical structures. Yet in the second and third periods, the theorists built firmly on the results of previous work. In fact, even the first period built on the laws of probability already established by the Bernoullis and others. The contributions of Gauss and others in the early 1800s were designed to estimate the paths of the planets across the skies from astronomical observations, although the methods were soon applied to a wide variety of scientific data and even to economic

Joseph W. Duncan is Vice President, Corporate Economist and Chief Statistician, The Dun and Bradstreet Corporation, 299 Park Avenue, New York, New York 10171. William C. Shelton is Retired Consultant, Statistical Policy Division, Office of Management and Budget, Washington, D.C.; his current address is 8401 Piney Branch Road, Silver Spring, Maryland 20901.

and social data. The early work of Karl Pearson and his colleagues, and a considerable part of their later work too, was designed to handle problems in genetics, anthropology and other fields of the life sciences. The work of Federal statisticians in more recent times was designed to obtain reliable current estimates of economic and demographic parameters for Government use in the Great Depression, World War II and later.

It is not surprising that such different problems required different conceptual and mathematical formulations. Rather, it is surprising that these different mathematical formulations fitted so well together into what we know today as the theory and practice of probability sampling.

Before passing to the contributions of U.S. Government statisticians in the third period, which is the subject of the present paper, one more point needs to be made to round out this historical introduction, namely, that Western civilization was in three very different phases of development in the three different periods, and this too had significant effects on the development of statistics.

1. The Gauss–Laplace period was part of a burst of intellectual energy which accompanied the French Revolution and the Napoleonic era and was one of a whole series of breakthroughs in the physical sciences at that time.
2. The Pearson–Fisher period accompanied the intensification of industrialization (which was probably the most important application of progress in the physical sciences) in Europe and its spread to the rest of the world, but its relationship to this development was not very close; its association to the impressive contemporary developments in the life sciences and agriculture was far closer.
3. The Hansen–Hurwitz period accompanied the assumption by the U.S. Government of far-reaching new powers to (1) recover from the Great Depression, (2) fight and win World War II and (3) manage the post-war reconversion and assume a position of world leadership which was thrust upon the nation after the war.

As will be seen, probability sampling and statistical analysis made significant, though hardly decisive, contributions to all of these needs.

PROBABILITY SAMPLING

There were few examples of probability sampling in the Federal Government before 1933, although a modest amount of common sense or intuitive sampling was done. Three stages of statistical collection before probability sampling (stage 4) need to be distinguished:

1. compilation by complete enumeration
2. "intuitive" sampling forced because complete enumeration is not feasible, a form of sampling without knowledge of the laws of probability
3. "structured" sampling which, while not of a probability character, is made with some knowledge of the laws of probability and which looks for and attempts to eliminate major biases.

This third stage includes, as an example, most forms of "quota sampling" which are still widely used in marketing and public opinion polls. The distinction between stages 2 and 3 is often one of degree rather than of kind.

In 1933 when the Roosevelt New Deal instituted a massive program for recovery and reform, the status of sampling in Federal agencies was, briefly, as follows:

- The Bureau of the Census was mostly in stage 1.
- The Bureau of Foreign and Domestic Commerce was partly in stage 1 and partly in stage 2.
- The Bureau of Labor Statistics was partly in stage 2 and partly in stage 3.
- The Bureau of Agricultural Economics and the Federal Reserve Board were rather clearly in stage 3.
- The State agricultural experiment stations (jointly funded by the Federal Government) were in stage 4, at least nominally.
- The Bureau of Internal Revenue in publishing *Statistics of Income* was in stage 4, using a simple probability sample (all returns with $5,000 or more of net income plus an 8% sample of all other returns).

Most other Federal agencies were in the compilation stage (Education, Interstate Commerce Commission, Public Roads) or were dealing with quite inadequate sample data (Public Health, Weather). At this time, not only censuses but statistical studies of all kinds in the Federal Government were carried out by rather large numbers of clerks and clerical supervisors, directed more by administrators than by professionals. Statisticans and subject-matter professionals with formal statistical training were few.

Franklin Roosevelt interpreted his election to the Presidency in November 1932 as a call "for action and action now." Within three months of his swearing in on March 4, 1933, he had not only reopened nearly all of the nation's banks (all of which had been closed for at least one week), but Congress in special session had passed more than a dozen major pieces of legislation aimed at recovery and reform. These and later laws established a national industrial recovery program, an agricultural adjustment program, an emergency relief (welfare) program, a regulation of securities and exchanges and insurance systems for bank deposits and home mortgages, to mention only a few programs.

Probability Sampling Takes Root in the Federal Government, 1933-1972

The needs for figures in these programs were tremendous since they involved not only statistics but the administrative figures required to establish and enforce production quotas and enforce other regulations for individual enterprises. A new and different look at Federal statistics was obviously necessary. The most fundamental contributions to Federal statistics in 1933 and 1934 were not made by a Federal agency but by the Committee on Government Statistics and Information Services (COGSIS).

COGSIS was sponsored jointly by the American Statistical Association (ASA) and the Social Science Research Council (SSRC), and it was financed for 18 months, June 1933 to December 1934, by the Rockefeller Foundation, but it was definitely a part of the Roosevelt New Deal. It grew out of an initiative by Secretary of Labor Frances Perkins, and its services were offered to and accepted by the Secretaries of Agriculture, Commerce, Labor and Interior. There was a crisis atmosphere in Washington, and the leaders of COGSIS, Edmund E. Day, Meredith B. Givens, Frederick C. Mills and Stuart A. Rice saw that they had a unique opportunity to do the following:

1. initiate badly needed statistical programs
2. make solid improvements in existing programs
3. establish a small permanent agency to coordinate Federal statistics.

COGSIS achieved all three of these results, and two additional results turned out to be extremely important, eventually perhaps even more important than the three intended ones: (1) COGSIS provided professionally trained personnel to direct some of the operating statistical work of Federal agencies, and (2) COGSIS stimulated research activities and innovative thinking which eventually had really revolutionary results in the development of the theory and applications of probability sampling, in establishing the national income and product accounts and in greatly extending the use of analytical methods.

COGSIS gave much attention to probability sampling. In order to put survey methods on a sounder basis, a Civil Works Administration (CWA) project was set up in late 1933 specifically to test different sampling techniques for measuring unemployment in three Northern cities. COGSIS, the Bureau of Labor Statistics and the Central Statistical Board provided professional direction; CWA hired relief workers as interviewers, and Census tabulated the results on punch card machines. Three types of samples were used: households, buildings and blocks.

Study of the results by a committee of the Central Statistical Board led to the conclusion that the random sampling of households was the most accurate of the available methods under the most favorable conditions, whereas the most feasible procedure on a nationwide basis was the counting of complete townships in rural areas and the block segment method in cities. Under the block segment method, representative blocks are cut into smaller segments by the use of city directories, and sparsely populated blocks may be combined, to achieve more uniform population per unit. It was also concluded that preparatory work for such a survey would be expensive and the attainment of accuracy would depend on having the highest quality supervision and enumeration.

This was apparently the first significant attempt to study the methods by which the theory of sampling could be used under the practical conditions faced in Federal economic and social statistics, and three important conclusions were correctly drawn from it. First, the conclusion was affirmative that sampling could be used as a practical method. Second, the smallest sampling unit (that is, the household) gave the smallest sampling error, and it was concluded that the sample designer should move to a larger unit only if practical problems required it, and then only to the smallest practical unit. Third, the use of samples requires much preparatory work and carefully controlled execution because practical problems are many and errors are multiplied, sometimes a thousand-fold, in going from a sample to an estimate of the parameter. All three of these conclusions are as true today as they were in 1934, though the principles are now spelled out in far greater detail.

This groundbreaking CWA study was never published, but a good account of it appears in the final COGSIS report, *Government Statistics*, published by the SSRC as Bulletin 26 in 1937, and its effect on the development of probability sampling in the Federal Government over the next five or more years was quite direct because the professionals who directed it, Calvert L. Dedrick, Samuel A. Stouffer and Frederick F. Stephan, and the administrators who reviewed it, Isador Lubin, Commissioner of Labor Statistics, among others, figured prominently in many later projects.

There were many Federal sample surveys in various subject-matter fields in the middle and latter 1930s, and most of them attempted to be of a structured or a probability nature. (The difference between these two stages of sampling was not clearly recognized at this time.) But nearly all were directed toward the resulting estimates, and contributions to the development of methods were mostly incidental. Sampling ideas were also applied to current reporting systems but less successfully.

Acceptance of Sampling. Acceptance of sampling by administrators and the public increased fairly rapidly but irregularly. The Literary Digest and other public opinion polls before the 1936 presidential election between Roosevelt and Alfred Landon initially had a negative effect. The Literary Digest poll was by far the largest, but telephone directories were used as the basic list and it was a voluntary mail survey. It predicted a large majority for Landon, when in fact he carried only two states. Smaller polls more carefully selected showed Roosevelt majorities. Gradually, the message got through that the distribution of the sample needs to be in direct correlation with the distribution of the votes and that this is much more important than the size of the sample except for very small samples. It has to be said, however, that by the late 1930s probability sampling had not yet proved itself in the Federal Government in the sampling of human populations or businesses.

Jerzy Neyman's Influence. What became the foundation article making the proving possible was published in the *Journal of the Royal Statistical Society* for 1934 by Jerzy Neyman. It is entitled "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection." First, he gave the formula for the sampling error of the mean from a stratified random sample for a finite population in which every stratum is sampled, and he showed that the sampling error can be minimized by allocating the sample among the strata in proportion to the $N_i\sigma_i$, where N_i is the number of elements in ith stratum and σ_i is the standard deviation of these elements.

Second, he pointed out mathematically in this article that the method of purposive selection, in which elements believed or calculated to be typical are selected for the sample and other elements are omitted, is likely to give biased results and that it provides no sound basis for an estimate of error. He, therefore, opposed its use except in special circumstances.

This article opened the door to accurate sampling error calculation and hence to near-optimum sample design for a host of statistical problems, especially in the Federal Government. W. Edwards Deming, who organized the teaching of statistics courses at the U.S. Department of Agriculture Graduate School in 1936 and who also held informal seminars on statistical methods from time to time, arranged for Neyman to come to Washington in April 1937 to hold a conference entitled "Sampling Human Populations." This conference had direct influence on Federal statisticians, and even more importantly it called attention to Neyman's article in a forceful way.

In the six years 1937 to 1942, four important applications of probability sampling were carried out in the Federal Government, each of which successfully applied a Neyman-type method to a particular important problem in order to (1) reach significant results and (2) develop the method by making it precise in particular directions. Estimates of sampling errors were published in some of these reports:

1. *The 1937 Enumerative Check Census of Unemployment.* This (a) demonstrated that total unemployment in November 1937 was about 11.0 million, rather than the 7.8 million persons who had registered, and (b) published sampling errors with a much more detailed description of methods of calculation than in prior studies. (Professional credit goes to Calvert L. Dedrick and Morris H. Hansen for this work.)
2. *Sampling in the 1940 Census.* The first use of sampling in the U.S. population census (a) permitted including many more questions in the census without increasing cost, (b) permitted earlier publication and more detailed cross-tabulation and (c) developed and applied a method of eliminating line bias. (Professional credit goes to W. Edwards Deming, Morris H. Hansen and Frederick F. Stephan.)
3. *Monthly Sample Survey of Unemployment.* The Work Projects Administration (WPA) brought its sample survey of unemployment (ancestor of the present Monthly Report on the Labor Force) to publishable quality in 1940 to (a) provide for the first time reliable estimates of unemployment and (b) introduce cluster sampling in a reasonably efficient way. (Professional credit goes to J. Stevens Stock and Lester Frankel.)
4. *Family Spending and Saving in Wartime.* The Bureau of Labor Statistics (BLS) and the Bureau of Home Economics (BHE) carried out this survey in 1942, and it (a) showed that family spending and saving patterns by income level had not changed as much as previously thought up to 1941 but had changed sharply in the first quarter of 1942, and (b) improved efficiency in nation-wide sample surveys by deriving and applying the formula giving minimum variance for a given total cost assuming that it cost X dollars to add an extra field office and Y dollars to add an extra household and actually estimating X and Y. (Professional credit goes to Jerome Cornfield.)

By about the time the United States entered World War II, probability sampling had taken root in the Federal Government. It was no longer an esoteric subject; it was a well-recognized method of obtaining information, and agencies were beginning to depend upon it. Most Federal administrators still thought of sampling as a cheap substitute for a complete count rather than considering probability sampling as a method in its own right which, as was later proved, is sometimes more accurate than a complete count. But between early 1933 and early 1942, not only had important developments been made in the techniques of applying the principles of probability sampling to the practical conditions faced in Government operations, but (1) many Federal administrators had been convinced that if they needed information, they should consider the possibility of taking a sample rather than making a complete count, and (2) a few Federal administrators had been convinced that instead of taking a controlled sample eliminating extreme cases (purposive selection), they should take a probability sample, allowing the extremes to enter approximately in the proportion in which they exist in the population.

In any case, by 1942 there was enough confidence in sampling to provide a modest budget for probability sampling experts at several Federal agencies, especially the Bureau of the Census.

During and soon after World War II, Government statisticians made many more contributions to probability sampling, and the most influential results were those achieved by Morris H. Hansen and William N. Hurwitz at the Bureau of the Census, and it is their 27 years of unique collaboration which must now be summarized because of their outstanding results and lasting influence.

Hansen and Hurwitz at the Bureau of the Census, 1942-1968

The Bureau of the Census was slower than some other Federal Statistical agencies in adopting probability sampling before World War II. In 1933 when the New Deal began, the Bureau was at a low ebb as a result of the Hoover Administration economy acts and the operation of the seniority system.

Reconstruction at the Bureau of the Census. William Lane Austin, the new Director, who had worked his way up through the career service, knew little of sampling, and COGSIS personnel were able to get Stuart A. Rice, a founder of COGSIS and a believer in sampling though not himself a mathematical statistician, appointed as Assistant Director, the number two position. Rice made slow progress in selecting a cadre of academically trained statisticians during the three years before he left to become Chairman of the Central Statistical Board (CSB). He brought Dedrick over from CSB as Assistant Chief of the Statistical Research Division (SRD) in 1935, and he became Division Chief in 1938. Dedrick's first appointment was Morris H. Hansen, who transferred from the personnel office in 1936, and they brought in William N. Hurwitz in 1940. Dedrick, with help from outside the Bureau, led the fight for the 1937 Enumerative Check Census of Unemployment and for the use of sampling in the 1940 census. After Dedrick left for civilian service with the War Department in 1942, Hansen became division chief.

Hansen had majored in economics, specializing in accounting at the University of Wyoming, and he obtained an M.A. in statistics at American University in 1940 with some courses at the Graduate School of the U.S. Department of Agriculture. Hurwitz had been a student of Hotelling at Columbia, and he developed a remarkable facility for solving problems in mathematical expectation and minimizing variance in complex sample designs. These two men were both in their early thirties, and they developed an unusual working relationship. They worked together on most problems with little specialization, although, as Hansen says, he never developed Hurwitz's facility in handling complicated mathematical formulations (see *Journal of the American Statistical Association*, 1969, p. 1123).

By 1942, the use of probability sampling was fully accepted at Census. Austin, who had opposed its use, retired in 1941, and the new Director, J. C. Capt, knowing that it had been used successfully in the 1940 census and finding that it was fully accepted by the statistical community and most Government administrators, became a strong supporter of it.

The earliest theoretical work of Hansen and Hurwitz was directed toward straightening out concepts and studying the relative efficiencies of sampling units of different sizes. Classical probability sampling theory from Gauss to R. A. Fisher had assumed an infinite population. Neyman, followed by Cochran, and Yates and Zacopany, had begun the consideration of sampling from finite populations, but no one had constructed straightforward realistic models of structured finite samples from finite populations. Hansen and Hurwitz did this first in connection with a proposed annual sample population census which never materialized.

A published result of this work is a rather brief but significant article by Hansen and Hurwitz in the 1942 *Journal of the American Statistical Association* entitled "Relative Efficiencies of Various Sampling Units in Population Inquiries." It uses a direct finite population approach, and it proves that the increases in sampling variance from increasing the size of the sampling unit or cluster while leaving constant the number of elements sampled is directly related to the intraclass correlation coefficient. Although there is usually a loss of efficiency in increasing the size of the sampling unit, if the intraclass correlation coefficient is a large negative, there can be a gain. For example, in estimating the sex ratio (proportion of males in the population), it is more efficient to use the household as the sampling unit than to use the individual because most families include at least one male and at least one female, thus greatly reducing the variance of the sex ratio.

This striking example drove the point home, but the greatest contributions of the article were (1) to provide a method of measuring loss of efficiency from using clusters and (2) to point out that the loss was very different for different characteristics, for different sizes of cluster and even for different kinds of sample design. Thus, multiple characteristic sampling cannot be highly efficient for every characteristic.

Monthly Report on the Labor Force. In 1942, WPA was abolished, and the *Monthly Sample Survey of Unemployment* was transferred to the Bureau of the Census. Because of the broader purpose which it was now to serve, its name was changed to *Monthly Report on the Labor Force* (MRLF). The sample design for the MRLF was naturally assigned to SRD, and Hansen and Hurwitz, assisted by John N. Webb and Lester Frankel, who transferred from WPA with the survey, got their first major opportunity to apply the theory of finite sampling which they had been developing.

Hansen and Hurwitz, with the assistance of Frankel, spent about a year in (a) analyzing the MRLF and revising the sample design and (b) using it as a vehicle to develop the theory of sampling from finite populations.

One significant bias which they uncovered was the fact that in open country areas, beginning the sampling process by selecting houses which were closest to a point (in this case, the southwest corner of a given one-square-mile section) meant that isolated houses had a greater chance of being selected than houses located close together. The newly devised sampling procedure avoided this bias through a strict application of probability methods, and the results showed at least one substantial difference from those of the WPA sample. This was a reduction of nearly 20% in the estimate of agricultural employment, other differences being much smaller.

The Hansen and Hurwitz article "On the Theory of Sampling from Finite Populations," which appeared in the *Annals of Mathematical Statistics* for 1943, seems to be the first well-rounded general article published by Federal employees on the subject of sampling from finite populations, and it is one of the first articles of this type published by anyone. It treated the MRLF as an instructive example of the principles which it discussed, and it estimated reductions in sampling variances of as much as 50% and even 75% in some national estimates by the use of improved methods without increases in cost (except overhead cost).

Its two most important contributions to theory were (a) a discussion of ratio estimates and the circumstances under which they are superior to best linear estimates and (b) the use of probability proportional to size. Both principles are discussed and applied in a complex multistage sample design. Neither of these concepts was new (for example, Cornfield had used both concepts in designing the sample and the estimating procedure for *Family Spending and Saving in Wartime*), but the exposition of the theory was new. And the mathematical formulation of sampling error and the bias of various alternatives which are presented in the article permitted a discussion in quite practical terms of the advantages of these two recommended procedures, of the conditions under which they would be superior and approximately by how much they would be superior.

Recognition at the Bureau of the Census. The progress in sample surveys, both theory and practice, made in 1942 and 1943 was recognized at the top in the Bureau of the Census as well as elsewhere, and in 1943 Hansen was made Statistical Assistant to the Director of the Census and a member of the Executive Staff. The organizational structure and titles changed a number of times in the next 25 years, but the team of Hansen and Hurwitz remained intact. With support from Director Capt and his successors, first the use of sampling and later increased control of other forms of survey error were extended to one subject-matter division after another in the Bureau.

The Census of Agriculture in 1945, including what became known as the Master Sample of Agriculture, and some other jobs provided different combinations of the problems Hansen and Hurwitz had faced, and sometimes problems which had been minor became major and could no longer be ignored. But the next group of problems which required a major new approach arose in the establishment statistics as distinct from the individual and household estimates, and the first case in which a frontal attack was made on sampling nonfarm establishment statistics efficiently was in the estimates of retail trade, especially retail sales.

Sample Survey of Retail Stores. Shortly after World War II, Hansen and Hurwitz looked into this problem, realizing that monthly estimates would be needed by some kind of business and that public attention would be focused on month-to-month and year-to-year changes. As it became clear how great the problems were, especially for the independents (that is, excluding chain stores), Max A. Bershad was the staff member designated to take the lead in finding solutions.

The biggest reason establishment statistics require a substantially different approach from households is skewness; that is, a small percentage of establishments have a large proportion of sales. But instead of measuring skewness by the third moment, as Karl Pearson had done, Hansen and Hurwitz found it much more useful to calculate what they called "rel-variance," that is, the square of the ratio of the standard deviation to the mean (the square of the coefficient of variation). In simple random samples, the rel-variance is proportional to the size of the sample necessary to obtain results within, say, a 1% error. The rel-variance of retail stores classified by sales in the 1939 Census of Retail Trade was nearly 50. By excluding stores with more than $300,000 of sales (8,000 stores with 21% of total sales), the rel-variance was reduced to less than 1. This calculation made it obvious that an efficient sample must include all of the large stores and thereby restrict the sampling error to the small- and medium-sized stores.

The second big problem in sampling retail stores is the high turnover rate. This was known qualitatively before the sample design work was begun, and a survey carried out during this work estimated it at an overall average of 15% per year. This fact indicated that an area sample would be necessary to eliminate the new-store bias which had plagued the previous identical store sample. An available sample structure was the set of 68 areas used for the *Monthly Report on the Labor Force*. Checks with the 1935 and 1939 Censuses of Retail Trade showed that using the MRLF weights for combining retail sales by kind of business in these 68 Primary Sampling Units (PSU's) into national totals gave figures close to the census totals in both years.

Hence, the major part of the problem was reduced to designing a sampling procedure for retail stores within the 68 PSU's. It was necessary to obtain and keep up to date a complete list of large stores for these areas. This was done with considerable care, starting with a list from the Old Age and Survivors Insurance System (OASI). For retail stores not on the specified list of large stores, a sample of approximately 2% of all retail stores in the stratum (not the PSU) was drawn by quite a complicated procedure of selecting area segments. The large stores reported every month; and the stores in the sample were divided randomly into 12 groups, and they reported one month during each year, giving sales for the current month and the preceding month.

For estimating nation-wide retail sales by kind of business in any month, two methods were available. The estimate for each PSU could be blown up to an estimate for the stratum (and the strata summed) either by multiplying by the ratio of 1940 population (the reciprocal of the factor originally used as probability of selection) or by multiplying by the ratio of 1939 retail sales. The first of these is the simple unbiased estimate; the second is a ratio estimate. Sampling errors were calculated for both estimates, and, as expected, the ratio estimate had a smaller sampling variance. It was, therefore, used for the published estimates.

For its year of establishment, 1948, there were the usual problems of getting started, but after the 1948 Census of Retail Trade became available and mandatory reporting authority was enacted by Congress in 1949, the program became quite successful. Thus, an efficient method was developed to use probability sampling to estimate national aggregates on a current basis for an industry having millions of establishments, high skewness and rapid turnover – all three! For the many practical problems encountered and how they were solved, see Bureau of the Census Technical Paper No. 1, *The Sample Survey of Retail Stores*, published in 1953.

The Book – *Sample Survey Methods and Theory.* In 1949 and 1950, Hansen and Hurwitz were chiefly concerned with establishing the *Annual Survey of Manufacturers* and the sample designs, postenumeration surveys and other work for the *1950 Census of Population, Housing and Agriculture*.

Along with William G. Madow, they had obtained a contract with Wiley and Sons to write a book, which finally appeared in 1953 in two volumes under the title *Sample Survey Methods and Theory*. As the work on the 1950 census became less intensive for them, they spent more time at night working together on what soon would become recognized as the standard work on survey sampling.

Volume I, *Methods and Applications*, is a complete treatment except that it does not contain derivations and proofs, which are in Volume II, *Theory*. Hansen and Hurwitz wrote virtually all of Volume I, but Madow was very much involved in Volume II. The book had the full support of the top officials of the Bureau of the Census and active encouragement from Howard C. Grieves, the Assistant Director. These officials realized, as Government administrators do not always do, that these new methods should be widely applied by other governments and by business and that the theory would be immensely useful to the statistical profession.

Much of the importance of the book lies in its systematic treatment of sampling from finite populations. Deming produced a book three years earlier than Hansen, Hurwitz and Madow under the title *Some Theory of Sampling*, and it contains some parts of what is in the later book. This title was well chosen because Deming's book is not confined to finite sampling but includes considerable parts of biometrics, quality control and what was once called the combination of observations. Deming's book was of great use before the Hansen, Hurwitz and Madow book appeared, and it is still of use in a more general context. There was so much professional interaction among these four people, and to a lesser extent with Cochran and others outside of Government, that it is dangerous to attribute specific ideas to any one person without painstaking research. But because Hansen and Hurwitz faced the practical problems of finite sampling in many different frames and had a competent staff to assist them, their contribution was preeminent in the field.

Total Survey Error. During their last 15 years of working together, Hansen and Hurwitz continued to contribute to the theory and methods of finite sampling, but their most innovative work was in the broader field of improving efficiency and controlling total survey error. Efficiency was improved, for example, by increased use of administrative records. Lists and data from the OASI and from Internal Revenue were used (a) to provide lists of firms, in which use they replaced the field canvass, and (b) for data for small establishments, in which use they reduced or eliminated the need to send out questionnaires. Also during this period, Hansen continued to be important in the development of computers and their applications to statistical work, especially FOSDIC for the automatic handling of input.

Total survey error includes, in addition to sampling error, (a) data processing errors, which Deming had made a large contribution to controlling in the 1940 Population Census, (b) data collection errors and (c) errors of concept and definition, which are primarily the responsibility of the subject-matter specialist but for which the statistician can often make important practical suggestions. Hansen and Hurwitz instituted the Post-Enumeration Surveys (PES) first in connection with the 1945 Census of Agriculture and on a full scale with the 1950 census, and they proved to be especially useful in measuring data collection errors.

A major part of data collection error is interviewer error, and Hansen and Hurwitz and their staff studied this problem intensively in the decade of the 1950s, including a large Enumerator Variance Study. An important change in census procedure from this work was a major switch from interviewer enumeration to self-enumeration.

Appraisal. In assessing the effect of Hansen and Hurwitz (they really cannot be considered separately) on the development of the theory and methods of probability sampling, it seems clear by now that their contribution is of the same order of magnitude as that of each of their three great modern predecessors, Karl Pearson, R. A. Fisher and Jerzy Neyman. Differences in problems worked on and approaches taken are so great that it is foolish to attempt direct comparisons. Since perhaps the most important question is the extent of innovation, it is interesting to contrast the work of Hansen and Hurwitz with that of R. A. Fisher. (These are the two of the four who have written comprehensive books presenting their methods.)

Six of the major differences introduced by Hansen and Hurwitz are as follows:

1. They assume a finite population.
2. They make great use of ratio estimates instead of best linear unbiased estimates.
3. With stratification by size, they insist on complete coverage of large establishments.
4. They deal with sampling for multiple characteristics.
5. They often use multistage sampling procedures.
6. They apply their methods to current reporting systems, where month-to-month and year-to-year changes are frequently more important than absolute level.

Developments Outside of the Bureau of the Census After 1942

Bureaus other than Census also made great strides in probability sampling after 1942. Many of these developments were influenced or even assisted by the Bureau of the Census, but some of them were independent of it. The Division of Statistical Standards (DSS) at the Bureau of the Budget made a serious effort to spread probability sampling to all Government statistical agencies, and Deming was responsible for this function part-time from 1942 to 1947 and full-time from then until 1953. There is space in this article to mention only a few of the most notable achievements by these agencies, concentrating on those which faced different problems and contributed to methods and theory.

Agriculture. The Department of Agriculture continued, improved and greatly expanded its sampling work. Led by Iowa, the State agricultural experiment stations, jointly funded by the Federal Government, were using the R. A. Fisher methods to some extent in the 1920s and 1930s, but it was the 1940s before these and other experimental designs and methods became standard practice generally, with a tremendous payoff in research results and farm productivity.

In the 1960s, probability sampling was successfully applied to monthly crop forecasts. Research work on objective measurement of crops by randomly selected small plots for observation of growth and, finally, harvest measurement had begun as early as the 1930s, and in the 1950s this work had been expanded on the recommendation of O. V. Wells, Chief of the Bureau of Agricultural Economics. The special and most difficult problem in this application of sampling is the objective translation of actual measurements at a particular stage of growth into an estimate of yield assuming "normal" weather and treatment between time of measurement and harvest.

With substantial increases in appropriations for research beginning in 1960 and a reorganization setting up a Statistical Reporting Service (SRS) at bureau level in 1961 under Harry C. Trelogan, the back was broken on this and related problems by 1966. By that year, objective sampling methods were in use in all 48 contiguous states. Earl E. Houseman, Bruce W. Kelly and Walter A. Hendricks deserve much of the credit for working out and applying the statistical techniques used. See *The Story of U.S. Agricultural Estimates*, U.S. Department of Agriculture miscellaneous publication No. 1088, April 1969.

Family Expenditure Surveys. At the Bureau of Labor Statistics (BLS), the successive family expenditure surveys of 1950–51, 1960–61 and 1972–73 built on the earlier work of Cornfield and made use of probability methods increasingly, not only in the selection of cities and households but also in the selection of items to be priced and establishments at which the prices were to be obtained and for general methodological purposes.

A peculiarly important problem in family expenditure studies, whether for revision of the Consumer Price Index or for broader uses, is the fact that most consumers do not keep accurate records of expenditures. In the expenditure surveys from 1934–36 to 1960–61, BLS depended primarily on the "recall method." Work in foreign countries depended much more on recordkeeping (diaries), and a study by Joseph Waksberg and John Netter at the Bureau of the Census in 1964 (Census Technical Paper No. 13) tended to confirm the rather large errors in the recall of infrequent and irregular expenditures over a period as long as a year. Recordkeeping may involve errors of a different sort; for example, families tend to draw down inventories during the record period and thus reduce postponable expenditures. Hence, the problem of measurement is complicated and is one which cries out for careful testing by probability methods.

In the 1972–73 expenditure survey, for the first time, serious efforts were made, in spite of high dollar cost, to combine recall and record methods in order to minimize errors for all expenditures. Robert B. Pearl and Julius Shiskin, working both at BLS and at the Office of Management and Budget, were the central figures in pushing through a complicated probability design for this purpose, with the sample design and field work being done at Census and at Westat, Inc. This procedure, though costly, proved quite successful and is still used in the family expenditures survey, which is now continuous rather than periodic. The best summary reference for this work is in "A History of the U.S. Consumer Expenditure Survey: 1935 to 1988," by Eva Jacobs and Stephanie Shipp, which appeared in the *1990 Proceedings of the Social Statistics Section of the American Statistical Association* (pp. 233–238).

Sequential Analysis. During World War II, a notable contribution to probability sampling theory was the development of sequential analysis by Abraham Wald. This work was done at the Statistical Research Group at Columbia University financed by the War Department. It was given a security classification during the war, but Wald's book, *Sequential Analysis*, was published in 1947. Instead of calculating probabilities for a fixed sample size, as was done in earlier sampling theory, repeated samples of a given size are drawn. Then instead of simply accepting or rejecting a hypothesis after each sample, a third conclusion is also permitted, namely, that the sample is insufficient to draw a conclusion and further samples must be taken.

Wald developed the mathematical theory for optimizing the decision rules under different conditions and assumptions. He considered chiefly the applications of sequential analysis to quality control in the physical sciences, but applications to social sciences and commercial research have also been made.

Satellite Sensing. The availability of satellite sensing has opened up a new field for the application of probability sampling. In the 1970s, both acreage and crop estimation from U.S. satellites became possible not only for the United States but for the world, and it is actually being used. Also, in natural resources, energy and the environment, the Earth Resources Observation System (EROS) was begun. This is a program for acquiring, processing, distributing and applying remote sensor data collected from aircraft and spacecraft toward the solution of resources and environmental problems.

A large number of other applications of probability sampling have been made in the Federal Government, and quite a few of them were very important for government operations. The National Health Survey, the National Crime Survey, the censuses of transportation, housing starts and other sample construction estimates and major improvements in the IRS's *Statistics of Income* and the OASI's *Continuous Work History Sample* are examples. In many cases, their methods as well as results were published, and there no doubt are cases where they have enriched sampling theory in significant ways.

STATISTICAL ANALYSIS

Broadly speaking, growth in the use of statistical analysis in the Federal Government has paralleled the development of probability sampling except that there was no person or pair of persons who provided a unifying framework of methods and theory. Of course, there could hardly be because statistical analysis overlaps and intermingles so much with analysis in the subject-matter fields in which statistics is applied and these fields cover most of scientific and social knowledge. The techniques used in different fields, although often similar or at least analogous, do not provide a strong unifying force.

Federal statistical agencies were doing occasional analytical work at least as early as the last part of the 19th century. In the 1920s, the leaders in this work were the Bureau of Agricultural Economics (BAE) and the Federal Reserve Board, but until 1933 there was comparatively little analysis done in other Federal agencies. The first important book to be issued as a byproduct of Federal statistical activity was *Methods of Correlation Analysis* by Mordecai Ezekiel (Wiley, 1930) of BAE. This book included an exposition of the method of graphic correlation analysis, which had been worked out by Louis H. Bean, another leading statistician in BAE, and published in his article in the *Journal* *of the American Statistical Association* for December 1929. The method is both rapid and not restricted to linearity. A very important contribution of the book was the clarity and simplicity of its presentation, making it well suited to statisticians with limited mathematical background. The book was used as a text in the U.S. Department of Agriculture Graduate School and elsewhere for many years, and it went through three editions.

The New Deal approach to governmental problems, and especially the influence of the Committee on Government Statistics and Information Services referred to under Probability Sampling, brought about a rapid spread in the use of statistical analysis to many Federal agencies during the 1930s. The simpler types of analysis, including seasonal adjustment and simple regression lines, became commonplace and, in some agencies, standard operating procedure. Multiple correlation and regression were still not often used for publication, though they were frequently a part of the underlying analysis. The professionalization of Federal statistics which took place in this period is very evident in even a cursory comparison of Government statistical publications around 1940 with those around 1930.

National income estimates and the statistics used in preparing them made much progress during the Great Depression both in soundness and in currency, and much of this work was done in the National Income Division of the Department of Commerce. One of the most important uses of them was in fiscal analysis. Many economists subscribed to the economic theory of inadequate purchasing power first under the slogan "pump priming" and later "compensatory spending." Increasingly, they used national income as the analytical framework for their estimates and recommendations. After the appearance of John Maynard Keynes's book, *The General Theory of Employment, Interest and Money*, in 1936, and the onset of the 1937-38 recession, this work was intensified. The Roosevelt Administration never specifically supported this approach, but there is little doubt that its policy decisions were influenced by it. Lauchlin Currie and Martin Krost at the Federal Reserve were leaders in this work.

The Victory Program and the Feasibility Dispute

During the defense build-up of 1940–41 and the ensuing years of World War II, the analytical framework of the national income was modified specifically to meet the needs of wartime economic planning by the introduction of the gross national product (GNP) framework. This effectively separated the final product side from the income side of what are now called the national income and product accounts (NIPA). This change was central to clear thinking on the extent to which economic resources could be mobilized for defense production. It made possible the establishment of the Victory Program announced in January 1942 by President Roosevelt as the maximum military production program that could be mounted by the United States at that time. It also led to the rejection by the War Production Board (WPB) of the armed services' request for a further increase in the military program later the same year in what is known as the feasibility dispute.

Robert R. Nathan and Simon Kuznets were the key analysts at WPB in these historic decisions. Both were primarily economists, but each had many years of experience in the statistical construction and analysis of national income accounts. It was Milton Gilbert, chief of the National Income Division at the Department of Commerce, who codified the logic in a paper presented at the December 1941 meeting of the American Statistical Association and printed in the June 1942 issue of the *Journal of the ASA*.

Operations Research Groups

Also worthy of note during World War II, even in the present brief account of statistical analysis in the Federal Government, are two operations research groups in the armed forces. The first was the Antisubmarine Warfare Operations Research Group (ASWORG) under Professor Philip M. Morse of the Massachusetts Institute of Technology which dramatically reduced allied shipping losses. The second was the Naval Ordnance Laboratory Operations Research Group (NOLORG) under Ellis A. Johnson and Walter Michels which designed the highly successful mining of Japanese-controlled waters. Many other groups in all three services, however, made valuable contributions. This work was done chiefly by physical scientists, but it involved substantial amounts of statistical analysis.

In the three decades following World War II, statistical analysis made progress in so many different subject-matter fields in the Federal Government and in such different ways that it would be quite impossible to cover them in a meaningful summary. The only recourse seems to be to select a few outstanding and heterogeneous developments which have had significant effects on academic work, business and finance, and the general public. The four developments selected are as follows:

1. Linear programming, a method an important part of which was developed in the U.S. Air Force.
2. *Business Cycle Indicators*, developed by the Department of Commerce to present current business and economic statistics in a form which emphasizes timing relationships as a basis for business cycle forecasting.
3. The report, *Smoking and Health*, by the Advisory Committee to the Surgeon General which directly faces the question of statistical causation.

4. Attempts to measure social welfare, which have so far crystallized in two editions of *Social Indicators.*

Linear Programming

Linear programming (LP) is a method of mathematical and statistical analysis which rates, along with sequential analysis and many other forms of probability sampling already discussed, as one of the principal contributions which the Federal Statistical System has made to academic theory and methods. LP methods were being worked on in the years after World War II by the U.S. Air Force, by Koopmans and others at American universities and by experts in the USSR. The simplex method of solving LP problems systematically was developed by George B. Dantzig in the U.S. Air Force, and this was the breakthrough which made possible the rather rapid spread of LP theory in academia and LP practice in industry.

Mathematically, the linear programming problem can be stated as that of minimizing a linear form (what statisticians usually call a product sum) subject to a system of linear inequalities. For example, a person has daily nutritional requirements for known amounts of calories, protein and various minerals and vitamins, etc., and if he or she wants to meet these requirements at minimum cost, the linear form is total food expenditure, which is the sum of products of the (known) price multiplied by the (unknown) quantity for each food consumed.

The minimization has to be subject to a series of conditions specifying that total calories shall be at least equal to a given figure, total protein at least equal to another given figure and similarly for each mineral and vitamin. The nutritive content of each food is assumed known.

Early procedures worked out by Dantzig began by finding a feasible solution, that is, a set of quantities which provided adequate nutrition but which was not necessarily at minimum cost, and then proceeded to reduce the expenditure by making substitutions in a prescribed order to arrive at the least-cost solution. The rules prescribed form an algorithm, and, when written out in computer language, they become a program. This became known as the simplex algorithm and, with additions and refinements, as the simplex method.

When Dantzig wrote his first paper on this in February 1948 (published after changes in 1951), calculation was often tedious, but with the invention of computers and their development in the 1950s and 1960s, the algorithm was programmed on computers, and this was such a major time saver that it made the method far more useful.

Dantzig's book *Linear Programming and Extensions* (Princeton, 1963) not only sets forth the simplex algorithm but gives procedures for obtaining a first feasible solution or bypassing this step by introducing certain artificial variables. At the Air Force, where the simplex method was developed, linear programming has been used for a few problems, a notable example being allocating the use of transportation routes to users.

The major contribution of the simplex method is that it provided the first general solution to one specific form of a rather complicated constrained optimization problem. Its use has been widespread rather than concentrated, including, for example, production and allocation problems in industry. Other methods of minimizing linear forms have been developed in recent years, but the simplex method is still much used.

Business Cycle Indicators

The National Bureau of Economic Research (NBER) gave high priority to the objective study of business cycles from its founding in New York at the end of 1919. In the late 1950s, Arthur F. Burns was Chairman of the Council of Economic Advisors in Washington, and he, Raymond J. Saulnier and others decided that the methods worked out at NBER could be useful in the Federal Government. In 1957, Saulnier requested Julius Shiskin, chief economic statistician at the Bureau of the Census (who had worked at NBER) to prepare a regular monthly report on this subject. The first published report appeared under the title *Business Cycle Developments* (BCD) in October 1961.

The method of analysis on which BCD is based consists of five steps. The first step is to assemble a broad selection of economic series, making a special effort to include series which lead (that is, have their turning points earlier than) the business cycle. The second step is to analyze each individual series making working day and seasonal adjustments in an attempt to exclude variations other than time trend and cycle. The turning points in each series (specific cycles) are then marked off in the seasonally adjusted data. The third step is to date the business cycles by collating the specific turning points in series which represent important aspects of aggregate economic activity. The turning points of the business cycles are referred to as "reference peaks" and "reference troughs." The fourth step is to classify each series according to its usual timing in the business cycle as leading, roughly coincident, lagging or unclassified, and then to select a list of the most widely representative of the series as a three-part list of indicators. The fifth step is to develop various analytical and derived measures. These include composite measures, which combine leading, coincident and lagging indicators, and also a number of other measures.

The purpose of BCD was to provide a sound and up-to-date statistical basis on which forecasts can be made without relying on any particular theory, Keynesian or not. The presentation of economic series agreed upon was in both tables and charts, but without interpretive text other than explanatory notes.

BCD soon came to be used quite widely for current business cycle analysis. The dating of the reference cycles and in fact much of the professional underpinning was left with the NBER. The national income and product accounts continued to be relied on as current measurements and as a framework for forecasts both cyclical and long term.

BCD was cosmetically improved with the November 1968 issue, and its name was changed from *Business Cycle Developments* to *Business Conditions Digest*. A new section was added to present the national accounts in a reasonable amount of detail, and subsections were added for anticipations and intentions and for Federal Government activities. Composite indexes for leading, coincident and lagging indexes were also added. In practice, the composite leading indicator became quite widely used and still is, but the other two are used much less. The real GNP itself continues to be the most widely used coincident indicator in spite of the fact that it is available only quarterly.

The selection of individual indicators paid little attention to whether the series were in value, quantity or price terms, and with the increase in inflation in the late 1960s and the sharper increase in the late 1970s, there was an obvious need to reexamine the indicators. The changes in quantities and real values had to be given primary attention. After BCD was transferred to the Bureau of Economic Analysis in 1972, this was done. On the recommendation of Victor Zarnowitz of the University of Chicago, with the assistance of Charlotte Boschan of the NBER, the changes were made during 1975 and 1976 under the direction of Feliks Tamm of BEA and the Statistical Policy Division at OMB.

The principal contribution of BCD was to provide the current economic indicators in a framework which highlights usual timing relative to the business cycle. (It was discontinued as a separate publication in 1990 by OMB, but a current summary is included monthly in the *Survey of Current Business* as a separate section called "business cycle indicators.") Two of its other contributions to business cycle analysis and forecasting are (1) its assembling of cyclically sensitive series promptly and presenting them in properly digested and convenient form and (2) the objectivity with which this is done, subscribing to no particular theory of what causes business cycles and to no particular appraisal of what the most important factors are in the immediate situation.

Smoking, Health and Statistical Causation

The controversy over the relationship between smoking and health has contributed greatly to the development of the concept of statistical causation. Indeed, it is beginning to appear that the publication by the Federal Government in early 1964 of *Smoking and Health* may be a landmark on the following question: Under what circumstances can statistical association be considered to be of a causal nature?

Early in the 20th century, Karl Pearson, who was a philosopher for many years before he became a statistician, and his followers focused on the question of statistical causation but offered no simple answer. It gradually became clear, however, that (1) in the physical sciences, where errors of observation are the main reason why known relationships do not hold precisely, the distinction between correlation and cause is seldom critical, (2) in the life sciences, where the units being studied are born, grow and die, the distinction is essential and (3) in the social sciences, the relationship is unclear and sometimes even tenuous.

The U.S. Public Health Service first became officially engaged in an appraisal of the available data on smoking and health in 1956–57 when Surgeon General Leroy E. Burney had a scientific study group established, and, after it concluded that there was a causal relationship between excessive smoking of cigarettes and lung cancer, he issued a statement that "The Public Health Service feels the weight of the evidence is increasingly pointing in one direction: that excessive smoking is one of the causative factors in lung cancer."

By this time, there was widespread empirical evidence relating smoking to cancer in general, diseases of the heart and arteries and many other health conditions. Studies took the following forms: (1) animal experiments, (2) clinical and autopsy studies and (3) population (epidemiological) studies of two types, retrospective and (a few) prospective. Experimentation on humans has not been considered feasible. Much of this work was done under National Institutes of Health (NIH) research grants, but much was done by many other research agencies world-wide.

The Berkson and Fisher Attacks. But in 1958, a scientific attack against the conclusion that smoking causes serious health problems and often kills was led by Joseph Berkson, chief of medical statistics for the famed Mayo Clinic in Rochester, Minnesota, and by R. A. Fisher. Both pointed to the lack of randomized clinical trials, arguing that without such objective data, no results could be considered dependable. Berkson's main point was that the negative effects of tobacco were too pervasive to be convincing and that there were no known mechanisms for relating tobacco smoke to such a wide variety of conditions. He suspected biases in sample selection, and he seemed, therefore, to doubt even the association between heavy smoking and ill health. Fisher did not deny the association, but he insisted that association did not prove cause. The principal mechanism which he offered as an alternative to the causal one was that persons with a hereditary predilection to smoking may also have a hereditary tendency for various diseases.

The attacks by these two respected scientists and also by some others attracted widespread attention and forced researchers in this field to defend their position. During the late 1950s and early 1960s, literally hundreds of research projects on the questions raised were carried out, and the NIH alone financed dozens of these. A notable paper in 1959 with six authors, of whom three were government employees, summarized the situation regarding lung cancer. It concluded that the magnitude and the consistency of epidemiological and experimental evidence strongly supported the causal relationship between smoking and lung cancer. It also pointed to "serious inconsistencies in reconciling the evidence with other hypotheses which have been advanced."

In 1962, with the approval of the Secretary of Health, Education and Welfare and of President Kennedy, Surgeon General Luther L. Terry established a blue-ribbon committee to make "an objective assessment of the nature and magnitude of the health hazard" from tobacco after reviewing "critically all available data." In addition to Terry, who acted as chairman, there were ten members, of whom eight were noted doctors of medicine, one a statistician, William G. Cochran, and one a chemist, Louis J. Fieser.

Report of the Advisory Committee to the Surgeon General. The Advisory Committee to the Surgeon General met first on November 9–10, 1962 and worked hard for about a year. Their report was printed in early 1964. From the point of view of statistical causation, the crucial chapters are "Criteria for Judgment" (Chapter 3) and "Summaries and Conclusions" (Chapter 4).

Criteria for assessing epidemiological evidence had been discussed in the profession for some time. Those used by the committee were the following:

a. consistency of the association
b. strength of the association
c. specificity of the association
d. temporal relationship of the association
e. coherence of the association.

The last section of Chapter 3 carries the heading "causality," and it is worth quoting extensively:

> 2. When a relationship or an association between smoking, or other uses of tobacco, and some condition in the host was noted, the significance of the association was assessed.
>
> 3. The characterization of the assessment called for a specific term. The chief terms considered were "factor," "determinant," and "cause." The committee agreed that while a factor could be a source of variation, not all sources of variation are causes. It is recognized that often the existence of several factors is required for the occurrence of a disease, and that one of the factors may play a determinant role, i.e., without it the other factors (such as genetic susceptibility) are impotent. Hormones in breast cancer can play such a determinant role. The word "cause" is the one in general usage in connection with matters considered in this study, and it is capable of conveying the notion of a significant, effectual relationship between an agent and an associated disorder or disease in the host.
>
> 4. It should be said at once, however, that no member of this committee used the word "cause" in an absolute sense in the area of this study. Although various disciplines and fields of scientific knowledge were represented among the membership, all members shared a common conception of the multiple etiology of biological processes. . . . [T]he end results are the net effect of many actions and counteractions.
>
> 5. Granted that these complexities were recognized, it is to be noted clearly that the committee's considered decision to use the words "a cause" or "a major cause" or "a significant cause" or "a causal association" in certain conclusions about smoking and health affirms their conviction.

The committee then proceeded in Chapter 4 to draw the general conclusion that "Cigarette smoking is a health hazard of sufficient importance in the United States to warrant appropriate remedial action." The conclusions by individual diseases begin with a statement on lung cancer: "Cigarette smoking is causally related to lung cancer in men: the magnitude of the effect of cigarette smoking far outweighs all other factors. The data for women, though less extensive, point in the same direction." On the other hand, for heart disease, the conclusion was that "male cigarette smokers have a higher death rate from coronary artery disease than nonsmoking males, but it is not clear that the association has causal significance."

In the years since 1964, a tremendous volume of research in the Federal Government and elsewhere has buttressed the conclusions of the report of the Advisory Committee to the Surgeon General and permitted much stronger conclusions on many points than appear there. These have been set forth in annual reports entitled *The Health Consequences of Smoking*.

Statistical Causation. From the point of view of making statistical inferences from data, the 1964 report broke new ground in at least two ways. First, it applied the term "cause" in a statistical context and defined it as "a significant, effectual relationship." Although it did not use the term "statistical causation," it implicitly distinguished the problems it faced from those faced by, for example, the physicist and the astronomer in deriving the law of gravity, which in this context may be thought of as "simple causation." (It needs to be noted in passing that the article on "causation" in the *International Encyclopedia of the Social Sciences* uses the terms "deterministic causation" and "probabilistic causation" for what are here called "simple causation" and "statistical causation," Volume 2, p. 352.)

The report referred to "variable man" and to tobacco as "a complex agent," and it also stated that no member of the committee used the word "cause" in an absolute sense. It recognized that, although a majority of men who smoke heavily all their adult lives never develop lung cancer, nevertheless, for men who do develop lung cancer, cigarette smoking is the cause in the large majority of cases. This is different from simple causation, in which all bodies having mass are attracted to all other bodies having mass by an immutable formula. It is the same as with many other medical phenomena, however; for if many people are exposed to, say, concentrations of tuberculosis germs, some will contract tuberculosis and some will not.

Also, the report insisted, again without using the term, that knowledge of mechanism is necessary before causation can be inferred from association. It carefully assessed experimental and clinical results, and it insisted that closeness of association by itself is not adequate grounds for inferring causation. These two basic principles seem likely to stick in government medical work, and they may have applicability more broadly.

Social Indicators

The paramount concerns of the Federal Government during the 20th century were economic or military until about the middle 1960s, but social statistics were not ignored. Demographic statistics were collected in censuses from the founding of the republic. Education, health, vital statistics, labor, welfare, housing and some other fields were gradually added to published Federal statistics before World War II. But interest in societal assessment, which is the analytical problem addressed here, really dates from the "Great Society" programs instituted in the middle 1960s.

At the direction of President Johnson in 1966, the Secretary of Health, Education and Welfare, John W. Gardner, appointed a distinguished group of 41 social scientists and assembled a small staff under the direction of Alice M. Rivlin. The January 1969 pamphlet of a little more than 100 pages, prepared by Rivlin and her staff, and issued by Wilber J. Cohen, who had become Secretary of HEW, was entitled *Toward a Social Report.*

After discussing the recent anti-Vietnam War demonstrations and the ghetto burnings of 1968, it proposed a social report or a set of social indicators or preferably both. They would (1) "give social problems more visibility and thus make possible more informed judgments about national priorities" and (2) by measuring social progress in relation to changes in public programs, "ultimately make possible a better evaluation of what public programs are accomplishing." The report then went on to summarize what was known and what was needed to be known in seven subject-matter fields, namely, (1) health and illness, (2) social mobility, (3) our physical environment, (4) income and poverty, (5) public order and safety, (6) learning, science and art and (7) participation and alienation.

During the Nixon Administration, work was begun in the Statistical Policy Division (SPD) of the Office of Management and Budget on the first comprehensive social indicators report. *Social Indicators 1973*, which was issued in February 1974 by one of the authors (Duncan), was factual, and its presentation was primarily in a set of charts, together with the figures on which the charts were based; the text was largely confined to explanations of how the figures were derived and what they meant. The work was carried out by Daniel B. Tunstall under the immediate direction of Robert B. Pearl and the overall direction of Julius Shiskin, before he moved from OMB to become Commissioner of Labor Statistics.

In this publication, eight subject-matter fields are examined. Four of these—health, public safety, education and income—were included in *Toward a Social Report*; the other four were new, namely, population, employment, housing, and leisure and recreation. The indicators selected were intended to measure output, rather than input—for example, educational achievement or attainment rather than school budgets or number of teachers. They measure individual and family rather than institutional or government well-being, and most are time series. There was considerable breakdown to show the disadvantaged segment of the population for each subject. There was no summary, and there was very little information on people's attitudes. The presentation in four colors was striking.

Social Indicators 1973 was well received both within the government and by private researchers and the public, although there was some feeling that it promised more than it could deliver. There was clearly a need for an updating at periods of from one to several years.

In July 1974, Denis F. Johnston joined the staff of SPD to prepare the second issue under the immediate direction of George E. Hall and the overall direction of Joseph W. Duncan. With assistance from the Bureau of the Census, this report was essentially completed by the end of 1976 and issued in early 1977.

Johnston considers it necessary to distinguish three broad types of indicators: descriptive, analytical and programmatic. Analytical indicators involve interpretation or impute a causal relationship. Thus, death rates by age and sex are descriptive, but life expectancy tables and years of working life lost because of industrial accidents can be considered analytical. Programmatic indicators are those needed for judging the degree of success of a program, and they thus include measures of input as well as output. The indicators included in *Social Indicators 1976* are nearly all descriptive.

The 1976 issue includes three new chapters on the family, social security and welfare, and social mobility and participation, in addition to the eight included in the 1973 issue. There is also an introduction, which presents some ethnic detail and selected summary information on public perceptions drawn from public opinion polls.

It is clear that the Federal Government was still at an early stage in turning *Social Indicators* into an analytical framework. There is a considerable measure of agreement on a set of descriptive indicators in ten or a dozen social fields. But there was no consensus on the relative importance of different fields and how far the government should attempt to go in each. Public discussions of economic growth are now usually stated in terms of annual growth rate in real GNP. There is as yet no comparable consensus on the manner of stating social objectives. The development of a set of programmatic measures to balance cost and benefit for each program is equally badly needed. Finally, it is to be hoped that the development of analytical measures will go hand in hand with progress on discovering relationships between social causes and effects.

The variety of these four post–World War II examples indicates the growth in the use of statistical analysis in the Federal Government. Dozens of other examples could be described, and some of them would be just as important in their fields as those presented here, but space is limited.

FEDERAL STATISTICS SINCE 1976

The preceding sections of this article, though reoriented, are largely drawn from a publication, *Revolution in United States Government Statistics, 1926–1976*, which was written by the authors of this article and issued in October 1978 by the Office of Federal Statistical Policy and Standards in the U.S. Department of Commerce. In the decade and a half since the end of the half century covered by the *Revolution*, there have been a number of significant events and developments which merit review and comment.

It is difficult to be objective about a period so recent and about developments which are administrative and even political as well as professional. In striving for objectivity, it will be best to take a broader view and to consider the Federal Statistical System as a whole, rather than confining the discussion to probability sampling and statistical analysis.

The *Revolution* emphasized four major themes: (1) probability sampling and its applications, (2) the national income and product accounts and their uses, (3) mechanization and computers and (4) coordination and building a statistical system. This history was intended as a companion document to *A Framework for Planning U.S. Federal Statistics for the 1980's*, which was a comprehensive review of the Federal Statistical System in the middle 1970s and an analysis of Federal data needs and improvements for the 1980s and 1990s. Unfortunately, there has proved to be limited interest in improving existing statistics or in developing new statistics; hence, the framework document received very little attention or follow-up.

The purpose of this last section is to provide a brief overview of selected themes in Federal statistics for the post-Revolution period, 1977–92. Within the space of this article, it is not possible to review all Federal statistical activities since 1976. Instead, the focus will be on statistical reorganizations, the budgets for statistics, the controversy over the census adjustment and the quality of economic statistics. These are all topics of special interest to the authors and the readers of the *Revolution* and the *Framework*.

Statistical Reorganizations: Statistical Policy Function

After 1976, a number of significant developments arose regarding how Federal statistics are organized. The most notable was the creation of the Office of Federal Statistical Policy and Standards (OFSPS), which was brought about by President Carter's efforts to streamline the Executive Office of the President. Since 1939, the responsibility for planning and coordinating the Federal statistical system had been carried out by the Bureau of the Budget and its successor, the Office of Management and Budget (OMB). With Executive Order No. 12013, dated October 7, 1977, most of the Statistical Policy Division of OMB was transferred to become the newly created OFSPS in the Department of Commerce (which already had two major statistical agencies, the Bureau of the Census and the Bureau of Economic Analysis). The OFSPS was assigned the responsibility for ensuring the integrity, accuracy and timeliness of Federal data, as well as developing and enforcing statistical standards and guidelines. The Carter Administration viewed the relocation of the statistical planning and coordination function to the Department of Commerce (clearance authority for reports remained with OMB) as temporary, pending a final evaluation by the President's Reorganization Project for the Federal Statistical System—an evaluation which was formally announced in May 1978.

The assignment of statistical policy coordination to one of the cabinet departments with statistical agency responsibility was controversial since other cabinet departments did not wish the Secretary of Commerce to have a say in their direct line activities. Initially, this was overcome by a letter of agreement signed by the Director of the Office of Management and Budget which delegated OMB functions (like budget and legislative advice) to the Commerce unit. However, the effectiveness of this delegation gradually deteriorated as OMB divisions began to operate independently from the OFSPS.

Eventually, the President's Reorganization Project recommended, and President Carter appproved in mid-January 1980, establishment of an Office of Statistical Policy in the Executive Office of the President for the coordination of the Federal statistical system – a return to the system in effect from 1933 to 1939. This office would be a separate agency in the Executive Office reporting to the President and accountable to the Congress. The report leading to this recommendation is known as the *Bonnon Report* because the task force that started work in 1978 was chaired by Prof. James Bonnen.

The Administration favored establishing a separate agency to strengthen institutionally the statistical policy function, but this proposal was never implemented due to the passage of the Paperwork Reduction Act of 1980. Signed into law December 11, 1980, the Paperwork Reduction Act consolidated a number of information policy activities (including statistical policy) into the newly created Office of Information and Regulatory Affairs (OIRA) in OMB. It also required the transfer back to OMB of the statistical policy function – a transfer which was effected by Executive Order No. 12318, dated August 21, 1981, after the Reagan Administration took office.

The return of the statistical policy function to OMB did not end the controversy over how best to implement that responsibility. Difficulties began when OMB insisted upon reducing the staffing level of the statistical policy group from 29 to 8. Further, the function was downgraded from a Division of OMB to a Branch. This downgrading triggered the resignation (in November 1981) of Joseph W. Duncan, one of the co-authors of this review.

Many in the statistical community were critical of OMB for this action, especially when the *Statistical Reporter* publication (which had been in existence since July 1940) was terminated with the December 1981 issue. Even more devastating to the function was the reorganization of OIRA in May 1982 which resulted in the elimination of the Statistical Policy Branch.

These events, coupled with 1981–82 budget cuts for statistical programs, enabled the statistical community to gain the attention of OIRA's congressional oversight committees. Congressional hearings were held, and OMB was charged with overemphasizing its paperwork and regulatory responsibilities to the neglect of its statistical policy responsibilities under the Paperwork Reduction Act. Finally, in June 1983, OMB appointed a chief statistician – Dorothy Tella – and a Statistical Policy Branch was reestablished in OIRA.

To this day, the statistical policy function remains understaffed – with only five professionals, including the current chief, Dr. Herman Haberman – and undervalued as an activity within OMB. This is the lowest level of staffing since the function was returned to OMB in 1981. Each time OIRA comes up for reauthorization, Congressional critics and users of Federal data try to get more detail written into the legislation directing how OMB should execute its statistical policy responsibilities. These efforts are continuing as this article goes to press.

The 1986 Paperwork Reduction Act amendments, for example, required OMB to appoint a chief statistician who was a trained and experienced professional statistician and to report annually to Congress on the state of major statistical programs. The current reauthorization bills that are still pending would further specify how OMB should carry out its statistical policy functions.

During recent reauthorization hearings, some have resurrected the idea of moving statistical policy out of OMB again, but this time making it a separate office. Without having immediate access to the OMB budget and legislative processes, as well as the benefit of using the paperwork clearance and regulatory processes to enforce statistical policy and standards, it is not clear how this function could be strengthened by locating it outside OMB. It is the opinion of the authors that the reports review, legislative review and budget review are central to effective coordination policies.

Other Reorganizations

Internal reorganizations and the creation of new statistical agencies have also occurred since 1976. During 1977, in the Department of Agriculture, the Statistical Reporting Service (SRS), the Economic Research Service (ERS), the Farmer Cooperative Service and the Economic Management and Support Center were combined into the new Economics, Statistics, and Cooperative Service (ESCS). Shortly before the end of 1980, the Department of Agriculture formed the Economics and Statistics Service by removing from ESCS those activities that provided technical assistance for cooperatives. In October 1981, SRS and ERS were designated as separate agencies, and in October 1986 SRS was renamed the National Agricultural Statistics Service.

In 1977 the National Center for Health Statistics (NCHS) was transferred from the Health Resources

Administration to the Office of the Assistant Secretary of Health. In April 1987, NCHS was transferred to the Centers for Disease Control.

Several other statistical agency developments should be noted. With the formation of the Department of Energy in 1977, the Energy Information Administration was created with broad authority and responsibility in the collection, analysis, evaluation and dissemination of data related to energy. In December 1979, the Bureau of Justice Statistics was created as the principal agency responsible for the collection, analysis, and dissemination of state, local and Federal statistics on crime and the criminal justice system. In early 1980, a separate Department of Education was created, and the National Center for Education Statistics became part of the Office of Educational Research and Improvement.

Most recently, the Intermodal 1 Surface Transportation Efficiency Act of 1991 mandated the creation of the Bureau of Transportation Statistics within the Department of Transportation. This bureau will be a separate agency, but the statistical programs within the Department of Transportation will not be centralized; rather the bureau will serve as a coordinating body. As this article goes to press, the bureau has not yet been formally established.

While these restructurings of statistical agencies were going on, one agency remained stable—the Bureau of Labor Statistics (BLS). The BLS remained under the able leadership of one person, Janet Norwood, throughout the period from 1976 until her resignation at the end of 1991. Norwood served as a unique pillar of strength in the federal statistical system, with a clear concept of objectives and a sound plan for allocating resources to key activities.

Funding for Statistics

In 1989, the National Association of Business Economists issued a special report on the state of Federal economic data. This report examined the budgets for 1976–88 for the Bureau of Economic Analysis, the Bureau of Labor Statistics, the Bureau of the Census and the Internal Revenue Service. The effort, which was chaired by one of the authors (Duncan), used a constant program approach to evaluate the funding for economic statistics. Selected major economic statistics programs were analyzed over time, and an index of real spending was developed.

The constant program index calculated by the NABE Statistics Committee showed that the constant dollar budget for the programs that generate the major economic indicators declined from 1976 (the end of the *Revolution*) through 1988 in all agencies except the Bureau of Labor Statistics, with the statistical program of the Internal Revenue Service being most severely reduced. Essentially, there were no real gains in resources devoted to basic economic statistics for more than a decade even though the total economy and the federal budget expanded greatly, both in size and complexity.

With limited resources, statistical agencies have been faced with the problem of trying to maintain their eroding data bases and having to defend the quality of the data they produced for formulating public policy. Austere budgets did not permit agencies to undertake statistical research and development essential to modernizing their methodologies for measuring a changing society and economy.

For example, the October 1977 *Gross National Product Data Improvement Project Report*, chaired by the current Federal Reserve Board Chairman, Alan Greenspan, contained numerous recommendations for improving the quality of the data used in producing the GNP estimates. However, because of limited resources, the Bureau of Economic Analysis could only implement some of the improvements. The quality of these estimates would be further debated during the 1980s and are still the subject of considerable attention.

Recently, under the leadership of the current chairman of the Council of Economic Advisors, Michael Boskin, there has been a significant effort to rebuild and improve several key economic indicators. This is discussed below.

Controversy over the Decennial Census

Because population data from the decennial census are used for determining how many seats a state has in the House of Representatives and for distributing Federal funds to state and local governments, the accuracy of the count has come under increased scrutiny, beginning with the 1970 census. For the 1980 census, lawsuits were filed charging that there had been an undercount of the population. The Census Bureau supported research by the Committee on National Statistics of methods for adjusting the population figures to approximate more closely the actual population. A consensus was developing among statisticians that an adjustment was both technically sound and feasible. The Census Bureau announced in May 1987 that it would develop and test procedures for adjusting the 1990 census.

Politically appointed officials in the Department of Commerce, it is alleged, decided against adjustment, and the bureau abandoned its work on adjustment methodology. Pursuant to a legal challenge, the Department of Commerce agreed to develop guidelines for the relevant technical and nontechnical statistical and policy grounds for a decision on whether or not to adjust the 1990 decennial population counts.

Along with the controversy over adjustment, there was also much public debate over the content of the decennial census long and short questionnaires, as well as how large a sample of the population should receive the long form. This debate was brought about by OMB trying to reduce the burden of the census. This chapter is not finished at the time this manuscript is being prepared, but it is safe to predict that more lawsuits will be filed, any adjustments that are made will be challenged because there is controversy about the methodology to be used and, if adjustments are not made, political and legal pressures will mount.

Of particular concern to the authors is the fear that the professional, legal, political and public controversy will create conditions that will make the Census for the year 2000 an especially difficult task. Comparability of data to the 1990 census, a necessary prerequisite for researchers, will probably be a secondary objective. A planning program is already underway at the Bureau of Census to confront this future challenge, but it will undoubtedly take strong and creative leadership in the statistical community to assure a sound census at the beginning of the 21st century.

Quality of Economic Statistics

During the roughly 65 years of Federal statistics described in the *Revolution* and in this article, a small group of statisticians, economists and other professionals succeeded in taking theories and concepts developed in the academic world and applying them to the data problems the Federal Government was facing. This was a time when probability sampling was being used in the development of many surveys still in existence (e.g., the Current Population Survey) and when the National Income and Product Accounts were being developed. These statistical developments helped the Federal Government to recover from the Great Depression, to fight and win World War II and to manage the post-war reconversion. They also contributed to statistical theory. At the end of World War II and later, the Federal government—the statistical agencies and the scientific agencies—played a major role in the invention and early development of the digital computer.

During the decade of the 1980s, the users of Federal data became quality conscious and questioned how good the numbers really were from the surveys developed in the previous decades. The Joint Economic Committee held a number of hearings on the quality of economic statistics. Many articles appeared in the press, trade publications and academic journals questioning the accuracy of major statistical series, including the GNP estimates and the Index of Leading Indicators. In an effort to assess the quality of their data, the statistical agencies contracted with the National Academy of Science's Committee on National Statistics to review their data and make recommendations for improvement in such areas as foreign trade, occupational safety and health statistics, education, health and immigration, and for the Survey of Income and Program Participation.

Because of the continued budget constraints during the 1980s, the statistical agencies have fallen behind in applying technological advances to their data collection, processing and dissemination activities. Further, as noted earlier, there have not been resources to undertake research on the needed conceptual and data definition issues that are associated with a rapidly changing global economic system.

As resources permit, new technology is being introduced to make statistics more widely and rapidly accessible electronically, but the introduction of new delivery systems has badly lagged behind private sector capabilities. Because of noncompetitive salary levels, agencies also have difficulty recruiting highly qualified information management and data processing personnel, and even if salaries were competitive, highly skilled people would probably rather work in a state-of-the-art technological environment. Further, the Federal statistical system as a whole has no program or resources for supporting university research and development of data quality problems of interest to the Federal Government.

The public controversy and the Congressional hearings about the quality of economic statistics elevated these issues to the attention of the Economic Policy Council, which formed a Working Group on the Quality of Economic Statistics in the spring of 1986. The Working Group issued a report in April 1987 that focused on the need to make improvements in five high-priority areas: (1) accuracy of the GNP estimates, (2) adequacy of merchandise trade statistics, (3) adequacy of service sector statistics, (4) quality of business lists used to collect economic data and (5) a system of user charges for statistical data products.

In the February 1989 *Building a Better America*, President Bush expressed his support for a "sound, balanced program to collect and disseminate comprehensive and accurate statistics on America's population and its economy." President Bush established a Working Group on the Quality of Economic Statistics to the Economic Policy Council. Chaired by Michael J. Boskin, Chairman of the Council of Economic Advisers, this working group includes representatives of the major producers and users of economic statistics in the Federal Government.

Based on the recommendations of the Working Group, the President has approved a multi-year initiative to improve economic statistics in three major areas of policy concern: (1) productivity, output and prices; (2) investment, saving and wealth and (3) employment, income and poverty. The President's budget request for fiscal year 1991 included additional funds to begin implementation of some of the recommendations. The "Boskin Initiative" has been widely discussed, and it has been effective in gaining new funding for the Bureau of Census to improve service sector coverage, to improve construction statistics, to strengthen corporate financial data and to do research on emerging industries and to measure the underpricing of exports. Funding has been directed to the Bureau of Economic Analysis to adopt the UN system of Standard National Accounts (SNA), improve the quality of the national income and product accounts and enhance the quality of balance of payments and international investment and services data. The Burea of Labor Statistics participates in the program with resources to improve the accuracy of employment estimates, expand service sector coverage, improve business population lists and develop automated data collection techniques.

While much more work needs to be done, the leadership of Boskin has made a significant contribution to halting the decline that began in the 1970s in governmental statistical programs. His leadership in the effort to improve economic data emphasizes not only the severity of the problem, but also temporarily fills the void created by the lack of leadership of the Statistical Policy Office in OMB in carrying out its coordinating and long-range planning responsibilities in this area. Even OMB's appointment several years ago of a professionally qualified statistician has not resulted in any movement toward issuing the final OMB circular on guidelines for statistical activities, any improved long-range planning for meeting Federal data needs or any improvement within OMB of the status of the statistical policy activity.

In summary, the 50 years of statistical activity described in the *Revolution* was a period of building the Federal statistical system, its surveys, methodologies and products, as well as its planning and coordinating function. Since 1976, the resources available for Federal statistics have not permitted the agencies to continue building upon previous accomplishments. Much research needs to be undertaken to examine statistical methodology and definitions to assure their adequacy for collecting data about our changing economy and society. Advances need to be made in applying technology to Federal data activities. Making government salaries for statisticians more competitive is needed to attract the talent needed to produce quality data. In particular, the coordination function described in Chapter 5 of the *Revolution* has never recovered from the blow it received in 1977. It needs to be reconstituted, to organize and lead progress along all of these lines.