

USES OF ADMINISTRATIVE DATA AT THE SOCIAL SECURITY ADMINISTRATION

by Jennifer McNabb, David Timmons, Jae Song, and Carolyn Puckett*

The Social Security Administration (SSA) collects a wealth of data in its role as administrator of two large national entitlement programs. Linking SSA's administrative data with survey data yields a broader set of demographic and socioeconomic information and also improves the quality of the survey data. The agency uses these data to produce analyses and research on policy initiatives for its programs and on the earnings of the working and beneficiary populations. SSA studies how these programs and potential changes to them affect individuals, the economy, and program solvency, and develops models to project demographic and economic characteristics of the current working population into the future. The agency also produces public-use microdata files that are available to outside researchers, as well as a variety of research and statistical publications to inform policy-makers and the public.

Introduction

The Social Security Administration (SSA) is an independent agency of the federal government. Its mission is to deliver Social Security services that meet the changing needs of the public. SSA is responsible for one of the largest federal entitlement programs: Old-Age, Survivors, and Disability Insurance (OASDI), commonly referred to as Social Security. As the name suggests, OASDI provides monthly benefits to qualified retired and disabled workers and their dependents, and also to survivors of insured workers.

Eligibility and benefit amounts are determined by the worker's contributions to Social Security. There is no means test to qualify for benefits, although for those under the full retirement age there is a limit on income earned from working while receiving benefits.

Today, more than 163 million people work and pay Social Security contributions, and more than 50 million people receive monthly Social Security benefits (Board of Trustees 2008). During 2006 approximately 162 million employees and self-employed workers, along with employers, contributed \$626 billion to the OASDI trust funds, from which benefits are paid (SSA 2008). Workers and employers each contribute 6.2 percent of covered earnings (up to \$106,800 in 2009)

and self-employed workers contribute 12.4 percent of covered earnings. In December 2006, total benefits paid by the OASDI program exceeded \$46 billion each month (nearly \$546 billion annually). According to the 2008 Social Security Trustees Report, these cash benefits made up 4.3 percent of the nation's gross domestic product.

Social Security benefits are essential to the economic well-being of millions of individuals. Benefits are paid to about 90 percent of the U.S. population aged 65 or older. Social Security is the major source of income (providing 50 percent or more of total income) for 66 percent of the beneficiaries. It contributes 90 percent or more of income for one-third of the

Selected Abbreviations

| | |
|-----|--|
| CMS | Centers for Medicare and Medicaid Services |
| CPS | Current Population Survey |
| DMF | Death Master File |
| FEM | Financial Eligibility Model |
| HRS | Health and Retirement Study |
| IRS | Internal Revenue Service |

* Jennifer McNabb was with the Office of Retirement and Disability Policy (ORDP), Social Security Administration (SSA), when this presentation was written. David Timmons, Jae Song, and Carolyn Puckett are with ORDP, SSA. This article is adapted from remarks presented by Linda Drazga Maxfield at the International Seminar on the Use of Administrative Data for Economic Statistics and the Register-based Population and Housing Census, held May 19–20, 2008, in Daejeon, Republic of Korea.

Selected Abbreviations—*Continued*

| | |
|-------|--|
| MBR | Master Beneficiary Record |
| MINT | Modeling Income in the Near Term |
| NCHS | National Center for Health Statistics |
| OASDI | Old-Age, Survivors, and Disability Insurance |
| ORDP | Office of Retirement and Disability Policy |
| ORES | Office of Research, Evaluation, and Statistics |
| SIPP | Survey of Income and Program Participation |
| SSA | Social Security Administration |
| SSI | Supplemental Security Income |
| SSN | Social Security number |
| SSR | Supplemental Security Record |

beneficiaries. Social Security reaches almost every family, and at some point will touch the lives of nearly all Americans (Fisher 2008b).

SSA also administers Supplemental Security Income (SSI), a needs-based program that provides financial support for aged, blind, and disabled adults and children with limited income and resources.¹ In 2006, 7.2 million people received monthly SSI benefits totaling \$38 billion, with an average benefit of \$455 (SSA 2007).

SSA is headed by a Commissioner and has a staff of approximately 60,000 employees. The Agency's central office is located in Baltimore, Maryland, but the vast majority of the staff serves in a decentralized field organization with 10 regional offices, 6 processing centers, approximately 1,300 field offices, and over 140 hearing offices. The agency issues Social Security numbers (SSNs) to nearly all legal U.S. residents, maintains detailed earnings records for covered workers, keeps recipient records current and accurate, and determines eligibility for Medicare health insurance. SSA also provides support to the Railroad Retirement program, the Food Stamp program, and the Medicaid health insurance program for those with limited income. Because of these broad responsibilities, SSA collects and maintains a substantial amount of program-related data on current and potential beneficiaries residing in the U.S. and abroad.

With program administration as its primary function, SSA as a whole is not a statistical agency. The most prominent government agencies with a primary

statistical function include the Census Bureau, the Bureau of Labor Statistics, the National Agricultural Statistical Service, and the National Center for Health Statistics (NCHS). However, the Office of Management and Budget, which oversees policies and procedures for all U.S. statistical programs, includes one component of SSA under its statistical purview: the Office of Research, Evaluation, and Statistics (ORES) within the Office of Retirement and Disability Policy (ORDP). ORES uses the Agency's administrative data to produce a wide range of research and statistical publications, as well as other products that inform the public about the beneficiary population and the operation of Social Security programs. ORDP develops and maintains a series of detailed statistical databases for research, evaluation, and analysis.

This article discusses the advantages and limitations of using administrative data for research, examines how linking administrative data to survey results can be used to evaluate and improve survey design, and discusses research studies and SSA statistical products and services that are based on administrative data.²

SSA Administrative Data

Data Systems

SSA maintains numerous administrative data systems. The four most commonly used are:

Numident file. The Numident file is a record of applications for Social Security cards. Unique, life-long SSNs are assigned to individuals based on these applications. A full record of all changes to the information (such as change of name) is also maintained. To obtain a card, the applicant must provide documented identifying information to SSA. Through the "enumeration at birth" program, children can be issued a Social Security card when they are born. Examples of data elements on a Numident record include name, date and place of birth, parents' names, and date of death.

Master Earnings File. The Master Earnings File contains the individual lifetime records of wages and self-employment earnings. The file's primary sources of information are the W-2 form (for wages) and electronic files of form 1040, schedule SE (for self-employment income) from the Internal Revenue Service (IRS) in the Department of the Treasury. The most frequently used data elements are the individual's SSN, annual total wages (1978 to present), annual

self-employment earnings, annual earnings used for OASDI contributions (1951 to present), and report year.

Master Beneficiary Record (MBR). The MBR is used to administer the OASDI program and contains beneficiary and payment history data. An MBR record is created whenever an individual applies for benefits and SSA adjudicates the application as an award, a denial, an abatement, or a withdrawal. Information maintained in the MBR includes the primary worker's SSN, the beneficiary's own SSN, benefit application date, benefit entitlement date, and type and amount of benefit.

Supplemental Security Record (SSR). The SSR contains information on individuals applying for SSI payments. SSA uses the income, resources, disabling condition, and living arrangement information from the application and other sources in determining eligibility for and administering the needs-based SSI program. SSR data elements include SSN, date of claim, citizenship status, income, resources, eligibility code, payment code, and payment amount.

Advantages and Limitations of Administrative Data

Because administrative data are used for determining eligibility and benefit amounts for social insurance programs, they are subject to stringent quality control procedures. However, because these data are typically limited to information required for program administration, they are restricted in scope and do not include broader variables of interest to the research community. For example, focusing on individual eligibility and participation, they often lack economic and demographic variables (such as total family income or marital status) that are critical to programmatic evaluations. In addition, administrative records alone cannot be used to address all analysis questions since they typically contain no information about nonparticipants who could be affected by a proposed program policy change. SSA researchers need this information to project policy change impacts on program costs, as well as potential distributional effects on different demographic or economic groupings. Survey data can provide this information.

Benefits of Supplementing Administrative Data with Survey Data

The federal government conducts numerous large surveys that produce key information to support decisionmakers and to document economic and social trends. Surveys conducted by statistical agencies ask

a broad range of questions on a wide variety of topics. Survey results often include extensive demographic information and are typically representative of the civilian noninstitutionalized population.

Analogous to administrative data limitations, survey data are limited in that they do not typically contain enough program-level detail to compute or model the features of program eligibility. In addition, survey data are subject to various sampling and non-sampling errors—the latter often resulting in incomplete or inaccurate responses due to the respondent's inability to recall accurately or report demographic or economic information.

SSA takes advantage of the enhanced analytic potential afforded by linking survey and administrative data. In fact, SSA has been linking its administrative data with survey results for over 40 years. Some of these linkages are with surveys that SSA commissioned to study specialized populations, such as the Social Security New Beneficiary Survey, the National Survey of Supplemental Security Income Children and Families, and the National Beneficiary Survey. However, SSA's administrative data are more often linked with ongoing surveys conducted by other federal agencies. Linking survey and administrative data allows SSA to produce otherwise unavailable demographic estimates of the current beneficiary population and to develop models to project demographic and economic characteristics of the current working population into the future.³

Survey Information Used in Data Linkage

SSA's biggest data-linkage partner is the Census Bureau. Two of the Census Bureau's major survey efforts are the Current Population Survey and the Survey of Income and Program Participation. These surveys vary in sample size, amount of detailed information collected, and periods covered.

Current Population Survey (CPS). CPS is a monthly survey of 50,000 households. It collects data on employment, unemployment, earnings, income, and hours of work. It also has data elements covering a variety of demographic characteristics, including age, sex, race, marital status, and education. Monthly CPS supplements provide additional demographic and social data. The Annual Social and Economic Supplement, fielded in March of each year, focuses on income and poverty in the United States. CPS is the source of official unemployment rate and poverty rate statistics.⁴

Survey of Income and Program Participation

(SIPP). SIPP provides considerably more detailed information on income and program participation than the CPS. It also features recurring modules focusing on special topics. SIPP data elements include income from all money and nonmoney sources (including public assistance programs and employer-provided benefits), financial assets, and family characteristics (including size, composition, income, and education of household members). The survey uses a “panel” design. Each panel consists of a set of respondents interviewed every 4 months for 32 to 48 months (Census Bureau 2009).

Linking administrative and survey data combines the completeness and accuracy of SSA administrative records with the range and scope of survey results, maximizing the strengths and minimizing the limitations associated with each. With the information on program participation and benefits from SSA administrative records, analysts are able to correct misreported values in survey files, yielding more accurate underlying data and improving statistical estimates.

Linking survey data and SSA administrative records also significantly expands research opportunities beyond those provided by either source alone. Survey information provides detailed background information on demographic, income, self-reported health status, and other characteristics of Social Security program participants and nonparticipants. Administrative records supplement this information by providing individuals’ lifetime work and earnings histories, as well as accurate Social Security program participation histories. Researchers can use matched data to study work and earnings dynamics of survey respondents before, during, and after their interviews. Furthermore, the linked survey data allow for the construction of detailed profiles of individual and family characteristics at the time of program participation, as well as detailed information related to program dynamics.⁵

There are substantial methodological benefits to linking administrative and survey data. One major advantage involves the accuracy of the respondent’s recollection of past program participation and income receipt. When comparable data are collected in both an administrative file and a survey, statisticians and policy analysts are able to evaluate the extent of underreporting or overreporting attributable to the respondent. For example, survey respondents often confuse one of SSA’s programs (OASDI or SSI)

with the other when reporting benefits or payments received. Further, survey responses matched to administrative data can document the benefit amounts that the recipient reports in the survey and compare them with the actual dollar amounts distributed. Another methodological benefit of matching administrative and survey data involves asset income. In 2002, only 55 percent of CPS respondents aged 65 or older reported any asset income, down from 69 percent of comparable respondents in 1990. The Census Bureau and SSA are linking CPS data with Social Security benefit and earnings data, and also with IRS income files, to investigate whether asset income among the elderly is actually declining or is merely unreported.⁶

Matching administrative and survey data also provides operational efficiencies. Rather than collecting its own information, an agency can tap into a source of information that is already being collected and validated by another government agency, saving both time and money. This is a real research concern, as people increasingly decline to participate in voluntary surveys because of identity fraud and privacy concerns.

Obstacles to Linking Administrative and Survey Data

Before linking its administrative records to survey data, SSA verifies the identity of the survey respondents to make certain that the survey record is matched to the correct administrative record. Because the SSN is the most commonly used unique identifier in the United States, it is the key variable used to link data. The SSN, name, date of birth, and gender from the survey files are matched with information in SSA’s Numident file, the master file of SSN assignments. SSA uses an algorithm called the Enumeration Verification System for this validation. Certain tolerances are applied: For instance, the system checks for transposed digits in the SSN and tries variations of compound surnames. Only records that pass the validation check are linked.

Historically, to permit the linkage of individually collected survey data and administrative records for statistical research, the Census Bureau asked its survey respondents directly for their SSNs. For survey respondents who voluntarily provided a SSN, the bureau sent the SSNs and accompanying identifying information to SSA, where the information was validated through the Enumeration Verification System. Once SSNs were verified, SSA extracted the appropriate data from its administrative data files and sent the

data extracts to the Census Bureau for linkage with its corresponding survey record. The Census Bureau then removed the SSNs from the linked data and replaced them with unique survey identification numbers to protect the respondents' privacy.

Regrettably, survey respondents have become increasingly reluctant to provide their SSNs to survey data collectors. Because SSNs are widely used as a universal identifier, widespread access to them from non-SSA sources has provided individuals with the opportunity to commit identity theft. Respondents refusing to provide SSNs to SIPP interviewers increased from 12 percent to 35 percent between the 1996 and 2004 panels. Those refusing to provide SSNs in CPS increased from approximately 10 percent in 1994 to almost 23 percent by 2003. Declining response rates threatened the utility of linked survey and administrative data. One problem was that missing SSNs meant smaller and smaller proportions of the sample could be matched to administrative records. Additionally, differing rates of SSN nonresponse could instill potential bias into subsequent analyses if respondents who provided SSNs differed in some systematic, nonrandom way from those who did not.

Reacting to an expanding SSN nonresponse problem, the Census Bureau has stopped directly requesting a SSN. Instead, under a new methodology, a respondent is informed that the survey data will be matched with other federal data for research purposes. Unless the respondent opts out, the Census Bureau then combines SSN application information from SSA's Numident file with address records from the IRS, SSA, and other sources to determine the respondent's correct SSN. Once a match is found, survey and administrative data for the respondent are linked. Using this methodology, match rates have increased from about 60 percent in 2001 to 79 percent in 2004.

Data Sharing Authority and Procedures

As provided under the Privacy Act (5 U.S.C. § 552a), SSA is responsible for safeguarding the information maintained in its administrative files against an invasion of an individual's personal privacy. Other legal protections of the information SSA maintains or links to are provided by the Social Security Act and regulations, the Confidential Information Protection and Statistical Efficiency Act, Title 13 of the United States Code governing the Census Bureau, and the Internal Revenue Code covering earnings data that are considered to be tax return information.

SSA policy is to share identifiable data only with those having the legal authority to access data for a particular purpose, and only if identifiable data are required to accomplish a research or statistical purpose. The requestor must submit a proposal, a data protection plan, and confidentiality agreements. A Memorandum of Agreement must be approved by SSA's Office of the General Counsel. The user must guarantee to keep the data secure, not redisclose the data, and restrict the use of the data to the approved purpose. Access to SSA data that have been linked to Census Bureau data is subject to additional restrictions imposed by Title 13 of the U.S. Code, such as requiring users to obtain Special Sworn Status and permitting access only for Census-approved purposes and at a Census-approved site. Census Bureau procedures and regulations dictate how survey data can be used. SSA is not authorized to grant access to matched CPS or SIPP data. Additionally, the Internal Revenue Code provides its own restrictions, such as limiting access to earnings data to certain individuals and for certain purposes.

Economic Analysis and Modeling

Linked administrative and survey data are of vital importance in developing predictive modeling systems that enable SSA and policymakers to understand the broad impact and distributional effects of current program regulations and reform proposals. To address this need, SSA has developed microsimulation models to analyze the current status of its programs, the scope and impact of those programs in the future, and the effect of proposed changes to the Social Security system. Model outputs describe the impact of SSA programs on our economy, society, and beneficiary populations, and provide detailed demographic and economic information on beneficiaries and covered workers. Those products are used by government planners and policymakers and also by actuaries, economists and other social scientists, the media, and the public to analyze Social Security programs and their impacts. As such, these models are powerful research tools. Two significant examples are briefly described below.

Modeling Income in the Near Term (MINT). MINT is the most prominent model used in OASDI analysis. MINT is a microsimulation dataset that links household data from Census Bureau surveys with SSA administrative records to obtain information on earnings, benefit receipt, and date of death. It covers

individuals born between 1926 and 1972, with a core population consisting of individuals born between 1931 and 1965. The most recent MINT dataset contains more than 350,000 observations.⁷

MINT is used to estimate the effects of a variety of policy and other program changes. It tracks the experiences of survey respondents and projects their income and other characteristics into the future, adjusting for expected demographic and socioeconomic changes. Accordingly, MINT projects the major pillars of retirement income: Social Security benefits, pension benefits, income from assets, earnings (for working Social Security beneficiaries), and SSI. In addition, MINT simulates events such as marital outcomes, age at first benefit receipt, and year of death, as well as the characteristics of former, current, and future spouses.

Because many of the parameters in the MINT data system can be altered by the analyst, the model has numerous uses in potential policy evaluation. For example, MINT has been used to examine cross-cohort differences in the sources of retirement income, and to assess the impact of Social Security benefit reforms on the level of benefits, expected retirement income, and expected poverty rate among future retirees. With its detailed demographic information, MINT enables examinations of economic well-being in retirement by sex, race, education, marital status, and birth cohort. MINT is also used to analyze the effects of proposed or hypothetical policy reforms.

Financial Eligibility Model (FEM). SSA also regularly models eligibility and participation in the SSI program. SSI is the income source of last resort for individuals who are elderly or severely disabled. Eligibility is restricted to individuals with limited resources, and the payment amount is reduced as the recipient's income rises. Information from SIPP is matched to SSA administrative data to model SSI eligibility and participation. SIPP collects detailed information on sources and amounts of income, as well as assets, which are vital in determining eligibility under SSI program rules. The fact that SIPP asks respondents about program participation and provides income data on a monthly basis is also critical to modeling SSI eligibility, which can vary from month to month.

FEM simulates the effects of potential changes to SSI eligibility criteria on the number of eligible individuals, the number of participants, the distribution of SSI benefits among participants, and poverty status under various policy regimes.⁸ However, FEM is limited in the area of behavioral modeling.

The core SIPP demographic characteristics, as well as household composition, are important factors in determining SSI eligibility. Other characteristics such as race, sex, ethnicity, educational attainment, and health insurance coverage are not directly used in the SSI eligibility determination, but are important descriptors that can be used to model SSI participation. Information on disability and work limitations can be used to estimate whether an individual meets the disability criteria for SSI eligibility, while data on assets are used to estimate resource eligibility for SSI.

Incomplete surveys and administrative data can affect the accuracy of modeling estimates. It is particularly critical to use the correct program participation information and benefit amounts in the FEM because these values are used to estimate model parameters. For modeling in particular, the linking of administrative and survey data maximizes the robustness of the model's base information. Modeling efforts benefit from having the wide range of survey data items (often with incomplete or inaccurate respondent reporting) supplemented by the complete and accurate data from program administrative records.

SSA Public-Use Information Products

For research and statistical purposes, SSA develops a wide range of information from linked data that is shared with other researchers, policymakers, and the public. One way SSA disseminates information is by creating public-use versions of its administrative data. Public-use microdata files are beneficial for conducting statistical analyses and research studies that could not be performed using other publicly available data.

SSA has two strategies for producing public-use files. One involves working with other agencies to develop a synthetic file, which has all of the statistical properties of the original dataset, but is artificially generated so as not to breach the confidentiality of survey results. This methodology is the outcome of a joint research project of the Census Bureau, SSA, IRS, and Congressional Budget Office, in which SSA benefit and longitudinal earnings records are linked to SIPP data. To prevent disclosure of individual identities, especially through linkages of previously released SIPP public-use files, synthetic data are generated based on models prepared using the actual underlying data sets. Two criteria must be satisfied before this file can be publicly released: protection of the confidentiality of the source data and the analytical validity of the synthetic data. Testing has confirmed that the data file

meets all privacy protection criteria. Still in progress is an evaluation of the quality of the data resulting from this new methodology.

SSA has used a different methodology to produce three more traditional public-use microdata files based on its administrative data. The agency took a number of steps in developing the public-use files to ensure that individuals cannot be identified, including removing information such as SSN, name, address, and exact date of birth; topcoding (removing extremely high values and substituting a ceiling value); and rounding benefit and earnings amounts. The files were also reviewed by a Disclosure Review Board, using a detailed checklist on disclosure potential, looking in particular for unique records and for overlap with other publicly available data. Approval was obtained by the Office of Public Disclosure in SSA, and by the IRS for the file containing earnings information.

The first of these three traditional public-use files, released in 2003, is based on 2001 data for the OASDI program. It consists of approximately 460,000 records—a 1 percent sample of SSA's MBR—and can be used to study the beneficiary population and the effects of current and proposed legislative and program provisions. Because of its size, it can also be used to study relatively small subpopulations. It includes such detailed information as type and amount of benefits received, timing of benefit receipt, benefit reductions resulting from early retirement, and benefit increases resulting from delayed retirement.

The second public-use file, also released in 2003, is based on 2001 data for the SSI program. It consists of a 5 percent sample of the master record of SSI applicants and beneficiaries. It includes approximately 320,000 records and provides a number of programmatic variables concerning the SSI population, such as disability diagnosis code, living arrangements, and non-SSI income.

The third SSA public-use file, released in 2005, uses a 1 percent random sample from the MBR. It consists of approximately 470,000 records representative of beneficiaries who were entitled to receive Social Security benefits for December 2004. This file consists of two separate but linkable subfiles—one with benefit information and the other with longitudinal earnings information. This public-use file is significant since it is the first public release of longitudinal earnings records drawn as a representative sample of the beneficiary population. Because of the importance of earnings histories for calculating benefits, this file has

broad appeal to outside researchers studying Social Security-related issues.

SSA also maintains a record of deaths called the Death Master File (DMF), a version of which is available to the public through the Department of Commerce's National Technical Information Service.⁹ As of December 2008, the DMF contained more than 83 million records. The information for each decedent consists of SSN, name, date of birth, date of death, state or country of residence (for records added before February 1988), ZIP code of last residence, and ZIP code of lump-sum death payment. The public version of the DMF does not include data from certain states that restrict SSA's redisclosure of their death information. This file has been used for research to determine the vital status of subjects in longitudinal studies, to evaluate age reporting in other data sources, as a sampling frame for long-lived individuals, and for genealogical purposes.

In addition to public-use data files, SSA produces a wide array of publications and related products that range from ORES research monographs in support of policy analysis to recurring statistical publications. Monthly or annual publications provide statistics on the operation and beneficiaries of the OASDI and SSI programs and on the earnings of the working and beneficiary populations. SSA publications can be categorized as research and analysis publications, statistical publications and chartbooks, publications that cover the OASDI and SSI programs, publications on the income of the aged, and special topic publications.¹⁰

Additional Statistical Linkages and Services

Linkages between SSA administrative files and Census Bureau surveys are discussed above. SSA also links its administrative data with survey data from NCHS' National Health Interview Study, the principal source of information on the health of the civilian population. In addition, SSA collaborates with data collection efforts of nongovernmental research institutions. For example, the University of Michigan conducts the Health and Retirement Study (HRS), which is in part supported by SSA. Every 2 years, this survey collects socioeconomic and health-related information on more than 22,000 Americans over age 50.¹¹ Because of limitations under the Privacy Act, SSA can share its administrative data with the University of Michigan only if the survey respondent has signed a release.

SSA also matches its administrative data with the administrative data of other federal, state, and local agencies for internal research purposes, as well as for external researchers on a cost-reimbursable basis. For example, SSA's benefit and earnings records have been matched with files identifying homeless people compiled by the New York City Department of Homeless Services. SSA used the linked data to produce statistics showing the impact of benefits and earnings on the homeless population's use of shelters. In addition, there is a long-standing agreement with the Centers for Medicare and Medicaid Services (CMS) to match CMS and SSA data for internal research projects and contract-based research.

Because SSA's administrative data cover virtually the entire U.S. population, Congress directs the agency to provide vital status information to epidemiologists when such projects are determined to support the national health interest. For instance, members of the National Cancer Registry provide lists of cancer patients to SSA, or industry epidemiologists provide SSA with industry-specific lists of former employees. These files are used to check SSA's death records, beneficiary rolls, and earnings files to ascertain if the persons have died, or can be presumed alive.¹²

On request, SSA also provides tabulations of its data to Congress and to executive branch entities to answer policy questions and to better inform policymakers about characteristics of the worker and beneficiary populations.

Conclusion

Data and data accessibility lie at the heart of social science and policy-related research. SSA collects a wealth of data in its role as administrator of two large national entitlement programs. SSA and SSA-approved research organizations use these data to produce a wide variety of information that is vital to developing social insurance policy.

Linking SSA's administrative data with survey data yields a broader set of demographic and socioeconomic information and also improves the quality of the survey data. These data are used to produce analyses and research on policy initiatives for the OASDI and SSI programs, and on the earnings of the working and beneficiary populations. SSA studies how these programs and potential changes to them affect individuals, the economy, and program solvency. The agency develops models to project demographic and economic characteristics of the current working population into

the future. SSA also produces public-use microdata files that are available to outside researchers, as well as a large variety of research and statistical publications to inform policymakers and the public.

SSA administrative data are a great benefit not only to those administering the Social Security programs, but also to the wider statistical, research, and policy analysis community.

Notes

¹ For characteristics of program participants, see DeCesaro and Hemmeter (2008).

² See Haines and Greenberg (2005) for more detail.

³ Some examples of studies using linked survey and administrative data include Rupp, Strand, Davies, and Sears (2007) and Powers and Neumark (2001).

⁴ See Census Bureau (2008), Koenig (2003), and Fisher (2008a) for more detail.

⁵ Examples of studies using linked data include Olson (2002), Huynh, Rupp, and Sears (2002), and Neumark and Powers (2004).

⁶ See Fisher (2008a) and Butrica (2008) for more detail.

⁷ Some examples of the many studies that use MINT are Butrica, Iams, and Sandell (1999), Butrica and Iams (1999), Butrica, Iams, Moore, and Waid (2001), and Burtless, Bosworth, and Sahm (2001). For more detail on MINT, see Toder and others (2002) and Shoffner, Biggs, and Jacobs (2005).

⁸ An example of a study using FEM simulations is Davies and others (2002).

⁹ See www.ntis.gov/products/ssa-dmf.aspx.

¹⁰ A comprehensive list of publications can be found at www.socialsecurity.gov/policy. More information on public-use files and other SSA data is available at www.socialsecurity.gov/policy/docs/microdata/index.html and www.socialsecurity.gov/policy/data_subject.html.

¹¹ Some examples of studies using HRS data include Gustman, Mitchell, Samwick, and Steinmeier (1997), Cunningham and Engelhardt (2002), Gustman and Steinmeier (2005), and Engelhardt and Kumar (2006).

¹² Epidemiologists can request vital status information from SSA at www.socialsecurity.gov/policy/about/epidemiology.html.

References

[Board of Trustees] Board of Trustees of the Federal Old-Age and Survivors and Disability Insurance Trust Funds. 2008. *2008 annual report of the federal Old-Age and Survivors and Disability Insurance Trust Funds*. Washington, DC. Available at www.socialsecurity.gov/OACT/TR/TR04.

- Burtless, Gary, Barry Bosworth, and Claudia Sahn. 2001. The trend in lifetime earnings inequality and its impact on the distribution of retirement income. CRR Working Paper No. 2001-03. Chestnut Hill, MA: Center for Retirement Research at Boston College.
- Butrica, Barbara A. 2008. Older Americans' reliance on assets. Opportunity and Ownership Facts No. 10. Washington, DC: Urban Institute. Available at www.urban.org/publications/411632.html.
- Butrica, Barbara A., and Howard M. Iams. 1999. Projecting retirement income of future retirees with panel data: Results from the Modeling Income in the Near Term (MINT) project. *Social Security Bulletin* 62(4): 3–8.
- Butrica, Barbara A., Howard M. Iams, and Steven H. Sandell. 1999. Using data for couples to project the distributional effects of changes in Social Security policy. *Social Security Bulletin* 62(3): 20–27.
- Butrica, Barbara A., Howard M. Iams, James H. Moore, and Mikki D. Waid. 2001. Methods in Modeling Income in the Near Term (MINT I). ORES Working Paper No. 91. Washington, DC: SSA, Office of Research, Evaluation, and Statistics. Available at www.socialsecurity.gov/policy/docs/workingpapers/wp91.pdf.
- Census Bureau. 2008. Current Population Survey (CPS). Washington, DC. Available at www.census.gov/cps.
- . 2009. Survey of Income and Program Participation (SIPP). Washington, DC. Available at www.census.gov/sipp.
- Cunningham, Christopher R., and Gary V. Engelhardt. 2002. Federal tax policy, employer matching, and 401(k) saving: Evidence from HRS W-2 Records. *National Tax Journal* 55(3): 617–645.
- Davies, Paul S., Minh Huynh, Chad Newcomb, Paul O'Leary, Kalman Rupp, and Jim Sears. 2002. Modeling SSI financial eligibility and simulating the effect of policy options. *Social Security Bulletin* 64(2): 16–45.
- DeCesaro, Anne, and Jeffrey Hemmeter. 2008. Characteristics of noninstitutionalized DI and SSI program participants. Research and Statistics Note No. 2008-02. Washington, DC: SSA. Available at www.socialsecurity.gov/policy/docs/rsnotes/rsn2008-02.html.
- Engelhardt, Gary V., and Anil Kumar. 2006. Employer matching and 401(k) saving: Evidence from Health and Retirement Study. NBER Working Paper No. 12447. Cambridge, MA: National Bureau of Economic Research. Available at www.nber.org/papers/w12447.
- Fisher, T. Lynn. 2008a. Estimates of unreported asset income in the Survey of Consumer Finances and the relative importance of Social Security benefits to the elderly. *Social Security Bulletin* 67(2): 47–53.
- . 2008b. Measuring the relative importance of Social Security benefits to the elderly. *Social Security Bulletin* 67(2): 65–72.
- Gustman, Alan L., and Thomas L. Steinmeier. 2005. Imperfect knowledge of Social Security and pensions. *Industrial Relations* 44(2): 373–397.
- Gustman, Alan L., Olivia S. Mitchell, Andrew A. Samwick, and Thomas L. Steinmeier. 1997. Pension and Social Security wealth in the Health and Retirement Study. NBER Working Paper No. 5912. Cambridge, MA: National Bureau of Economic Research. Available at www.nber.org/papers/w5912.
- Haines, Dawn E., and Brian Greenberg. 2005. Statistical uses of Social Security administrative data. Paper presented at American Statistical Association Joint Statistical Meetings, Minneapolis, MN. Available at www.oecd.org/dataoecd/58/13/36237686.pdf.
- Huynh, Minh, Kalman Rupp, and James Sears. 2002. The assessment of Survey of Income and Program Participation (SIPP) benefit data using longitudinal administrative records. SIPP Working Paper No. 238. Washington, DC: Census Bureau. Available at www.census.gov/dusd/MAB/wp238.pdf.
- Koenig, Melissa L. 2003. An assessment of the Current Population Survey and the Survey of Income and Program Participation using Social Security administrative data. Paper presented at the Federal Committee on Statistical Methodology Research Conference, Arlington, VA. Available at www.fcsm.gov/03papers/Koenig.pdf.
- Neumark, David, and Elizabeth T. Powers. 2004. The effect of the SSI program on labor supply: Improved evidence from Social Security administrative files. *Social Security Bulletin* 65(3): 45–60.
- Olson, Janice A. 2002. Social Security benefit reporting in the Survey of Income and Program Participation and in Social Security administrative records. ORES Working Paper No. 96. Washington, DC: SSA, Office of Research, Evaluation, and Statistics. Available at www.socialsecurity.gov/policy/docs/workingpapers/wp96.pdf.
- Powers, Elizabeth T., and David Neumark. 2001. The Supplemental Security Income program and incentives to take up Social Security early retirement: Empirical evidence from matched SIPP and Social Security Administration files. NBER Working Paper No. 8670. Cambridge, MA: National Bureau of Economic Research. Available at www.nber.org/papers/w8670.
- Rupp, Kalman, Alexander Strand, Paul Davies, and Jim Sears. 2007. Benefit adequacy among elderly Social Security retired-worker beneficiaries and the SSI federal benefit rate. *Social Security Bulletin* 67(3): 29–51.
- Shoffner, Dave, Andrew Biggs, and Preston Jacobs. 2005. Poverty-level annuitization requirements in Social Security proposals incorporating personal retirement accounts. Issue Paper No. 2005-01. Washington, DC: SSA.

[SSA] Social Security Administration. 2007. *SSI Annual Statistical Report, 2006*. Washington, DC. Available at www.socialsecurity.gov/policy/docs/statcomps/ssi_ast/2006/index.html.

———. 2008. *Annual statistical supplement to the Social Security Bulletin, 2007*. Washington, DC. Available at www.socialsecurity.gov/policy/docs/statcomps/supplement/2007/index.html.

Toder, Eric, Lawrence Thompson, Melissa Favreault, Richard Johnson, Kevin Perese, Caroline Ratcliffe, Karen Smith, Cori Uccello, Timothy Waidmann, Jillian Berk, Romina Woldemariam, Gary Burtless, Claudia Sahm, and Douglas Wolf. 2002. Modeling Income in the Near Term: Revised projections of retirement income through 2020 for the 1931–1960 birth cohorts. Washington, DC: Urban Institute. Available at: www.urban.org/publications/410609.html.