

Social Security Statistical Data, Social Science Research, and Confidentiality

by **JOSEPH STEINBERG** and **HEYMAN C. COOPER***

AS A RESULT of the significant involvement of the American people with the Social Security Administration, either directly or through third parties, large masses of information are recorded for the purpose of administering the program. Collateral uses of social security data for statistical research and analysis have begun to have increasing value. The potential of these data has increased in the last decade, as coverage under the OASDHI program has expanded to include all earners except Federal employees and some other groups. Information collected is available to the Social Security Administration through applications for social security numbers, the reports of earnings, and other types of administrative actions such as the filing of claims for cash benefits.

The new service benefits in the health area have made available significant amounts of demographic data concerning aged persons not otherwise beneficiaries under the cash benefits program. For all persons aged 65 and over, data on utilization of hospital services, as well as information on providers of the services, are part of the ongoing information flow for administration. For persons enrolled for supplementary medical benefits, data on significant uses of doctors' services have become available.

Earnings are reported to the Social Security Administration for about 80 million persons each year under OASDHI. About 1 person in 8 either receives a cash benefit or is entitled to receive health service benefits. The social security number is used to assure that earnings are recorded for the holder of the number. When cash or service benefits under the program are claimed, the social security number becomes an essential identifier. In day-to-day activities, it has also become commonplace to use the number for a variety of other purposes. For most people this number is used as an identifier for data to be made available by banks and corporations to the Internal Revenue Service, as well as on their tax forms.

At an early point in the history of the Social

Security Administration, policies on the confidentiality of identifiable data were established. Because this area is of such significance in relation to the availability and use of the data, a substantial portion of the article is devoted to the topic.

GENERAL OVERVIEW OF SOCIAL SECURITY STATISTICAL DATA

Social security statistical data are collected and recorded for individuals as well as for business establishments and providers of health services. Though the data are integrated in relation to specific administrative missions of the Social Security Administration, it may be instructive to discuss the statistics for each of the following general categories: (a) demographic and economic data for individuals, (b) economic data for business establishments, (c) health service utilization data for individuals, and (d) data on characteristics of hospitals and other providers of health services.

Demographic and Economic Data for Individuals

Basic information about individuals comes from the application for a social security number.¹ Date of birth, sex, race, and place of birth are the primary information of statistical interest on this form. The data on the application initiate the establishment of a machine-readable record for the individual.

¹ Primarily, Form SS-5—Application for Social Security Number. For a short span of time, IRS Form 3227—Application for Account Number, was used in cooperation with the Internal Revenue Service in providing numbers for taxpayer identification; that form made no provision for recording race. U.S. Civil Service Commission Form BRI 49-262 was used for a special group for whom social security numbers have been assigned and is the source of basic demographic information. Other persons in groups aged 65 and over for whom information is available supplied such information to the Railroad Retirement Board for their administrative use or to the Social Security Administration in connection with application for a health insurance claim number.

* Office of Research and Statistics.

As the individual works under covered employment,² his employer uses the social security number as the numerical identifier for reporting quarterly wages. For agricultural workers, data are reported annually. These earnings reports³ are the primary source of data for accumulating information on taxable wages in the record of the specified employee. In a similar fashion, information on earnings for self-employed persons becomes available through information filed by the individual with his Federal tax return as part of his Schedule C or Schedule F.⁴

Earnings data, summarized for each calendar year, are part of the machine-readable record for each account number. Quarterly data, together with codes for county of employment and industry, are also maintained for a 1-percent statistical probability sample, the Continuous Work History Sample (CWHHS). It is selected on the basis of the social security number with constant selection pattern since the start. This sample design provides a basis for within-the-year as well as over-the-years linkage of the more detailed information that is maintained for statistical purposes.

Economic Data for Business Establishments

Program administration provides a variety of economic data for business establishments. Some of these data are basic classifiers; others are data from current program operations.

Every employer with a worker whose earnings are covered under the social security program is required to obtain an employer identification number (EI number). The application for employer identification number,⁵ provides basic classification information on the location and nature of the business. A copy of this application is used by the Social Security Administration as part of its administrative and statistical systems.

² *Social Security Bulletin, 1965 Annual Statistical Supplement*, page 111.

³ For nonagricultural work, IRS Form 941—Employer's Quarterly Federal Tax Return; for agricultural employment, IRS Form 943—Employer's Annual Tax Return for Agricultural Employees; for domestic employment, IRS Form 942—Employer's Quarterly Tax Return for Household Employees.

⁴ IRS Form 1040, Schedule SE, U.S. Report of Self-employment Income.

⁵ IRS Form SS-4—Employer's Application for Identification Number.

Additional information on multi-unit companies is developed in connection with the statistical activity.

The earnings reports from the employer's quarterly Federal tax return provide current data for the employer entity. Information on the number of employees during the reporting period and on aggregate taxable payrolls for each unit can then be combined with the information on nature of business and geographic location. Agricultural, State and local government, and household employer entities also report earnings. A basis for classification by specific industry and county of employment for multi-unit companies is provided in the establishment reporting program (ERP)⁶ of the Social Security Administration that permits identification of employees by unit. A self-employed person reports on his tax return the portion of his self-employment earnings that will bring his total earnings recorded in his social security records up to the taxable limit. Some information permitting classification on nature of business becomes available for these business earnings as part of the total set of economic data for establishments available at the Social Security Administration.

Health Service Utilization Data for Individuals

Administration of the health insurance program for the aged provides data on the utilization of health services for persons aged 65 and over.⁷ Records for the hospital insurance part of the program show the number of days of care received by each aged person in a hospital or extended-care facility and the number of home health visits received during specified periods that relate to the program concepts of "deductible" amounts and "spell of illness."⁸

Each episode of hospitalization provides data on length of stay, discharge status, charge and

⁶ Social Security Administration, *The Establishment Reporting Plan*, August 1965.

⁷ Howard West, "Health Insurance for the Aged: The Statistical Program," *Social Security Bulletin*, January 1967; Aaron Krute, Newton Dikoff, and Jack Scharff, *Medicare, the Statistical Elements*, 1967.

⁸ For a full description of the provisions of the health insurance program, see Wilbur Cohen and Robert M. Ball, "Social Security Amendments of 1965: Summary and Legislative History," *Social Security Bulletin*, September 1965.

payment data, primary diagnosis, surgical procedures, and preoperative and postoperative lengths of stay. Diagnostic and surgical information is coded by the Social Security Administration for a 20-percent sample of the hospital insurance beneficiaries. The diagnostic code is for the primary discharge diagnosis reported by the hospital, and the surgical procedure code relates to the primary discharge diagnosis. For outpatients, coding is for a 40-percent sample; all home health bills and extended-care facilities bills are coded for diagnostic information.

Data available on utilization of physician and related medical services cover time and place of each service, the exact procedure or service provided, the condition treated, the physician's or supplier's charge for the specific service, and the charge allowed for the same procedure or service. For nonsurgical medical services, there are some descriptive data available on the type of services provided by the physician at each visit. For surgical cases, data on the surgical procedure, the diagnosis, and the overall charge for the entire procedure are available. The data become available through the administrative requirements for submittal of bills (at least \$50 of services during the year is required). These data on services under the medical insurance program are centrally recorded for a 5-percent sample of all enrolled persons (a subsample of the 20-percent sample for the hospital insurance diagnostic coding).

Characteristics of Hospitals and Other Providers of Health Services

The hospital insurance program covers reimbursement for services of providers such as hospitals, home health agencies, extended-care facilities, and independent laboratories. Each provider, in applying for participation, has initially supplied (and will update periodically) a wide variety of detailed information, including data on the number of beds, type of control, and the major type of services provided. Also available are data on staff characteristics, such as the number of physicians, registered nurses, qualified special therapists, licensed practical nurses, home health aides, and other skilled medical care personnel. Information on the

volume of the provider's services is also supplied—for example, the annual total of adult admissions and discharges, the number of patient days, the number of persons served, and the current reimbursement rate.

SOCIAL SECURITY DATA AND SOCIAL SCIENCE RESEARCH

The data available from the operations of both the cash and service benefits programs of the Social Security Administration probably represent one of the richest single aggregations of current demographic and economic information in terms of the number of people involved. The records are an important current source of information about business establishments.⁹ *County Business Patterns*, first published in 1946, provides statistical data on first-quarter employment and taxable payrolls by county and industry of reporting unit. Some possible uses, general and specific, of social security data are discussed below to illustrate their variety.

The data have primary importance as a research resource, not only because they provide information on covered workers, but also because the information is longitudinally linkable. Some analytic studies have taken advantage of this longitudinal characteristic of the recorded information.¹⁰ The cumulation of information on episodes of hospitalization and their characteristics and on the utilization of medical services provides an important resource for studying health conditions of the aged.

The interrelationship of variables available from other sources linked with social security administrative record information (under appropriate conditions to insure confidentiality) has become the basis for additional analysis (beyond that based solely on data directly available to the Social Security Administration). Followup of economic characteristics of worker cohorts¹¹ or of health characteristics of aged cohorts are additional important research potentials.

⁹ Bureau of the Census, *County Business Patterns* (various years).

¹⁰ See Lowell E. Gallaway, *Interindustry Labor Mobility in the United States, 1957 to 1960*, Office of Research and Statistics, Research Report No. 18, 1967.

¹¹ Social Security Administration records are being used to follow up a cohort of workers laid off in the shutdown of operations of a large industrial corporation.

The general area of economic analysis of the labor force involves analytic consideration of the dynamics of labor supply. Social security data are an important resource for many types of analysis in this field. Studies of labor mobility, whether between industries or between different geographic locations within a given industry, are a part of such analysis.¹² Differences in earnings patterns among groups are an important analytic interest. Varying patterns of lifetime earnings of various cohorts, defined in terms of available demographic variables such as age, mobility patterns, and industry may help in understanding the relationship between the processes and effects of labor mobility and economic policy questions. They also have a bearing on the concept of earnings replacement through social insurance. The nature of the processes of labor mobility may be considered a significant element for the study of the efficiency level of labor-market mechanisms in allocating human resources in the economy.

Cohorts defined by membership in a specific program are important study populations. A parallel control cohort may be studied, in addition to those of the particular program, if the experimental design elements have been identified in advance. Social security data may play a role in examining the effects of Government training and education programs to ameliorate existing disadvantages for a specified population group. The effect of the Job Corps or of other programs of the "war on poverty" may be studied in a variety of ways. Longitudinal records of earnings and other demographic patterns of the various subgroups exposed to a program may be studied prospectively through the use of records.

Cohorts defined as a result of major economic incidents, because of attachment to specified types of industrial activity, or because of specific characteristics such as those of Selective Service rejectees can be followed prospectively. Studies of the effects of programs for attracting industries to disadvantaged areas on the retention of population in those areas or analysis of the effects of programs for retraining disadvantaged people within economically disadvantaged areas are further illustrations of some uses of social security data. The impact of technological changes

in agriculture on acceleration of migration from rural to urban areas may also be studied. The relationship of patterns and time sequences of migration into and out of depressed areas with time sequences of programs for dealing with such areas and with depressed industries may provide useful data for analysis in this area.

Data collected for a given administrative purpose often exclude many variables of significant research interest that may have been omitted because of policy or public interest considerations. Data on age or race are often in this category. In many programs, either of these (or other) variables may be deemed to have a possible detrimental effect if the data are immediately available in conjunction with other data being reviewed for a program decision. Race is not recorded on any claims forms for social security benefits, for example. Linkage with other data from Social Security Administration records may facilitate statistical analysis of subgroups otherwise not identifiable. The differential effects of various types of changes in income-tax policies on various age cohorts of taxpayer units can be studied through linkage of samples to social security data.

To the extent that analyses of differences between a worker's industry of major job in a given period and other industries of employment are of interest, both within a year and longitudinally over time, such data are available for overall cross-sectional analysis as well as for specific cohorts. Since such data are the product of the ongoing administration of the Social Security Act, certain important variables are not available, however, except when special studies are made. Thus, occupational mobility and its interaction with other aspects of labor mobility or education are not ordinarily subjects that can be studied. From social security data, some limited information on the number of shifts between employers within a given industry in a year or longer can be considered as an added variable for its potential in helping to define groups with differing labor-force mobility. The possible effects of private pension plans and the extent of vesting, if any, in relation to labor mobility and retirement patterns may be studied if data on private pension plans are available to the researcher.

When greater detail is available from another

¹² Lowell E. Gallaway, *op. cit.*

source, more intensive analysis becomes possible. If, for example, data on farm operators by detailed type of farm operation are available, the phenomenon of changes in patterns of the use of the agricultural work force may be studied more continuously.

The wealth of data becoming available from the health insurance programs represents a new resource for social science research—limited of course to data for persons aged 65 and over. The longitudinal patterns of hospital utilization and medical care utilization can be studied. Cross-sectional analyses as well as cohorts defined in terms of available variables within the program (geographic location, age, sex, race) may be studied. Additional program variables such as diagnosis or procedures and patterns of earnings or labor mobility may be considered. As noted, cohorts defined on the basis of information from other programs can be studied.

Information may be explored on groups defined in terms of their specific previous medical care practices where data on individuals exist (Blue Cross compared with non-Blue Cross members, for example). Comparison of patterns of health care use for specified cohorts for whom data are available before and since the advent of Medicare also becomes possible. Answers to policy questions concerning the nature of the upgrading or shifts in use of health services can be developed. Rates of change among groups over a period of time can now be studied without the problems of retrospective recollection or continuous survey processes, subject only to the limitations of the program requirements and the data collection processes of the records systems.

Many suggestions for research can be developed. A few are sufficient to illustrate some of the possibilities. The data available solely from the Medicare program may provide some basis for answers to social researchers' interests in areas such as the following: (1) the patterns of utilization of medical services by people who migrate during the year; (2) the patterns of differences in use of services between areas of residence and of service; (3) the measurement of expected numbers of preoperative and postoperative visits for specific surgical procedures; and (4) changing relationships between use of extended-care facilities and hospital utilization.

When data are also available from records outside the health insurance program, the impact of the program on specific groups such as veterans, welfare recipients, the poor (identified by general geographic location as well as levels of earnings) can be measured over time. Data about individuals with or without other forms of health insurance can be studied in regard to utilization and financing of medical services.

A number of useful studies of methodological interest are possible with the wide variety of statistical data from other data sources. Reported data on age, race, or earnings in the records system data can be compared with those in other studies for the same individuals and measures of differences made available. Differentials between nonrespondents and respondents in a survey can be studied with the available variables from the social security data.

CONFIDENTIALITY OF SOCIAL SECURITY DATA AND DISCLOSURE

Steadfast determination to keep a pledge of confidentiality for records on individuals and a conscientious desire to make information available for planning and research in the public interest may appear, at first glance, to be strange bedfellows. In the administration of the Social Security Act, the pledge of confidentiality was made at the beginning and has been stoutly maintained ever since. This fact is clearly visible at every stage in the 30-year history of the program. There is nonetheless adequate evidence that the program's principal executives understood and expected that in time the information being collected and stored for proper administration would begin to serve other useful purposes beyond the specific program purpose.

During most of its existence the social security program's administrators not only focused on the essentials of program planning and execution for effective and efficient operations but also gave thought to future needs for growth and improvement. Steps were taken to serve the immediate public interest beyond the narrow purpose of program administration. A philosophy of intention and interest to contribute to such aims was clearly formulated.

This section of the article examines the nature

of the Social Security Administration's confidentiality policy and the principles that govern the disclosures that may or may not be made from information in its possession. Such an examination seems appropriate at this time since it is now possible to see what the agency's collection of records is and what it is likely to be. It is possible to see and appreciate both present and future problems of accessibility. It is possible to explore the Administration's continued linking of a firm confidentiality policy and a positive interest in contributing through the agency's facilities to public interest social science research.

Statute and regulations protect the confidentiality of records and information in the possession of the Social Security Administration. The reasons for the statute and the regulations relate directly to "the public interest and efficient administration."¹³ These reasons have held consistently throughout the three decades of existence of the Federal social security program. Interpretation of these broad reasons has been dynamic and their meaning has broadened with program growth, but there has been no deviation from the fundamental confidentiality position.

The confidentiality policy is concerned with prohibition against the disclosure of information of a personal nature about individuals and their activities in the possession of the Social Security Administration. Both the statute¹⁴ and the regulations¹⁵ protect the rights of privacy of those who give information, either voluntarily or pursuant to law, with the understanding that such information is to be used only for social security and related purposes. There is no question or issue of Government secrecy about the operations of the Federal social security program. From the beginning, the Social Security Administration has published statistical summaries of all its program activities. There is no question or issue, either, of holding back information that affects the rights of individuals under the program. The Public Information Act of 1966, as amended in 1967, explicitly continues the guarantee of protection of such information.

¹³ Regulation No. 1 of the Social Security Board, filed with the Division of the Federal Register, June 16, 1937.

¹⁴ Section 1106 of the Social Security Act, as amended.

¹⁵ Regulation No. 1 (as amended), *Disclosure of Official Records and Information* (part 401, chapter III, title 20, Code of Federal Regulations), Social Security Administration, 1964.

The significance of a strong position on confidentiality of records was clearly perceived by the Social Security Board¹⁶ well before operations began under the original Social Security Act on January 1, 1937. Since the Board was also responsible for administering a program of grants to the States for public assistance, it was deeply concerned with the possibility that access to public assistance records might be sought for political purposes by some State and local elected or appointed officeholders. In addition, before the start of the old-age insurance program, there were many rumors and news stories predicting the use of the records by employers to obtain information about the age of employees or for blacklisting purposes about a worker's previous work history. There were fears of infringement of civil liberties and the creation of a "police state."

It had been clear from the start that operation of the social insurance program would call for the collection and maintenance of considerable personal and private information from both employees and employers. To obtain the needed cooperation of both the working force and the employers, the Board publicized the types of information that would be needed for operation of the program. To allay public fears, a public pledge was made that the records would be used only for social security purposes and otherwise treated as restricted. A press release, issued on November 23, 1936—the day before the availability of applications for social security numbers—informed the public that:

The Social Security Board announces that . . . information required of every worker on this form will be regarded as confidential within Government sources. Only the worker himself, or his immediate family, or Government employees having official responsibility in connection with the social security files will have access to this information.¹⁷

¹⁶ Independent Federal agency, placed under the Federal Security Agency (now the Department of Health, Education, and Welfare) under Reorganization Plan No. 1 approved by the Congress to become effective July 1, 1939. The Board was abolished on July 16, 1946, by an Executive Order that created in its place the Office of Commissioner of Social Security.

¹⁷ In the face of continuing sharp reaction, a second press release was issued on December 10, 1936, reiterating the Board's pledge of confidentiality: "The Board will, at all times, regard the information received from an employee as confidential. The files will be open only to those who have a legitimate interest in the administration of the Social Security Act."

The basic need for program operations in 1937 was to start and subsequently to maintain earnings records. Social security numbers were assigned to individuals.¹⁸ Employers were required to obtain identification numbers. Earnings identified by the social security number were required to be reported by employers on an established periodic basis.

For operation of the cash benefits aspects of the program, information needed from individuals in applying for the benefits has been of a highly personal nature, including such matters as identity, immediate family relationships, documents giving evidence of age, details of marital history, and questions relating to children or to the support of dependent aged parents—matters most people consider very private.

The initial and continued maintenance of confidentiality of data by the social security system¹⁹ related clearly to the need to assure effective and efficient administration. The validity of the earnings records and of all the other information necessary to reach a determination to pay a cash benefit had to be assured at all times. Any relaxation of the restrictions against disclosure, to any material degree, carries with it the serious problem of the effect on the proper care of the money in the trust funds. The question that is all important is whether relaxation of the restrictions might impel some persons to furnish incomplete or even false information to avoid disclosures they would consider contrary to their best interests.

The social security program has been guided in its conceptualization and its day-to-day administration by the basic concern with individual human beings. The records, over the years, of legislative hearings and debate sit side-by-side with the records of administration to bear out the concerns that serve to maximize each person's rights in matters of individual equity, the rights of individuals to know precisely how the program affects them, as well as the assurance of accuracy and timeliness in paying benefits to every entitled person.

¹⁸ More than 171 million social security numbers have been issued since November 24, 1936.

¹⁹ This article is concerned, principally, with the social security program as it relates to old-age, survivors, disability, and health insurance. References to related programs such as public assistance are made only when pertinent to the immediate point in the article.

Implicit in the acquisition of personal information are the obligations that go beyond the visible horizons of administrative need. Basic obligations are those that involve the freedoms that are valued by people. A major freedom that is germane to this discussion is concern with the rights of an individual to privacy. The basic philosophy of the Social Security Administration is that, as some encroachments on privacy necessarily occur in requiring people to supply the information needed for valid administration, there must be full protection of the data obtained against availability for unintended purposes. Disclosure of confidential information has been permitted only for social security administration and related program purposes, with few exceptions.²⁰ Disclosures may be authorized by an individual with respect to his own record, however.

Guardianship of the confidentiality of its vast set of records is for the Social Security Administration a large responsibility. One of the first essentials involves setting up internal safeguards to prevent improper disclosures. Procedures have been established to make certain that the confidentiality requirements set forth in statute,²¹ agency regulations,²² and policy are carried out. There are thousands of points of contact with the public in the program's far-flung national operations and numerous situations in which individual judgments must be made. Continuing vigilance is necessary to make certain that practices and procedures are within the established policies. The efficacy of the rules must often depend substantially, however, upon the degree to which the employees are well-informed on the agency's policies and penalties and on their individual obligations.

Another essential of major importance is the promulgation of regulations governing the limits within which the disclosure policies may operate. Public knowledge of the content of the regulations bolsters the trust individuals feel in their right to privacy when they supply personal information to the Government for a particular purpose.

²⁰ Exceptions are made in cases involving national security and in providing information concerning aliens as requested by the Attorney General, pursuant to section 209 (c) of the Immigration and Nationality Act of 1952, 8 U.S.C. 1360 (c).

²¹ Section 1106 of the Social Security Act, as amended.

²² Regulation No. 1, *op. cit.*

Public knowledge serves as a deterrent for those who would seek improper access to this information,²³ and such knowledge is, of course, very helpful to the prospective legitimate researcher.

The responsibility that goes with custody of a large body of information about individuals is more than a matter of devising rules identifying disclosure exceptions and of maintaining a system of security with respect to employee practices. Much of the agency's concern relates to the need to carry out obligations that do not necessarily always complement each other. There is, clearly, an underlying obligation to protect the confidentiality of records. A second obligation is the need to accomplish the agency's basic mission, clearly spelled out in statute. A third obligation is to put to use or permit to be put to use, by proper means, the information obtained either compulsorily or by voluntary cooperation on the part of the public. The third obligation involves increasing pressure as data accumulate and as the public becomes aware of the nature of these data.

The Social Security Administration has attempted, over the years, to fulfill each of its obligations to each part of the public concerned, making it clear, at the same time, that no element of confidentiality implies any secrecy concerning the nature of the program operations. In regard to the third obligation, the Administration has appreciated the fact that information on earnings and other data can be useful in economic and social research. The use of social security data by other Federal agencies, such as the Bureau of the Census,²⁴ has served to reduce reporting burdens upon employers.

Under a provision in the Social Security Administration Regulation No. 1, "statistical data or other similar information not relating to any particular person which may be compiled from records . . . may be disclosed when efficient administration permits." Information, not neces-

sarily in the form of statistical tabulations, contained in or compiled from reports submitted by employers (again, not relating to identified or identifiable individuals) similarly may be disclosed to any other Federal agency "for use in its statistical and planning work only."²⁵ The word "planning" as used here was intended, and it is so interpreted, to refer to normal planning purposes such as budgetary planning, workload estimates, and the like, but not for use in regulatory, enforcement, or compliance investigations.

Congressional confidence in the manner in which the social security records are safeguarded is evident from the authority, inserted in the statute, to charge for information or services furnished, subject to the agency's limiting regulations and determination of administrative feasibility.²⁶

The Social Security Administration's disclosure regulations and policies represent a balance between, on the one hand, allowable situations in which it is deemed unnecessary to obtain the individual's permission before disclosing data and, on the other hand, valid understandable prohibitions designed to protect the confidentiality of the records. The permissible situations relate principally to the administration of the social security program or of the other programs encompassed in the original Social Security Act—including public assistance and unemployment compensation.

Subsequently, these instances were expanded to include the administration of other Federal programs related to social security such as pension, retirement, and other benefit programs. Use of the information by other Federal agencies is for specified purposes and subject to the penalties applicable to violations of the statute.²⁷

In the broadest sense, the disclosure policy attempts to recognize obligations of the Administration. The difficult task is to meet them in a consistent way that can rationalize contribution to the public interest and maintain the integrity of individual records without impairment to the agency's basic mission. As a matter of public policy, the agency makes special contri-

²³ It is interesting to note that administrators of the social security program have consistently preferred dependence upon regulations instead of specific statutory disclosure prohibitions. Regulations permit greater flexibility in recognizing situations that can then be appropriately administered. In the 1950 legislative deliberations of amendments to the Social Security Act, the Senate version of H.R. 6000 would have inserted a very stringent nondisclosure provision, barring disclosure even by regulation.

²⁴ Bureau of the Census, *County Business Patterns*.

²⁵ Regulation No. 1, section 401.3(k), as amended.

²⁶ Section 1106(b) of the Social Security Act, as amended. This provision was initially enacted as part of the 1950 amendments.

²⁷ Section 1106 (a) of the Social Security Act, as amended.

butions to the program needs of Federal agencies. In an official statement of its objectives,²⁸ the obligations of the Administration are set forth as follows:

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.

Disclosure policy can have a significant impact on a wide range of research interests within other government agencies or on outside research groups. The policy acknowledges the Administration's obligations, which involve the need to publish data useful both to program administrators and to social scientists. Then, there is the need to encourage the linkage of data with other bodies of statistical information. There is the need to make data available for outside statistical and research uses. Further, in its publicly stated objectives, the Social Security Administration has expressed its continuing interest in stimulating and helping to support research in the general area of income maintenance and the broad area of social security to be carried out at universities and other nonprofit research centers.²⁹ A major accompanying concern is the development of sound bases upon which to make data available without detriment to any individual.

The applicable statute and regulations provide a basis for specific rules concerning the kinds of information that in most cases may or may not be released. The rules apply to information for administrative uses, as well as to the availability of statistical data not relating to any particular person and compiled from the agency's records. A special regulation³⁰ permits the release of statistical data contained in or compiled from

employer reports to other Federal agencies for use in statistical and planning work only. The usual considerations of administrative feasibility and reimbursement of costs apply.

The statute and regulations and the associated disclosure policy provide a framework for decisions concerning availability of data for research purposes on an ad hoc basis. Five general conditions have pertinence and must be met: (1) the project must be in the public interest; (2) no private individual or organization may derive exclusive benefit from the research; (3) the information or services requested must be available only through the Social Security Administration facilities; (4) it must be demonstrated that no prejudice or impairment to the interest of any persons involved will ensue, directly or indirectly; and (5) it must be determined that it is feasible to provide the information or services requested.

When these conditions appear to be met, the next step is to obtain a formal review and an ad hoc approval or denial.³¹ This procedure flows from constraints within the regulation, since general disclosure is not specifically authorized by the regulation itself. Actual practice has been carried forward on a case-by-case basis. Formal procedures involve legal, policy, and technical review before the final review of the recommended action. Decisions (numbered and dated) carry special authorization to disclose or to deny in each case. Although individual case decisions play an important part in evaluating subsequent user requests, special authorization still must be obtained for each new case. There have been no instances in which decisions on release of data relating to individuals have been determined as precedent to proceed without the ad hoc special authorization.

Several general types of cases may illustrate usefully the nature of requests and results of the Administration's decision process. Many persons and organizations have asked to use the facilities of the Social Security Administration for tracing or finding individuals of interest in a research project. Such researchers are interested because of the belief that the social security records have current addresses. Many fail to realize, however, that the records do not now

²⁸ Social Security Administration, *The Objectives of the Social Security Administration*, chapter III, "Responsibility to the Government as a Whole," August 1965.

²⁹ *The Objectives of the Social Security Administration*, chapter II, "Responsibility for the Program" (Objective No. 13).

³⁰ Regulation No. 1, section 401.3 (k).

³¹ Regulation No. 1, section 401.1.

contain and never have contained a home address for other than a small fraction of the social security number holders. Unless the individual sought is receiving a cash benefit payment at the time, the Social Security Administration does not systematically have a reasonably current residence address. Sometimes, a request is made that the agency attempt to reach the individual by writing to him in care of his last reported employer. Neither the statute nor the regulations require or prohibit cooperation in such situations.

The question of cooperation in forwarding letters to an individual is a complex one. The confidentiality of personal information, including possible location, is very clear. There are related aspects, however, that may not be as clearly perceived yet are fundamentally part of the principle of an individual's rights to privacy. There are the implications of his current or recent employers' knowledge and participation in this process.

There are burdens produced by a letter that is not related to social security purposes. There is concern with the possible feeling of invasion of privacy arising from the individual's seeing it as a threat—even if the threat is felt rather than real. There is the risk of making the employer aware of some characteristic of the worker that may be detrimental to him. There is a burden upon his employer to seek him out. Under no circumstance can the employer's identity or location be divulged. Since with these necessary limitations response is likely to be minimal, the Social Security Administration discourages this practice.

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. Usually, the researcher hopes to obtain a random selection from the records that will represent a definable group or for some geographically defined area such as a city or several communities, expecting that he will be able to write to the sample group to urge participation in his study. Such proposals obviously infringe on the privacy of the individual by making known to a third party one or more privileged pieces of information.

Sometimes, understanding that these records are restricted, the requester would have the Social Security Administration select a sample according to his specifications and forward letters to the individuals to ask their voluntary participation. Again, to maintain the confidential nature of the records, cooperation with such proposal is denied. The propriety of direct participation by the agency in the identification of respondents—an action that well may be considered an invasion of privacy—has also been a matter for concern.

METHODS FOR ACCESS TO SOCIAL SECURITY DATA

Substantial summary and analytic data from the large volume of economic, demographic, and health data available to the Social Security Administration are available, in the first instance, in regular publication media. The basic data are published regularly in the *Social Security Bulletin* and its *Annual Statistical Supplement*, in special releases of the Office of Research and Statistics (such as *Health Insurance Statistics*), and in special reports.

It is recognized that the statistical records and files constitute a source of data of great potential value for other research uses. The policy of the Social Security Administration is to make its data resources available, while insuring the confidentiality of information on individuals and reporting units, and at the same time to keep them subject to administrative feasibility, while carrying forward its basic administrative mission.

In making its data resources available, the Social Security Administration staff has devoted considerable effort to the development of appropriate mechanisms for carrying out the general research policy. The primary efforts have been directed toward developing procedures for the content of the program that has been longest in operation—detailed as well as summary earnings records. Planning in this area has forecast some of the ways in which data in the newer areas of health statistics also may become available.

A number of rules have been established as a framework for the operation of these mechanisms. These basic rules are:

1. No data will be made available to a general researcher that will disclose information identifiable to an individual or a reporting unit.

2. Data for a comparatively small cross-section probability sample of social security records (1 percent or less) may be made available, if all identifiers to individuals are removed. (An illustration of the application of this rule is discussed more fully below—in discussing how the Continuous Work History Sample tapes are being made available.) There must also be essential assurance that no combination of data made available will permit simple identification of any person or unit by use of the data. For many useful research efforts, each record might be assigned a “case number” (not related to any known identifier) in order to permit association of information for identical units over time. Any data made available under such a policy needs to be available to all researchers. The responsibility of the Social Security Administration is solely to ensure the confidentiality of the records through appropriate review before issuance of the research data and to make available to all users full information on the source and accuracy of the data. Each researcher would have access to detailed information concerning the general nature of inputs, meanings of codes, and technology used in the reduction of the data of the social security record form.

3. To the extent administratively feasible the Administration will carry out such special tabulations or cross-tabulations with data supplied by others as will be consistent with the other policies discussed. Standard tabulation programs will be used, as far as possible. Cross-tabulations with other data need to meet the test that research be in the public interest. Thus, an adjunct of this test is that general availability of cross-classified information may be assumed to be possible at an appropriate and reasonable time, to serve the public interest and all specific interests involved when any social security data are used as an element in cross-tabulations.

4. All public references in oral or printed form to unpublished social security data made available under any of the policies described should be accompanied by the following general disclaimer: *Certain data used in this publication (or paper, report, talk, etc., whichever is appropriate) were derived from statistics furnished by the Social Security Administration. The author(s) did not at any time have access to any information relating to specific individuals or reporting units. The Social Security Administration assumes no responsibility for whatever analysis, interpretation, inference, conclusion or recommendation is developed from the statistics by others.*

The Social Security Administration, mindful of the important research potentials that might need to be served, has begun to develop a few alternative mechanisms for data access, described below. In developing these alternatives, the Administration has sought the counsel and expert judgment of many people in a variety of professional fields. Some experimental efforts are also being pursued. In general, the additional policies used as guidelines in developing the operational

mechanisms, as well as for the experimental activities, have required that:

1. Research efforts involving use of or access to individual record data must be in the public interest and be within the setting of a research mission as opposed to any possible regulatory or investigative mission of an organization.

2. Access to any individual data must be confined to specified, limited research use with appropriate control mechanisms for insuring limited access. Audit by social security staff to insure that control mechanisms are operative is to be part of the control mechanism. Access for the defined purpose is further limited only to the time needed for the specified purpose.

Two general procedures for data access have been developed. Both relate to the records of lifetime covered earnings. One procedure makes data available from the 1-percent Continuous Work History Sample. These data, made available on tape, show no identification linkable to an individual but permit linkage with other social security data for the identical individual through the use of common case numbers. These data, developed to serve some basic research needs of the Social Security Administration, are formatted in several different tape files.

Of primary interest, to serve many of the research needs discussed briefly earlier in the article is the file identified as the “Employee-Employer 1-percent Sample.” This file covers one calendar year at a time. For each sample individual, the file combines information relating to a given employer for each employee. Each record includes coded information for the following variables: month and year of birth, sex, race, taxable wages for each quarter and for the year, estimated total nonfarm wages for year, and State and county of employment and industry. A document to provide good understanding of the sources and factors affecting accuracy of the data is being prepared. Limited access by researchers to this resource is being made through a tape file that contains no identifiers but includes a case number to permit linkage over time.

One file, which serves additional types of research needs, records data for a 1-percent sample of the self-employed. Another file shows some longitudinally linked annual earnings data for the past 15 years or so. Other statistical files also are used for 0.1-percent samples.

A second approach that has been developed

permits a variety of data to become available, provided suitable tests of "nondisclosure" are satisfied. This approach deals with the interest of many researchers in relating data for a specified cohort for which they have information to the earnings data and other demographic characteristics available from the data on lifetime covered earnings. Most research interests previously discussed with Social Security Administration staff appear to be met by a plan of cross-tabulation of social security earnings data with other data, which can be structured into a standard table format.

This special tabulation program, developed to serve the interests known to the Social Security Administration, reduces substantially the technical problems for the agency's cooperation in compiling such data. Details of the method, conditions, and considerations for researchers are described in a general guide.³² In general, the researcher interested in cross-tabulation possibilities through this special tabulation program would provide the Social Security Administration with data in a standard punchcard format. He would specify the extent of each field of his own data, the scope of records to be included in the standard tabulation, the stub classification scheme and the reference year of social security earnings data. Certain other factors affecting the degree of accuracy or potentials for violating the confidentiality of the standard tabulation would need to be established. These elements would include a decision on whether the researcher wishes the Social Security Administration to investigate for a correct social security number if his data is in error or incomplete with respect to that number. Other elements are described in detail in the guide.

Of importance to note in this type of access to social security data are some administrative considerations, if basic conditions concerning confidentiality are assured. These elements include reimbursement of costs by the researcher and consideration of the feasibility of performing the necessary operations without impairment to the administration of the social security program. From time to time, program and operational requirements related to major amendments

to the Social Security Act are of such a magnitude that the Social Security Administration may be unable to meet a researcher's request on a timely basis. This is a factor that should be kept in mind in relation to the development of study plans using this means for access to data from the earnings records.

Both available methods have been designed to provide reasonable use of the research potential of social security data, within the clearly recognizable needs for safeguarding confidentiality and administrative feasibility. Some experiments in other means for data access have been undertaken from time to time as Social Security Administration policy has evolved over the years. In each case, a determination is then made as to the further possibilities for continuing any one of them beyond the immediate experimental effort. In the past, a few special statistical tabulations have been provided, some of which have linked social security data with data collected by other researchers, always within the constraints previously mentioned. The cost and the time required to meet special purpose interests have been a significant element, however, in affecting the availability of data to researchers. These handicaps have indicated the need to initiate and continue to explore methods for making more of the data available, in more timely fashion, and at lower costs.

CONCLUSION

In view of the great volume of information contained in the Administration's records, which generally is not available from any other source, the potential interest in the use of such records is virtually unlimited. A substantial amount of worthwhile research is under way in medical and social science fields that may benefit from access to social security data. Substantially more will be undertaken. It is the policy of the Social Security Administration to do all that is possible to encourage, stimulate, and cooperate in the furtherance of research. It is not within the province of the Social Security Administration, as a public agency, to appraise the merits of any researcher's plan. It is necessary, however, to insure that access to social security data does

³² Social Security Administration, Office of Research and Statistics, *Some Statistical Research Resources Available at the Social Security Administration, 1967.*

not contribute to an intrusion into the rights of individuals to privacy. The Social Security Administration does not, however, maintain that research in fields related to the social security program, or any that could be aided by the information or experience of the Social Security Administration, should be exclusively that of the public agency.

Many significant policy and research interests can be served through access to social security statistics. It is essential that the research community be fully aware of the potential of this information, as well as the constraints. The significant involvement of social scientists in helping in the development of methodology beyond that already available would be very valuable. Improved processes of data collection,

storage, and retrieval are inevitable. As resources become more generally available, the momentum of demand will increase. Methods for meeting the demand depend not only on publication and processing by the Social Security Administration but on exploration and development of additional alternatives that may fit within the general policy considerations and that will create even greater and more timely access for serving the public interest. Some study of the characteristics of data access through micro-aggregation³³ may lead to adoption of this approach. Other possibilities certainly need to be and will be explored.

³³ E. L. Feige and H. W. Watts, *Protection of Privacy Through Micro-Aggregation*, University of Wisconsin, Social Systems Research Institute, June 1967.