THE IMPACT OF NEW LEGISLATION ON STATISTICAL AND RESEARCH USES OF SSA DATA

Thomas B. Jabine, Social Security Administration

INTRODUCTION

Balancing the requirements of the public's need for information against those of protecting the confidentiality of information about identifiable individuals has always been a matter of importance to the Social Security Administration.[1] Records maintained for social security programs are a rich and unique source of information about labor force participation, earnings history, mobility, recipt of benefits, use of health services and even mortality. Combined with data from other sources, usually on a sample basis, this information can be used for research on important economic and social issues. Such uses, however, cannot be allowed to affect individual data subjects in ways that are incompatible with the purposes for which the data were initially made available by them or on their behalf.

These matters have received even more attention than usual during the past 2 years, starting with the release, in July 1973, of the report of the Secretary of HEW's Advisory Committee on Automated Personal Data Records [8], sometimes referred to as the Ware Report, and proceeding through the consideration and passage of the Privacy Act of 1974 [15c] and the 1974 Amendments to the Freedom of Information Act [15b]. These events are having a profound effect on all of the functions and operations of the Social Security Administration. This paper will present a preliminary description and assessment of their impact on SSA's statistical and research activities. Readers or listeners are cautioned that the paper is being written during a period of rapid change. Operational experience in complying with provisions of statutes and regulations, and issuance of new regulations and guidelines may eliminate some of the problems and uncertainties described and may create others. This caveat applies particularly to all discussion of the Privacy Act of 1974.

To provide the background needed to understand the impact of new legislation on SSA statistical and research activities, it is necessary first to summarize the scope and objectives of these activities, their placement within the SSA organization, and the statutory and regulatory framework within which they are carried on.

SSA STATISTICAL AND RESEARCH ACTIVITIES

The Office of Research and Statistics (ORS) is the principal research arm of the Social Security Administration. Its main function is to conduct program-oriented studies in the areas of income maintenance and health insurance. Research covers topics such as the redistributive effects of social security benefits; the relation of public and private income maintenance programs; and the impact of existing and alternative health benefit systems on the cost, availability and quality of health care.

ORS also tries to make SSA data resources available to outside users to the greatest extent possible, consistent with its primary mission. This is done in part by publishing results of its own studies, in part by providing access to statistical data bases and in part by providing access to program records. In the last instance, ORS usually acts as an intermediary between the outside user and SSA's Bureau of Data Processing, which maintains the operating files. In all cases, appropriate restrictions are imposed to preserve the confidentiality of information about individuals.

ORS program-oriented research takes several different forms, depending on objectives and on the sources of information available. The main sources of information are surveys, experiments or demonstration projects, and program records. Some discussion of each source follows.

Surveys - Surveys may be directed at beneficiary populations, or at target populations, i.e. the populations of individuals defined as being potentially eligible for specified programs. An example of the former is the Current Medicare Survey, a continuing panel survey of aged and disabled Medicare beneficiaries. An example of the latter is the 1971 Survey of Recently Disabled Adults, a survey of persons reported in the 1970 Census of Population as not disabled but identified as disabled in a subsequent screening survey.

For most of the large-scale national surveys, the Census Bureau has acted as SSA's collecting agent. Private contractors have been used in a few instances. Data processing and analysis have been done primarily by ORS, although more recently some of these functions have been carried out by contractors.

Some of the surveys are longitudinal. The Survey of the Low Income Aged and Disabled was conducted in two rounds, the first just prior to the initiation of the Federal Supplemental Security Income program in January 1974, and the second a year later. The Longitudinal Retirement History Survey is designed to follow a cohort of individuals through the pre-retirement and early retirement years. Survey data were collected in 1969, 1971 and 1973, and additional rounds are scheduled for 1975, 1977, and 1979.

In order to improve the accuracy of the survey data and to provide additional information for analysis, program data on earnings and benefits are often extracted from the operating files and merged with the survey data for the persons in the sample.

For some more narrowly defined studies of special populations, all data collection, processing, and analysis functions are carried out by contractors or grantees, and SSA receives only the final report.

Experiments and Demonstration Projects - Various amendments to the Social Security Act in recent years have authorized experiments with alternative methods of providing health services under Medicare, Medicaid and maternal and child health programs, and with the methods of reimbursement to the providers of health services for existing and alternative types of services [7, Sec.1875, p.481]. The usual procedure is to have one contractor carry out an experiment or demonstration and to have a second contractor evaluate the results. Alternative methods are evaluated for feasibility and for their effects on the cost and quality of health care.

Most of these experiments do not make use of data on individual recipients of health services, but there are some exceptions, e.g., a project concerned with the provision of ambulatory surgical services and one designed to study the effects of coinsurance, copayments and deductibles in drug insurance programs.

Program Records - ORS has developed several valuable data bases for program research and evaluation by extracting samples of records from major SSA operating files, such as the Summary Earnings Records, the Master Beneficiary Records, the Medicare Enrollment and Payment Record Files, and others. Typically, data for the same sample individuals are brought together from several sources to provide a comprehensive record of participation in one or more SSA programs.

The best known of these systems is the Continuous Work History Sample (CWHS), a one-percent longitudinal sample of individuals which combines data on earnings from employer reports and reports of self-employment, demographic information from the individual's initial application for a social security number, and industrial classification and place of work from the employer's application for an identification number. Since over 90 percent of the labor force works in employment covered by Social Security, the CWHS, in addition to providing an important data base for studies directly related to SSA programs, has been widely used for national and area studies of employment, earnings, migration and related subjects. Recently, the development of a 10-percent sample providing the same kind of data based on employer reports of earnings for the first quarters of 1971 and 1973 has substantially increased the amount of area detail available.[5]

Some data bases have been developed which combine data from SSA administrative records with administrative and/or survey data from other Federal and State agency sources. Often, both SSA and the other agencies involved are interested in using the results, so a joint project is established.

The most complex of these projects, and probably also the one with the greatest potential benefits is a series of studies involving the linkage of data from Census, the Internal Revenue Service and SSA. These studies, which are jointly undertaken by the three agencies, have several research objectives, which may perhaps be summed up as the development of more accurate data on the distribution of individual and family income and its components, along with related demographic and socio-economic variables. Experience over many years with the Census Bureau's Current Population Survey and other surveys has shown that the survey process alone cannot produce income data sufficiently accurate for all types of analysis. This alternative approach has been developed and tested over a period of more than 10 years, always with the greatest of care to observe all confidentiality requirements of the agencies involved.[4,12, 13] It is now being used to develop a data base for the evaluation of current and projected future effects of existing and proposed transfer and tax programs.

Other such interagency joint projects include one with the Vocational Rehabilitation Service for evaluation of the relationship between vocational rehabilitation and subsequent earnings and SSA benefits, and one with the Civil Service Commission to study the relationships between retirement and survivors benefits under the SSA and Civil Service programs. Also, arrangements have recently been made with the Office of Tax Analysis, Treasury Department, to merge SSA earnings data with data from individual income tax returns for a sample of taxpayers. The merged data file will be used to study tax-transfer policy issues.

Non-SSA Uses of SSA Data - In an official statement of its objectives [10], SSA included the following:

> Objective No.14 - Contribute to Government-wide and community planning for the aging, the widowed, the disabled, and children.
>
> Objective No.15 - Contribute to the effectiveness and economy of operations Government-wide.
>
> Objective No.16 - Contribute through the effective development of our unique record resources to the improvement of the Federal statistical system and to the sources of knowledge of the economy.

These stated objectives are achieved, in part, by making SSA data available to outside users in various forms. It is important to note here that while we are speaking of non-SSA uses in a formal sense, some of the activities described below have produced and continue to produce results which are of substantial value to SSA for various types of policy analysis.

1. Publications. Regular program statistics and articles based on survey results and other sources of information are published monthly in the Social Security Bulletin. An Annual Statistical Supplement to the Bulletin contains general time-series data on social security and the economy. Most large-scale surveys and administrative record sample data systems provide the basis for a variety of publications, ranging from articles and brief notes to comprehensive survey reports and monographs.

2. Frames for surveys. Employer lists, including name and address, standard industrial classification and, in some cases, a size code based on number of employees, have been made available to other Federal agencies for use in statistical surveys only. The Bureau of the Census has been

the principal user, relying on SSA as its primary source of information on new employers for use in economic censuses and surveys. Other agencies that have received lists of selected types of employers are the Department of Agriculture's Statistical Reporting Service, the Office of Education, the Bureau of Economic Analysis, the Public Health Service, the Federal Trade Commission, and the Federal Reserve Board.

Lists of individual participants in SSA programs have not been made available to non-SSA users for surveys.

3. Microdata. Tape files of individual data records, with identifiers removed or scrambled, and with deletion or restriction of other items to avoid disclosing the identity of specific individuals, are released to users in order that they may tabulate these records and perform statistical analyses appropriate to their own research interests. Files from the Continuous Work History Sample (CWHS) have been widely used in this way. Record files from several surveys have been or are in the process of being released as "public use" files.

Other kinds of releases of microdata files are made on a more restricted basis. A CWHS file with social security numbers and employer identification numbers included was released to the Census Bureau for its use on an experimental basis to study the relationships between place of residence and place of work, and to prepare tabulations combining data for individuals with data on the characteristics of their employers. Limited data, including name and address, for a sample of persons enrolled in Medicare were made available to the Census Bureau for its use in evaluating the completeness of coverage of persons aged 65 and over in the 1970 Census.

There have been other releases of SSA program data for identifiable individuals to selected Federal agencies for statistical and research purposes. SSA earnings data for participants in manpower training programs have been released to the Manpower Administration, Department of Labor, for use in the statistical evaluation of these programs. Data on earnings, benefits and, where applicable, the fact and circumstances of death, for identified individuals have been supplied to the National Institute for Occupational Safety and Health and to the Atomic Energy Commission for use in studies of the effects of radiation and other occupational hazards on employed persons.

4. Special tabulations. Special-purpose tabulations from various ORS data systems may be prepared at cost for non-SSA users, subject to restrictions of administrative feasibility. An important subset of such cases is that where the user wishes to combine SSA data, usually on earnings or benefits, with data that he has obtained from other sources for individuals in a population or sample he is studying. The standard procedure is for the user to give SSA a file containing his data, including social security numbers and other identifiers. SSA locates the earnings, benefits and other data specified, merges this information with the user's data for these individuals, and prepares tabulations as specified by the user, taking care to avoid the release of small data cells which could allow the user to identify SSA information for specific individuals.

This technique has been employed most commonly in the evaluation of various training and rehabilitation programs, and in epidemiological studies of persons exposed to various kinds of occupational and environmental risks.

Some users for this type of study would prefer to have access to the merged microdata files containing SSA data and their data for the study population, in order to carry out statistical analyses which they feel cannot be fully specified in advance, but rather require a sequential or interactive approach. Such a file, without identifiers, was released to one academic research group under very strict conditions on use and publication of results; however, we do not expect to continue this practice because of the theoretical possibility which exists for researchers to identify individuals in the file by sorting and matching on the basis of data items previously in their possession. As an alternative, we are now experimenting with techniques that will permit a researcher remote access to the merged microdata files in order to undertake statistical analyses in an interactive mode, but with constraints that deny access to any individual records.

STATUTORY AND REGULATORY FRAMEWORK FOR SSA STATISTICAL AND RESEARCH ACTIVITIES

The statistical and research programs described above operate under conditions imposed by several different statutes and regulations. The authorization for most ORS programs is contained either in Section 702 or in Section 1875 of the Social Security Act.[7]

Section 702 assigns to the Secretary of HEW "the duty of studying and making recommendations as to the most effective methods of providing economic security through social insurance, and as to legislation and matters of administrative policy concerning old-age pensions, unemployment compensation, accident compensation, and related subjects." Section 1875 carries similar requirements with respect to health care of the aged and disabled. For programs undertaken jointly with other agencies, of course, authorization comes in part from the statutes of those agencies.

Another part of the Social Security Act that bears heavily on these activities is Section 1106, which deals with "Disclosure of Information in Possession of Department." Section 1106, and SSA Regulation 1 [9], issued by SSA under the authority of Section 1106, have, in the past, prohibited the disclosure of any kind of information received by SSA in the course of carrying out its functions, except for certain purposes expressly defined in Regulation 1, or as otherwise expressly authorized by the Commissioner of Social Security. The latter provision has been referred to as the Commissioner's "ad hoc" authority to disclosure information. It has been exercised by means of a "Commissioner's Decision" process, which requires a formal review of any proposed release by all interested components of SSA prior to a decision by the Commissioner.

Some of the disclosures of information for statistical and research purposes described earlier have been made under specific provisions of Regulation 1, without the need for a Commissioner's Decision. These include the publication of statistical data (always with the condition that data in small cells which might disclose information about an individual beneficiary, employer, or provider be suppressed), the preparation of special tabulations on a reimbursable basis, and the release of microdata files without identifiers (in the broad sense). Other kinds of disclosures, such as the releases to the Census Bureau of the CWHS file with unscrambled SSN's and of a sample of Medicare enrollees with their names and addresses have been made on the basis of specific Commissioner's Decisions.

Over the years, Regulation 1 has been amended several times, to provide for specific types of disclosures considered essential for new programs established under the Social Security Act or other legislation. A basic revision of Regulation 1 became effective in July 1975 [9], at which time its scope was restricted primarily to data about individuals, in the sense of natural persons, with questions concerning the release of other kinds of information to be governed by the Freedom of Information Act (FOIA)[15b] and its implementing regulations [9]. Thus, most information not relating to individuals is now available to the public unless exempted under FOIA and the implementing regulations.

SSA statistical and research activities are also affected significantly by the provisions of the Federal Reports Act (15a), which provides for the coordination of Federal statistical activities under the direction of the Statistical Policy Division of the Office of Management and Budget (OMB). All Federal statisticians having responsibility for the collection of information from the public in surveys or by other means are familiar with the forms clearance requirements, by means of which OMB exercises its duty to minimize the burden on respondents, to avoid duplicative collection activities, and to restrict content to that which is necessary "for the proper performance of the functions of the agency or for any other proper purpose."1/ In addition, the Federal Reports Act contains provisions concerning the conditions under which one Federal agency may release information to another.

Under the authority of the Budget and Accounting Procedures Act of 1950 (31 U.S.C. 18b), OMB has issued Circular No.A-46 (Revised, May 3, 1974) on the subject of "Standards and Guidelines for Federal Statistics"[3]. The standards and guidelines in this circular are intended "...to improve the gathering, compiling, analyzing, and publishing of statistical data." Exhibit A of Circular A-46, "Standards for Statistical Surveys" includes a section on "Relations with the public," which contains most of the requirements subsequently incorporated in Subsection e(3) of the Privacy Act of 1974 (which covers any kind of collection of data about identifiable individuals, not limited to surveys) for informing individual respondents about authority and purposes of the survey, and whether response is mandatory or voluntary.

COMPLIANCE WITH NEW LEGISLATION

The Freedom of Information Act

Although the FOIA was passed in 1967, its effect on SSA statistical and research activities was minimal prior to the February 17, 1975 effective date of the 1974 amendments to FOIA. In fact, as late as 1974, the Office of Research and Statistics reported to SSA's Freedom of Information Officer that it had not yet received any requests for information under FOIA!

The discussion immediately following the passage of the 1974 amendments to FOIA made it clear that many of the inquiries being received and handled routinely by ORS fall within the scope of FOIA and must be handled in accordance with its provisions and those of the implementing regulations of HEW and SSA.

The definition of "record" used in FOIA is broad, covering publications, internal reports and memoranda, questionnaires, and information recorded on tape or microfilm or computer printouts. The Act as amended also provides that "any reasonably segregable portion of a record shall be provided...after deletion of the portions which are exempt under this subsection." This raises the possibility of a requester asking for copies of microdata tapes developed for statistical and research purposes, with obvious identifiers and other information which could lead to the identification of individuals deleted. Whether such tapes would be useful to the requester would depend on a number of factors, including the amount of documentation available, the extent to which the tapes had been edited to minimize the effects of response and processing errors, and the extent of deletions required to meet confidentiality requirements.

Within SSA's Office of Research and Statistics, response to the 1974 Amendments has been of two kinds. First, there has been a thorough effort to inform all employees about the requirements of FOIA, so that they will recognize FOIA requests when they are received, and handle them quickly and properly. Guidelines prepared specifically for ORS staff have been circulated to all employees. Several orientation and training sessions have been held for employees, and the compliance requirements have been discussed at length in executive staff meetings. A member of the staff with both statistical and legal training has been identified as a "resident expert" on FOIA and is available for consultation on cases where the appropriate response is not obvious.

Second, ORS policies for the release of microdata without identifiers have been thoroughly reviewed and updated to make them consistent with both freedom of information and privacy legislation. It is the intention of ORS to release microdata files to users to the greatest extent and on as timely a basis as possible, subject to confidentiality requirements, and without interfering with its primary obligation to conduct program-oriented research for SSA. Several files have already been made available with suitable documentation, and others are now being made ready. No conditions will be placed

on the use of files for which it is ascertained that all requirements are met for protecting the confidentiality of data about individuals.

Two specific requests that ORS has handled under FOIA may be of interest to statisticians. The first relates to one of our demonstration projects in the health area. A contractor for SSA is operating an ambulatory surgi-center, i.e., a facility in which minor surgical procedures are performed and the patient does not remain overnight. A second SSA contractor is conducting an evaluation to determine how the costs and quality of health services are affected by the availability of this type of facility.

The evaluation contractor submitted a detailed research design and plan of analysis to SSA for approval. A request was received for a copy of this document. The project officer and his superiors in ORS wanted to deny the request on the grounds that availability of the detailed evaluation plan to those directly involved in the demonstration might condition their behavior in ways that could bias the results of the evaluation.

The Office of the General Counsel was of the opinion that the document requested did not come under any of the exemptions in FOIA, so it was given to the requester.

The other request of particular interest was for information relating to a validation study of hospital accreditations. By law, hospitals may be certified for participation in the Medicare program in one of two ways--either by accreditation received from the Joint Commission on the Accreditation of Hospitals (JCAH), a private organization, or through surveys conducted specifically for this purpose by State agencies for SSA. Recent legislation required SSA to conduct a validation study of the JCAH accreditation process. This was done by selecting samples of hospitals recently surveyed by JCAH and arranging for State agency survey teams to survey these hospitals, using the regular Medicare certification survey procedures.

Following release of some of the initial results of this study, including the identification of hospitals which failed to meet Medicare certification standards, a letter was received from a State hospital association which requested, among several other items relating to the validation study, full details of the sample selection methodology used as follows:

> Details of the methodology used by the SSA Office of Research and Statistics in selecting participating accredited hospitals to be surveyed on a sample basis. This should include the specifics of the statistical sampling technique; the total hospitals included in each universe (identified by name and location of institution), and other information which would permit validation of the sampling technique with respect to any individual hospital surveyed on a sample basis. If a computer program is involved, we would like copies of any flow charts, narratives, and any other documentation available to permit us to identify and interpret the algorithm used. [6]

Since no data about individuals were included in the request and there were no other obvious reasons for a denial, all of the information requested was made available. The point of this story is simply to make it clear that government researchers and statisticians who become involved in studies that are controversial or have significant policy implications will more than ever be operating in full view of the public. They must fully document their methods and procedures, for if they do not, FOIA requests can demonstrate that they have not and thus cast doubt on results. They must also be prepared to defend the methods and procedures they have adopted.

The Privacy Act

The efforts required for compliance with the Privacy Act of 1974 are of a completely different order of magnitude from those called for by FOIA. Since this paper is being presented prior to September 27, 1975, the effective date for most provisions of the Privacy Act, the full impact of the Act's requirements on statistical and research programs has yet to be felt, and we can talk only about our experience with preparations for compliance.[2/]

Our compliance activities to date have been based on the provisions of the act itself, on the Privacy Act Guidelines issued by OMB[3/], on the recommended format for public notice issued by the Office of the Federal Register[4/], and on instructions issued by SSA officials responsible for these activities. For a number of specific implementation questions we have been unable to obtain clear answers from any of these sources. Since deadlines imposed by the implementation schedule have not allowed us to wait for answers, our policy has been to do what we think is intended by the act and proceed with the knowledge that we may have to revise our procedures later.

Review of the provisions of proposed legislation made it obvious to us, even prior to the passage of the Privacy Act in December 1974, that an inventory of ORS record systems was essential. We conducted a pilot inventory in one division in December 1974, revised the questionnaire and instructions based on that experience, and inventoried all ORS record systems in March and April 1975.

For the inventory, each division was asked to complete a relatively simple two-page form for each record system under its control, following as closely as possible the definition of record system used in the Privacy Act. The content of the inventory form was largely determined by the public notice requirements of the act; however, in addition we obtained information on

- Non-statistical uses of information in ORS record systems (important in connection with possible future exemptions from the access provision of the Privacy Act).
- Plans for additional data collection subsequent to September 27, 1975. This

identified the systems where public use forms would have to be reviewed and modified to meet provisions in subsection (e)(3) of the Act for informing respondents about authority for the collection of data, purposes and uses, whether response is mandatory or voluntary, and the effects on them of responding or failing to respond.

- Current or planned releases of data about individuals outside of SSA, with or without identifiers. All releases expected to continue after September 27 would have to be reviewed for compliance with Privacy Act requirements.

The initial returns from the inventory identified 79 record systems. In the process of preparing public notice statements we found it appropriate to delete some systems which by September 27 would no longer meet the definition of a record system, and to consolidate others. As a result, we published notices on 54 systems, of which 52 are used solely for statistical and research purposes and 2 have limited nonstatistical uses.

The review of active public use forms has been completed and provisions have been made in all cases for informing individual respondents of the reasons for requesting the information, the uses to which it will be put, and the consequences, if any, of not providing the information. Similar requirements for future data collection forms have been incorporated into the normal forms clearance procedures. Cooperation rates for the ongoing programs will be monitored for any deterioration which may result from the new respondent notification requirements.

In one survey for which we expect to begin collecting data later this year, most of the interviewing will be by telephone. The study population will consist of persons who have visited SSA district offices or called teleservice centers for assistance, and it is important that the survey interviews be conducted as soon as possible after the contact is made. The Privacy Act and the OMB Guidelines are not explicit on how the subsection (e)(3) requirements can be handled in telephone surveys. Our present plan is to print the required statement on the survey form and have the telephone interviewer read it to the respondent, and then offer to send him a copy if he would like to have it.

In another area of compliance with the Privacy Act, we have undertaken a thorough review of disclosures of SSA data about identifiable individuals and of disclosures to SSA of such data by other Federal agencies for research and statistical purposes. Under the Privacy Act, there are 4 ways in which individual records can be disclosed for research and statistical purposes.

1. With the consent of the individual to whom the record refers.
2. Within the agency, as needed by employees in the performance of their duties. The "agency" for this purpose, is the Department of Health, Education and Welfare.[5/]
3. To the Census Bureau for "planning or carrying out a census or survey or related activity pursuant to the provisions of Title 13."
4. For a "routine use," defined in the act as use for a purpose compatible with the purpose for which the information was collected.

As of this writing, we see no legal obstacle under the Privacy Act to the continuation of the following types of disclosures of records about identifiable individuals.[6/]

- Between SSA and Census in projects of joint interest, such as the match studies of income distribution described earlier. Two-way transfers are necessary in these studies because of the impracticability of transferring complete earnings or benefit files from SSA to Census. Census agents in SSA receive a "finder tape" with identifiers for individuals in the study population. This tape is used to locate and extract earnings and benefit information for these individuals which is then returned to the Census Bureau for merger with survey and IRS data.
- Between SSA and contractors in connection with research and statistical functions performed for SSA.
- Between SSA and the Census Bureau in connection with surveys for SSA in which Census undertakes data collection and/or processing functions on a reimbursable basis.
- To the Census Bureau for Title 13 purposes, e.g., for its use in checking the completeness of coverage in the Population Census.
- Between SSA and other Federal agencies such as the Civil Service Commission, the Railroad Retirement Board, and the Treasury Department's Office of Tax Analysis for projects of joint interest to the agencies involved. Such transfers may take place under routine uses established by the agencies involved.

Other kinds of disclosures, especially those deemed not to be directly related to the administration of the Social Security Act, may be discontinued. Continuation of these disclosures would require both a legal determination that they are permissable under the routine use provision of the Privacy Act, and a policy decision by SSA to establish routine uses for these purposes.

Operationally, by far the most complex and burdensome requirements imposed on ORS by the Privacy Act are those related to the access provision, which gives any individual the right to be notified whether a record system contains a

record for him, and, if it does, to receive a copy of the record in a form he can understand. The act provides that exemption from this provision can be established by agency heads for record systems "required by statute to be maintained and used solely as statistical records;" however, statutes governing the establishment and use of ORS records systems do not contain such a provision and the exemption is not presently available to us. Pending any legislative action which would make it available, ORS must be prepared to comply with the access and related provisions of the act.

Unlike SSA program records, the files used by ORS for research and statistical activities have not been designed for rapid and inexpensive retrieval of records for specific individuals identified by name, address or Social Security number. To prepare for access is, therefore, requiring a very substantial systems effort. We do not expect to receive large numbers of requests for access to ORS records systems, because, with 2 exceptions, they are never used for determinations about individuals. However, we could not assume that there would be no requests, and so we have had to draw heavily on resources ordinarily used for other purposes[7/] to prepare for compliance with the access provision.

This discussion of our activities in preparing to comply with the Privacy Act of 1974, while rather lengthy, only begins to give an idea of the steps required to comply with this complex legislation. We have estimated that during the current fiscal year (July 1975 through June 1976) about 2 man-years will be needed for compliance activities, exclusive of the efforts required to fulfill the access requirement. Preparing for access will require a very substantial initial systems and programming effort, and will add significant costs to each new program, unless some kind of relief is obtained.

THE IMPACT OF NEW LEGISLATION

General Remarks

The rest of this paper is a preliminary, personal assessment of observed and potential effects of privacy and freedom of information laws on statistical and research uses of SSA data. While some problem areas have already been clearly identified, there are undoubtedly others which we have not yet fully discerned. Much depends on how the requirements of the Privacy Act are spelled out in the implementing regulations of Federal agencies, and on how certain key phrases of the act, such as "relevant and necessary" and "compatible with the purpose for which it [a record] was collected" are interpreted. New legislation at the Federal and State levels, and amendments to existing legislation, such as HEW's proposed amendment to Section 7 of the Privacy Act, dealing with the use of social security numbers, may place further constraints on some kinds of statistical and research activities.

Because this legislation has such serious implications for producers and users of statistics, and because the effects cannot be clearly seen today and will not be for at least a few months yet, it would seem appropriate and useful to plan for a followup session to this one at the 1976 annual meetings of ASA.

I don't want to imply a belief on my part that these laws have had only bad effects on Federal statistical and research activities. Quite the contrary. Some results are universally accepted as beneficial; other outcomes are more controversial among Federal statisticians, but are supported by many as being desirable. In this evaluation I will discuss both the good and bad results as I see them.

Freedom of Information Act

A good result of the 1974 FOIA amendments, which imposed time limits for responses to requests, has been an effort to handle all requests for information on a more timely basis. While one might argue that such an effort could interfere with ORS's primary functions in serving SSA, we seem to be adapting to this requirement without undue problems.

Some problems may arise in connection with requests for preliminary tabulations and reports of various research projects. What is an administrator to do, for example, when he gets an FOIA request for a report on a controversial subject, and the staff member who prepared the report has used statistical techniques which the administrator or another of his staff members thinks are incorrect or inappropriate? Recent court decisions under FOIA suggest that factual portions of such materials must be released, but that opinions expressed as part of the internal decision-making process may be withheld. How do we separate fact from opinion in statistical analyses? If a report based on sample data contains a number of apparently factual statements, but these have not been checked for statistical significance because the sampling errors have not been computed yet, must it be released? Similar issues abound.

Tabulations requested under FOIA must, of course, be carefully reviewed to be sure that they do not contain any information that might permit recipients to identify specific individuals. Release of such information would, in general, be prohibited both under Section 1106 of the Social Security Act and under the Privacy Act. The same restriction would apply in the case of a request for microdata with identifiers removed. However, where this restriction is met, there is not any obvious basis for not releasing tabulations or microdata obtained at any stage of a survey or research project.

This possibility of a "forced release" of microdata files or tabulations has been a matter of considerable concern to researchers in SSA's Office of Research and Statistics. This is partly because it raises the possibility of irresponsible use of data files which have not

been sufficiently reviewed for consistency and accuracy prior to analysis and which have not been fully documented. Fortunately, most requesters appear to be willing to wait for data files that have been adequately reviewed and documented.

At the moment, the most serious adverse consequence of FOIA to statistical and research programs would appear to be the threat of invalidation or reduction of the value of certain kinds of research by the premature release of preliminary results or of detailed research designs. The problem is illustrated by the example given earlier of the release, in response to an FOIA request, of the detailed evaluation plan for an ambulatory surgery demonstration facility. Medical researchers in other parts of HEW have expressed great concern about this kind of problem and have called for further amendments to FOIA to resolve it.

To convince Federal administrators and legislators that there is a danger in this area, it will not be enough to describe the problem in a general way. It will be necessary to cite specific instances showing how premature release of information has or could invalidate important research efforts. Only then will it be possible to solve the problem through legislation or by other means.

Privacy Act

There have been some benefits from the compliance activities already described. Some data tapes no longer needed have been blanked, and some questionnaires and other records destroyed. Identifiers are being removed from records where they are clearly not needed any longer; and managers of new projects will undoubtedly plan to minimize the length of time for which identifiers are retained with data about individuals.

ORS, in keeping with SSA legislation and policy and OMB requirements, has always taken seriously the need to inform respondents to surveys fully about the conditions of their participation, and to protect the confidentiality of data provided by individuals in surveys or by means of their participation in SSA programs. Nevertheless, the increased emphasis on these requirements in the last two years, culminating with the passage of the Privacy Act, has led us to undertake a very careful review of relevant policies and procedures, especially with respect to interagency transfers of data for statistical and research purposes, and also with respect to steps taken to avoid disclosure of information about identifiable persons in connection with the release of tabulations and microdata files to users outside of SSA. While we believe that the possibility of significant harm to individuals resulting from unintended disclosure of the kinds of data collected in most ORS record systems is minimal, we nevertheless feel that in order to maintain our standards of compliance with confidentiality requirements at a high level, we must keep the probability of any such disclosures very close to zero. Toward this end, we are now undertaking a very careful review of the CWHS public use files to determine whether it may be desirable to restrict the content of these files or to "contaminate" the data in a way that will minimize the probability of unintended disclosure.

On the debit side, Privacy Act compliance activities have diverted, at least temporarily, a significant part of resources normally used by ORS in pursuit of its basic objectives. As explained earlier, the main effort will be that of complying with the access requirement. Accounting for disclosures will also be burdensome. If a legislative solution can be found, i.e., by provisions of law that would require record systems established by SSA for statistical purposes to be maintained solely for such purposes, then this diversion of resources will be largely temporary, and the records in these systems will be better protected than they now are. If such a solution is not found, the need to comply with access and accounting requirements will impose a substantial continuing tax on all of our research. Access to ORS record systems can be of little if any value to individuals, since these records are not used to make any determinations about them. Requests for such information, for the most part, can only result from idle curiosity, a desire to test the system, or misunderstanding of the nature and purposes of records in these systems.

One unfortunate feature of the Privacy Act has been its lack of specificity and, in at least one instance, its internal contradictions. This has caused several difficulties in the development of ORS compliance procedures, the principal ones being

- Lack of specificity as to the precise meaning of the term "routine use" as it applies to the disclosure of records to other agencies or to persons outside the agency, and failure to explain clearly whether intra-agency uses of records must be described as routine uses in public notices of record systems and in informing respondents.
- Failure to address the problem of how to inform respondents to telephone surveys of the conditions under which the information is being requested, particularly in the case where the initial contact is by telephone.
- Inclusion of the requirement that an agency maintain in its records about individuals information that is "..relevant and necessary to accomplish a purpose of the agency required to be accomplished by statute or executive order.."[15c, subsection (e)(1)]. While all ORS record systems are relevant to the objectives set forth in Sections 702 and 1875 of the Social Security Act, as described earlier, we would be hard put to classify some of these as necessary and others as not necessary for the accomplishment of these objectives. The problem is that the words "necessary" and "research" are not really compatible. This provision, if taken literally, could eliminate all statistical and research programs except those specifically required by statute or executive order,

such as the head count in the decennial census, or the maintenance of a consumer price index.

- Section (b)(5) of the Privacy Act says that records that are not "individually identifiable" shall not be disclosed unless the recipient "...has provided the agency with advance adequate written assurance that the record will be used solely as a statistical research or reporting record..."[15c] This is internally contradictory because a "record" is earlier defined as an item or collection of information that includes some type of identifier. It also contradicts section (b)(2) which permits disclosure without restriction when required under the Freedom of Information Act. We are thus left in substantial doubt as to how we should handle the release of microdata without identifiers.

No doubt these issues will be resolved one way or the other during the next several months by means of agency regulations, and legal opinions provided with respect to specific compliance issues. In the meantime, the orderly planning of statistical and research projects has been disrupted, and the start of some activities delayed as a result of these uncertainties.

Over the longer term, the Privacy Act may be expected to cause a reduction in the amount and kinds of statistical data which the Office of Research and Statistics will be able to provide both for internal SSA purposes and for outside users. The data that are provided will be less timely and of poorer quality. The bases for this prediction are as follows:

- The requirements for informing respondents (section (e)(3)), especially if interpreted in a formal and legalistic manner, will reduce cooperation rates in voluntary surveys.
- Several more steps have been added to the already complex clearance procedures for initiating and conducting surveys.
- Access and accounting requirements will cause delays in normal processing activities, and will significantly increase the overall cost of developing and maintaining a statistical data base.
- Many kinds of interagency transfers of data for statistical and research purposes will be eliminated. Alternative methods of obtaining the desired information, e.g., relying solely on the survey approach, will be more costly, place a greater burden on respondents and produce less reliable data.

A further word is in order about interagency transfers of data. The Privacy Act establishes irrational and arbitrary limitations on such transfers. This may be explained by the following model.

Consider a statistical study in which it is desired to merge data about individuals from Federal agency A and Federal agency B (neither agency being the Bureau of the Census). The results may be of interest to both agencies or only to one. As a practical and economic matter, in some cases the merger can be accomplished by a transfer of identifiable data in one direction only; in other cases, transfers in both directions may be necessary. The following chart shows in which of these circumstances the merger can be accomplished by using the "routine use" provision of the Privacy Act.

Merger is of interest to (compatible with mission of)	Direction of Transfer		
	A to B	B to A	Both directions
A only	Yes	No	No
B only	No	Yes	No
Both Agencies	Yes	Yes	Yes

Thus, if the study is compatible with the record maintenance purposes of both agencies, the merger can be accomplished by transferring the data in either or both directions. However, if it is "compatible" only for agency A, it can be accomplished only by having agency A transfer its identified records to agency B so that agency B can merge the records and transfer statistical summaries or merged records without identifiers[8/] back to Agency A. This procedure will often not be feasible, because it puts the main data processing burden on agency B, which is the one that is not interested in the study.

Even those kinds of merges for statistical and research purposes that are legally possible under the Privacy Act may be difficult to realize in practice, because many agency administrators will be reluctant to declare "routine uses" beyond those which they consider absolutely essential for operational reasons.

CONCLUSION

Most of the problems described in this paper which affect the statistical and research uses of SSA data apply more or less equally to all Federal agencies engaged in statistical and research activities, with the exception of the Census Bureau, which received special treatment under the Privacy Act. Thus, the predictions made apply with equal force to nearly all of the Federal Statistical system.

I have tried to demonstrate that the objectives of privacy and freedom of information legislation can be achieved only at a certain cost, expressed in terms of the adequacy and timeliness of statistical data and research findings needed or desired by society. Members of ASA, both as providers and users of data, are urged to consider these tradeoffs and make their views known to the Ad Hoc Committee on Privacy and Confidentiality, which is considering these issues.

Footnotes

1/ The phrase "for any other proper purpose" is significant. The Privacy Act places further restrictions on the collection of information by a Federal agency.

2/ To the extent that the Privacy Act incorporated the recommendations of the Ware Report, HEW has had some experience with compliance as the Secretary in February 1974 directed all units of the Department to do nothing contrary to those recommendations without his express approval.

3/ Draft issued March 7, 1975; final version on July 9, 1975[2].

4/ Draft issued April 22, 1975; final version on June 19, 1975[16].

5/ This provision will probably be interpreted as also permitting release to contractors of the agency and may also cover release to a second agency acting for the first agency on a reimbursable project.

6/ While the Privacy Act does not prohibit such disclosures, their continuance will depend on several factors, such as the willingness of agencies to establish the necessary routine uses, the willingness of Census to continue using the "census agent" provision of Title 13, and the absence of new legislative restrictions.

7/ Individuals requesting records may be charged fees for making copies, but not for the costs of searching for and retrieving their records.

8/ It must also be noted that it is theoretically possible, at some cost, for an agency receiving merged records without identifiers to match these records against its own source files and restore the identifiers.

Bibliography

1. Cooper, Heyman C. and Joseph Steinberg, "Social Security Statistical Data, Social Science Research, and Confidentiality," Social Security Bulletin, October 1967.
2. Executive Office of the President, Office of Management and Budget, "Privacy Act Guidelines-July 1, 1975," Federal Register, Vol.40, No.132, July 9, 1975, pp.28949-28978.
3. Executive Office of the President, Office of Management and Budget, "Standards and Guidelines for Federal Statistics," Circular No.A-46, Revised, Washington, D.C., May 1974.
4. Herriot, Roger, "Discussion" (of papers in session on "The Role of the Social Security Number in Matching Administrative and Survey Records"), Proceedings of the American Statistical Association Social Statistics Section, 1974, pp.155-156.
5. Hirschberg, David A., "The Continuous Work History Sample," Proceedings of the American Statistical Association Business and Economic Statistics Section, 1974, pp.54-56.
6. Illinois Hospital Association, letter to Assistant Commissioner for Public Affairs, Social Security Administration, April 25, 1975.
7. U.S. Congress, Senate, Committee on Finance, The Social Security Act (As Amended Through January 4, 1975) and Related Laws, U.S. Government Printing Office, Washington, D.C., February 1975.
8. U.S. Department of Health, Education and Welfare, Records, Computers and the Rights of Citizens. Report of the Secretary's Advisory Committee on Automated Personal Data Systems. U.S. Government Printing Office, Washington, D.C., July 1973.
9. U.S. Department of Health, Education and Welfare, Social Security Administration, "Law and Regulations Issue No.209." Includes Regulation No.1, "Disclosure of Records and Official Information," and Regulations No.22, Subpart E, "Availability of Information and Records to the Public." July 1975.
10. U.S. Department of Health, Education and Welfare, Social Security Administration, The Objectives of the Social Security Administration, Chapter III, "Responsibility to the Government as a Whole," August 1965.
11. U.S. Department of Health, Education and Welfare, Social Security Administration, The Research Program of the Social Security Administration. Report of the 1967 SSA Advisory Committee on Research Development, U.S. Government Printing Office, Washington, D.C., 1968.
12. U.S. Department of Health, Education, and Welfare, Social Security Administration, Office of Research and Statistics, "Some Observations on Linkage of Survey and Administrative Record Data," by Joseph Steinberg, Studies from Interagency Data Linkages, DHEW Publication No.(SSA)74-11750, Washington, D.C., August 1973.
13. U.S. Department of Health, Education and Welfare, Social Security Administration, Office of Research and Statistics, "1973 Current Population Survey Exact Match Study: Initial Stages", by Roger Herriot and Frederick J. Scheuren, Studies from Interagency Data Linkages: Report No.4, U.S. Government Printing Office, Wash., D.C., 1975.
14. U.S. Department of Health, Education and Welfare, Social Security Administration, Office of Research and Statistics, Work Plan, Fiscal Years 1975 and 1976, ORS publication No.008-74, Washington, D.C., August 1974.
15. U.S. Laws, Statutes, etc.
    a. The Federal Reports Act (as amended), Public Law 90-620, 44 U.S.C.§421, 90th Congress, October 22, 1968.
    b. The Freedom of Information Act (as amended), Public Law 93-502, 5 U.S.C. §552, 93rd Congress, H.R.12471, November 21, 1974.
    c. The Privacy Act of 1974, Public Law 93-579, 5 U.S.C. §552a, 93rd Congress, S.3418, December 31, 1974.
16. U.S. Office of the Federal Register, "Publication Guidelines for the Privacy Act of 1974, "Federal Register, Vol.40, No.119, Part V, June 19, 1975, pp.25988-25997.