

---

# Access to Social Security Microdata Files for Research and Statistical Purposes

by Lois A. Alexander and Thomas B. Jabine\*

---

This article focuses on the characteristics of SSA microdata files and on the development of a disclosure policy aimed at serving the public interest while protecting the privacy of individuals and the confidentiality of research and statistical information. Several dimensions of the disclosure question are explored: A description of the persons (both living and dead) and other entities that are data subjects, the ability of users to associate known data subjects with information about them, the sources of data, expectations as to recontact of data subjects, and the terms and conditions under which microdata are released to users outside SSA. The factors controlling the decision whether or not to release microdata are also discussed. The factors range from those intended to protect the data subject—the criteria specified by law for maintaining confidentiality for example and the principles applied by SSA in assessment of disclosure risk—to those more concerned with agency function, such as financial cost to the agency, and interference with its primary mission. Some particular practices are described to illustrate application of present policy principles. Brief attention is given to future implications of certain current developments such as the Privacy Act, the Sunshine Act, and the Tax Reform Act of 1976.

---

This article presents a broad view of the conditions under which the microdata files of the Social Security Administration are available to researchers. Microdata, as used here, are those files containing records with information about individual data subjects from a defined study population. For the most part, the data subjects are persons (both living and dead). In some files, however, they may be other entities, such as employers or providers of services under Medicare. Most of the files discussed include information only for a sample of the data subjects in the study population.

A key element in determining conditions of access to a particular file is whether or not the records for individual data subjects include identifiers such as name, address, social security number, or employer identification number. Since microdata files with identifiers are available for research purposes only on a very restricted

basis, most of this article will be about access to files without identifiers.

The Social Security Administration has a relatively liberal policy of making research data available. It is easier to generalize about broad availability, however, than to identify simply criteria on which to base routine practical decisions about particular releases of microdata. The first and rather formidable hurdle in the way of simple policy statements is the complex set of legal requirements that must be met. Restrictions arise in a number of quite different statutes, with different purposes and contexts and with varying degrees of interaction.

The Social Security Act prohibits the disclosure of any information except as provided by published regulations. In the past, the Social Security Administration regulations have permitted release of anonymous statistical information "not relating to any particular person" and have also allowed release to other Federal agencies of identifiable employer information for use only in statistical and planning work of those agencies. Disclosure of certain information about deceased persons was

---

\* Office of Assistant Commissioner, Office of Research and Statistics, Social Security Administration. Adapted from a paper presented at the Workshop on Policy Analysis with Social Security Research Files, Williamsburg, Va., March 1978.

permitted routinely by the regulation, and ad hoc authority was used occasionally to release information to other agencies (primarily the Bureau of the Census) for statistical use

That situation has been somewhat changed by the passage of the Privacy Act in 1974. This legislation narrowed the limits of discretionary release under the Freedom of Information Act (FOIA) passed in 1966. Now releases may be made only if they are either authorized by an express provision of the Privacy Act or are required by the FOIA. The recent "sunshine" legislation has added an element of uncertainty on what releases are required, as distinguished from merely being permissible, under the FOIA. The latest legal complication was introduced by the Tax Reform Act of 1976, which includes in its expansive definitions of information under the control of the Internal Revenue Service the earnings reports for persons in covered employment under the Social Security Act. The full implications of this change have not yet been clearly identified.

Added to these complexities is the fact that the FOIA provides penalties for improper withholding of information requested and that other statutes provide penalties for improper disclosure. The limitations on disclosure placed by these statutes are discussed more fully later in the article.

A second obstacle to simplicity comes from the considerable differences in the needs and circumstances of the data users. Some requests involve only minor programming to obtain records from SSA's existing statistical data systems, such as the Continuous Work History Sample. In other situations—when, for example, the researcher provides input in the form of identifiers (social security numbers) and characteristics of a particular study population and wants the Social Security Administration to merge information from program records—the task is greater, both in terms of computer access and in terms of editing for nondisclosure of individual information.

Even if the sample covers only a few hundred individuals, the process of extracting earnings or beneficiary information, for example, would require retrieval from one or two large computerized systems, each of which contains many millions of individual records. Once the records are compiled, the Office of Research and Statistics (ORS) of the SSA must then review the microdata, in the same manner as it would review tabulations, to assure that the researcher could not identify particular items of information with particular individuals known to the requester.

A third consideration in framing access policy is the relationship of the outside research to SSA's own program mission. Like all Federal agencies, SSA is required by law to recoup the cost of providing information to non-SSA users, and most ORS work for outsiders

is consequently performed under reimbursable agreement. SSA does not have a formal statistical service function, however, and its work for outsiders is performed as a byproduct of its own research activities. By law, ORS is not permitted to allow its outside work commitments to interfere with SSA program-related activities, in particular with ORS's own work plan.

Finite budget and personnel limits are placed on the statistical resources available for performing reimbursable statistical work, and there may be hard choices to make in setting workload priorities. A project undertaken by SSA jointly with another agency, for example, may take precedence over outside contract work without SSA program relationship. In other situations, the size and timing of the outside project, the amount of programming involved, the possibility of "piggybacking" on a regularly scheduled pass through a computer file, all affect the job schedule. At times, the imperatives of other Federal programs and of court orders to produce data for use in litigation have to be accommodated.

At a given time, ORS may have requests from a university researcher to link employment history in a manufacturing industry with cancer mortality, from the Equal Employment Opportunity Commission for statistical tabulations showing racial composition of a workforce for litigation in a discrimination case, or from the General Accounting Office for statistics to assist in evaluating the earnings effect of a Department of Labor manpower project. Numerous projects like these and others compete for the time of a small number of ORS statistical programmers and analysts.

## Statement of General Policy

Despite all the impediments to simple policy formulation mentioned above, the basic policy principle that guides ORS in making decisions about access to microdata is one of making the microdata as widely available as possible, subject only to necessary legal and operational constraints. Considerations of two kinds set the limits to the information available. One limitation is SSA's unwavering commitment to protect confidential information entrusted to it. The other is the practical necessity of minimizing interference with SSA's program functions.

## Benefits of Release

A synergistic aspect to the release of microdata to non-SSA users is present that enhances the benefits to be derived from SSA data bases. Whether or not the particular project serves an immediate SSA purpose, this aspect is present.

**Benefits to SSA programs.** The wealth of information amassed administratively under the SSA programs—such as social security retirement, disability,

supplemental security income, aid to families with dependent children, and Medicare—is enriched through SSA's own surveys and through linkage with other agency files. The uses that can be made of this information by ORS itself are obviously limited by the number of its researchers and the time, computer facilities, and other resources available. These in turn are governed by budget and work plan constraints and, ultimately, by the necessity for accountability to the social security trust funds. With its priorities shaped by current and anticipated needs of SSA policymakers, ORS can only scratch the surface of all the interesting and potentially valuable research its data could support.

Thus SSA gains greatly in knowledge from having other researchers, both inside and outside the Federal Government, mine its rich lodes and share their findings with SSA researchers. Part of the benefit to SSA comes from an increase in the number of researchers working on problems of interest to SSA and the exploration of promising new avenues of knowledge that may in future prove their value though they may not currently justify SSA budget support. In addition, the feedback from outsiders—criticism, verification, or supplementation of ORS findings—helps ORS to develop new techniques, broaden its perspectives, and improve the quality of its own research data bases. The CWHS provides a leading example of the benefits of such feedback from outside users of SSA microdata files.

**Benefits to researchers and the general public.** The flow of benefits from the release of microdata is not one way, of course. Outside researchers benefit from the availability of microdata files that no individual, and probably no nongovernment institution, could hope to compile—both because of the cost and because of the various special incentives for employers and individuals to provide SSA with the information it collects. The public also benefits in ways that are too diffused to measure. The benefits to the public from ORS research itself are enhanced by the cross-fertilization that takes place in the relationship between SSA and non-SSA researchers. Improved quality and quantity of information, greater sophistication in analysis, and better tools for information processing are a few of the rewards from the release of ORS statistics and the resulting interactions between ORS and outside researchers.

## Types of Microdata Releases

### Identifiability

**Records with identifiers.** The Privacy Act, which has a dominant influence on Federal record policy with respect to personal information, defines the existence of records about individuals partly in terms of retrieval by personal identifiers. For SSA, the social security number (SSN) is clearly the principal mechanism for

identification, as well as for retrieval. For purposes of internal linkages, the use of the SSN as an identifier is an operational necessity, both because of the vast numbers of the basic program records and because the number appears on most of these records. In addition, Federal policy has required that, to the extent that other Federal agencies use personal identifiers, they should use the SSN rather than create new identification numbering systems for individuals whose records they keep. Thus the SSN is also the principal vehicle for linkage of files of other agencies. The name is also an obvious identifier, although not a unique one, like the SSN or the name plus SSN.

The Privacy Act places stringent restrictions with specific rules for the release of records with identifiers to users for any purpose, including purposes of research and statistical use. These Privacy Act rules pertain only to records containing information about individuals, as noted, and not to records containing information about other entities.

Records relating to employers are also maintained internally by SSA with identifiers—principally the firm name and the employer identification number (EIN) assigned by IRS to aid the keeping of records on employment and contributions of covered workers. The disclosure status of the EIN has been uncertain since passage of the Tax Reform Act of 1976, because the information on which it is based is collected in connection with earnings information supplied to IRS. Similarly, the standard industrial classification (SIC) code is used as an industry descriptor. When that number is assigned by the Bureau of the Census, its application to an identified employer cannot be divulged by SSA except to other Federal agencies for their research or planning purposes. Thus, statutory limitations have been placed on the release of employer records with identifiers, depending on the source and content of the records.

**Records without identifiers.** In addition to the name and SSN, other personal characteristics—such as sex, date of birth, place of work—combine uniquely to identify or verify the identity of a given individual. Information on characteristics, together with other information about an individual, may permit a user to identify the individual even if obvious identifiers are purged from the file. Potential risk of identification depends both on the content of the file and on the pertinent information outside the file that is available to the user.

The Privacy Act requires that transfer of a statistical file containing information about individuals must be in a form that is not individually identifiable by the recipient. Furthermore, the preparation of a file for public use requires considerable exercise of judgment as to the availability of other information—not only that generally available to the public at large, but also that in the hands of particular data users—that could be matched against the records to identify particular individuals.

Similar possibilities of identification arise in connection with files containing information about other entities. Identification of firms may be possible on the basis of characteristics, even though the names and identification numbers are removed from the files. Combination of location and industry classification, for instance, may make the identity of certain employers obvious, simply on the basis of common knowledge about the area and industrial activity. Depending on the content of the file and the source of the information in it, disclosure avoidance procedures may be necessary.

## Data Subjects

**Individuals.** By the nature of the income- and health-maintenance programs provided under the Social Security Act, the bulk of SSA's computerized records contain information about individuals—that is, about natural persons.<sup>1</sup> Many of these individuals are living persons who are currently paying contributions into the trust funds as workers, are drawing basic disability or retirement benefits on their own account, or are receiving supplemental security income payments. Some are drawing benefits on the accounts of others, as survivors of deceased persons whose earnings and benefit records are also kept in SSA files. In addition, although the Medicare program is now administered by another component of HEW—the Health Care Financing Administration—the beneficiary records, containing information on both living and dead persons, remain in the custody of SSA.

**Other data subjects.** Employers form a large class of “legal persons” identified by the EIN who file reports on covered workers' earnings and contributions (formerly on a quarterly basis, now on an annual basis).

Corporations, of course, are the legal entities employing and filing earnings reports for the majority of workers. Other employers include State and local governments, much of whose information is a matter of public record. Apart from the location and industry classification of individual establishments, little information is ordinarily collected by SSA about the employers.

Another important and diverse group of data subjects includes the providers, carriers, and intermediaries for the Medicare and Medicaid programs. The providers may be sole or group medical practitioners, nursing homes, or hospitals, the carriers and intermediaries are usually insurance companies that contract with HEW to manage Medicare claims and disbursements. These organizations are accountable to HEW, and information about them is used by the Health Care Financing Ad-

ministration for both management and research purposes. The computer records generated from the Medicare program, however, continue to be kept by SSA and remain a source of statistical information.

Finally, it should be mentioned that some of the “legal person” entities—such as physicians who are sole practitioners or householders who employ domestic help—may also have information about themselves as natural persons in SSA files. When both personal and business types of information are contained in the same files—in a survey of physicians containing both demographic and practice information, for example—complicated questions arise with respect to the agency's obligation to protect confidentiality.

## Sources of Data

SSA microdata files of interest to researchers contain information about data subjects obtained in many different ways. “Ultimate,” or original sources of information include

### The data subject—

**Examples:** (1) An individual applying for an SSN, (2) an employer applying for an EIN, (3) an individual applying for disability benefits under title II of the Social Security Act, (4) an individual respondent to a survey conducted by or on behalf of SSA.

### “Representative” of the data subject—

**Examples:** (1) A parent applying for an SSN for his or her child, (2) a family member or other representative applying for social security benefits on behalf of an individual, (3) a “household” or “proxy” respondent to a survey conducted by or on behalf of SSA.

### Employer of the data subject—

**Example:** An employer who has filed a report on covered earnings and social security taxes paid on behalf of his employees.

### Records of determinations about the data subject made by SSA—

**Examples:** (1) Information in SSA files about benefit amounts paid to individuals, (2) statistical classifiers assigned on the basis of information provided by or on behalf of the data subject—type of impairment for a disabled person, for example, or standard industrial classification (SIC) code for an employer.

Often the information passes through other hands before reaching SSA. Some significant examples of such “intermediate” sources include

**The Internal Revenue Service.** Applications for EIN's are filed with IRS. In the past, most information on individual earnings in covered employment has

<sup>1</sup>The term “individual” is defined in SSA's Regulation No. 1 as meaning only a natural person. The Privacy Act definition is more restrictive, being limited to natural persons who are living and who are U.S. citizens or lawfully admitted aliens.

been submitted to IRS by employers and self-employed persons and then transmitted to SSA by IRS<sup>2</sup>

**Outside researchers seeking SSA information for specified persons in their study populations.** Types of information commonly sought by researchers include earnings, benefits, age, sex, race, and mortality. The researcher submits microdata with identifiers to SSA. The desired SSA data are extracted from operating files and merged with the researcher's file. Data from the merged file are normally returned to the researcher either in the form of tabulations or as microdata without identifiers.

The sources of SSA microdata affect decisions on what can be released primarily in two ways. First, where the data come to SSA from another agency, such as IRS or the Bureau of the Census, releases of such data by SSA are subject to whatever statutory and policy restrictions apply to the data. Restrictions of this kind are discussed later under statutes and regulations (pages 9-11).

Second, it is very important that every effort be made to honor any pledges or guarantees given to the original sources concerning the uses to be made of the data. Releases for statistical and research purposes must not materially increase the risk that pledges of confidentiality will be violated, even inadvertently.

## Types of Uses by Recipients

In discussing availability of records, it is important to examine the purposes for which the information is to be used. Some particular purposes may affect decisions as to availability.

**Uses involving recontacts of data subjects.** When a researcher's data needs are not served by microdata files extracted from SSA records, even for a sample population tailored to the researcher's specifications, another dimension is added to the disclosure problem. On some occasions, researchers have asked ORS to draw a sample of names and addresses in order that persons in the sample might be contacted to provide new information in an interview.

An invasion of personal privacy occurs when an individual is contacted by a researcher on the basis of the person's relationship to SSA. As a general principle, recontact by non-SSA researchers is not considered to be within the reasonable expectation of persons whose records are kept by SSA in administering its various programs. Indeed, even recontact for ORS research interviews related to the assessment and improvement of SSA's own programs is kept to a necessary minimum,

<sup>2</sup>Under the annual reporting system now in effect, most information on wages and salaries in covered employment will come directly to SSA on the Form W-2. See page 16 for discussion of annual reporting.

and procedures are carefully exercised to avoid reaching the same individual for more than one research project in a given year. ORS has not considered it proper to identify an SSA sample to non-SSA researchers, regardless of the objectives of their research projects.

**Other research and statistical uses not involving recontact.** A variety of purposes of the outside researcher can be served with data in SSA files, appropriately edited to preserve confidentiality. ORS generates some public-use (or quasi-public-use) microdata files that may be made available to the researcher, perhaps subject to some confidentiality conditions. These files represent samples preselected by ORS, and researchers can work with data for the full sample or can subsample on the basis of file content and documentation, depending on their purposes. File merging, however, is possible only to the extent that the file itself is a longitudinal one. The ORS Survey of the Disabled is an example of a microdata file that can be analyzed statistically but cannot be matched to other files or linked longitudinally. Repetitive survey files, on the other hand—such as the Retirement History Study—can be used for longitudinal matching and analysis.

Alternatively, the researcher may specify characteristics for ORS to draw a sample or may provide ORS with a sample of names and SSN's. ORS can have information extracted from one or more of SSA files containing information about the selected persons. After purging identifiers and editing to avoid individual disclosure, the resulting files may be provided in microdata form.

## Types of Controls on Recipients

The microdata records discussed here are in SSA's possession by virtue of its legal authority to collect and use information and, by the same token, are subject to legal controls on access, use, and disclosure. ORS employees have freer access than outside researchers to SSA data, based on their need to know in performance of their duties. They are subject to direct administrative and management controls on access, use and safeguard requirements, with both civil (reprimand or dismissal) and criminal (fine or imprisonment) sanctions for improper disclosure.

**Contracts.** The user whose relationship to SSA most nearly parallels that of the employee is the researcher who works under a contract with SSA to perform a project to ORS specifications. Depending on the performance requirements of the particular project, the contractor may be given access to certain SSA data needed to carry out his duties to the agency—he may, for example, need the names and addresses of a sample of elderly beneficiaries whom he is employed to interview. In this type of relationship, the contract researcher stands in the shoes of an SSA employee and is subject to the same legal obligations as the employee to protect

and prevent improper disclosure of the data in his hands—both the data from SSA records entrusted to him and those he is employed to collect. Moreover, the contractor performing for ORS is subject to the obligations imposed on SSA by the Office of Management and Budget (OMB) with respect to clearance of interview forms, etc. OMB reviews the forms-clearance package from the point of view of type of data collected, assurances of confidentiality to the respondent, anticipated response rate, and length of interview (with the last item related to respondent burden). Finally, the contractor is ordinarily required to return to SSA the files received or produced under the contract and to purge all identifiers from any copies of records he is authorized to retain.

**Interagency agreements.** When the non-SSA researcher is another Federal agency, the release of SSA data is made in accordance with interagency agreements, which are the practical equivalents, in terms of obligations and conditions, of contracts with researchers outside the Federal Government. When SSA enters into an interagency agreement to have another agency collect interview data for a sample population selected by ORS, SSA provides names, addresses, and possibly other information on a need-to-know basis. Historically, the Bureau of the Census is the agency performing this function for ORS. Under a recent amendment to the statute governing the collection of Census data, the information collected by the Bureau under interagency agreements is subject to the protection of that statute (title 13 of the U.S. Code). Before this change, information collected by the Bureau of the Census for other agencies on a reimbursable basis was regarded as the information of the sponsoring agency and subject to that agency's protections on confidentiality, rather than to the Bureau's protection. The SSA has not contracted with Federal agencies other than the Bureau of the Census to collect data for statistical and research purposes.

When SSA enters into a reimbursable or cooperative agreement with another agency to produce matched microdata files, the resulting records ordinarily become subject to the combined controls of the statutes of both agencies. The agreements entered into by ORS provide for interagency consultation (in the case of the Bureau of the Census through a continuing interagency committee) and advance written authorization by the other participating agency before either can release resulting merged files. When public-use microdata files are produced, if the disclosure criteria set forth in the applicable statutes and rules of the participating agencies differ, the files must meet the strictest of these criteria.

When other Federal agencies obtain microdata from SSA for purposes of their own research, with no ORS involvement or official interest in the project, the conditions set forth in the agreements usually parallel those in reimbursable contracts with nongovernment users, described below. Ordinarily, microdata files containing

information about individuals (natural persons) are available to other Federal agency researchers only in unidentifiable form. (The exception is the Bureau of the Census, which, because of the restrictive nature of its confidentiality statute, can receive identifiable records for its own use.)

With respect to information about entities other than natural persons, other agencies have been able to receive the microdata in a form that identifies the entity (an employer, for example) for their research and planning purposes. The interagency agreements governing the other agencies' use of the information have limited such use to the specified research purpose and have prohibited redisclosure in identifiable form. They also have required the return of the microdata to SSA when the particular project has been completed. In addition, the data released by SSA may contain information obtained from other agencies and the conditions of use in such situations will incorporate any requirements imposed by those agencies. An example is the SIC code when it is obtained from the Bureau of the Census. In that case, its disclosure by SSA is limited to other Federal agencies. Another example relates to data defined as "return information" by the Tax Reform Act of 1976—a term that now includes SSA's earnings information. Release of such information is subject to the provisions of the Tax Reform Act and is governed by IRS regulations. In identifiable form, it can only be released to those agencies expressly entitled to receive it under the Tax Reform Act—the Treasury Department's Office of Tax Analysis, for example, or the Commerce Department's Bureau of Economic Analysis (corporate return information only).

**Grants.** Another class of user is the recipient of a Federal grant. SSA has not funded grants extensively in recent years. The Social and Rehabilitation Service (the HEW component, formerly responsible for welfare matters) traditionally funded considerable grant research. With the research activities in the program of aid to families with dependent children and related areas now melded into ORS, this grantee research will in future be performed as an SSA function.

Unlike the contract by which an agency purchases a defined research product, the grant is, conceptually at least, in the nature of a gift, allowing the research recipient (often a university) to pursue its own line of inquiry. In such situations, the controls attached are for purposes of establishing fiscal accountability, rather than for imposing requirements related to Federal agency mission. As a consequence, the grantee is not regarded as an agency employee, as the contractor is, and has no special access to SSA data. When the grantee receives microdata from SSA, it is provided in nonidentifiable form, and the grantee is subject to the same conditions of use and nondisclosure as any other outside recipient.

**Other conditions of use.** When a researcher contracts

with SSA under a reimbursable agreement by which SSA prepares data to the researcher's specifications, somewhat different conditions apply than for a contractor's performance of an SSA task. These users are not privy to SSA's identifiable records, and, since SSA does not disclose confidential information to them, they are not generally restricted in the use of the information or the publication they may make of their resulting files. Nevertheless, to the extent that incidental disclosure could occur because the researcher has some external knowledge enabling him to recognize individuals whose data are provided to him by SSA, the contract typically prohibits any efforts on the part of the researcher to make such identification. This prohibition clearly precludes the possibility that the researcher can match individual SSA files with other individual files in his possession (No restrictions are placed on statistical matching.) Where the potential for identifying any SSA data subjects in the file exists, the researcher is prohibited from making the identification himself, as well as from making any disclosure of the data to others in a form that would carry the risk of individual identification.

Conditions of use agreements are also made in connection with those microdata files originating in ORS research activities that are not based on an outside researcher's sample population. Even a microdata record purged of obvious identifiers—name, SSN, EIN, etc.—may nevertheless include particular data subjects that could be recognized if the data about them were matched to information about location, characteristics, association, etc., external to the file. When this possibility is significant, ORS places conditions on the use of the microdata file.

If the information is about individuals and is subject to the Privacy Act, certain requirements must be met. Information can be released to a recipient who has provided the agency with advance written assurance that the record will be used solely for research and statistical purposes. To such a recipient, the record can be transferred only in a form that is not individually identifiable. ORS has interpreted this requirement to mean that it covers microdata files for which there is some significant risk that the recipient may have information from other sources that might permit the identification of some individuals whose records are in the file, even when the file is stripped of standard identifiers. For this type of file, the ORS agreement meets the special requirements of the Privacy Act. In addition, it requires assurance of the recipient that no effort will be made to match records or to make any effort to identify individuals in the file on the basis of other knowledge or information.

When the files contain confidential or restricted information about entities other than individuals subject to Privacy Act controls, the agreements must also prohibit

efforts to match records or otherwise identify these subjects.

**No controls.** Finally, there are microdata records for which the risk of individual identification is negligible, and for which no controls are needed to protect confidentiality. This low risk may rest on various considerations—for instance, large population, small sampling fraction, nonspecific geography, etc. For such files, controls on use would not only be redundant, they would usually be inconsistent with the rules of availability under the Freedom of Information Act. Thus, when files have been thoroughly reviewed and found not to have any measurable disclosure risk, they are released on a public-use basis. In the recent past, a number of files have been put into this category, including some of the Census/SSA/IRS match files.

## Factors Affecting Decisions to Release Statutes and Regulations

**Freedom of Information Act (FOIA).** The Freedom of Information Act declares a public-policy mandate of openness, making all Federal Government records available to the public unless there is a specific statutory basis for withholding. For ORS, there are three applicable exemptions to the FOIA on the basis of which requests may be refused.

The first applies to information required by statute to be withheld. Examples of such statutes are the confidentiality provisions of the statute relating to Census data (title 13 of the U.S. Code) and the Internal Revenue Code. FOIA requests for information that SSA may hold would have to be denied if it is subject to one of these statutes. Examples of such data would be Census/SSA match files containing identifiers while they were being processed or the earnings records of individual taxpayers.

Before the FOIA amendment of 1977, SSA considered the Social Security Act to be a statute that protected any information about individuals collected under its authority and thus no disclosures of personal information could be commanded under FOIA. That interpretation has been ruled out by the FOIA amendments, however, and SSA must now apply other FOIA rules in deciding whether particular information can be disclosed.

A second FOIA exemption applicable to data about individuals—one applied by SSA—is the provision that protects personal privacy. Under this provision, SSA tests each request for data about individuals against the amorphous concept of a "clearly unwarranted invasion of personal privacy." Although one can make intuitive judgments as to what constitutes an invasion of personal privacy, it is difficult to frame objective criteria or to measure relative degrees of inappropriate invasion. The

idea of privacy itself is a highly subjective one, and the distinctions between warranted and clearly unwarranted invasions of privacy depend on complex factors and circumstances

In resolving questions of invasion of privacy, the courts apply a balancing test, by which they weigh the public interest in disclosure against the private interest in nondisclosure. The balancing principles tend to produce an unstable equilibrium, however, with varying weight given to particular facts in different situations—such as availability of the information from alternative sources, the commercial or noncommercial use to which the information will be put, and the strength of the individual's interest in keeping the information private.

The third FOIA provision applicable to release of SSA data protects trade secrets and certain confidential business information. This provision was more significant for ORS in the past when Medicare surveys and experiments were conducted within SSA, but it continues to be important to HCFA where such information continues to be collected.

**Privacy Act.** Before passage of the Privacy Act, an agency could exercise a certain amount of discretion in deciding to release information about individuals, where the release was not mandatory under the Freedom of Information Act but appeared to the agency to be in the public interest. An example is the release of information from social security records for valid research of another Federal agency unrelated to SSA's programs. Under the Privacy Act, that range of discretion has been considerably narrowed and conventionalized.

Except for information about individuals that fails the "clearly unwarranted invasion" test and thus must be released under FOIA, the Privacy Act prescribes highly structured criteria governing authorized disclosures and prohibits releases not meeting these criteria. The Privacy Act includes two provisions for release of individual data for research: (1) An authorization to disclose to the Bureau of the Census for its title 13 purposes, without restriction on the type of data, and (2) the restrictive provision previously mentioned, permitting disclosure of data only in unidentifiable form and with written assurance of the recipient that the use will be limited to statistical purposes.

Two other provisions are relevant for the transfer of information on an individually identifiable basis for statistical purposes. One is the provision for interagency use on a need-to-know basis. The other is the provision authorizing agencies to make disclosures for a "routine use." The Act defines routine use as a use "compatible with" the purpose for which the record was collected. SSA has provided for some transfers of this type as a routine use and has interpreted the compatibility requirement narrowly, as meaning closely related to the administration of the Social Security Act. Present policy is to limit routine-use transfers of identifiable data for

research and statistical uses to agencies administering income- or health-maintenance programs. Moreover, the policy is to provide data in aggregate or anonymous form to the greatest possible extent, with release of identifiable data the exceptional case.

**Social Security Act and Regulations 1 and 22.** The Social Security Act contains a provision (section 1106) that declares that all information obtained in administering the Act is confidential and can be disclosed only as provided by agency regulations. Historically, SSA has regarded this as a blanket statutory prohibition on disclosure that served as a basis for denying requests for information under the Freedom of Information Act. In publishing its regulation to provide for authorized disclosures, as section 1106 permitted, the guiding principle was that information about individuals would be released only as needed to administer the Social Security Act, or to comply with other laws, or in special circumstances to benefit the subject of the information. Thus Regulation 1 has specified the classes of authorized recipients, the particular types of data available to each class of user, and the particular applicable purposes and conditions of use.

As SSA programs have been altered and expanded, most notably with the addition of Medicare, the changed composition of the information collected by SSA has necessitated some changes in the disclosure principles. Moreover, under the amendments to the Freedom of Information Act included by Congress in the 1977 Government in the Sunshine Act, section 1106 is no longer a statute that provides exempt status under FOIA. Consequently, Regulation 1 now protects information about natural persons (including deceased persons), and it protects that portion of the information for which the disclosure would either be a clearly unwarranted invasion of personal privacy or would be prohibited by another statute than section 1106. Disclosure rules for information about entities other than natural persons are included in Regulation 22, which applies the provisions of the FOIA to such information.

**Internal Revenue Code.** The old-age, survivors, and disability insurance programs administered by SSA are financed through taxes collected by IRS from employers and self-employed persons. Before 1976, IRS was regarded as a conduit through which information on earnings and contributions was channeled to SSA for program administration purposes. Confidentiality of the information was expressly provided for in section 1106 and Regulation 1. Some ambiguity existed as to which agency actually controlled the information.

The Tax Reform Act of 1976 has defined most of this earnings-related information as "return information" subject to the Tax Reform Act, which is incorporated in the Internal Revenue Code. Thus, for most of the information reported to IRS, the Tax Reform Act has reduced that aspect of ambiguity. By providing for

disclosure of return information to SSA for administering relevant titles of the Social Security Act, however, the law introduced a new ambiguity with respect to the uses of the information that SSA can make and the rules governing disclosure by SSA of the "return information" it obtains. Although the law went into effect over a year ago, a number of basic issues are still unresolved, and SSA has felt obligated to discontinue many releases until interagency agreement can be reached.

The Continuous Work History Sample (CWHS) is the data set for which this situation creates the greatest concern to ORS. The CWHS is made up principally of earnings information ("return information") filed by employers on behalf of employees and by self-employed persons. In addition, certain employer characteristics (SIC and location codes) are contained in the CWHS files. Although the question is not yet settled, these characteristics may ultimately be found to be "return information." If that is the case, substantial impediments to continued release of the CWHS could arise, even though the files are in anonymous format.

Until the question can be resolved, SSA has discontinued release of any files containing IRS source data received by SSA since the effective date of the Tax Reform Act (January 1, 1977). In effect, release of any 1976 earnings data is thus precluded, since, as a practical matter, it is not possible to determine which 1976 files are contaminated with data received after that date.

**Title 13** The Census Act—title 13 of the U.S. Code—places the most stringent restrictions on disclosure of data identifiable to the data subject. Only employees of the Bureau of the Census have access to Census records. For this reason, when joint agency projects have involved the merging of Census and SSA records, the SSA employees involved directly have had to be sworn in as Census "agents" before they could process records containing Census information. When Census sample lists of individuals are used, these files are subject to strict security precautions and are returned to the Bureau as soon as the process of extracting and adding SSA record information is completed. The Bureau performs the merging of data and removes identifiers from matched files that it returns to ORS.

The matched files are edited to meet Census disclosure avoidance rules and to put them in public use form. Until this editing process is finally completed, the files, even without identifiers, are available only to those ORS employees who are Census "agents" for this project.

## Disclosure Risk

When a microdata file is released without any restrictions on its use or further release, a careful assessment of the risks of statistical disclosure before release is a must. Statistical disclosure resulting from the release of

a microdata file occurs when a recipient of the file is able to identify one or more individuals whose records are in the file and, by so doing, to learn more about those individuals.

It must be understood that there is no such thing as a zero-risk release of a microdata file. The task of the source agency is to assess the risk of statistical disclosure for each potentially releasable file in order to decide whether and under what conditions it can be released.

Factors affecting the risk of statistical disclosure in microdata files have been discussed in detail in a working paper recently issued by the Office of Federal Statistical Policy and Standards.<sup>3</sup>

The factors that must be evaluated are

**File content.** How much information is included for each person? Do some variables have extreme values that stand out, such as unusually high incomes for individuals? To what extent are individual records likely to be unique? A key question is the level of geographic information in the file. The risk of disclosure is much higher if the records show the localities in which individuals live or work.

**Sampling fraction.** Does the file contain all members of some defined population group or only a sample? If the latter, what sampling fraction was used? In general, the smaller the sampling fraction, the less the risk of disclosure.

**Availability of matching data from other sources.** A person having access to a microdata file cannot relate records in the file to known individuals unless he already has some matching information for those individuals from another source. Such matching information may come from personal knowledge and observation, or from publicly available records, such as birth records, *Who's Who*, lists of public officials, etc. The likely existence of matching data from such sources must be taken into account.

A special case occurs when SSA data are merged with files provided by outside researchers. Although the identifiers may be removed from the merged file, if it is released to the researcher, he is known to have or can be assumed to have matching information for the same individuals. Such merged data files can therefore only be released under severe restrictions. (See the limitations on release of microdata files containing merged data, page 12.)

## Feasibility and Cost

The function of the Office of Research and Statistics (ORS) is to carry on research and statistical activities

<sup>3</sup>Subcommittee on Disclosure-Avoidance Techniques, Federal Committee on Statistical Methodology, *Statistical Policy Working Paper 2 Report on Statistical Disclosure and Disclosure-Avoidance Techniques* (Department of Commerce, Office of Federal Statistical Policy and Standards), May 1978. See especially Chapter IV.

pertinent to the administration of social security programs. Such a mission does not preclude the release of microdata files to outside researchers. On the contrary, SSA recognizes that it cannot fully exploit the basic data files and that one way to realize their potential in a better manner is to make them accessible to researchers, many of whom are interested in the same problems as SSA.

For this reason, it has become the policy to make microdata files from SSA's major data systems accessible to outside users, subject to necessary confidentiality requirements. Examples include the Continuous Work History Sample and the Retirement History Study. Substantial returns have already been received from the resources devoted to preparing such files for release.

No doubt it will never be possible to satisfy users completely as to the format, documentation, and timeliness of these files. Most improvements in these areas can only be made at the cost of delay or cutbacks in other activities. ORS will continue to try to strike a reasonable balance.

Most user requirements will be met by making available to them multipurpose microdata files of the type just described. Requests requiring the creation of special files are much more difficult to handle and must often compete with in-house projects for available staff, so that compliance with the request is problematical even when the user is willing to pay the full cost. Nevertheless, special microdata files are occasionally created or assistance is given with their creation on a reimbursable basis. In deciding when to do such work, both the amount of staff resources needed and the potential benefits to SSA of the research to be carried out are taken into account.

Potential users should keep in mind one important question that affects the cost and feasibility of creating special files: The question is, "Are the data accessible in a computerized mode?" Employment-history data such as place and type of employment for a defined study population, for example, are only available at a prohibitively high cost because they require reference to microfilm records. On the other hand, data on amounts of covered earnings for designated persons are fairly readily available on computerized systems designed for ready access to such data.

## **Some Specific Policies and Procedures**

### **Public-Use and Restricted-Use Microdata Files**

SSA has adopted what might be called a "two-tier" system for the release of microdata files with identifiers removed.

Designated as public-use files are those microdata files for which, in SSA's judgment, virtually no chance

exists that users will be able to identify specific individuals and obtain additional information about them from the records in the file. No restrictions are made on the uses of such files. Users are urged to take into account sampling errors and other limitations of the data in conducting and reporting on their analyses, to cite the sources of the data in their reports, and to send SSA copies of publications based on analysis of SSA data. SSA does not automatically review drafts of reports and other publications based on SSA microdata files. If staff are available and the subject is of special interest to SSA, however, such review is sometimes possible.

Typically, the public-use files are based on national samples with small sampling fractions (usually less than 1 in 1,000) and the files contain no geographic codes or at most regional and/or size of place designators. Each file proposed for designation as a public-use file is reviewed carefully to ensure that all data that might make it possible to identify specific individuals are removed. Most of the microdata files released to outside users by SSA are public-use files.

Those microdata files considered as carrying a disclosure risk greater than is acceptable for a public-use file are released only under conditions of restricted use. These conditions are set forth in user agreements, signed by the user and by a representative of ORS. Conditions of use normally include limitation of use to purposes specified in advance by the user, agreement not to try to identify specific individuals or link data from other sources, a prohibition on rerelease without SSA approval, and agreement not to release listings of individual records or tabulations in a form that might reveal information about individuals.

The only restricted-use files currently being released to nongovernment users are those derived from the CWHS system. A copy of the current user agreement is shown as Exhibit 1. These agreements have been in use since mid-1976. When the use of these agreements began, other steps were taken to reduce disclosure risks:

Restriction of the content of released CWHS files to the specific data items needed by each user. Some items, such as day and month of birth, were eliminated entirely.

Use of improved techniques for the encryption of social security and employer identification numbers (the encrypted numbers are included in some cases to allow users to update their files periodically).

Introduction of noise into the earnings data.

### **Limitations on Release of Microdata Files With Merged Data**

As mentioned earlier, the release of files containing SSA data merged with data for the same individuals provided by the researcher can only be done under severe constraints. It is assumed that most researchers

# Exhibit 1

## Conditions for Release of Continuous Work History Sample Microdata Files by the Social Security Administration

A I, \_\_\_\_\_, representing \_\_\_\_\_, agree to observe the following conditions of use of the CWHS microdata files released to me or derived from such files

1 The files will be used only for the following purposes

Any other uses will be subject to prior approval by SSA

2 None of these files, or any files extracted or derived from these files will be released to any other organization or individual without SSA approval

3 No attempt will be made to identify any specific individuals, employers or establishments for whom records are included in these files

4 No attempt will be made to link information from any other source to records for specific individuals, employers or establishments for whom records are included in these files. This provision, however, does not preclude statistical matches of individual records, i.e., those matches carried out by linking records for persons with similar characteristics, without attempting to ascertain that both records do, in fact, pertain to the same person

5 \_\_\_\_\_, will be designated as custodian of these files, and will be responsible for observance of all conditions of use, and for establishment and maintenance of security arrangements to prevent unauthorized use. If the custodianship is transferred within the organization, SSA will be notified promptly

6 No listings of information from individual records, with or without identifiers, will be published or otherwise released by the holder of these files. No statistical tabulations or research results will be released which reveal information about identifiable individuals, employers or establishments

7 Subject to conditions 2 and 6, statistical and research results derived from these files may be published. (Freedom to publish such results is not denied by the part of Condition 4 (Form SSA-1034) which reads "The requesting agency is free to use such information itself but not to publish it, in whole or part, or to furnish it to anyone else, without written approval.")

8 Authorized representatives of the Social Security Administration will, upon request, be granted access to premises where the files are kept for the purpose of inspecting physical security arrangements

B (This provision to be used only for organizations receiving CWHS files with SSN's encoded by the new system)

The \_\_\_\_\_ does not retain, nor does it have access to any CWHS-based files that contain or are linkable to identifiers (SSN and/or EIN) that are not encoded or that were encoded by the system used prior to the one currently in use

C I have received and read Section 1106 of the Social Security Act and SSA Regulation 1. I am aware that any person violating any provision of Section 1106 may be punished by a fine not exceeding \$1,000 or by imprisonment not exceeding one year, or both

\_\_\_\_\_  
(representative of organization)

\_\_\_\_\_  
(custodian of files, if different)

DATE  
\_\_\_\_\_

are not interested in obtaining SSA data about specified individuals for nonstatistical purposes. Nevertheless, the fact cannot be ignored that it would be possible for the researcher to associate SSA data with identified individuals if the merged file contained any individuals with unique combinations of the characteristics provided by the researcher.

Therefore, the general rule is that any subset of the merged file that can be defined in terms of combinations of characteristics in the file submitted by the researcher must contain at least five individuals. It is recognized that this requirement may severely limit the content of the file. If such a file does not meet the user's needs, alternative approaches may be explored in which SSA prepares the desired tabulations and analyses and releases them to the user only after reviewing for disclosure potential.

### Information on Mortality

Mortality information provides an exception to the general rule that SSA information for identifiable persons may not be released. Records for deceased persons are not covered by the Privacy Act of 1974, and release of information concerning fact and circumstances of death has long been permitted under SSA Regulation 1.

Information in SSA records on mortality for the general population is not as complete or as detailed as that available in death certificates. It does have the advantage of being available from a single source. Especially since access to IRS information on mortality has been closed off as a result of the new confidentiality provisions in the Tax Reform Act, SSA is often the logical place for researchers to start followup studies of persons whose current or recent addresses are unknown.

The National Center for Health Statistics is now developing plans for a National Death Index, which is expected to cover all registered deaths from 1979 on. Using this index, researchers will be able to determine which members of their study populations have died and which State registrars to contact for information needed from the death certificates such as cause of death.

Since the National Death Index will cover only deaths from 1979 on, both sources of information will be important to researchers for some time to come. Chart 1 compares these two sources of mortality information—SSA records and the National Death Index. Exhibit 2 is a statement prepared by ORS's Division of OASDI Statistics on release of SSA death information to outside researchers. This division handles requests involving fewer than 1,000 persons, larger requests are handled by SSA's Bureau of Data Processing.

Mortality information in SSA operating records can fairly readily be added to the CWHS system, thus permitting analysis of differential mortality for covered workers by industry. The 1-percent sample data do not

permit in-depth analysis of detailed industry groups but can be quite useful in providing a first cut at the question to guide the application of further studies to particular industry groups. Some preliminary results have been reported by Goldsmith and Hirschberg.<sup>4</sup> Much more use could be made of this potential data source, especially if resources were made available to add cause-of-death information to the file.

### Use of SSA Records As Sampling Frames for Surveys

This topic is somewhat tangential to the main theme of the article because it necessarily involves the use of lists with identifiers. Nevertheless, it merits discussion here because it is a very live issue today. The Social Security Administration does not maintain current residence addresses in connection with its earnings records, so use of these records as a survey sampling frame, even for surveys conducted by or on behalf of SSA, is not practical.<sup>5</sup>

Current addresses are maintained, of course, for beneficiaries of federally administered SSA programs—old-age, survivors, and disability insurance, Medicare, and supplemental security income. The beneficiary records for these programs have been used on several occasions to select samples for surveys and research studies conducted by or on behalf of SSA. In designing these samples, care is taken to ensure that no beneficiary will be contacted in more than one survey or research study within a short time interval. Potential survey respondents are notified that participation is voluntary and that response or nonresponse to the survey will not affect their benefits in any way.

Many survey researchers outside SSA would like to have access to these sampling frames. They offer good coverage of certain population groups, especially persons aged 65 and over, and the lists have current addresses. Many surveys that these researchers undertake have obvious potential benefits to society, and in some cases the knowledge obtained might be helpful in administering social security programs.

Against these obvious benefits, however, must be placed the clear threat to the privacy of the beneficiaries if these lists were made available for surveys not sponsored entirely or in part by SSA. No matter what criteria might be used to limit such access, the number of qualified applicants would almost certainly be large, and it would be difficult to justify giving access to some and

<sup>4</sup>J. R. Goldsmith and D. A. Hirschberg, "Mortality and Industrial Employment," *Journal of Occupational Medicine*, March 1976, pages 161-164.

<sup>5</sup>Individuals could be reached through their most recent employers, but, because of the obvious privacy implications, SSA has not chosen to do so. The possibility of obtaining current residence addresses from IRS for this purpose has not been specifically tested.

**Chart 1.**—Comparison of the SSA beneficiary and earnings records and the proposed NCHS national death index as sources of death information for use in followup studies

Characteristic	Proposed National Death Index	Master Beneficiary Record (includes data from Summary Earnings Record)
1 Years covered	Deaths beginning with 1979	Deaths starting with 1956 See also item 2
2 Completeness	Expected to be very high—99+%	Relatively high for persons 65+ and males 35+, improving over time Study now under way will provide reliable estimates of coverage for persons 35+
3 Conditions for access to system	In general only for uses relevant to public health activities	Unrestricted
4 Data needed to enter system	Must have at least one of the following data element combinations (1) Name and date of birth (2) Name and social security number (3) Name and mother's maiden name (4) Name and maiden name	Searches can be made on social security number only Name is also desirable as a basis for verification of matches Searches for small files of cases without SSN's but with other identifiers may be made at considerably higher cost
5 Data provided for matches		
a Fact of death	a Yes	a Yes
b Registration area	b Yes	b No, but gives best available county of residence information
c Death certificate number	c Yes	c No
d Date of death	d Yes	d Yes, or date of claim if date of death not available
e Information on possible matches	e Yes	e No
6 Cost per case submitted	Not available	For fewer than 1 000 persons, \$1 00 per person Costs per person are smaller for larger numbers
7 Turn around time	Not available	1 to 5 months depending on volume

not to others. The demands on ORS staff to review requests and to assist users with the development of appropriate selection procedures would be heavy.

For these reasons, SSA policy has consistently been to deny all such requests. An earlier BULLETIN article describes this policy clearly.

Under the confidentiality rules, the position of the Social Security Administration is unequivocal on requests for use of social security records in selecting and identifying a sample of individuals to be contacted for research purposes. Uniformly, such requests must be denied to preserve the confidentiality of the Social Security Administration records. . . . Such proposals obviously infringe on the privacy of the individual by making known to a third party one or more privileged pieces of information.<sup>6</sup>

The authors explain that some researchers ask SSA to select samples according to their specifications and forward letters to the individuals asking for their voluntary participation in a survey. Cooperation with such proposals is also denied.

## Possible Future Developments

### Recommendations of Privacy Protection Study Commission

The Privacy Act (Public Law 92-579) created a Privacy Protection Study Commission (PPSC). The Com-

mission had a 2-year mandate to study and report on the strengths and weaknesses of the Act and on the successes and failures of Federal agency implementation.

An important recommendation of the PPSC in its final report<sup>7</sup> was that statutory barriers should be created to prevent the availability or use of individually identifiable research and statistical data for making decisions affecting individual data subjects. The thrust of this and corollary recommendations in the chapter on research and statistical studies was to create a safe environment in which both survey and administrative records can be compiled and used for research and statistical purposes, without risk that they will return in individually identifiable form into the stream of information for decision-making about the data subjects. An additional goal of these recommendations was to increase agency accountability to data subjects for the use of those records.

One recommendation warrants separate attention. The PPSC recommended that the National Academy of Sciences, in consultation with the interested research community, be asked to develop and promote techniques for disclosure avoidance in connection with statistical and research uses of information about individuals. The purpose would be to render the records less vulnerable to individual identification.

If enacted into law, the Commission's recommendations on research and statistical records would make some changes in internal agency practices and procedures and would formalize procedures affecting the

<sup>6</sup>Heyman C. Cooper and Joseph Steinberg, "Social Security Statistical Data, Social Science Research, and Confidentiality," *Social Security Bulletin*, October 1967.

<sup>7</sup>U.S. Privacy Protection Study Commission, *Personal Privacy in an Information Society, Report of the Privacy Protection Study Commission* (Chapter 15, Research and Statistical Studies), 1977.

content and availability of microdata files for outside research users. In general, ORS considers its practices compatible with the Commission's recommendations.

The PPSC proposal for statutory restrictions on compelled disclosure of individual information to courts by researchers is of particular significance. SSA has long had a policy of resisting requests for individually identifiable research information for purposes of litigation. Instead, it has provided assistance, when requested, in the form of statistical information prepared on a reimbursable basis. SSA does not, however, have express statutory protection of research information from litigation or other decisionmaking use. Statutory protection would reinforce the assurances of confidentiality given to data subjects, with anticipated improvement in both rate and quality of response to surveys.

## Annual Reporting

The Social Security Amendments of 1975 (P L 94-202) made changes in the Social Security Act and the Internal Revenue Code, giving the Departments of the Treasury and HEW the authority to exchange information needed to permit conversion of social security tax reporting by employers from a quarterly to an annual basis.

Under "annual reporting," as the new system is called, the employer who used to file five reports a year for each employee (four quarterly reports on Form 941-A and one annual report on Form W-2) now is able to file a single consolidated annual wage report for each employee showing both his total earnings for the year and the additional information needed to determine quarters of coverage and benefit amounts under the social security program.

The annual reporting procedure applies to earnings received in 1978 and thereafter. The W-2 will be the vehicle for the consolidated annual report, and the forms submitted by employers will be processed initially by the Social Security Administration. The processing operation has been designed to provide both IRS and SSA with the information necessary to carry out their respective responsibilities as efficiently as possible.

This change did not provide for what is sometimes called "true annual reporting," which would require no information on covered employment by quarter. The provisions of P L 94-202 were modified, however, by the Social Security Amendments of 1977 (P L 95-216). Under that law, employers will be required to report on the W-2, for social security purposes, only the total covered earnings of each employee during each calendar year. Quarters of coverage will be deemed on the basis of total covered earnings for the year, initially at the rate of one quarter for every \$250 of earnings.

Annual reporting has several important consequences for the CWHS system.

**Potentially improved coverage of the labor force.** W-2's will contain data on both covered and non-covered earnings, and SSA will be processing all W-2's, including those for persons with no covered earnings. The word "potentially" is used here because all earnings information on W-2's is considered "return information" and therefore subject to the new confidentiality provisions in the Tax Reform Act of 1976. The extent to which these provisions will allow internal use of such information for statistical purposes by SSA, and the release to others of microdata files (without identifiers) containing such information, has not yet been fully determined.

**More nearly complete earnings information.** Currently, total earnings are estimated for each person whose earnings exceed the maximum for covered earnings during the year, now the W-2's will show both covered and total earnings. Possible effects of the Tax Reform Act on availability and use of the information are also a consideration here.

**Files available and timing.** For years after 1977, of course, a first-quarter CWHS file will no longer be possible. An attempt will be made to speed up the availability of the annual employer-employee file.

**Effect on the Establishment Reporting Program (ERP).** The ERP permits assignment of industry (SIC) and geographic codes on an establishment or reporting-unit basis for employers with more than one establishment. Voluntary arrangements are made with these multiunit employers to report information on covered earnings of their employees separately for each reporting unit. The ERP arrangements will continue under annual reporting. The accuracy of SIC and place of work codes in CWHS has deteriorated, however, in recent years because of lack of resources for ERP quality assurance and maintenance activities.<sup>8</sup> The switchover to annual reporting may aggravate the problems in this area.

**Effect on migration analysis.** Most migration studies based on CWHS have used first-quarter employer-employee files for different years. Since these files will not be available for years after 1977, the migration studies will, of necessity, be based on year-to-year changes in the annual files. Although these files are more complete (quarterly files have excluded farm workers and the self-employed), the kinds of migration rates studied and their interpretations will be different from those used in earlier studies.

Clearly, CWHS users are faced with new opportunities as well as with actual and potential problems in getting access to and using files from the system. Subject to the operational and statutory constraints described, efforts will be made to maintain and improve it.

## Controlled Remote Access to Microdata Files

Some researchers feel that their needs for detailed microdata cannot be met by the SSA files available to

<sup>8</sup>David W. Cartwright, *Major Limitations of CWHS Files and Prospects for Improvement*. (Presented at the Workshop on Policy Analysis with Social Security Research Files, Williamsburg, Va., March 1978.)

them, either on a public-use or restricted-use basis. The alternative of asking SSA to prepare tabulations or conduct statistical analyses for them is often unsatisfactory for a number of reasons—cost, timing, and especially the need of more sophisticated researchers to work with the files on an interactive basis.

It has been suggested by some that the solution to this problem is to allow researchers controlled access to the pertinent files. Under this approach, the researcher would have access to the microdata file via a remote terminal and would be able to manipulate the data in the file according to his specifications. The output would, however, be carefully controlled and monitored. The researcher would not be able to call for individual records nor to produce tables with fewer than a specified number of persons in a cell. Other restrictions, requiring either automated or manual review of proposed outputs to the researcher, would be necessary.

This mode of operation is theoretically possible, but the establishment of suitable software for monitoring output would call for a substantial investment of time by skilled systems designers and programmers for each microdata file thus made accessible. Such an investment would be justified only if no suitable alternatives exist and the expected payoff is high.

SSA has already had limited experience with a system in which an outside researcher submitted programs to be run for a nonreleasable microdata file. The program outputs were manually reviewed for disclosure avoidance by an ORS employee before release to the researcher. Monitoring of this arrangement took an unacceptably large amount of professional staff time.

## Exhibit 2

### Release of SSA Death Information to Outside Researchers

#### Background

Social Security Administration records have proven to be a valuable source of information to outside researchers for use in followup analyses in the conduct of various epidemiological studies, specifically, to obtain information on the mortality experience of worker cohorts. Besides the importance of just knowing that an individual is deceased, by having the place and date of death the researcher has the ability to extract valuable information from death certificates (e.g. cause of death) through State agencies.

We can supply death information for identified individuals for which SSA has received some form of death report. This information is obtained from the Master Beneficiary Record (MBR) and/or the Summary Earnings Record (SER) and consists of actual or presumed

date of death (month and year) and presumed place of death (city and State or county and State).

If requested, statistical data can be provided for the cases not identified as deceased, following the normal guidelines for insuring against disclosure of information relating to an identifiable individual.

#### Procedure

The requester must submit a file of finder records on magnetic tape or punch cards in a specified format containing social security number and if available, the first six letters of the surname. The finder record may also contain information from the requester that may be pertinent to the study.

The incoming file is matched to the MBR and SER and a linked record is generated. Depending on the volume, the combined information is maintained on paper for manual investigation or magnetic tape for electronic processing. The death information is returned to the requester in the form of annotations on listings or on the original punch card.

#### Criteria for Determining Death Information

##### Date of death

If MBR shows a date of death or a termination code for death in the benefit payment history field, then the date shown is given.

If MBR does not show date of death but shows survivors benefits being paid,

(1) and SER shows date of death same as initial entitlement to survivors benefits, this date is given.

(2) and SER shows date of death different than date of initial entitlement to child's benefits, then both dates are given.

(3) and SER does not show date of death, the date of initial entitlement to child's benefits is given. If there are no child's benefits then an interval is given based on the last year for which earnings are shown in the SER and the date of widows' entitlement.

If no MBR is present or no survivors benefits are being paid

(1) and SER shows date of death, this is given.

(2) and SER shows fact of death only, the last year with earnings is given as the open-ended interval (e.g. 1970 or later).

##### Place of death

If address is shown in MBR, city and State are given.

If no address is shown the State and county as determined by the State and county residence code are given.

If no address or State/county code is shown, the State of the servicing district office shown is given.

If no address, State and county code, or district office code is shown the State where the social security number was issued (as determined by the area position of the SSN) is given.