

RECORD OF SOCIETY OF ACTUARIES

1985 VOL. 11 NO. 4A

ROLE OF GOVERNMENT STATISTICS IN A DEMOCRATIC SOCIETY

Moderator: ROBERT J. JOHANSEN
Panelists: DANIEL G. HORVITZ*
FRITZ J. SCHEUREN**
LARRY WOODSON***
Recorder: MICHAEL H. TRENK

- o The outlook for U.S. government statistics after budget costs and other restraints
- o Statistics of income--compilations and analyses by the Internal Revenue Service
- o The outlook for linked studies--IRS, Census Bureau, Social Security Mortality, and Disability by income, by long-term occupational attachment
- o Plans for a 1987 National Medical Expenditures Survey
 - Possible comparisons with published and unpublished data from the 1977 survey
 - The need to get involved in planning for the survey
- o Use of government statistics in marketing studies including the experience of one company with census and other government statistics in planning its marketing strategy

DR. FRITZ J. SCHEUREN: I am going to talk generally about the role of statistics in government. Then I'll provide some background information about what we do in our statistical programs at the IRS. Finally, I will give you three applications from our work that should be of interest to you as actuaries.

Statistical Practice in Government

Government statistical practice is different from that practiced in academic life. It's much more of a management science than a mathematical

* Dr. Horvitz, not a member of the Society, is Executive Vice-President of the Research Triangle Institute.

** Dr. Scheuren, not a member of the Society, is Director of the Statistics of Income Division at the Internal Revenue Service.

*** Mr. Woodson, not a member of the Society, is Director of Research of the State Farm Insurance Companies.

PANEL DISCUSSION

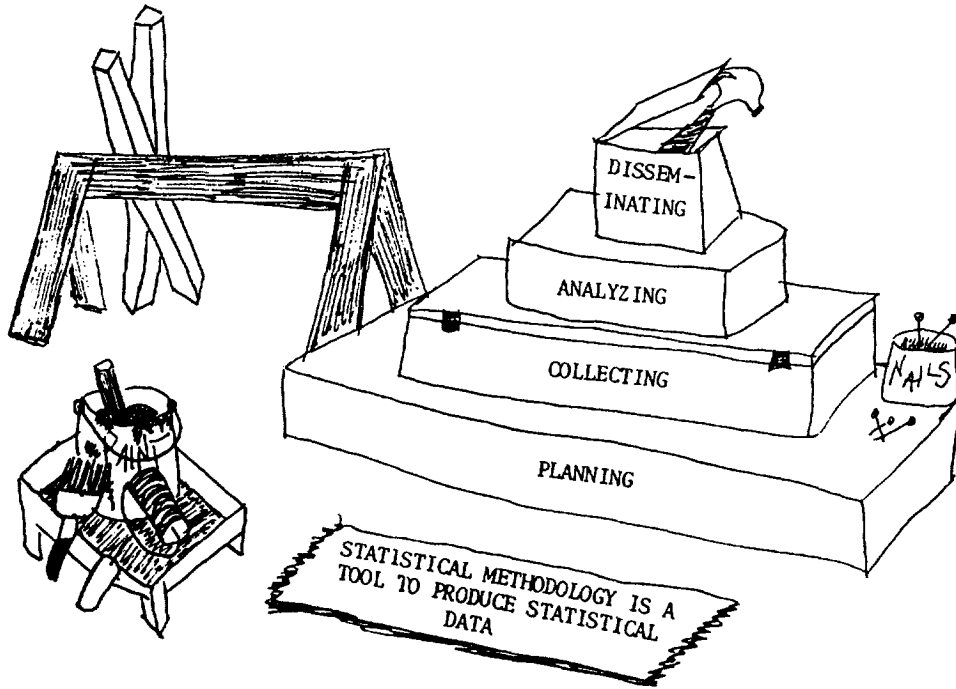
science. It's fair to say that when you look (see figure A) at the four kinds of activities we engage in--planning, collecting, analyzing, and disseminating--that the collecting phase is the one we spend by far the largest amount of our time and money on. In fact, we probably overdo it; that is to say, government statisticians and government statistical agencies may not do enough planning; they surely don't do enough analyzing, and probably in some agencies, they don't do enough data disseminating either.

Mainly, statisticians in government produce descriptive statistics. Sometimes the descriptive statistics can be used to make decisions, but those decisions are made usually by another group of people. Generally, government statistics are not produced as a result of a well-defined market; indeed normal market forces usually cannot be brought to bear, and it is unlikely that a pure market could ever be set up. Take, for example, the issues highlighted in figure B.

- o Priorities--The main problem in setting priorities is one of time lags (and attention spans). A producer of statistics operates perhaps with lead times of a year, two years, or even three years, while the user of statistics, perhaps a policy analyst, operates much more in a "real time" environment. One thing that is happening because of the technological change overtaking statistical agencies is that these horizons are shortening for the producers of statistics such as myself; however, I must say that the impatience of the user is growing apace, so the gap may continue.
- o Concepts--Defining concepts is an important, if little understood, part of the job of a government statistical organization. It turns out to be exceedingly difficult to come up with operationally meaningful concepts. This is especially true in an administrative world, like ours, where you're operating with a tax return which is constructed from legal definitions, and subject to change very quickly. The concepts that you get to measure and the concepts that you want to measure may not be very close at all.
- o Error Estimation--Error estimation is another housekeeping chore for government statistical agencies that is often far more difficult than it seems, especially if one looks at both sampling and nonsampling errors. This is an area, by the way, where government statisticians took the early lead. In fact, most of the basic theory underlying the estimation of sampling errors was invented by statisticians in the government in the 1940s and 1950s.
- o Quality--Quality is a big issue in government statistics, especially in the face of current budget problems. Quality may not be adequate for many purposes, and I see possible further erosion.

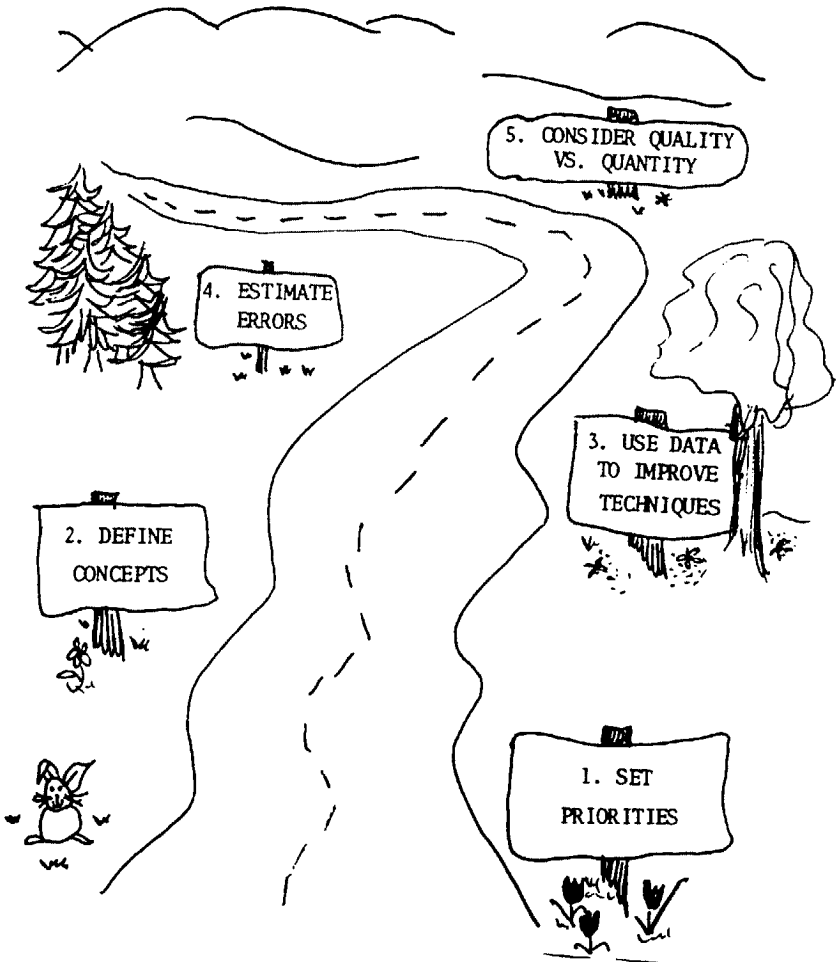
In regard to the dissemination of statistical information, honesty is essential. (See figure C.) Also essential is that we provide information about the data's limitations and how to interpret the results. There has to be an element of comprehensiveness and relevance. The federal government's statistics have not been as relevant as perhaps

Figure A. -- WHAT IS DIFFERENT ABOUT STATISTICS AS A PRODUCT RATHER THAN AS A METHOD?



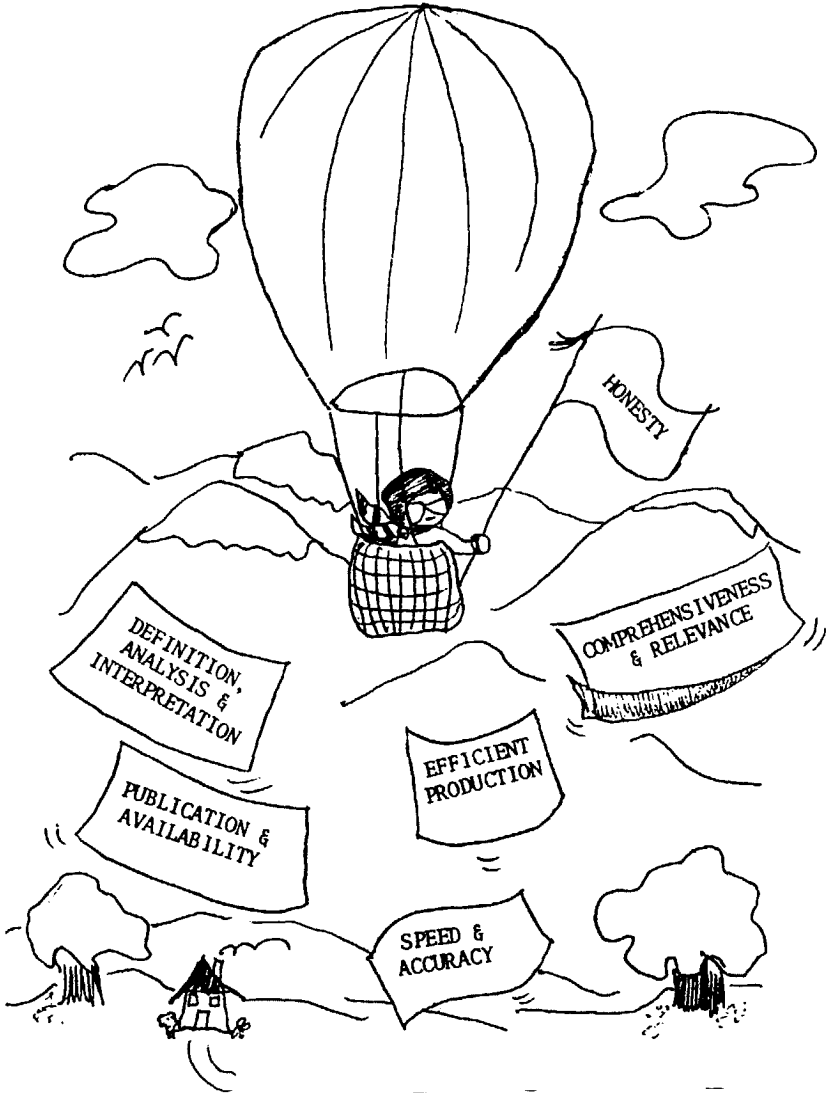
PANEL DISCUSSION

Figure B. -- CHARACTERISTICS OF PRODUCING STATISTICS FOR OTHERS



ROLE OF GOVERNMENT STATISTICS IN A DEMOCRATIC SOCIETY

Figure C.--CRITERIA FOR DISSEMINATION OF STATISTICAL DATA



PANEL DISCUSSION

they needed to have been. Given that we're in an information society, I would say that the federal statistics system has not kept up with the other parts of the information business in this society, and that hurts us both domestically and internationally. Speed, accuracy, efficiency--these are obvious too, but very hard to have (especially speed and accuracy together). In fact, as I've already indicated, we may not have them to the degree we need.

In my own environment and that of this administration, there have been massive budget cuts in statistical agencies. (See figure D.) Many of these cuts in traditional agencies like the Census Bureau and the Bureau of Labor Statistics have been restored at least partially; however, in an agency like mine, which is not primarily a statistical agency, these cuts have not been restored, and that increases the challenge of doing a better job.

I would say that the budget cuts have another kind of an impact, too. They lead one to use cheaper sources of information. Typically, a household survey interview will cost hundreds of dollars to collect the information, whereas from an already existing administrative source, if the same information is available, it can be obtained at a very modest cost--\$5 or \$10.

The policy decisions that have to be made in this environment are obvious. We must plan on making more productive use of existing administrative records. Six short-run goals I would propose are shown in figure E. One of these goals is that administrative records could be used to help conduct the census itself. In many European countries, this is already being done. Denmark is a notable example. If it's not done carefully and well, however, it can backfire as it did in Germany. There's no question that such as activity will not be undertaken for the 1990 census but might well be at least partially undertaken for the 2000 census.

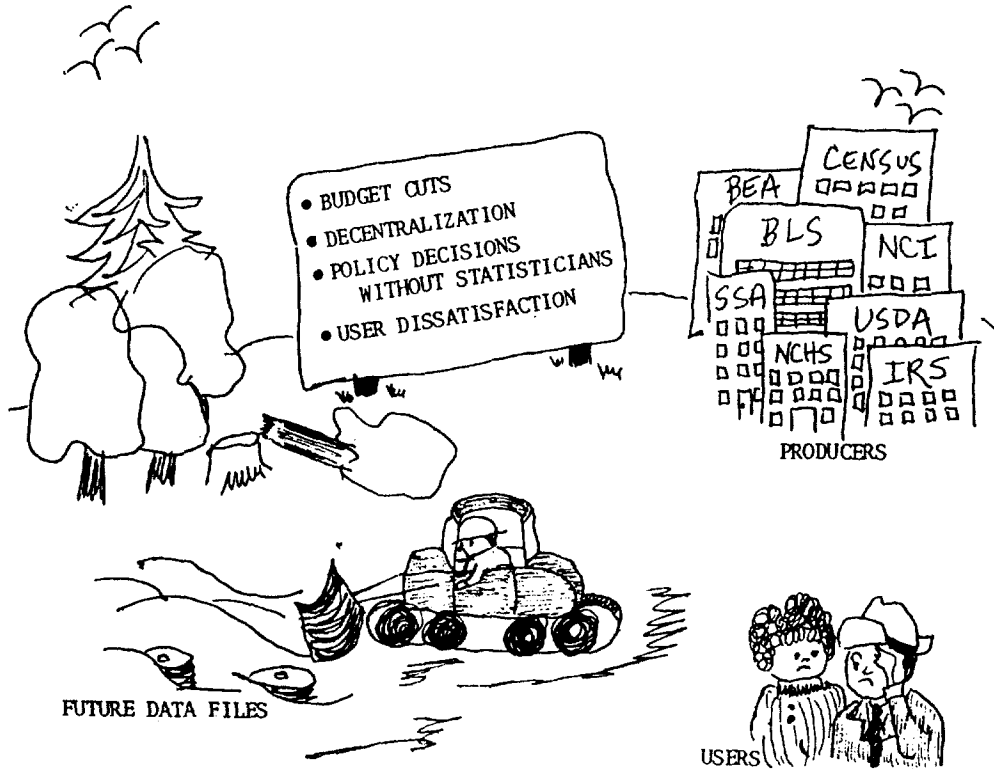
More samples of statistical records for many broad uses, particularly epidemiological uses, are needed and coming. Also coming are uses of administrative records to enhance household surveys and, of less interest to this audience, to deal with the business side of the statistical system in terms of building better sampling frames and providing better definitions of reporting units.

Statistics of Income (SOI) Program

About half of all the administrative records in the federal government are tax records. (See figure F.) We process about 180 million returns a year and over a billion information documents--that's W-2's, 1099's for dividends, interest, and so forth. Clearly we are drowning in data. However, we're not drowning in information. Converting data into information is not a trivial task, and in fact, the objectives of administering the tax laws and getting summary and other information from the returns are different enough so that quite a challenge exists for us.

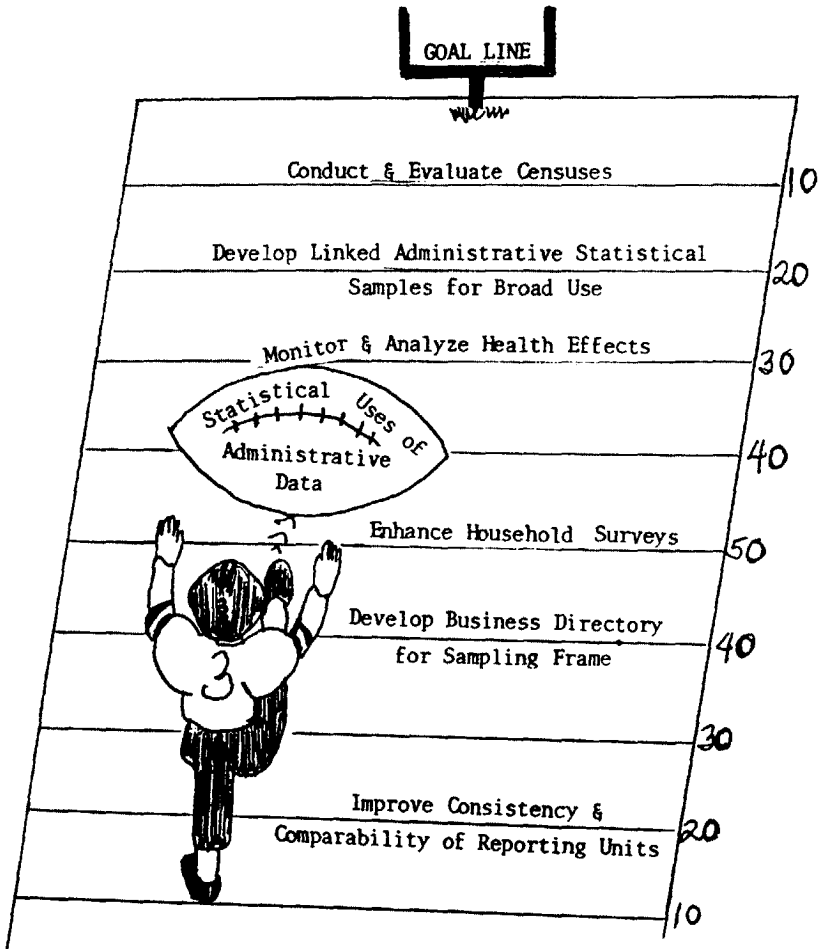
The SOI program, which I direct, is the oldest, major, purely administrative statistical program in the federal government. It was

Figure D.-- CHANGING STATISTICAL ENVIRONMENT



PANEL DISCUSSION

Figure E.--SHORT-RUN GOALS FOR USE OF ADMINISTRATIVE RECORDS



ROLE OF GOVERNMENT STATISTICS IN A DEMOCRATIC SOCIETY

Figure F.--FEDERAL ADMINISTRATIVE DATA



HALF OF ALL ADMINISTRATIVE DATA COLLECTED BY THE FEDERAL GOVERNMENT IS COLLECTED BY IRS -- THE OLDEST MAJOR STATISTICAL PROGRAM FROM ADMINISTRATIVE RECORDS IS THE STATISTICS OF INCOME PROGRAM.

PANEL DISCUSSION

started shortly after the reinstatement of the income tax in 1912. (There was an income tax during the Civil War but it was later declared unconstitutional.)

The main programs are for individuals and corporations, although we do a lot in the foreign area, too; the major users are the Congress, the Treasury, and the Bureau of Economic Analysis in the Department of Commerce (see figure G). Congress and the Treasury use the SOI data to understand the impact of changing the tax laws. Commerce employs SOI because many of the statistical summaries that come from tax data are used by them in helping to produce the national income accounts. Of course, there are also many private individuals and businesses who make use of them as well.

I'm now going to carry you through the processing of your individual income tax return (see figures H to K).

You began the process by mailing your tax return to one of the ten regional service centers that the IRS has. The return is examined. Obvious mathematical and other errors are corrected; the return items are key-entered into our computer system. The computerized data are verified and checked some more, put on tape, and the tapes are shipped to the National Computer Center in Martinsburg. (We do not telecommunicate the tapes, because of data security problems. We use an old-fashioned technology; it's safer.)

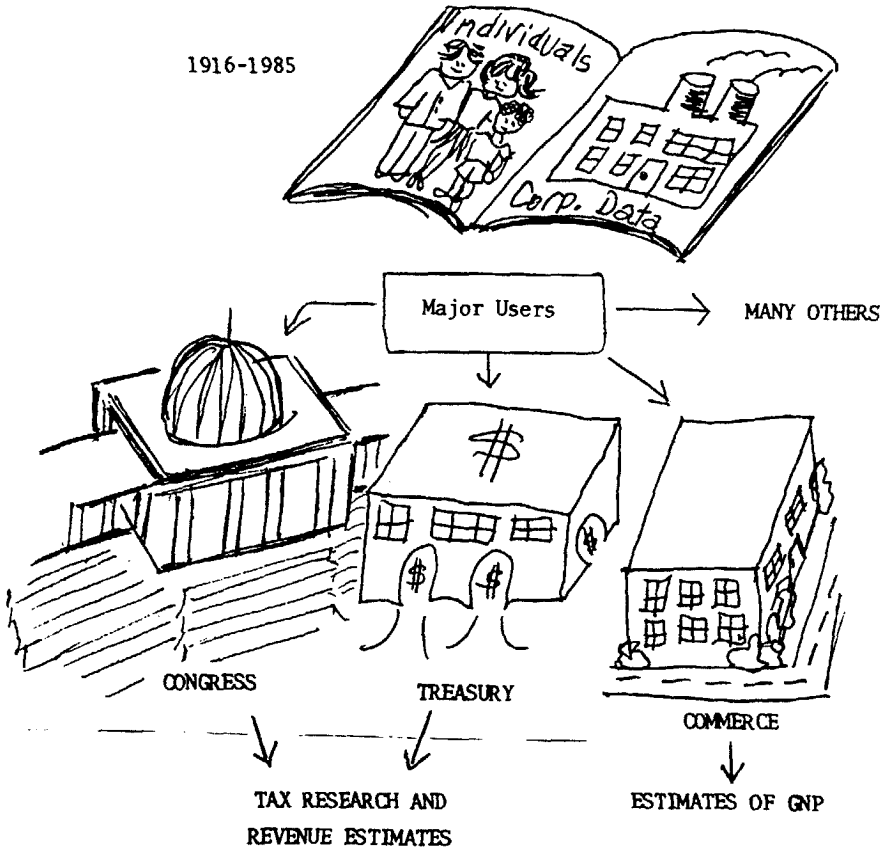
At the National Computer Center we take the tax return information and the phrase is, "We bump it up against your master file record." We keep a record of your filing over many years, and we adjust your account, credit or debit your account as needed, and then we do a lot of other activities that are conventional in any large-scale modern transaction-based accounting system. For example, we send you a refund, or a notice if you owe us money. We generate information which suggests that maybe we ought to look at your return a little more closely because, perhaps, you haven't fully complied with the tax laws.

It is at the National Computer Center that our samples for statistics are selected. The information needed to draw the sample is sent back to the service center where the returns are then pulled. For statistical purposes, we extract additional data, merge it with that already obtained, test it some more, ship it to another center (in Detroit), complete our processing, and tabulate the results.

We have a real horse-and-buggy system in a lot of ways; the technology is about 1970. We just purchased some new computer equipment for statistical purposes (3 VAX 780's), and we're going to be changing our technology and moving to an on-line system of direct data entry. In any event, we correct the statistical data, summarize it, tabulate it, and publish it. Right now we primarily publish on paper as we have for about 70 years. (Electronic publishing, I might add, has already arrived for some federal agencies, and we expect to be in that business in the next two or three years.)

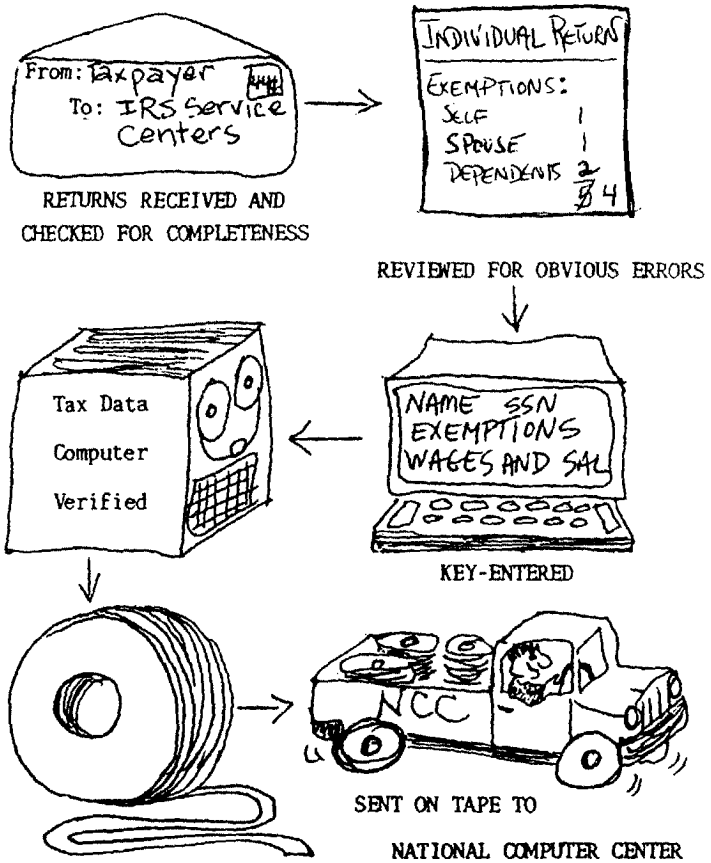
ROLE OF GOVERNMENT STATISTICS IN A DEMOCRATIC SOCIETY

Figure 6.-- STATISTICS OF INCOME (SOI) PROGRAM



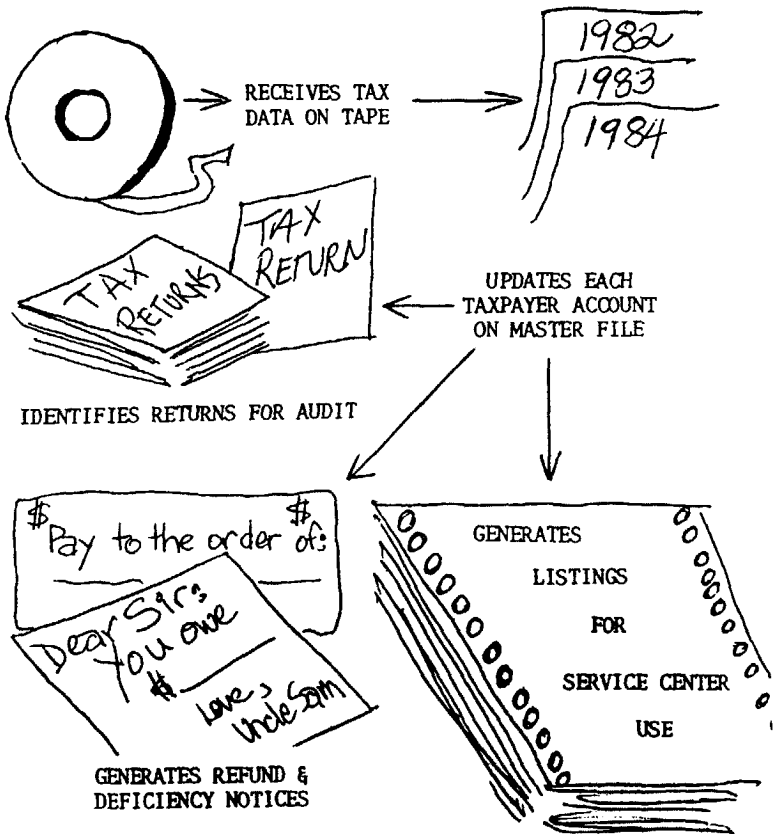
PANEL DISCUSSION

Figure H.-- IRS ADMINISTRATIVE PROCESSING



ROLE OF GOVERNMENT STATISTICS IN A DEMOCRATIC SOCIETY

Figure I.-- NATIONAL COMPUTER CENTER PROCESSING



PANEL DISCUSSION

Figure J.-- STATISTICAL (SOI) PROCESSING

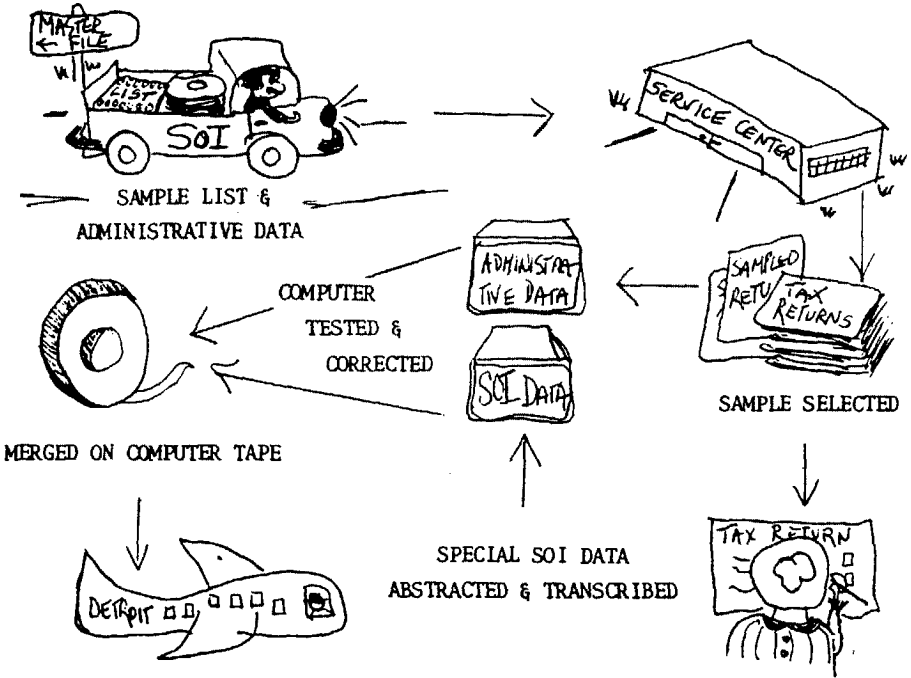
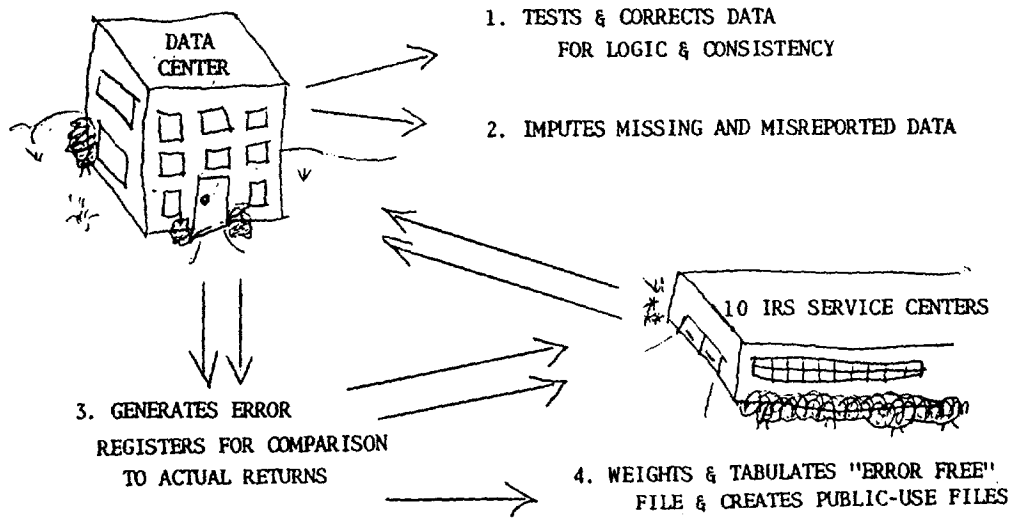


Figure K.-- SOI PROCESSING -- CONT'D



PANEL DISCUSSION

There are many methodological issues that have to be tackled by us in the course of doing our work. I don't know whether any of the items listed in figure L would be of interest to actuaries, but the one in particular that Mr. Johansen wanted me to talk about was data linkage.

Mainly we do what are known as "exact" data linkages where we match records together on social security number, name, and so forth. We are doing three particular studies: a study on occupational mortality and two studies comparing wealth and income by using data linkage techniques. In these three studies, virtually all of the basic data being linked are from tax records, although some data are being obtained under special arrangements with the National Center for Health Statistics and potentially the Social Security Administration.

In order to follow this part of the discussion, you have to know that there is a return called the Form 706 which is the federal estate tax return filed by executors for decedents who dies with wealth above a certain level. Until a few years ago, that wealth was about \$60,000. It has risen now to about \$400,000 and will rise to about \$600,000. (These decedents represent the top 1 or 2 percent of the population and may own perhaps 40 to 50 percent of all the assets held by households in this country.)

As an aside, while wealth in the United States is highly concentrated, it's not as highly concentrated here as it is in many other countries (for example, in Latin America). Concentration ratios (above a certain size) are usually predictive of social unrest, and the fact that our concentration ratio is reasonably low relative to other countries is probably a good indication of a healthy society.

Estate Collation Study

For this study, we took the estate tax return for a decedent, which has his assets, and matched it to his or her income tax return for the three years before death. Then we took the beneficiaries' returns, and we matched all their returns for the two years before the death of the decedent, the year of death, and the two years after the death of the decedent (see figure M).

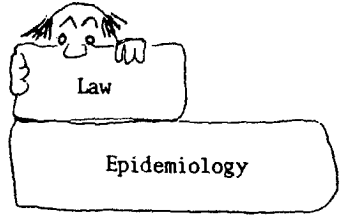
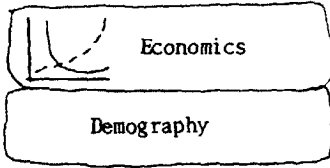
First, we looked at the rate of return to wealth, and we noted that the "income return to wealth" is much higher for people with small amounts of wealth (under \$100,000) than it is for people with a great deal of wealth (\$.5 million or more). What is going on, of course, is that as people move up in wealth, they shelter more of their income. More of their income is unrealized rather than realized, and the apparent rate of return goes down. That is, the realized rate of return goes down--not the real rate of return.

Now, when you look at the beneficiaries, you see a different pattern (see figure N). For beneficiaries there is little impact in terms of the rate of return to wealth, except for beneficiaries at the very bottom and at the top. At the very bottom you have people whose capital income is growing enormously, because of the new wealth that they have, and at the very top, you see people who are sheltering this income so fast that, even though their wealth went up, their income went down.

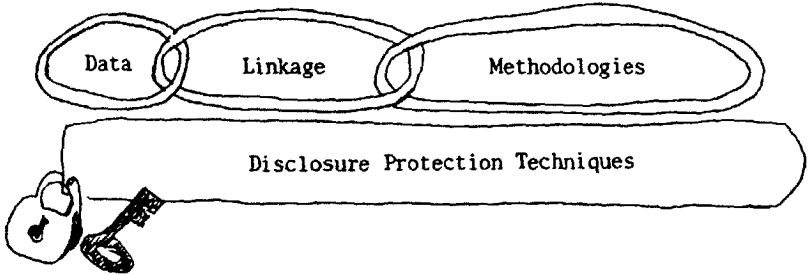
ROLE OF GOVERNMENT STATISTICS IN A DEMOCRATIC SOCIETY

Figure L. --GENERAL RESEARCH OPPORTUNITIES

1. ANALYSIS IN:

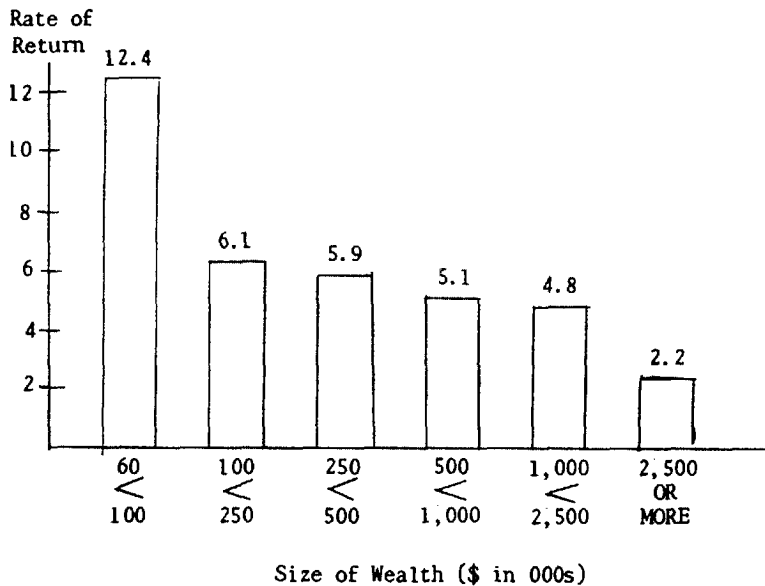


2. "HANDS ON" RESEARCH IN:



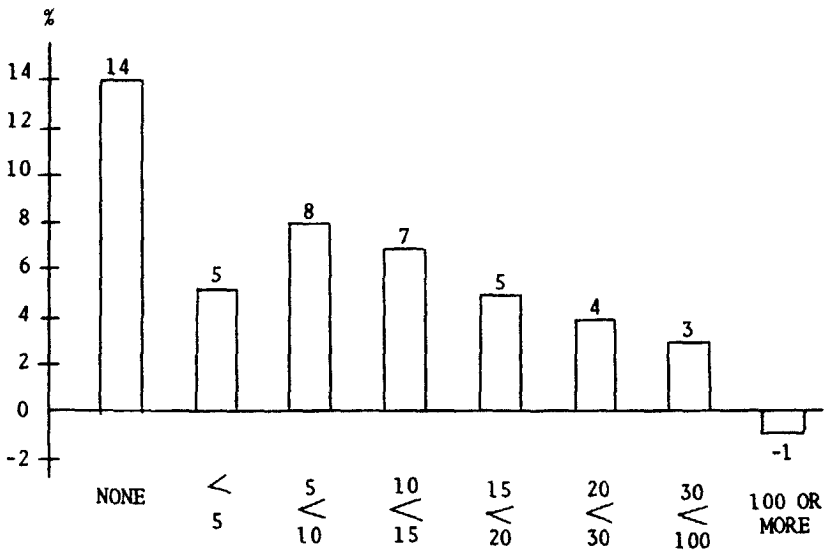
PANEL DISCUSSION

Figure M.-- 1976 DECEDENTS: RATE OF RETURN ON WEALTH
IN YEAR PRIOR TO DEATH



ROLE OF GOVERNMENT STATISTICS IN A DEMOCRATIC SOCIETY

Figure N.-- BENEFICIARIES OF 1976 DECEDENTS: CHANGE IN GROSS CAPITAL INCOME AS % OF INHERITANCE



1975 Gross Capital Income Prior to Inheritance (\$ in 000s)

PANEL DISCUSSION

Intergenerational Wealth Study

One of the most interesting studies we're pursuing examines the extent to which wealth is passed along from generation to generation (see figure 0). As a side benefit of this study, we expect to learn a great deal about the mortality of the very wealthy.

As you may know, estate tax returns have been filed since 1916. Estate taxes were put in place in preparation for World War I and have continued since then. Furthermore, the returns were never destroyed. They exist in federal record centers around the country. Most of them before 1950 are in Seattle. About a year ago we began to go to Seattle and dig through all the dusty files to look at these old returns because we wanted to process them before they were destroyed. This archive represents a major resource for learning about what has happened to the wealthy in this country over the last 70 years.

There is a well-known popular image of wealth accumulation in this country which goes something like this: shirt sleeves to shirt sleeves in three generations. The first generation earns it, the second generation spends it, and the third generation starts over again. Of course, it turns out not to be quite that simple; in fact, we don't understand exactly the nature of that process. Here, this project has much to offer.

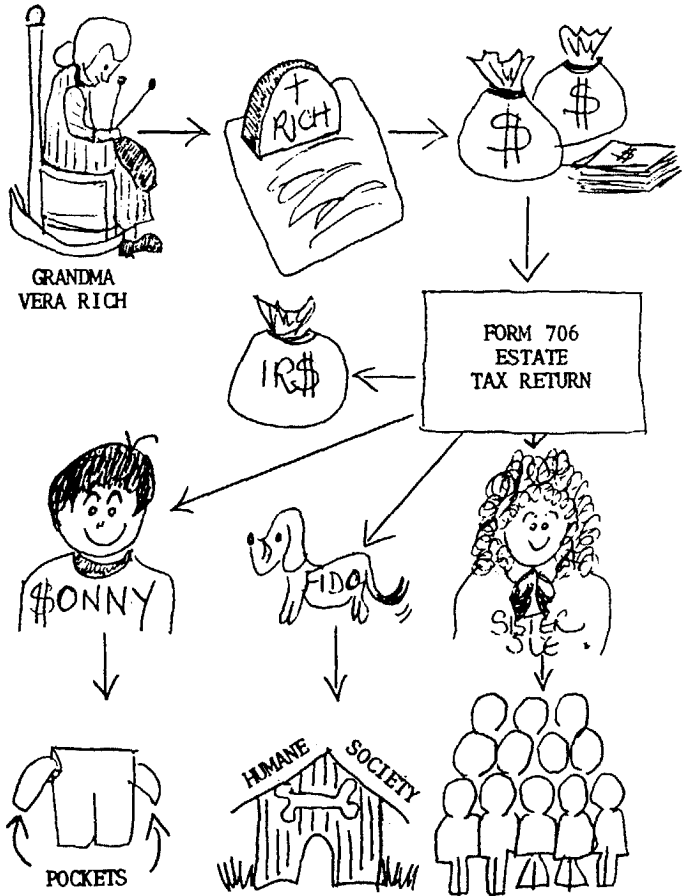
By looking at these old returns, we can learn a great deal about the composition of wealth of the rich in those early periods. The change in wealthholding between men and women would be one example. Typically 40 or 50 years ago, most of the top wealthholders were men. They earned it; they left it to their wives who held it for the children. That has changed. Well over half of the wealth of the very wealthy now is held by women, not only because they live longer, but also because they're changing their relationship to the labor force as well. There have been changes in the concentration of wealth over this period, too. Perhaps not surprisingly, the year 1929, before the stock market crash, showed the highest concentration of wealth by the top 1 percent in this country. Since then it's fallen off, although there is some evidence that it may be rising again.

A major factor in this study will be to look for beneficiaries from estate tax returns; then we're going to track the beneficiaries until they become decedents, and, hence, potentially estate tax filers as well. Our plan is to follow the beneficiaries through this 70-year period from the initial decedent through several generations (see figure P).

A typical pattern would be a man who dies in his 60s or 70s, leaves wealth to his wife who dies in her 70s or 80s, and then the wealth is left to the children and grandchildren. The children are about 50 years of age, the grandchildren in their 20s and 30s so that some of the wealth will begin showing up in 10 years when the wife dies, some more in 20 years when the children begin to die, and some more in 50 years when the grandchildren are going to die. This is obviously a complex pattern and, of course, there are certain transfers of wealth, for example, those in trust, where it is hard to track the wealth because it can skip generations.

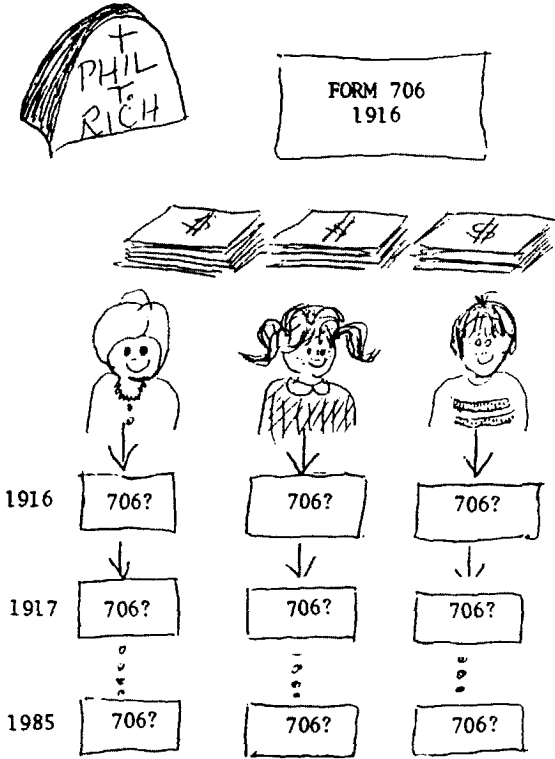
ROLE OF GOVERNMENT STATISTICS IN A DEMOCRATIC SOCIETY

Figure 0.--INTERGENERATIONAL WEALTH STUDY



PANEL DISCUSSION

Figure P.--INTERGENERATIONAL TRANSFER OF WEALTH



ROLE OF GOVERNMENT STATISTICS IN A DEMOCRATIC SOCIETY

In the later years of the study, it will be possible for us to match in with records that didn't exist in the early years. For example, we didn't have a social security system until the 1930s; furthermore, we didn't actually have the social security number on the estate tax return until the early 1960s. The death certificates, of course, are enormously valuable to us for looking at differential mortality and also for looking at differential causes of death.

Occupational Mortality Study

Another study that we're doing matches a sample of about 350,000 taxpayers in 1979 to the National Death Index (NDI). This gives us a way, using the occupation which is on your tax return and industry you worked in (which is available through your employer), to understand and to look at mortality differentials by industry and occupation. The 1979 study is a pilot for a much larger effort involving perhaps two million individuals. Thus we are talking about a scale of activity which rivals that involved in the landmark efforts of Iwao Moriyama and Lillian Guralnick in connection with the 1950 census.

As part of this study we are also looking at occupations on the estate tax returns (see Figure Q); we've already done a match to the NDI for about 1,000 returns with assets of \$5 million or more. Our plans are to begin shortly to write to the states for complete death certificate information.

If you want to ask me about how we're managing to do all of these data linking studies with all the various confidentiality and disclosure rules the IRS has, the National Center for Health Statistics has, the states have, and so forth, I'll be happy to go into the details later. Obviously, great care is being taken. We have observed a great number of precautions in order to guarantee that we comply with all the legal and other restrictions that have to be observed.

A Concluding Comment

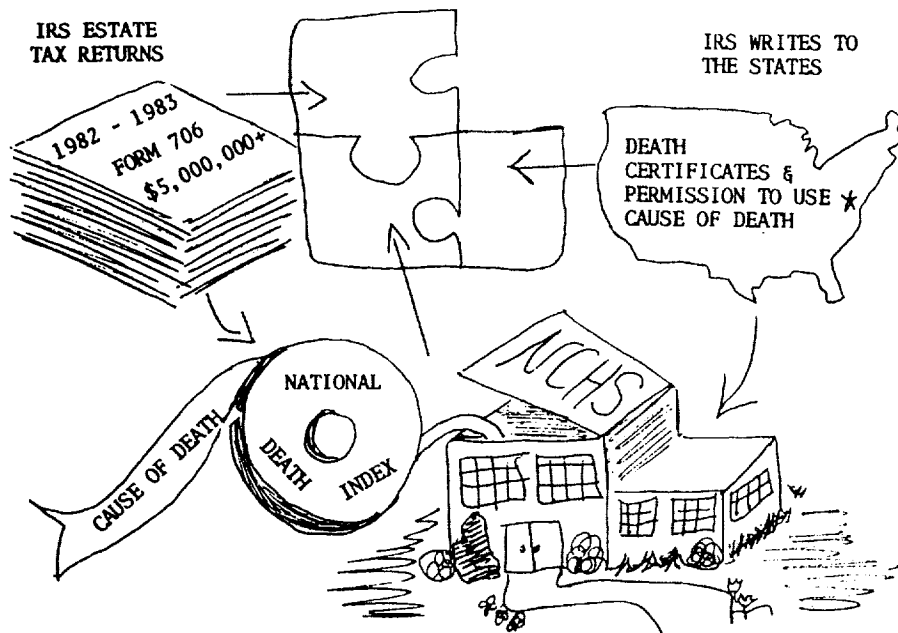
Let me now provide a very brief summary.

Right now federal government statisticians, as the Chinese curse goes, "Live in interesting times." Much more should and is being expected of us than ever before. For example, the public has become far more knowledgeable about statistical matters and is not willing to put up with some of the older practices.

The challenge to innovate and to better serve American society is made doubly difficult by the tight budget climate we are in. This is especially true in administrative agencies like IRS. Tax laws changes of recent years have been all too frequent and sweeping. In such a climate, it is hard to produce statistical information of high quality in a timely manner.

Despite its difficulties, I am not pessimistic about the future; just the opposite. While it would be inaccurate to say we're thriving in this environment, I can assure you that we're doing more than just coping.

Figure Q.--OCCUPATION/CAUSE OF DEATH STUDY



ROLE OF GOVERNMENT STATISTICS IN A DEMOCRATIC SOCIETY

SELECTED BIBLIOGRAPHY

Bentz, Mary. "Intergenerational Wealth Study: Prospects for Data Analysis and Methodological Research," a paper presented at the Conference on Tax Modeling, Lac. Ste. Marie, Quebec, Canada, September 17-19, 1985. (To appear in the Proceedings of the conference.)

Kilss, Beth and Alvey, Wendy (Eds.) Record Linkage Techniques--1985, International Revenue Service, 1985.

Jabine, Thomas B., Kilss, Beth and Alvey, Wendy (Eds.) Statistical Uses of Administrative Records: Recent Research and Present Prospects, Vols. 1 and 2, Internal Revenue Service, 1984.

Jabine, Thomas B. and Scheuren, Fritz. "Goals for Statistical Uses of Administrative Records: The Next 10 Years," Journal of Business and Economic Statistics, October 1985, Vol. 3, No. 4, pp. 380-404.

Martin, Margaret E. "Statistical Practice in Bureaucracies," Journal of the American Statistical Association, March 1981, Vol. 76, No. 373, pp. 1-8.

Petska, Thomas B. "Record Linkages Used in SOI Studies of the Business Sector," a paper presented at the Conference on Tax Modeling, Lac Ste. Marie, Quebec, Canada, September 17-19, 1985. (To appear in the Proceedings of that conference.)

President's Reorganization Project for the Federal Statistical System. "Improving the Federal Statistical System: Issues and Options," Statistical Reporter, February 1981, No. 81-5, pp. 133-221.

Sailer, Peter. "The Longitudinal Sample in the Study of Tax Related Issues," a paper presented at the Conference on Tax Modeling, Lac Ste. Marie, Quebec, Canada, September 17-19, 1985. (To appear in the Proceedings of that conference.)

Wilson, Robert A. "Statistics of Income--A Byproduct of the U.S. Tax System," a paper presented at the Conference on Tax Modeling, Lac Ste. Marie, Quebec, Canada, September 17-19, 1985. (To appear in the Proceedings of that conference.)

PANEL DISCUSSION

DR. DANIEL G. HORVITZ: There are at least two facets to the role of government statistics in a democracy. First, statistics provide public administrators with the data needed to carry on the business of government and to address policy issues of public programs. Second and more importantly, government statistics can be used to advantage by the private sector. Private sector access to and use of data generated by federal agencies has increased in recent years through the release of public use data tapes containing microdata collected at the individual and household level in a wide range of government surveys.

I will discuss both facets in the context of national health care and health insurance surveys. I will give some attention to how one might make use of the data collected in these surveys.

Rising health care costs have been a major health policy issue for at least the past decade. The structure of the health care delivery system and of private health insurance have also changed in recent years. Important changes are also occurring in the federal health care programs and in the demographic composition of the country, particularly in the growth of the elderly population. The search for feasible methods to contain health care costs, while assuring adequate access to health care, requires detailed information, including utilization rates of health services, charges, and who eventually paid the bill. To satisfy this need for information, federally funded surveys of medical costs were undertaken in 1977 and 1980. A third is planned for 1987. In addition to these ad hoc surveys, the National Center for Health Statistics conducts several national health surveys on a continuing basis, including the Health Interview Survey which collects data on health insurance coverage and morbidity. The Bureau of the Census has been conducting a large longitudinal survey, the Survey of Income and Program Participation (SIPP). SIPP also includes data on health insurance coverage. Public use data tapes are being issued for both of these surveys.

The 1977 medical expenditure survey referred to as the National Medical Care Expenditure Survey (NMCES) was jointly sponsored by the National Center for Health Services Research and the National Center for Health Statistics. Both of these agencies are in the office of the Assistant Secretary for Health. The 1980 survey, referred to as the National Medical Care Utilization and Expenditure Survey, was sponsored by the National Center for Health Statistics and the Health Care Financing Administration, the latter being responsible for the Medicare and Medicaid programs.

The proposed 1987 survey, which will essentially repeat the 1977 survey, is known simply as the National Medical Expenditure Survey. It has a number of sponsors. In addition to the National Center for Health Services Research, the National Center for Health Statistics, and the Health Care Financing Administration, it has the National Institute for Mental Health, the National Cancer Institute, the National Institute on Aging, and the Indian Health Service, all interested in data on medical care costs.

ROLE OF GOVERNMENT STATISTICS IN A DEMOCRATIC SOCIETY

The 1977 Medical Expenditure Survey was the most comprehensive study ever assembled on the health care experience of the U.S. population for a one-year period. The proposed 1987 survey undoubtedly will be even greater in size and complexity. It will repeat each of the three major component surveys which were carried out in 1977: (a) a survey of 14,000 randomly selected families across the United States tracing their total health care expenditures during 1977, with interviews six times over a 15-month period during 1977 and 1978; (2) a survey of the physicians and hospitals that provided care to the family respondents; and (3) a survey of the employers and insurance companies responsible for insurance coverage of the respondents.

The household survey component was the primary source of information. A core set of questions was asked of each of the 36,000 survey participants. In addition to standard demographic information, the core questions inquired about health insurance coverage, number of days of disability due to illness or accident, disabling conditions, details of each dental, doctor, clinic, and hospital visit, services received, charges, prescribed medicines, and finally, the source of payment--out of pocket versus an insurer or other third party, public or private.

Over 90 percent of the 1977 sample respondents agreed to let the project use data from their records. A sample of approximately 60 percent of the 22,000 different physicians and 4,500 different hospitals and clinics that were reported to have provided care to the survey participants were selected for the physician and hospital survey. This latter survey supplemented and verified the information obtained in the family survey.

Also, over 90 percent of the survey respondents signed permission forms which authorized and requested their employers, or their health insurance companies to verify their coverage, and to report the type of coverage, the premiums, the source of payment, and to provide a copy of the policy itself. Employers offering group plans were asked to provide data on each of the optional plans available to their employees. Finally, the employers of employed family survey participants who did not report holding a group health insurance plan were contacted to verify the absence of coverage for these specific employees. Approximately 80 percent of the 6,500 different employers and insurance companies provided the requested data.

In a special separate effort, some 6,900 employers of the employed family survey participants were queried regarding their total number of employees, total annual payroll, the number and categories of employees eligible for group health insurance coverage, the total annual premium for the health insurance benefits provided by the employer, and the employees' contributed share of the total annual premium. It's quite likely that exactly the same kinds of data will be collected in the 1987 survey.

One may ask, "Why bother with family respondent data at all if it is still necessary to carry out various record checks with medical providers, employers, and insurance companies?" In fact, the NMCES was done in this manner because of the uses to be made of the data.

PANEL DISCUSSION

Estimates and analyses were required for various subgroups of the U.S. household population defined by age, race, sex, family income, urban versus rural residence, and presence versus absence of health insurance. It would have been almost impossible to obtain the representative samples of families and individuals by sampling only the records of medical care providers and health insurers.

The health insurance policies collected were abstracted and coded in depth with respect to their specific coverage and benefits. These data, when merged with the individuals' 1977 utilization and expenditure data enabled a characterization of the general population with respect to the extent of coverage and benefits, not just a check on existence of coverage. It became possible to assess the adequacy of coverage in relation to utilization of health care services and distribution of health care costs.

The 1977 survey was huge, requiring about 25 data tapes. It's unfortunate that only a limited portion of these data have been made available to researchers through public use tapes. Various analyses have been carried out by the National Center for Health Services Research's intramural research staff, resulting in a series of at least 21 reports published by the Department of Health and Human Services. These reports, for example, include such titles as "Private Health Insurance Coverage of the Medicare Population," "Private Health Insurance Premium Expenditures and Source of Payment," "Private Insurance and Public Programs--Coverage of Health Services," and "Changes in Health Insurance Status--Full-year and Part-year Coverage in 1977." All four titles were released in 1984 or later, a rather long interval between the time of collecting the data and providing the reports. Still, the information in these reports is of considerable interest and would be of even greater interest, if updated.

The NMCES data have also been used to address health care policy issues by various federal agencies such as the Health Care Financing Administration, the Congressional Research Office and various Congressional committees. For example, the decisions to eliminate the health insurance income tax deduction and to raise the out of pocket health expenditure deductible from 3 percent of taxable income to 5 percent are undoubtedly due to analyses of the NMCES data. The data also showed that federal subsidization of health care costs of middle and upper income families through income tax deductions rivaled federal outlays for low income families through the Medicaid program.

The aging of the U.S. population, along with increases in the total costs of health care, make the financing of health services for the elderly a topic of major policy interest. It's quite possible that the 1977 NMCES provided data essential to the subsequent enactment of the 1980 Baucus Amendment to Title XVIII of the Social Security Act which set minimum standards for individual Medicare supplementary insurance. The 1977 NMCES provides essential baseline data on private plans held by the elderly Medicare population in 1977, showing which ones met the minimum benefit standards of the legislation. These data in combination with comparable data, collected subsequent to the Baucus legislation, can be used to assess the legislation's impact. The proposed 1987

ROLE OF GOVERNMENT STATISTICS IN A DEMOCRATIC SOCIETY

National Medical Expenditures Survey will meet the subsequent data requirements for such an evaluation.

The health care and expenditure data collected in 1980 did not include the same depth of information on health insurance as gathered in 1977. No detailed data on the types of plans, the coverage, or the benefits and premiums were collected. Copies of health insurance policies were not required. The data requested in the core questionnaire, however, were essentially the same as in the earlier survey. Analyses of the 1980 survey produced a number of published reports by the Health Care Financing Administration, including such titles as "Supplemental Health Insurance Among Aged Medicare Beneficiaries," "Outpatient Prescription Drug Utilization and Expenditure Patterns of Non-Institutionalized Aged Medicare Beneficiaries," "Dental Insurance, Medicaid and their Relationship to Utilization of and Expenditures for Dental Service," and "Factors Affecting Supplemental Health Insurance Coverage for Aged Medicare Beneficiaries." As with the 1977 survey, the 1980 data base has contributed useful and accurate information to various legislative committees concerned with health care policy issues.

The 1987 National Medical Expenditure Survey, as mentioned earlier, will essentially repeat the 1977 survey. Thus, a substantial investment will be made to gather detailed information on health insurance coverage, premiums, and benefits of the U.S. household population in 1987. As in 1977, the data on health insurance plans and premiums will be collected from both employers and insurers. Copies of health insurance policies will again be requested and coded in depth.

The 1987 National Medical Expenditure Survey will provide, in addition, up-to-date national estimates of health care use and expenditures. In addition to the civilian, noninstitutionalized population surveyed in 1977 and 1980, the proposed survey will also cover the population in nursing homes, psychiatric hospitals, and facilities for the mentally retarded, by means of a sample of 10,000 persons resident in these long-term care institutions at any time during 1987. A national probability sample of approximately 14,000 households will include oversampling of groups of particular interest, namely Blacks, Hispanics, the poor, the near poor, the elderly, and persons with functional limitations. There will also be separate samples of American Indians and Alaskan natives.

In what ways and to what extent can these national health care and health insurance surveys be of value to members of the Society of Actuaries? First, the 1987 National Medical Expenditure Survey data will provide an unusual opportunity to analyze the private health insurance market. While individual health insurance plans will not be identifiable by company in the data files, persons covered by Blue Cross and Blue Shield, by other private plans, and by public plans will be distinguishable. The detailed coverage and benefit information offers an almost unlimited and certainly unique opportunity for market analysis. Both national and regional market analyses should be possible. While the sample will not be adequate for analyses at the state level, it will be adequate for a wide range of subscriber groups defined by demographic characteristics: by types of health insurance

PANEL DISCUSSION

coverage; benefits and premiums; and by utilization, costs, and sources of payment for health care services during 1987.

Second, the 1987 data, together with the 1977 data, should provide excellent information on the major trends in the health insurance market during the intervening years, particularly with respect to employer-provided benefits. The data should also provide useful information on the influence of public program policy changes and of changes in health insurance plans and products on market behavior during the ten-year period.

Third, the 1987 data could be used to estimate the market potential for new private health insurance products, whether to serve the needs of the elderly or other special segments of the population. Since, for example, public programs do not provide benefits for hearing and vision services, for medicines prescribed out of hospitals, for home health care services unless provided by a skilled nurse, for routine dental care, or for physical therapy, it's possible that the 1987 survey data could assist in design of new health insurance products for the elderly.

Fourth, the 1987 survey should provide an excellent opportunity to compare the use of insured services by the U.S. population in general to that of subscribers to particular health insurance products.

I suspect that there are many other and more useful ways to use the data that will be gathered in the 1987 survey. The only remaining concerns are the availability of public use data tapes and your ability to process and analyze those tapes. It's not clear whether public use data tapes will be made available. Only a portion of the 1977 and 1980 surveys have been released in the form of public use tapes. The 1977 tapes were not released until 1982 and do not include the detailed health insurance policy benefit coding.

There is growing concern in the private sector about inadequate access to public data. Although efforts are under way to improve access, it's uncertain whether the 1987 National Medical Expenditure Survey tapes will be released for public use at all. Failure to release these data is, in my opinion, hardly in the best interest of the country. The cost of gathering the data and preparing the files is substantial, upwards of \$20 million. Maximum benefit will only be realized by releasing a complete set of public use tapes in a timely fashion. Such a release could give real added meaning to the role of government statistics in a democracy. I doubt that it will occur, however, without bringing pressure through the Congress or the Office of Management and Budget (OMB) on the Department of Health and Human Services. Those of you who are interested in such data may help in applying that pressure.

MR. LARRY WOODSON: There is a basic concept in marketing called the four "P's": price, product, place, and promotion. While actuaries are typically involved in the pricing and product development functions, federal statistics as a resource are most important in analyses of place and promotion. Pricing and product development are internally driven mechanisms, so I'm going to concentrate on place and promotion, since most of my energy goes into the market research function.

ROLE OF GOVERNMENT STATISTICS IN A DEMOCRATIC SOCIETY

The principal promotional technique in the industry is the friendly insurance agent. This is a volatile topic right now. Many people are saying they don't buy insurance today the way they bought it 20 years ago. Other people assert that the buying transaction hasn't changed. And then there are other people that say there is no buying transaction, there's a selling transaction. Much of the current research concerns the buying transaction and the promotional aspect of our business. Also, distribution channels are changing and broadening.

There's a lot of experimentation regarding how we sell our products to the consumer. The use of an agent might be characterized as a high touch methodology. It's a one-on-one transaction, and therefore isn't very subtle.

There are basically two types of data used to study these topics. Internal and primary data--primary data being data that we collect--are more important to us than external data, such as federal statistics. What we learn about the buying transaction is based largely on our own history and experience. We study our agency force, find the agents that do well, and learn from those agents.

An important part of the marketing process is to disseminate what we learn from the agents. That's one way that we enhance our promotional efforts. Another way is to conduct research. One of the techniques we use is focus group interviews; spending three hours with a group of folks asking them about their experience with us. We may ask questions like, "Did you buy insurance in the last two years?", "How did it happen?", "What did you think about our product?", or "What improvements would you like to see?" This type of research is called qualitative research. Most of our qualitative research is primary in nature. There's a limited role for federal statistics in the "promotion" aspect of our marketing process.

We make use of federal statistics in optimizing the place function for our marketing task. We need to have a thorough knowledge about the markets available to us. There are two methods by which we evaluate our marketing success: comparisons over time and comparisons of locations having similar characteristics. We're interested both in quantity and in quality of the market. Principally, my interest is based on a geographic perspective. I want to know the relative size of the market, our share of it, and how it changes over time. I like to know what's happening in the marketplace, and how I can attract the market's customers. I'm interested in concentration of wealth because wealth is a predictor for me. I'm also interested in size of purchase and in what the competition is doing. Not many federal statistics help me with that.

Because State Farm is already active in most markets, I look principally at selecting new customers within a marketplace. For instance, a State Farm agent in Hannibal, Missouri, is not interested in whether the company will enter into new geographical markets, but rather in getting support for selling to existing customers and in identification of new customers. I'm interested in evaluating local marketing performance over time.

PANEL DISCUSSION

Federal statistics are one of the important sources of the marketing data for managing this "place" function. We use the Federal Census Files, but have not made great use of the Current Population Surveys. The Census File is issued every ten years. Many of the items are based on 100 percent of the population. For economy's sake there are some items asked only of a sample of citizens. The Current Population Survey is taken annually. It surveys a sample of American households, collecting a broad spectrum of information.

An advantage of federal statistics is that they're cheap. A disadvantage is that they're slow to be released. Both of our other speakers alluded to the fact that the federal government collects data primarily to support governmental units needing the information, not to serve the private sector.

From the 1980 census, we use information at the zip code and county levels. Information is also tabulated at the tract level and the block group level. Tracts are defined as homogeneous sets of people--a useful concept in marketing, but tracts are not well recognized by our agents. Zip codes are larger than tracts; while they're not as homogeneous, field management understands them well.

We buy data principally in a computerized form and merge this information with computerized extracts from our corporate files to do our analyses. With respect to quantity, I'm interested in number of households, number of vehicles, and number of people. On the quality side, I'm interested in income indicators, both household income and dwelling values.

When budget constraints in 1980 prevented the Census Bureau from publishing census data by zip code, a consortium of about 15 companies, including State Farm, paid the Census Bureau about \$250,000 to produce the zip code file. The file was produced and is available today. The consortium got exclusive use of it for a couple of years; now you can buy it for a nominal cost from the Census Bureau. It's an important resource, made possible by private industry.

We're hoping that the Census Bureau will include the zip code file in the 1990 census, but if it does not happen, private industry may again fund its creation. It is important to the market research industry to have these base data at least every ten years. In 1981, there was talk about doing a five-year census, but I've given up this hope.

Market research vendors play an important role in the application of federal statistics. Data from the Census Bureau and other federal agencies are tailored to federal needs, not to private industry's. This I understand and accept. Currently, the 1980 census data are five years old--no longer accurate in marketplaces that have changed. Vendors are busy filling information voids by using data files from the IRS, local government agencies, state agencies, and private firms to provide updated estimates at the county and zip code levels. Having accurate estimates of this type is very important to me, and it's work that I could not do in my shop. Vendors can do the work once and sell it to many organizations, spreading the cost to make it affordable.

ROLE OF GOVERNMENT STATISTICS IN A DEMOCRATIC SOCIETY

The vendors process federal data by changing the format where needed, by making it readable on different hardware, by taking extracts for special topics (such as equal opportunity), and by doing actual analyses. Vendors also combine data from the census file with information gathered from other sources. Vendors are important conduits to the federal data statistics. I believe that the information available through research vendors is often better in quality and timeliness than raw federal statistics.

In the June 1985 issue of American Demographics there is an article on research vendors and products that they have developed from demographic data bases, primarily from the U.S. census. The article also talks about the qualitative aspect of these files. For research information dealing with the qualitative aspect of markets, these firms have useful products. Also, a November 1982 article in the same magazine, American Demographics, deals with cluster demographics, and talks about techniques for analyzing the quality side of market characteristics and buyer transactions.

MS. ANNA M. RAPPAPORT: With respect to the health care survey, is there a specific person or organization to contact if one is interested in obtaining 1977 data or influencing plans for the 1987 survey?

DR. HORVITZ: It would be appropriate to inquire through your Congressman about the availability of the 1977 NMCES data on health insurance plans. The survey was done under the auspices of the National Center for Health Services Research. One approach would be to pressure that agency, because they have not released in public use form the in-depth data on each health insurance plan for which they collected a health insurance policy document. But second, and more importantly, you can contact the OMB. The 1987 survey, for example, has to be approved through the OMB for both its content and funding. I know that individuals at the OMB are concerned about the fact that public use files were not prepared in a timely way and released for the 1977 survey, and they have expressed this in writing. It would be helpful to have some external pressure from the private sector applied to the OMB indicating awareness of the 1987 survey and interest in having such data made available.

MS. RAPPAPORT: I was curious whether there's been much comment on the introduction of Diagnostic Related Groups (DRGs) and the changes in the Medicare system. Are there federal statistics being published to show what's changed and by how much? How might one find such reports?

DR. HORVITZ: The Health Care Financing Administration supports some studies evaluating the impact of the DRGs. They are responsible for instituting DRGs within the Medicare program. Specifically, there is an Office of Research at the Health Care Financing Administration located in Baltimore, Maryland, adjacent to the Social Security Administration. The director of the Office of Research is Allen Dobson. You could probably inquire through him. I do not know of any specific study, but I feel fairly certain that there are such studies.

PANEL DISCUSSION

DR. SCHEUREN: About a month ago, the National Academy of Sciences held a series of meetings on the statistical needs of our aging population. Obviously, health is a major policy issue for the aging population. A contact at the Academy is Dorothy Gilford, former director of the National Center for Educational Statistics. She can be reached at the National Academy of Sciences, 21st Street and Pennsylvania Avenue, Washington, DC.

MR. ROBERT A. SUJECKI: Going back to the 1987 study, you mentioned that there will be policy information collected along with the health service information. Is there any intention to collect information on the type of delivery system, that is, whether the service was provided under a health maintenance organization (HMO) plan versus a preferred provider organization (PPO)?

DR. HORVITZ: In the previous study, HMOs were distinguished from other types of health insurance plans. I feel fairly certain that the new study will distinguish PPOs and other type of new organizations. I'm fairly confident that the 1987 survey will determine for each individual in the sample the kind of plan they have and that it will categorize the plans by type. The data on the file would be such that one could look at and compare the different delivery systems.

MR. PHILIP J. T. CERNANEC: Suppose I have a series of states, and I'm considering the population: the number of individuals who have insurance coverage, whether insurance coverage is provided through individual plans versus group plans, whether coverage is employer sponsored, whether it's through HMOs or PPOs, and so on. Where could I find the best information to help me evaluate the market potential for my products?

MR. WOODSON: I'm not aware of a source containing that information. It's easier to collect information at the state level than at lower levels.

DR. SCHEUREN: Supplements to the Current Population Survey have addressed your questions, although probably not in the depth that you've asked them now. The supplements do contain some information about types of insurance coverage. I'm not sure when the last one was done; it must have been two or three years ago. Wally Kolodrubetz of the Department of Labor might help you.

DR. HORVITZ: Current information about health insurance coverage is collected in the Health Interview Survey sponsored by the National Center for Health Statistics. You can ask for their latest publication on health insurance coverage. This is a national survey and probably will provide reliable information only at the regional level, or at best, the census division level. If your group of states is located in one part of the country, it could be helpful. It's possible that the changing health care delivery environment may lead to more detailed questions in future Health Interview Surveys. I believe HMO participation is being tabulated, but I'm not sure about PPOs. The National Center for Health Statistics's publications are very inexpensive investments and may provide you with some help.

ROLE OF GOVERNMENT STATISTICS IN A DEMOCRATIC SOCIETY

Another source is the SIPP. This survey also asks questions about health insurance coverages, but I don't think it goes into much depth. Public use tapes are being released for the SIPP. One could generate data for a group of states. One may also be able to separate data at the 11 census division levels--a step beyond the four regional units.

These are the only two studies that I know of that routinely collect information on health insurance coverage.

MR. ROBERT J. JOHANSEN: Perhaps we should suggest that the Health Insurance Section set up a project on a survey of available federal government statistics and perhaps also act as a conduit to make requests and suggestions known to the people who produce these surveys. Certainly trying to keep the federal governmental agencies up to date on different or new forms of delivery of health care and payment for health care seems like a worthwhile project from our viewpoint and also from theirs.

DR. SCHEUREN: Let me make an observation about the notion of asking for information from the federal government and being persistent about it. As an employee of the government, I can tell you that the weight of a professional society such as this one has a tremendous impact. If you ask for something, know why you want it, and if you persist in it, eventually you may get it.

MR. JOHANSEN: The Society has another avenue of access. The Society is a member of the Council of Professional Associations on Federal Statistics, an organization which has contact with people in the various agencies. This is another approach, along with writing to your Congressman, to OMB, and to individuals in the agencies as mentioned.

MR. HARRY C. BALLANTYNE: Dr. Scheuren made a comment on the IRS's income tax processing problems in 1985. Do you know of the effects these problems may have on the completeness or timing of processing your statistics of income?

DR. SCHEUREN: First of all, the IRS and the taxpayers are going through quite a traumatic experience. During the last several years, the IRS saved a lot of money by changing the way it operated. We cut back perhaps a little too far and, thus, were unable to cope as quickly as we should have with the start-up problems we've had with our new computer equipment. The impact of this problem on our Statistics of Income Studies has been serious. In February 1985, when we realized we were going to have a problem, we began to make contingency plans. Fortunately, we didn't have to use the worst of those plans. We're currently about one to two months behind schedule.

MR. BALLANTYNE: On your detailed data, how long is the usual lag?

DR. SCHEUREN: We published early estimates for tax year 1984 in the summer of 1985 (within three months after the filing deadline of April 15.) More detailed data for 1984, albeit still preliminary, will come out in January 1986; the final report for 1984 will appear late in the summer of 1986. The major effect of this year's computer problems will

PANEL DISCUSSION

be to cut back on the extent of the detail we'll have in our early preliminary reports.

MR. MICHAEL H. TRENK: This question is directed at the audience as well as the panelists. Should statistics be provided free by government or should users pay a fee that roughly matches their cost? Who should control the supply and the demand for statistics? How open should access to statistics be?

MR. WOODSON: By providing statistics essentially free, the government makes them available to everyone equally. Otherwise the people's data wouldn't be available to all the people. In that sense, I'm against differential pricing of these kinds of data.

By making statistics available to everyone, creative people in society can accomplish things that might not otherwise be accomplished. For example, somebody has created a product that uses census data to measure very small markets in terms of five or six different classifications for elderly people in the country, targeted at people in the business of elderly care facilities.

It's important to have these data disseminated among the public sector, because if the public sector can use the information to enhance their own wealth, then in turn the country's wealth is enhanced.

DR. SCHEUREN: It's debatable whether statistics are a public good or not. In some cases such as the zip code example, statistics are a quasi public and quasi private good. A partnership between the federal government and the private sector would be appropriate. I don't believe that the federal government should be in the business of producing private goods. But if a public good has differential value for different consumers and that differential value can be related to differential costs in the creation of that good, differential pricing may be possible. Personally, I believe that the market has a role in managing the federal statistical system and that some pricing of these goods is valuable.

The Canadians, for example, have enormously increased the price of their statistical products. Many European countries have done the same thing, especially in Scandinavia. The U.S. government spends more than a billion dollars a year on statistics. Outside help is needed to manage these resources, whether it's in the form of people writing their congressmen, or someone buying a product like the zip code example.

DR. HORVITZ: The federal government needs certain information to carry out its missions. If the federal government has to collect information in order to conduct its business, that information should then be released to the private sector, since the information has been generated by public funds.

I feel that the health care delivery system's efficiency depends upon the private sector having the necessary information to develop, finance, and carry out its plans. Withholding information from the private sector prevents the private sector from realizing its full potential. I

ROLE OF GOVERNMENT STATISTICS IN A DEMOCRATIC SOCIETY

feel strongly that statistics collected for the business of government should be made available immediately to the general public, so that the private sector can do its job in an efficient way. We would all benefit by it.

MR. JOHANSEN: Earlier this year, OMB issued a draft directive which would have severely limited the gathering and dissemination of federal statistics. Statistics would be gathered only if an agency was required to have them either for its own use, or as directed by law. Statistics would be analyzed only to satisfy agency needs and disseminated only as available. OMB received several hundred comments about the directive, with almost all of them objecting to it. The Council of Professional Associations on Federal Statistics now has a report from Mr. J. Timothy Sprehe of OMB indicating that the directive has been considerably changed in both tone and content. This episode shows that government can be responsive to public opinion on these matters.

Dr. Scheuren was talking about mortality of wealthy people. Does the IRS have any plans for linked mortality studies based on income?

DR. SCHEUREN: We are doing such a study right now for 1979 using occupation and income from the tax return and industry data from the Social Security Administration. We did a similar study for 1969 which used just the income data to analyze income differentials and mortality. Mortality was lower among the wealthy. The results are reported in a paper by Caldwell and Diamond which appears in the 1979 American Statistical Association Proceedings, Section on Survey Research Methods. Administrative records make these studies a lot cheaper, quicker, and easier to do.

MR. JOHANSEN: At a session this morning, mention was made of an insurance company information return called the Form 8390. Will your unit be handling those forms, Dr. Scheuren?

DR. SCHEUREN: Most of our budget money is spent on forms you've never heard of. Yes, we probably will be using Form 8390. We process nearly all of the corporate life insurance forms for statistical purposes.

