USES OF ADMINISTRATIVE DATA AT THE U.S. SOCIAL SECURITY ADMINISTRATION

Prepared for the
International Seminar
on the Use of Administrative Data
for Economic Statistics
and the
Register-Based Population and Housing Census

Korea National Statistical Office
May 19-20, 2008
Daejeon, Korea

Presented by
Linda Drazga Maxfield, Ph.D.
Office of Retirement and Disability Policy
Social Security Administration
United States of America

The United States Social Security Administration (SSA) is an independent agency of the United States federal government. Its mission is to advance the economic security of the nation’s people through compassionate and vigilant leadership in shaping and managing America's Social Security Programs. SSA is responsible for one of the United States’ (U.S.) largest entitlement programs—the Old-Age, Survivors, and Disability Insurance (OASDI) Program. OASDI is commonly referred to as Social Security. As the name suggests, OASDI provides monthly benefits to qualified retired and disabled workers and their dependents and 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

Today, more than 163 million people work and pay Social Security contributions, and more than 50 million people receive monthly Social Security benefits.[1] During 2006 approximately 162 million employees and self-employed workers, along with employers, contributed $626 billion to the OASDI trust funds—through which contributions are credited and benefits are paid. Workers and employers each contribute 6.2 percent of covered earnings (currently up to $102,000) and self-employed workers contribute 12.4 percent of covered earnings.[2] In December 2006, total benefits paid by the OASDI Program exceeded $46 billion each month (nearly $546 billion annually). According to the latest Social Security Trustees Report, these cash benefits made up 4.3 percent of the nation's gross domestic product.

[1] 2008 OASDI Trustees Report (http://www.ssa.gov/OACT/TR/TR08/trTOC.html)
[2] Annual Statistical Supplement (2006)

Social Security benefits are essential to the economic well-being of millions of individuals. Social Security pays benefits to about 90 percent of the U.S. population aged 65 or older. It 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 beneficiaries. Social Security reaches almost every family, and at some point will touch the lives of nearly all Americans.[3]

SSA also administers the Supplemental Security Income (SSI) Program, which is a needs-based program that provides financial support for aged, blind, and disabled adults and children with limited income and resources.[4] In 2006, 7.2 million people received monthly SSI benefits totaling $38 billion, with an average benefit of $455.[5]

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, and approximately 1,300 field offices. The Agency issues Social Security numbers (SSNs) to nearly all legal U.S. residents; posts earnings items to workers' earnings records; processes actions to keep recipient records current and accurate; and determines eligibility for Medicare health insurance. SSA also provides support to the Medicaid health insurance program for those with limited income, the Railroad Retirement Program, and the Food Stamp Program. Because of these broad responsibilities, SSA collects and maintains a substantial amount of program-related data on current and potential beneficiaries residing either in the U.S. or abroad.

---

[3] Fisher (2008)
[4] DeCesaro and Hemmeter (2008)
[5] SSI Annual Statistical Report (2006)

With its primary program administrative functions, SSA as a whole is not a statistical agency. The most prominent U.S. 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. However, the U.S. 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 SSA 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 statistics.

This paper discusses the advantages and limitations of using administrative data for research, how administrative data linked to survey information can be used to evaluate and improve survey data, research studies using administrative information, and information about SSA's statistical products and services that are based on administrative data.[6]

The Social Security Administration and Its Wealth of Data

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

---

[6] Haines and Greenberg

• The **Numident** file (Social Security Account **Num**ber **Ident**ification)

The Numident file is a record of applications for a Social Security Card, from which a unique life-long SSN is assigned to an individual. A full record of all changes to the information (such as change of address or change of name) is also maintained. To obtain a card, the individual must provide documented identifying information to SSA. Through the "enumeration at birth" program, children can be issued a Social Security Card when born. Examples of data elements on the Numident data include name, date and place of birth, parent's name and SSNs, and death date.

• The **Master Earnings File**

The Master Earnings File contains the individual lifetime records of wages and self-employment earnings. The file's primary source of information is the W-2 form of the Internal Revenue Service in the U.S. Department of the Treasury. The W-2 forms arrive at SSA (electronic submission is available) with the data entered into the earnings file. The most frequently used file data elements are: the individual's SSN, annual total wages (1978-present), annual self-employment earnings, annual earnings used for OASDI contributions (1951-present), and report year.

• The **Master Beneficiary Record**

The Master Beneficiary Record 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 the application is adjudicated as an award, denial, abatement, or withdrawal. Information maintained on the MBR includes the primary worker's SSN, the beneficiary's own SSN, benefit application date, benefit entitlement date, benefit type, benefit amount, and other payment data.

• The **Supplemental Security Record**

The Supplement Security Record contains beneficiary income, disability diagnosis, and payment information for determining eligibility and administering the needs-based SSI program. Some examples of data elements are SSN, date of claim, citizenship status, 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 of high quality and subject to 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/or demographic variables (such as, for example, 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 do not contain information about non-participants who could potentially be affected by a proposed program policy change. SSA researchers need this information to project cost changes in the program because of policy changes, as well as potential distributional effects of policy changes upon different demographic or economic groupings. Survey data can provide this information.

Benefits of Supplementing Administrative Data with Survey Data

The U.S. government conducts numerous large surveys to produce key information to inform decision makers and to document economic and social trends in the country. Surveys conducted by U.S. statistical agencies ask a broad range of questions on a wide variety of topics. They often include extensive demographic information and are typically representative of the civilian, non-institutional population.

Analogous to administrative program 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 and/or report demographic or economic information.

SSA uses survey data to take advantage of the enhanced analytic potential afforded by linking survey and administrative data. In fact, SSA has been linking its administrative data with survey data 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 and the National Survey of Supplemental Security Income Children and Families. However, more often, SSA's administrative data are linked with on-going 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.[7]

SSA's biggest data-linkage partner is the U.S. Census Bureau. Three of the Census Bureau's major survey efforts are the Current Population Survey, the American Community Survey, and the Survey of Income and Program Participation. These surveys vary in sample size, amount of detailed information collected, and time frame of data availability.

- The **Current Population Survey** (CPS)

  The 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, the term given to the annual CPS Supplement fielded in March of each year, focuses on income and poverty in the U.S. The CPS is the official source of statistics on the U.S. unemployment rate and the U.S. poverty rate. For more information, visit http://www.census.gov/cps.[8]

- The **American Community Survey** (ACS)

  The ACS collects a wide variety of demographic, social, economic, and housing information from three million U.S. residents each year. Collecting data every year reduces the cost of the official national decennial census, and provides throughout the decade more up-to-date information about trends in the U.S. population at the local community level. Questions cover such subjects as marital status and living arrangements, fertility, race, ethnicity and citizenship, work, income, disability, and mortgage indebtedness. For more information, visit http://www.census.gov/acs/www/index.html.

[7] Some examples of studies include Rupp, Strand, Davies and Sears (2007) and Powers and Neumark (2001).
[8] Koenig (2003) and Fisher (2007)

- The **Survey of Income and Program Participation** (SIPP)

  SIPP provides considerably more detailed information on income and program participation than does the CPS. It additionally includes topical modules that allow for thorough scheduled data collection on additional topics to better understand income and program participation. SIPP data elements include all sources of money and non-money income (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, where a panel is a set of respondents followed for 32 to 48 months with interviews occurring every four months. For more information, visit http://www.bls.census.gov/sipp.

The ability to use linked administrative and survey data to evaluate methodological reporting issues is an important tool for statistical analysis. Linking administrative data to survey information combines the completeness and accuracy of SSA administrative data with the range and scope of survey data—maximizing the strengths and minimizing the limitations associated with each of these two data sources. With the information on program participation and receipt from SSA administrative records, analysts are able to correct misreported values on survey files, yielding accurate underlying data and improving statistical estimates.

The linking of survey and SSA administrative records significantly expands research opportunities beyond that provided by survey or administrative data files alone. The survey information provides detailed background information on demographic, income, self-reported health status, and other program characteristics of Social Security program participants and non-participants.[9] The 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

[9] Olson (2002), Huynh, Rupp, and Sears (2002), and Neumark and Powers (2004)

for the construction of detailed profiles of individual and family characteristics at the time of program participation and detailed information related to program dynamics.

There are substantial methodological benefits to linking administrative data to survey data. One of the major advantages is the information gained from data reporting issues related to the respondent's recall of past program participation and income receipt. When comparable data are collected in both administrative data files and survey collection efforts, statisticians and policy analysts are able to evaluate the extent of underreporting or overreporting attributable to the respondent. For example, survey respondents often mistakenly attribute benefit receipt to one of the two Social Security Programs (OASDI or SSI) when in fact the benefit was received from the other program. Further, survey responses matched to administrative data can document the benefit amounts that the recipient reports in the survey and compare these respondent answers with the actual dollar amounts distributed. Another methodological data issue addressed by the matching of administrative and survey data involves asset income. In 2002, only 55 percent of the U.S. Census Bureau's CPS respondents aged 65 or over reported any asset income, down from 69 percent of comparable respondents reporting asset income in 1990. The Census Bureau and SSA are using U.S. Internal Revenue Service income and Social Security benefit and earnings data linked to CPS data to investigate the extent to which asset income among the elderly is actually declining or merely unreported.[10]

In addition to improving survey data, matching administrative data to survey data provides efficiencies. Rather than an agency having to fund the collection of information itself, that 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 in

[10] Sears et al. (2003), Fisher (2007), and Butrica (2008)

today's environment of identify fraud and privacy concerns, with increasing numbers of persons who do not elect to participate in voluntary surveys. This subject is a focus of the next section of this report.

Linkage of Administrative Data with Survey Data

Before SSA will link its administrative data to survey data, the identity of the survey respondents must first be verified to make certain that the identified survey record is being linked to the correct administrative record. Because the SSN is the most commonly used unique identifier in the U.S., where available it is the key variable used to link data. The SSN, name, date of birth, and gender from the survey data 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. The Enumeration Verification System applies certain tolerances in the match. For instance, the verification system will check for transposed digits in the SSN and will try variations of compound surnames. Only those records that pass the validation check are linked.

Historically, in order to permit the linkage of individually collected survey data and administrative records for statistical research purposes, the Census Bureau directly asked its survey respondents for their SSNs. For survey respondents who voluntarily provided a SSN to the Census Bureau, the Census Bureau sent the SSNs and accompanying identifying information to SSA, where the information was validated through SSA's Enumeration Verification System. In the case of ACS, 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 a unique survey identification number that ensured that the matched data protected the respondents' privacy guarantee. In the case of CPS and SIPP, SSNs are not available and publicly-releasable identification numbers are used.

Regrettably, survey respondents have become increasingly reluctant to provide their SSNs to survey data collectors. Because SSNs are widely used as a universal identifier in the U.S., widespread access to the SSN from non-SSA sources has provided individuals with an opportunity to commit identity fraud. Respondent refusals to provide SSNs to SIPP interviewers increased from 12 percent to 35 percent of respondents between the 1996 and 2004 panels, respectively. The percent 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, differential rates of SSN nonresponse could instill potential bias into subsequent analyses if those respondents who provide SSNs differ in some systematic, nonrandom way from those respondents who do not provide SSNs.

Reacting to an expanding SSN nonresponse problem, the U.S. Census Bureau stopped requesting respondents to provide a SSN directly during the survey. Instead, the Census Bureau developed, pretested, and adopted a new methodology that preserves respondent participation in surveys and maintains a high match rate. The Census Bureau informs respondents that SSNs will be determined and data will be matched with other federal data for research purposes. With the new methodology, the Census Bureau does not request the SSN from the respondent. If the respondent does not opt out of this provision, the Census Bureau uses a new system it has developed using SSN application information from SSA's Numident file combined with address information from the U.S. Internal Revenue Service, SSA, and other sources to determine the correct SSNs. Once a match is found, the data for the respondent is linked from the two data sources. Using this methodology, match rates increased to 90%, even surpassing those of the earlier decade.

Authority and Procedures to Share Data

As provided under the Privacy Act (codified into U.S. law under 5 U.S.C. § 552a), SSA has a responsibility to safeguard the individuals whose information is maintained in the SSA administrative files against an invasion of their personal privacy. Other legal protections of the information maintain or linked 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 has a policy of only sharing identifiable data with those with 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 is prepared and approved by SSA's Office of General Council. The user must guarantee to keep the data secure, not to redisclose the data, and to restrict the use of the data to the approved purpose. Users of SSA data that have been linked to Census Bureau data are subject to additional restrictions imposed by Title 13 of the U.S. Code, such as having to become a Special Sworn Census Agent and accessing the data only at a Census-approved site. The Census Bureau procedures and regulations dictate how survey data can be used. SSA does not have authority to grant access to matched ACS, CPS, or SIPP data.

Additionally, the Internal Revenue Code has its own restrictions, such as who can have access to earnings data for what purposes.

Models of Eligibility, Participation, and Policy Impacts

Linked administrative and survey data are of vital importance to SSA and policymakers in developing predictive modeling systems 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 obtain information on the current experience of its programs, the projected scope and impact of those programs in the future, and the effect of proposed changes to the Social Security system. Outputs from the models 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 as well as by actuaries, economists and other social scientists, the media, and the public to analyze Social Security programs and their impact on the nation.

Modeling Income in the Near Term (MINT) Model. The most prominent model used to analyze the OASDI program is the Modeling of Income in the Near Term (MINT) model. The MINT model is a microsimulation dataset that utilizes household data from Census Bureau surveys linked to SSA's administrative records to obtain information on earnings, benefit receipt, and date of death. It contains information on individuals born between 1926 and 1972, with the core population consisting of individuals born between 1931 and 1965. The most recent MINT dataset contains more than 350,000 observations[11].

Through statistical adjustments and forecasts of survey and administrative data, the MINT model is used to make estimates of the effects of a variety of policy and other program changes. The MINT model measures the experiences of survey respondents and statistically projects their income and 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, the earnings of working Social Security beneficiaries, and SSI. In addition, MINT simulates the year of death, marital histories, and the characteristics of former, current, and future spouses, as well as the age of first benefit receipt.

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

Financial Eligibility Model (FEM). SSA also regularly conducts modeling of eligibility and participation in the SSI program. The SSI program serves as an income source of last resort

[11] Some examples of studies that used the MINT model 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 the MINT model, see Toder et al. (2001) and Shoffner, Biggs, and Jacobs (2005).

for individuals who are elderly or severely disabled. Program eligibility is restricted to individuals with limited resources, and the benefit amount is reduced as the recipient's income rises. Information from SIPP data matched to SSA administrative data is used to model SSI eligibility and participation. The SIPP survey 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.

SSA's Financial Eligibility Model (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 the poverty status under various policy regimes. However, the model is limited in the area of behavioral modeling.[12]

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 efforts in

[12] An example of a past study is Davies, Huynh, Newcomb, O'Leary, Rupp, and Sears (2002).

particular, the linking of administrative data to survey data maximizes the robustness of the model's base information. Modeling efforts benefit from having the wide range of survey data items (with often incomplete or inaccurate respondent reporting) supplemented by the complete and accurate data from program administrative data systems.

Dissemination of SSA Products that Provide Information

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 files. Public-use microdata files are beneficial for conducting statistical analyses and research studies that could otherwise not be performed using the publicly-available aggregate information or survey data.

SSA has pursued two methodological strategies for the production of public use files. One involves development of a synthetic file, which has all of the statistical properties of the original dataset, but is artificially generated so as not to breach confidentiality of survey respondents. For this joint research project of the Census Bureau, SSA, Internal Revenue Service, and Congressional Budget Office, a linked file was created using SSA benefit and longitudinal earnings information and SIPP survey data. Using this linked file, a research program was established to produce synthetic data that integrates these three major data sources. To prevent disclosure of individual identities, especially through linkages of previously released SIPP public-use files, the data were synthetically 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 data and the analytical validity of the synthetic data. The data file has been tested and meets all privacy criteria. Still in progress is an evaluation project to assess the quality and reliability of the data file resulting from this new methodological strategy.

Using a second methodology that is more traditional, SSA produced three public-use microdata files based on its administrative data. A number of steps were taken in developing the public-use files to ensure that individuals cannot be identified, including removing information such as SSN, name, geography, and exact date of birth, and top-coding and rounding benefit and earnings amounts. The files were also reviewed by a Disclosure Review Board and by IRS, using a detailed checklist on disclosure potential and looking in particular for unique records and overlap of records in these files with other data already publicly available. Approval was obtained by the Office of Public Disclosure in SSA and by the IRS.

The first of these three traditional public use files, released in 2003, is based on the OASDI program, consisting of approximately 460,000 records—a one-percent sample of SSA's OASDI Master Beneficiary Record. This file 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 be used to study relatively small subpopulations. It includes such detailed information as types of benefits received and amounts, timings of benefits receipt, benefit reduction amounts resulting from early retirement, or increases in benefits due to delayed retirement.

The second public use file, also released in 2003, is based on the SSI program. It consists of a five-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 the diagnosis code for disability, living arrangements, and income other than SSI that affects benefit amounts.

The third SSA public use file, released in 2005, uses a one-percent random sample 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.

In addition to public use data files, SSA also produces a wide array of publications and related products that range from ORES' research in support of policy analysis, to recurring monthly or annual publications. The 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. Publications produced by SSA can be categorized as follows: research and analysis publications; statistical publications and chartbooks; OASDI publications; SSI publications; publications on the income of the aged; and special topic publications. A comprehensive list of publications can be found at http://www.socialsecurity.gov/policy. For more information regarding public use files and other SSA data, visit http://www.socialsecurity.gov/policy/docs/microdata/index.html or http://www.socialsecurity.gov/policy/data_subject.html.

Providing Additional Statistical Services

In addition to research and statistical products, SSA provides other services to the wider research community to the extent that resources are available.

The discussion above describes linkages between SSA administrative files with survey systems of the U.S. Census Bureau surveys. In addition, SSA links its administrative data with survey data from the National Center for Health Statistics (NCHS), part of the U.S. Centers for Disease Control and Prevention. NCHS' National Health Interview Study is the principal source of information on the health of the civilian population of the U.S. SSA also collaborates with data collection efforts of nongovernmental research institutions. For example, the University of Michigan conducts the Health and Retirement Study, which is in part supported by SSA. This survey collects socioeconomic and health-related information on more than 22,000 Americans over the age of 50, every two years. Because of limitations under the Privacy Act, SSA can share this survey's confidential information with researchers only if the survey respondent has signed a release form giving permission, the researchers have obtained approval from SSA, and the University of Michigan to access the matched data.[13]

As well as linking SSA's administrative data with survey data, SSA also matches its own administrative data with the administrative data of other U.S. and state agencies for internal research purposes and for external researchers on a cost-reimbursable basis. For example, SSA's benefit and earnings records were 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 the receipt of benefits and of earnings on the homeless population's use of shelters. In addition, there is a long standing agreement with the Centers for Medicare and Medicare Services (CMS) to match CMS and SSA data for internal research projects and contract-based research. CMS is the U.S. agency responsible for administering

[13] Some examples of past studies include Gustman, Mitchell, Samwick, and Steinmeier (1997), Cunningham and Engelhardt (2002), Gustman and Steinmeier (2005), and Engelhardt and Kumar (2006).

major health-related programs, most specifically the Medicare and Medicaid health insurance programs.

SSA, under Congressional direction due to the administrative data's coverage of virtually the entire U.S. population, also provides 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 lists of past industry-specific 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. Epidemiologists can request vital status information from SSA at http://www.socialsecurity.gov/policy/about/epidemiology.html.

SSA also provides tabulations of its data to the U.S. Congress and other components of the Executive Branch upon request to answer policy questions and to better inform U.S. policymakers about characteristics of the U.S. worker and beneficiary populations.

Conclusion

Data and data access 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 other SSA-approved research organizations use these data to produce a wide variety of information that is vital to developing social insurance policy.

By linking SSA's administrative data with survey data, a broader set of demographic and socioeconomic information is obtained, while also improving the quality of the survey data. SSA uses these data to produce analyses and research on policy initiatives for the Social Security Old-Age, Survivors, and Disability Insurance program, the Supplemental Security Income program, 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. SSA develops models to project demographic and economic characteristics of the current working population into the future. SSA also produces microdata public use files that are available to outside researchers, as well as a large variety of research and statistical publications to inform policymakers and the public.

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

References

Butrica, Barbara A. 2008. “Older Americans' Reliance on Assets,” The Urban Institute. Available at: http://www.urban.org/UploadedPDF/411632_relianceonassets.pdf.

Butrica, Barbara A. and Howard M. Iams. 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, 1999.

Butrica, Barbara A., Howard M. Iams, and Steven H. Sandell. Using Data for Couples to Project the Distributional Effects of Changes in Social Security Policy, *Social Security Bulletin*, 62(3):20-27, 1999.

Butrica, Barbara A., Howard M. Iams, J. Moore, and M. Waid (2001). “Methods in Modeling Income in the Near Term (MINT),” ORES Working Paper No. 93. Washington, DC: Social Security Administration, Office of Policy.

Burtless, Gary, Bosworth, Barry and Sahm, Claudia, "The Trend in Lifetime Earnings Inequality and its Impact on the Distribution of Retirement Income" (August 2001). Boston College CRR Working Paper No. 2001-03. Available at SSRN: http://ssrn.com/abstract=285982 or DOI: 10.2139/ssrn.285982.

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).

Davies, P. S., M. Huynh, C. Newcomb, P. O’Leary, K. Rupp, and J. Sears (2001/2002). “Modeling SSI Financial Eligibility and Simulating the Effect of Policy Options,” *Social Security Bulletin* 64 (2): 16-45.

DeCesaro, Anne and Jeffrey Hemmeter. Characteristics of Noninstitutionalized DI and SSI Program Participants, Research and Statistics Note, Office of Policy, Social Security Administration, 2008-02, 2008.

Engelhardt, Gary V. and Anil Kumar. 2006. "Employer matching and 401(k) Saving: Evidence from Health and Retirement Study," NBER Working Paper No. 12447.

Fisher, T. Lynn. 2007. “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.

Fisher, Lynn. Measuring the Relative Importance of Social Security Benefits to the Elderly, *Social Security Bulletin*, 67(2):65-72, 2008.

Gustman, Alan L. and Thomas L. Steinmeier. 2005. "Imperfect Knowledge of Social Security and Pensions," Industrial Relations, Vol. 44, No.2.

Haines, Dawn E., and Brian Greenberg. 2005. Statistical uses of Social Security administrative data. ASA proceedings of the Joint Statistical Meetings, pp. 1138-1145. American Statistical Association (Alexandria, VA).

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. U.S. Census Bureau.

Koenig, M. L. 2003. "An Assessment of the Current Population Survey and the Survey of Income and Program Participation Using Social Security Administrative Data," 2003 Conference Proceedings of the Federal Committee on Statistical Methodology.

Mitchell, Olivia S., Andrew A. Samwick, and Thomas L. Steinmeier. 1997. "Pension and Social Security Wealth in Health and Retirement Study," NBER Working Paper 5912.

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, Number 96, Social Security Administration, Washington, DC: Office of Research, Evaluation, and Statistics.

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," National Bureau of Economic Research Working Paper No. 8670.

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,* Vol. 67 No. 3.

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, Social Security Administration, Washington, DC: Office of Policy.

Social Security Administration. 2007. *Annual Statistical Supplement to the Social Security Bulletin, 2006*. Washington, DC: Office of Policy.

Toder, Eric, Lawrence H. Thompson, Melissa Favreault, Richard W. Johnson, Kevin Perese, Caroline Ratcliffe, Karen E. Smith, Cori E. Uccello, Timothy Waidmann, Jillian Berk, Romina Woldemariam, Gary T. Burtless, Claudia Sahm, and Douglas A. Wolf. Modeling Income in the Near Term: Revised Projections of Retirement Income Through 2020 for the 1931-1960 Birth Cohorts, Urban Institute, 2002. Available at: http://www.urban.org/url.cfm?ID=410609.