United States General Accounting Office

GAO

March 2003

RETIREMENT INCOME DATA

Improvements Could Better Support Analysis of Future Retirees' Prospects

GAO

Accountability ★ Integrity ★ Reliability

RETIREMENT INCOME DATA

Improvements Could Better Support Analysis of Future Retirees' Prospects

Why GAO Did This Study

Future demographic trends include a doubling of the nation's retiree population and only modest labor force growth, leading to concerns about retirement income adequacy for future generations. Credible projections of the effects of policy proposals on federal spending and future retirees' income are necessary. Because adequate data is critical to the analysis of retirement income and wealth, GAO was asked to identify data improvements that experts say are a priority for the study of retirement income and wealth, as well as factors limiting efforts to obtain the needed information.

What GAO Recommends

The Congress should consider directing Labor to obtain from plan administrators electronic filings of SPDs and summaries of material modifications and make them publicly available.
In addition, GAO recommends that the Secretary of Labor

- direct the Bureau of Labor Statistics to prepare a plan to improve data for analyzing retirement income and wealth in coordination with other agencies and
- obtain copies of summary plan descriptions in cases where analysts working on federally-conducted or sponsored research seek them for statistical purposes.

GAO recommends that the Internal Revenue Service publish tabulations of information filed in IRS forms 5498, 1099R and W-2.

www.gao.gov/cgi-bin/getrpt?GAO-03-337

What GAO Found

Experts consulted by GAO cited priorities for improving retirement data that fit into two broad categories: (1) obtaining better data from employers on employee benefits and (2) obtaining better data by linking more individual and household surveys with administrative data (such as employer records, and Social Security earnings history records). Information from employers, such as documents describing the features of their pension plans, would enable analysts to forecast future retirement income of pension holders, based on the specific features of their pension plans and the likely distribution of pension income and wealth for different segments of the population. Linking individual and household surveys with administrative data creates new information, such as the demographic characteristics of employees whose pensions are affected by the formulas that employers use to calculate contributions or pension payments.

Analysts attribute the shortcomings in retirement income data primarily to fragmentation of the responsibility for data collection and analysis, the burden of data collection on respondents, and confidentiality considerations that restrict access to these data. Fragmentation of responsibility occurs, in their view, because no single agency has a statutory mandate to collect or to analyze all the data needed to support a more comprehensive study of retirement income and wealth. With regard to respondent burden, some information on pension plans is no longer collected, in part, out of concern that it was an unnecessary burden on the firms having to submit it. Finally, certain kinds of data needed to make projections are not widely available to all analysts because of the confidentiality laws that authorize their collection.

Contents

Abbreviations

BLS	Bureau of Labor Statistics
CPS	Current Population Survey
EBRI	Employee Benefit Research Institute
EBSA	Employee Benefits Security Administration
ERISA	Employee Retirement Income Security Act of 1974
HRS	Health and Retirement Study
IRA	Individual Retirement Account
IRS	Internal Revenue Service
OMB	Office of Management and Budget
PBGC	Pension Benefit Guaranty Corporation
PSID	Panel Study of Income Dynamics
SCF	Survey of Consumer Finances
SIPP	Survey of Income and Program Participation
SPD	summary plan description
SOI	Statistics of Income

United States General Accounting Office
Washington, DC 20548

March 21, 2003

The Honorable Robert E. Andrews
Ranking Minority Member
Subcommittee on Employer-Employee Relations
Committee on Education and the Workforce
House of Representatives

Dear Mr. Andrews:

The nation's retiree population will double within the next few decades, while at the same time the labor force will grow only modestly, potentially stressing the national economy. In light of these demographic trends, policymakers have been moved to consider how the future economy can support the large retiree population, and whether retirement income levels will be adequate in the future.[1] To increase their understanding about issues related to retirement, policymakers need credible projections of the effects of their proposals on federal spending and on future retirees' income. Analysts will be better able to develop accurate projections if they have relevant, reliable, and timely data on patterns of saving and actual retirement income and wealth.

Because adequate data will be so important to analysis of retirement income and wealth for future retirees, including people in the "baby boom" generation and later generations, you asked us to assess the adequacy of data available for making such projections. In response to your request, as agreed, we identified (1) data improvements that experts say are a priority for the study of retirement income and wealth and (2) factors limiting efforts to obtain the needed information.

To address these topics, we conducted a Web based survey of nearly 200 individuals with retirement income expertise, held a daylong meeting with a diverse group of 11 retirement income experts, and interviewed retirement income analysts and officials of the Departments of Labor (Labor), Commerce, and Treasury. We conducted our work between

---

[1]For a discussion of standards for evaluating retirement income adequacy, see U.S. General Accounting Office, *Social Security: Program's Role in Helping Ensure Income Adequacy*, GAO-02-62 (Washington, D.C.: Nov. 30, 2001).

February and December 2002 in accordance with generally accepted government auditing standards. (For more details about our scope and methodology, see app. I.)

Results in Brief

Acknowledging that there is a great deal of missing data related to retirement income, experts we consulted cited priorities for data improvements that fit into two broad categories: (1) obtaining better data from employers on employee benefits and (2) obtaining better data by linking more individual and household surveys with administrative data (such as employer records and Social Security earnings history records). The kinds of information from employers that analysts indicated are missing included the documents employers provide to employees describing the features of their pension plans, such as the plan's normal retirement age and reductions for early retirement. This information would help analysts to forecast future retirement income of pension holders, based on the specific features of their pension plans and the likely distribution of pension income and wealth for different segments of the population. With regard to linking datasets, currently linkages between individual and household survey data and administrative data are limited. Linking data creates new information by matching survey data about individuals (using names, or taxpayer identification numbers) to a second set of records, such as administrative records on pension plans. There is inadequate information, for example, about which demographic groups have different types of pensions. Thus, while analysts may know the prevalence of certain formulas used to calculate employer contributions or pension payments, they generally lack reliable information about the demographic characteristics of the employees whose pensions are affected by these formulas.

Analysts attribute the shortcomings in retirement income data primarily to fragmentation of the responsibility for data collection and analysis, the burden of data collection on respondents, and confidentiality considerations that restrict access to these data. Fragmentation of responsibility occurs, in their view, because no single agency has a statutory mandate to collect or to analyze all the data needed to support a more comprehensive study of retirement income and wealth. For example, while the Department of Labor's Bureau of Labor Statistics (BLS) collects data on pensions, and the Census Bureau collects data on individuals' and households' income, neither agency is responsible for all of the data needed to project future retirement income and wealth. Other agencies such as the Department of Labor's Employee Benefits Security Administration (EBSA)—formerly called the Pension Welfare and Benefits

Administration—and the Social Security Administration also collect data, but the extent to which these agencies share data is limited. With regard to respondent burden, some information on pension plans is no longer collected out of concern that it was an unnecessary burden on the firms having to submit it, as well as concern about the Department of Labor's costs for storing the information. For example, the 1997 Taxpayer Relief Act (P.L. 105-34) ended the requirement that employers file with the Department of Labor copies of documents summarizing the features of the pension plans they offer. Finally, certain kinds of data needed to make projections are not widely available to all analysts because of the confidentiality laws that authorize their collection. The Census Bureau and others are exploring options for expanding access without compromising the confidentiality of the data. For example, the Census Bureau has established additional research data centers throughout the country where approved researchers with approved projects can work with confidential data and produce statistical summaries that meet strict disclosure requirements.

We are offering a Matter for Congressional Consideration and making recommendations to the Department of Labor and the Department of the Treasury that seek cost-effective approaches to help fill some of the data needs while taking into account respondent concerns about increased reporting burdens and agency concerns about maintaining confidentiality.

We provided a draft of this report to the Departments of Commerce, Labor, the Treasury and the Internal Revenue Service (IRS). We received technical comments from all four and incorporated their suggestions as appropriate. We also provided a draft of this report to the 11 members of our expert panel and modified the draft as appropriate in response to their comments. Commerce had no major comments on the report (see app. VI). Labor agreed on the need for access to accurate data but did not agree with our recommendations to the Secretary of Labor (see app. VII). Labor indicated that it did not have authority to require that summary plan descriptions (SPD) be filed electronically. Accordingly, we changed one of our draft recommendations to the Secretary of Labor into a Matter for Congressional Consideration. Labor also had concerns about the burdens our recommendations might pose. Although we acknowledge their concerns, nonetheless, we conclude that the need for improvements in retirement income data outweighs the likely costs involved. We therefore continue to address two recommendations to the Secretary of Labor and one to the IRS.

Background

Forecasting future retirement income needs—and how well they will be met through current savings, pension plans, and Social Security benefits—is a challenge, in part because of the many variables involved. Although Social Security is the primary source of income for many retirees, private pensions and other sources of income serve as key supplements and help retirees receive adequate income in retirement. In order to measure a person's current income and wealth, information is needed about many items, including the person's wage and nonwage income, home equity, pension, and nonfinancial assets and liabilities. In addition, to project a person's future income and wealth, researchers need information such as a person's earnings history, whether he or she chooses to participate in the pension plan offered by an employer, and how the person might respond to changing incentives for saving and investing for retirement. Other factors include whether the person chooses to accumulate savings apart from retirement plans, how long the person remains in one job, whether the person decides to cash out his or her retirement plan when changing employment, level of indebtedness, and the availability of health insurance during retirement. In addition, individuals' retirement funds depend on employer decisions, such as what kinds of pension plans and the availability and cost of retiree medical and long-term care insurance. To make estimates for people in different demographic groups, not just aggregate estimates for entire generations, analysts need to know how these factors vary based on individuals' demographic characteristics. Estimates of future income adequacy also rely heavily on projections of macroeconomic factors, including estimates of future rates of inflation, and rates of return on stocks and bonds and changes in home values. Furthermore, retirement income data needs keep changing, in part, as a result of trends in the pension industry and the labor force. The pension industry provides a growing variety of pension products with different features and legal structures. Forecasting pension income becomes more complicated with, for example, firms' converting of traditional pension plans into new pension hybrid products such as cash balance plans. These plans combine features of both defined benefit[2] and defined contribution[3]

---

[2]A defined benefit plan is a type of plan where the sponsor provides a guaranteed benefit generally expressed as monthly benefit based on a formula that generally combines salary and years of service to the company.

[3]A defined contribution plan is a type of pension that establishes individual accounts for employees to which the employer, participants, or both make periodic contributions. The benefits are based on employer and participant contributions to and investment returns (gains and losses) on the individual accounts.

plans, which adds to their complexity.[4] Changes in employment patterns, such as the decrease in the length of time employees spend in a single job, will also have an effect on pension income.

The federal government collects a great deal of data pertinent to the analysis of retirement income and wealth. Surveys of individuals and households collected by or sponsored by the Census Bureau, the Federal Reserve Board, BLS, and the National Institute on Aging are important sources of information. Similarly, surveys of businesses by BLS provide information about the pension and health care benefits firms offer their employees. A great deal more information pertinent to the analysis of retirement income and wealth is contained in the administrative documents that businesses and individuals must provide to state and federal agencies that administer governmental programs and enforce regulations. For example, many private employers must file, on a Form 5500, annual reports concerning their employee benefit plans for the IRS, Labor's EBSA, and the Pension Benefit Guaranty Corporation (PBGC).[5] Many pieces of information concerning pensions and retirement savings are also provided on forms that employers and financial institutions file with the IRS.

In addition to the federal government's collection of retirement income related data, several private entities also conduct surveys that provide useful information concerning retirement income. For example, the Health and Retirement Study (HRS) conducted by the University of Michigan's Institute for Social Research, is a longitudinal survey of adults over the age of 50. The information collected includes such topics as respondents' physical and mental health, insurance coverage, financial status, family support systems, labor market status, and retirement planning. The HRS is supported primarily by the National Institute on Aging, with additional funding over the years from the Social Security Administration, Labor's EBSA, the Department of Health and Human Services Assistant Secretary for Planning and Evaluation, and the state of Florida. Also, private benefit consulting firms and organizations, such as the Employee Benefit

---

[4]U.S. General Accounting Office, *Cash Balance Plans: Implications for Retirement Income*, GAO/HEHS-00-207 (Washington, D.C.: Sept. 29, 2000) and *Private Pensions: Implications of Conversions to Cash Balance Plans*, GAO/HEHS-00-185 (Washington, D.C.: Sept. 29, 2000).

[5]The IRS administers and enforces tax code provisions concerning private pension plans. EBSA enforces Employee Retirement Income Security Act (ERISA) pension requirements, and the PBGC insures the benefits of participants in defined benefit pension plans that are eligible for preferential tax treatment.

GAO-03-337 Retirement Income Data

Research Institute (EBRI), conduct surveys of employers concerning pension plans.

Retirement income analysts currently use available retirement income and wealth data to project future retirement income needs. For example, the Social Security Administration has developed a forecasting model, Modeling Income in the Near Term (MINT), to project demographic changes, retirement income, and Social Security benefits generally for persons born between 1931 and 1960. The Congressional Budget Office has also developed the Congressional Budget Office Long-Term model to study the range of possible outcomes for the balance of the Social Security trust fund.[6]

In a 1997 report, the National Research Council's Panel on Retirement Income Modeling reviewed the available sources of data on retirement income and recommended several improvements.[7] While the individual recommendations covered various topics, many involved the collection of additional information, as well as the establishment of an interagency task force for coordination purposes. While some of the panel's recommendations have been adopted, others have not. For example, as the panel recommended, the federal government has continued to support longitudinal studies. However, Labor has not acted on the panel's recommendation that it establish an interagency task force on employer data to plan collection of retirement income related data. In some respects, less data are available than was the case when the report was prepared. (See app. II for details.)

---

[6]For additional information on the Congressional Budget Office's model, see Congress of the United States, Congressional Budget Office, *Uncertainty in Social Security's Long-Term Finances: A Stochastic Analysis* (Washington, D.C.: Dec. 2001). The Social Security Administration and others such as EBRI have supported the development of other simulation models, such as the Social Security Policy Simulation Model (SSASIM) to study the effect of changes in the Social Security program and pension law. More recently the Social Security Administration and others have supported the development of the GEMINI model, a policy microsimulation model developed by the Policy Simulation Group. EBRI has developed its own model, the EBRI Retirement Income Projection Model.

[7]Constance F. Citro and Eric A. Hanushek, eds., *Assessing Policies for Retirement Income: Needs for Data, Research and Models* (Washington, D.C.: National Academy Press, 1997). The panel was sponsored by Labor's EBSA, the National Institute on Aging, PBGC, the Social Security Administration, and the Teachers Insurance and Annuity Association-College Retirement Equities Fund (TIAA-CREF) Institute.

Much of the data needed to assess retirement income and wealth are subject to federal laws protecting the confidentiality of information reported to the federal government. Laws limit the access to this information. Without access, it is not possible to pull together data from different sources to provide more complete information about individuals or organizations. The success of data gathering efforts by federal agencies and others relies on widespread trust that personal data will be kept confidential, protected from disclosure, and used only for specified purposes.[8]

Experts Cited Need for Better Data and Better Data Set Linkage

Experts we consulted cited priorities for improving retirement data that fit into two broad categories: (1) obtaining better data from employers on employee benefits and (2) obtaining better data from individual and household surveys by linking them with administrative data. The kinds of information concerning employers and employer-sponsored benefit plans that analysts sought included the features of their pension plans, such as minimum and maximum allowable contributions, or formulas for calculating benefits from defined benefit pension plans. Retirement income experts believed these data would allow them to more accurately measure or project retirement income and wealth now and in the future and estimate effects of potential retirement policy changes. Linkage of data from different sources creates new information by providing data about individuals matched to other data about the individuals from a second set of records, such as administrative records on pension plans. Analysts are able to use general data on some employer-sponsored pension plans and data on households and individuals. However, without linkages between these types of data it is difficult to obtain information about retirement offerings for specific households. As a result, analysis of pension offerings by demographic groups is limited.

Experts Place Priority on Improvement in Data from Employers

Most of the experts responding to our survey on retirement income data needs assigned better retirement income-related data from employers to the high or highest priority category. (See app. III for results of the survey.) In addition, the experts asserted that employers, rather than employees, could provide more accurate information about pension plans.

---

[8]Among other things, the *E-Government Act of 2002* (P.L. 103-347) permits the sharing of information concerning businesses among designated statistical agencies and provides for additional safeguards to protect the confidentiality of statistical data collected by all agencies.

Participants on our expert panel expressed interest in improving access to the mandatory pension plan disclosure reports, such as summary plan descriptions (SPD). (See app. IV for a summary of the expert panel's discussion.) Employers must provide these documents to employees who participate in a plan (such as a pension plan) in order to provide them with an understandably written description of their rights and benefits under the plan. To improve projections of future retirement income and the effects of policy options, panelists also expressed interest in data provided to the IRS on forms such as 1099-R and 5498.[9]

Employer Documents Are an Important Source of Information on Retirement Plans

Analysts responding to our survey, reported as their highest or high priority, the kind of information reported on some employer-submitted forms. Because pension plans vary widely, panel experts said they especially needed details of employee pension plans, such as the type of pension plan (defined benefit, defined contribution, or other), eligibility for participation and benefits, the plan's early retirement age, sources of contributions to the plan, and the method by which the amount of the contributions and benefits are calculated. Panelists believed this information would help analysts project future private pension benefits and the effects of proposed policy changes. Panel members also recommended that such pension data be obtained directly from employers, citing the need for accuracy as an important factor.

Inconsistencies have been found between employee and employer provided data. One study, for example, compared employees' reports concerning the employer-sponsored pension plans they were participating in, or that were available to them, with the information about those pension plans obtained directly from the employers.[10] It found significant discrepancies between the two sets of data, large enough for the authors to urge a great deal of caution in using the household survey data for reliable information about the actual characteristics of employer-sponsored pension plans. For example, among employees whose

---

[9]The IRS form titled "Distributions from Pensions, Annuities, Retirement or Profit-Sharing Plans, IRAs, Insurance Contracts, etc." is identified as IRS Form 1099-R. The IRS form titled "IRA and Coverdell ESA Contribution Information" is identified as IRS Form 5498.

[10]Alan L. Gustman and Thomas L. Steinmeier, *What People Don't Know About Their Pensions and Social Security: An Analysis Using Linked Data from the Health and Retirement Study* (Cambridge, Mass.: National Bureau of Economic Research, 1999). This paper discusses in detail some of the problems associated with having respondents provide details of their own pensions, as opposed to employers directly providing information.

employers reported that they were covered by a defined benefit plan, only 56 percent of the employees thought that they had such a plan. Likewise, there seemed to be high levels of error in other basic details about pension plans, such as the eligibility date for early retirement.

Using employer-submitted forms as a source of information was suggested as one way to increase the accuracy of pension data. Retirement income experts agreed that the Form 5500 is an important source of pension information available in government administrative records.[11] Sponsors or administrators of employee benefit plans subject to ERISA[12] must file this form annually. The Form 5500 was intended, in part, to measure employers' compliance with both the fiduciary and funding provisions laid out in ERISA legislation. The Form 5500 filed for pension plans contains useful aggregate information about plans. It provides information about the financial condition of the plan, annual amounts contributed by participants, and the plan's income on investments. The form also provides information on plan characteristics, such as plan type (defined benefit or defined contribution), method of funding, and numbers of employees, participants, and employees who are excluded from the plan for various reasons.

While the Form 5500 provides aggregate data about plans, it does not contain information useful for calculating individual's contributions or benefits. For example, it does not provide information on formulas for calculating employer contributions to plans or for calculating retirees' benefits.[13] In addition, it does not provide any data for pension plans outside the reporting requirements of ERISA, such as those for governmental employers, certain nonqualified plans for highly

---

[11]Form 5500 is a disclosure form that private employers with qualified pension plans are required to file with the IRS, Labor's EBSA, and the PBGC. Schedule SSA to the Form 5500, which is not publicly disclosable, identifies individuals who leave employment with deferred vested benefits. The Social Security Administration uses this information.

[12]The *ERISA of 1974* is a federal law that set minimum standards for pension plans sponsored by private employers. These standards govern the management, operation, and funding of the plan. Labor's EBSA enforces these ERISA provisions.

[13]U.S. Department of Labor. Pension and Welfare Benefits Administration, *Private Pension Plan Bulletin: Abstract of 1998 Form 5500 Annual Reports*, No. 11, Winter 2001-02.

GAO-03-337 Retirement Income Data

compensated people,[14] simplified employee pension (SEP) [15] plans, or savings incentive match plans for employees (SIMPLE).[16] Reporting and processing requirements also mean that data from Form 5500 reports may not be available to researchers for up to 3 years after the plan year. Finally, Labor officials find frequent errors in information submitted on the forms.[17]

Partly because of the limitations surrounding the Form 5500s, retirement income experts are increasingly interested in access to the SPDs, which are summaries of employers' pension offerings. The requirements of ERISA call for SPDs to include specific details about employee pension plans, including the type of pension plan, eligibility requirements, normal retirement age, vesting[18] and disqualification rules, sources of contributions to the plan, and method by which the amount of the contribution is calculated. ERISA required employers to file SPDs and documents, called "summaries of material modifications," describing changes to the plans with Labor. These were made publicly available at Labor's public disclosure room in Washington, D.C. The SPDs served many purposes: (1) they were a source of information to employees about the offerings included in their own pension plans, (2) they were a means of informing Labor about what types of plans a company was offering so Labor could perform monitoring and enforcement, (3) they also provided researchers with a high level of detail on pension offerings. EBSA officials noted, however, that the SPDs received by Labor were often out of date and that it was costly to store them. In 1993 we agreed that Labor should

---

[14]A qualified pension plan is an employer pension plan that receives preferential tax treatment in exchange for satisfying certain requirements established in the Internal Revenue Code of 1986. Employers or employees receive tax benefits for contributions they make to qualified plans within certain limits. A nonqualified pension plan is an employer-sponsored pension plan that does not meet these requirements.

[15]A SEP (402(h)) plan is a deferred compensation type retirement plan that allows employers and employees to make deductible contributions toward an employee's retirement fund. There are specific rules about contribution and deduction limits, which make the plan easier for a smaller employer to administer, but less attractive for a larger employer.

[16]A SIMPLE plan (401(k)(11)) is a deferred compensation type retirement plan that certain small employers (including self-employed individuals) can set up for the benefit of their employees.

[17]U.S. General Accounting Office, *Pension and Welfare Benefits Administration: Opportunities Exist for Improving Management of the Enforcement Program*, GAO-02-232 (Washington, D.C.: Mar. 3, 2002).

[18]Vesting refers to when a plan participant has earned a right to a benefit that cannot be taken away (i.e., a nonforfeitable right to the participant's accrued benefit).

stop collecting paper copies of SPDs, and instead provide access to electronic versions.[19]

Labor no longer collects SPDs and public access to them has become much more limited in the last 5 years. The Taxpayer Relief Act of 1997 (P.L. 105-34) amended ERISA so that employers no longer need to file SPDs with Labor. Instead, Labor was authorized to request SPDs from employers as needed, and uses this authority primarily to assist plan participants and beneficiaries in obtaining copies, though it has authority to request them for other purposes. However, since the law, Labor no longer requires that SPDs be filed, and SPDs prepared after the Taxpayer Relief Act are not publicly accessible, either for the general public or for researchers looking to model pension behavior.

Tax Information Returns Are a Source of Information on Retirement Plans

Our expert panelists noted that some of the pension details they need can be found in the administrative data provided by employers and others to the IRS. In addition to its collection of income tax returns, the IRS also collects "information" returns, such as the W-2, which contain details of employee information that provides valuable details relevant to future retirement income, such as wages, tax-deferred retirement contributions, lump sum distributions, rollovers, and retirement asset balances. In discussing pension data that the IRS has access to from its tax forms, experts from our panel reported that information from the Forms W-2, 5498, or 1099-R could provide important pension details. These forms provide detail on the extent of investments in retirement plans, such as the amount of contributions made to an Individual Retirement Account (IRA) or the amount of money contributed by the employee to deferred compensation plans.[20]

Form W-2 is a valuable source of pension data because it provides information on whether the employer provides some form of qualified retirement plan. The W-2 also includes amounts deducted from wages for contributions to pension plans, as well as codes that provide more detail on the different kinds of plans to which the contribution was made, such as whether the plan is a SIMPLE, SEP, or some other kind of deferred

---

[19] U.S. General Accounting Office, *Management Reform: GAO's Comments on the National Performance Review's Recommendations*, GAO/OCG-94-1 (Washington, D.C.: Dec. 3, 1993).

[20] IRAs authorized by ERISA allow workers to make tax-deductible and nondeductible contributions to an individual account for retirement savings.

compensation plan. Besides giving more detailed information on deferred compensation, the W-2 has another advantage in terms of the pension information it provides: employers are required to submit one for every employee. This requirement covers all workers for whom federal income tax or Social Security tax is withheld, including those who do not earn enough to be required to file individual income tax returns and those who are not covered by any pension plan.

Form 5498 is a form that financial institutions are required to file for all participants to report all contributions and the fair market value of their IRA accounts. It includes valuable pension information, including IRA contributions; rollover contributions;[21] and SEP, SIMPLE, and Roth IRA contributions.[22] Experts from the panel stated that the pension information from Form 5498 could be important in tracking an individual's retirement income balance with a specific custodial financial institution because it provides information on the fair market value of those assets held by the individual. While it does not give information on whether an employee who has separated from an employer converts the IRA into another tax protected IRA or pension account, this information should be reflected in the employee's federal income tax return.

Form 1099-R is another source of information for pension experts, which could provide additional information on pension resources. Form 1099-R is a statement filed by trustees concerning distributions to individuals from pensions, annuities, retirement or Profit-Sharing Plans, and IRAs. However, in many cases the form does not indicate whether the distributions are lump-sum distributions or rollovers into IRAs or other qualified plans.

The IRS makes some tax information publicly available through its Statistics of Income (SOI) program, which provides numerous tabulations and articles from its analysis of tax return data. From a sample of Form 1040s, the SOI currently provides aggregate tabulations of information, including taxable IRA distributions, deductible payments to an IRA, payments to a self-employed retirement (Keogh) plan, and taxable

---

[21]A rollover contribution is a direct transfer of pension benefits received as a lump-sum payment to another tax-qualified retirement plan or an IRA free of taxes. In many cases, however, the IRS forms do not indicate whether or not distributions were rolled over.

[22]A Roth IRA is a type of individual retirement plan that is similar to a traditional IRA except that contributions are not tax deductible, and that qualified distributions are tax free.

pensions and annuities.[23] This information is provided as a sum total of amounts accrued for the entire U.S. population that filed income tax returns, and is also broken out in detail by income level. For example, for the year 2000, the SOI provides taxable IRA distributions for about 9 million returns, with distributions totaling about $100 billion. While SOI breaks down these totals by income brackets, the tables do not provide other useful pension information such as pension accrual amounts by race or ethnicity. These demographic characteristics could be added to the data if individual tax return records were linked to the Census Bureau's detailed household survey records.

The SOI staff are preparing an article on retirement related data available from the IRS and, in doing so, will make some retirement related aggregate data, including information from Forms 5498 and W-2, publicly available for the first time. They told us that they are considering making this information available as a part of regular SOI releases, but they currently have no formal plans to do so. The SOI tabulations being prepared include the fair market value of pension plans, broken down by 10-year cohorts. Our expert panelists said that these kinds of aggregate data from IRS forms could help them ensure that their analyses reflect accurate information about retirement assets, such as the fair market value of IRA accounts.

Experts Place Priority on Linking Data on Individuals to Administrative Data

Given different possible options for improving retirement income data, retirement experts showed the greatest interest in increasing the availability of, and expanding researchers' access to, data sets on individuals or households linked to administrative data sets. (For characteristics of selected survey data sources, see table 3.) Eighty-one percent of respondents assigned this as a high or highest priority. Data linkage creates new information by matching data about individuals (using names or taxpayer identification numbers) to a second set of records, such as administrative records on pension plans, which provide additional information. The linked data are then preserved as a new data set, with personal identification information removed, that can generate new, fuller information on the population. Linking existing sources of data can provide detailed information with no additional respondent burden and at a much lower cost than is associated with collecting new survey data.

---

[23]Keogh plans are retirement plans for self-employed workers, authorized by the *Self-Employed Individuals Retirement Plan Act of 1962* (P.L. 87-792).

Our expert panelists made several suggestions for linking individual and household survey data with administrative data sources to help improve the analysis of retirement income and projections of the effects of policy changes on future retirement income. For example, they suggested expanding linkage between the Survey of Income and Program Participation (SIPP) and administrative data sources.[24] SIPP is a survey of households conducted by the Census Bureau providing wide-ranging demographic information, including different age cohorts, which makes it an attractive source of information when linked to administrative records. It provides, for example, information about benefits households receive from government programs. One component of this survey deals with retirement and pension plan coverage, in which several pension and retirement related questions are asked. However, it lacks definitive information on features of respondents' pension plans. Linking SIPP information to administrative data sets is a powerful way to expand information about how different groups will be covered in retirement. For example, linked data sets can indicate the extent to which the availability of various pension features differ for people in different age and demographic groups. Such linkage can also indicate how much knowledge respondents have about their pension plans and their retirement savings options. Panelists said that linkage to SPDs, Form 5500 data, and Social Security earnings and benefit histories would be especially valuable for projecting retirement income for different demographic groups.

Survey data on individuals have already been linked to administrative data sources in order to improve retirement income data. The Census Bureau is already linking SIPP records with administrative data related to retirement income, including Social Security earnings and benefit records. However, many potential linkages are hampered by lack of access to needed administrative data. For example, the University of Michigan Institute for Social Research has linked information gathered from HRS survey participants to SPDs gathered either from the employers or by Labor prior to 1997. Unfortunately, this process was made more difficult, and the information less satisfactory, because the HRS researchers could only obtain about 50 percent of the SPDs they were seeking, in part, because employers chose not to provide the SPDs. HRS researchers have also experimented with linking respondent surveys to Social Security earning

---

[24]The SIPP data have, for example, been linked with Social Security earnings records except in cases where respondents were unwilling to provide their Social Security numbers to the Census Bureau.

histories for two-thirds of the respondents who permitted SSA to provide the records.

Many Factors Limit Needed Retirement Income and Wealth Data

Experts we consulted believed that data needed for the study of retirement income are not collected or made available because of factors that include the fragmentation of data collection responsibility, the burden of data collection on respondents, and confidentiality considerations that restrict access. Fragmentation stems from no single agency having a statutory mandate to collect and analyze all the data needed to support a more comprehensive study of retirement income and wealth. Some information is no longer collected out of concern that it was an unnecessary burden on the firms having to submit it and because it was only being used to a limited extent by the government. Finally, certain kinds of data needed to make retirement income projections are not made available because they contain information that must by law be carefully protected against unauthorized disclosure and misuse.

Fragmented Retirement Data Responsibilities Contributes to Data Shortcomings

Although many federal agencies are involved in collecting and analyzing retirement income and wealth data for a variety of different purposes, no single agency is responsible for these activities. Instead, some agencies—including the Bureau of the Census, Labor, Federal Reserve Board, IRS, and SSA—collect data needed for their specific purposes. For example, the Census Bureau collects retirement income and wealth information on individuals and households using the Current Population Survey (CPS) and SIPP. Labor collects data on pensions for both ERISA enforcement purposes and to track pensions for statistical purposes. The IRS collects data from individuals and firms for tax enforcement purposes. Table 1 shows the major agencies involved in retirement income and wealth data collection, the purpose of their data collection, and a brief description of the information collected. None of these agencies is charged with coordinating all retirement income data collection efforts or planning improvements needed in data collection and analysis.

Table 1: Examples of Federal Agency Retirement Income-Related Data Collection

Agency	Data collection program	Purpose of data collection	Examples of data collected
Census Bureau (Department of Commerce)	Survey of Income and Program Participation (SIPP)	Provide information about income and government program participation	-Work experience -Earnings -Program participation -Benefits received -Property Income -Demographic characteristics
Bureau of Labor Statistics (Department of Labor)	National Compensation Survey	Provide information on wages salaries and benefits	-Occupational earnings -Compensation trends -Benefits offered -Benefit participation -Detailed plan provisions
Employee Benefits Security Administration (Department of Labor)	Form 5500	Enforce ERISA pension requirements	-Type of plan -Number of employees -Number of participants -Plan amendments -Plan financial position -Actuarial assumptions -Employer & employee contributions
Social Security Administration	SSA earnings and benefit records	Administer the Social Security benefit programs	-Earnings histories -Social Security benefit histories -Supplemental Security Income benefit histories
Internal Revenue Service (Department of the Treasury)	IRS tax records	Administer/enforce the tax code	-IRS individual 1040 returns -IRS Information returns -Account balances -Withdrawals

Source: GAO's analysis of data from the U .S. Census Bureau, the U.S. Department of Labor, Social Security Administration, and the U.S. Treasury Department.

Agencies involved in the analysis of retirement income and wealth data often limit their analysis to a portion of the retirement income and wealth information. Panel members noted that many of the agencies place little or no priority on a comprehensive analysis of retirement income and wealth data. Instead, agencies focus their data collection and analysis on data needed to address their mission. For example, Labor's EBSA collects information on compliance with ERISA regulations, including Form 5500 submissions. Although EBSA's strategic plan includes retirement income data analysis efforts, its principal focus is on enforcement of ERISA. Similarly, the Bureau of the Census collects retirement related information in the SIPP and other surveys but its analysis of this information is primarily in the context of its income and poverty measurement mission. The SSA's MINT model uses an estimate of workers' pensions based on SIPP data from the Census Bureau, which generally relies on survey responses, has been criticized for using inaccurate estimates of

nonearnings income, including retirement income. Because the Census Bureau data provided to the SSA for the MINT was not gathered with it in mind or coordinated for its use, the MINT model has had to use simplifications and assumptions of these data which makes its final model less useful for policy makers.[25]

Despite this fragmentation, some agencies have attempted to increase communication between federal agencies concerning data collection efforts related to retirement income. For example, the BLS' strategic goals include improving the accuracy, efficiency, and relevancy of U.S. economic statistics and enhancing coordination with other agencies. The Census Bureau's strategic goals likewise include an emphasis on providing accurate, timely, and accessible information on the U.S. population and economy, and to maintain relationships with agencies compiling administrative record data. Both the Census Bureau and the BLS are members of the Federal Interagency Forum on Aging Related Statistics,[26] which has a goal of improving both the quality and the use of aging related data. In addition, BLS is one of three agencies that share responsibility for leading the Inter-Departmental Committee on Employment-Related Health Insurance Surveys. This committee of a dozen members was created in 1998 to improve coordination and reduce respondent burden by reducing redundant requests for information.

Respondent Burden Considerations Limit Available Data

Members of our expert panel reported that respondent burden, as well as requirements set up to limit respondent burden, could hinder agencies' efforts to obtain retirement-related information. Panel members noted that answering survey questions about retirement income and wealth could be a substantial burden on respondents. They acknowledged that asking for too much information could result in partial responses, erroneous responses, or, in some cases, a reduction in the overall response rate.

---

[25]SSA statistically matched defined benefit pension plan characteristics from the Pension Benefit Guarantee Corporation to the survey responses.

[26]The Federal Interagency Forum on Aging Related Statistics was established in 1986, with the goal of bringing together federal agencies that share a common interest in improving aging related data. Member agencies include: National Institute of Aging, National Center for Health Statistics, Census Bureau, Administration on Aging, Agency for Healthcare Research and Quality, BLS, Centers for Medicare and Medicaid Services, Department of Veterans Affairs, Office of Management and Budget (OMB), Office of the Assistant Secretary for Planning and Evaluation in the Health and Human Services Administration, and SSA.

However, efforts to reduce respondent burden may also limit the collection of retirement information. Legislation requires OMB to review surveys before they are used to collect data. The Paperwork Reduction Act of 1995 (P.L. 104-13) and similar previous legislation[27] is designed to minimize the paperwork burden on the public while at the same time recognize the importance of information to the successful completion of agency missions. The act requires OMB to approve all existing and new collections of information by federal agencies. In approving agency collection efforts, OMB must weigh the burden to the public against the practical utility of the information to the agency. Panel members noted that agencies were reluctant to propose additional data collection unless they could clearly establish that the benefit outweighed the perceived burden. Panel members noted that efforts to reduce existing data collection requirements sometimes result in a loss of information. For example, the 1997 elimination of the requirement that firms routinely submit SPDs was connected to an effort to reduce the respondent burden on employers.

Privacy and Confidentiality Concerns Limit the Collection and Use of Retirement Data

While restrictions on the collection and use of retirement data are critical for the protection of personal information, some panel members noted that these restrictions also limit the availability of such information. Several laws exist to protect individuals' rights to privacy and protect the confidentiality of personal and proprietary information held by federal agencies. For example, laws set strict requirements to protect data collected by the Census Bureau and to limit the use of taxpayer data.[28]

Implementing such laws requires federal agencies to restrict access to data they collect. For example, the Census Bureau's data set that links the SIPP with earnings and benefit records from SSA is not available to the public. Protecting the confidentiality of such linked data sets is particularly crucial because linked data sets may be more detailed or more sensitive than the component data sets were before they were linked. Agency officials must remove personal identifiers such as Social Security numbers, names, and addresses. Even without these personal identifiers,

---

[27]Previous legislation includes the *Paperwork Reduction Act of 1980* (P.L. 96-511) and the *Paperwork Reduction Reauthorization Act of 1986* (P.L. 99-500).

[28]See 13 U.S.C. 9 and 26 U.S.C. 6103. An exception in 26 U.S.C. 6103(j) authorizes the furnishing of return information to Census "for the purpose, but only to the extent necessary in the structuring, of censuses and…conducting related statistical activities authorized by law."

as more and more information is linked, the risks that individuals could be identified increases. The computer files for these nonpublic data sets are available only at a limited number of secure research data centers to approved analysts working on approved projects.[29] Outside analysts working with these data sets must be sworn Census Bureau officers and their work must serve, at least in part, to support the Census Bureau's mission. If analysts permitted to use these data combine any other data with the restricted data, the combined data are subject to the Title 13 protections. Analysts are not allowed to make copies of the data or remove data from the secure data center.[30] Before taking any of the results of their work from the center, Census Bureau staff must review the results to ensure that they meet the agency's requirements for confidentiality protection. Thus, the results that can be taken out of the center are limited to statistical results that do not disclose data for specific individuals. Also, the external researchers must agree that the results of their work will be available in the public domain and not maintained as proprietary information.

However, despite the government's efforts to protect the information they collect from misuse, surveys of the public and Census Bureau interviewers indicate that people are apprehensive about the government's use of personal information. In a survey conducted both before and after (or just prior to) the 2000 Census of Population and Housing, about half of the respondents (51 percent before and 50 percent during) indicated they thought the Census Bureau's promise of confidentiality could be trusted, down from about 79 percent in 1990. A smaller, but still substantial proportion of the workers conducting census interviews and providing those promises to respondents also indicated a lack of trust in the Census Bureau's assurances. A 1998 study indicated that 16.7 percent of Census Bureau interviewers and 32.2 percent of non-Census Bureau interviewers believed that the Bureau would give individual survey data to government agencies such as the Federal Bureau of Investigation, the Central Intelligence Agency, the Immigration and Naturalization Service, and the IRS. Federal law prohibits the Census Bureau from sharing information about individuals with these agencies. Census Bureau information about

---

[29]The Census Bureau's research data centers are located in Washington, D.C.; Boston, Massachusetts; Pittsburgh, Pennsylvania; Los Angeles, California; Berkeley, California; Durham, North Carolina; Ann Arbor, Michigan; and Chicago, Illinois.

[30]Individuals with access are subject to penalties, including fine and imprisonment if they disclose any confidential information.

GAO-03-337 Retirement Income Data

particular individuals or businesses is only available for statistical purposes, not for law enforcement purposes.

Public distrust of federal agencies' use of their personal information can undermine people's willingness to participate in federal surveys, potentially making the information collected less reliable.[31] An article in the Journal of Official Statistics noted that in the 1990s the rate at which people refused to participate had risen for six surveys conducted by the Census Bureau.[32] For questions about types of income in the March 1999 CPS survey, for example, the percentage of respondents providing data for particular items ranged from 33 percent to 78 percent. Low response rates can undermine analysts' statistical projections if the individuals who choose not to respond differ in important respects from those who provide data. If the remaining respondents are dissimilar to the population being surveyed, conclusions about the population may not be reliable. For example, if those who chose not to respond have higher incomes, estimates of the populations' income may not be reliable. Statisticians make adjustments that can mitigate this problem, but the lower response rates are, the more uncertainties remain.

Federal agencies and researchers continue to explore options to maximize data usefulness without compromising respondent privacy and confidentiality.[33] For example, the Census Bureau has received permission from the IRS to link the survey records to additional items from IRS and SSA records.[34] In addition, the Census Bureau has recently increased the

---

[31]See Eleanor Singer's study "Public Perceptions of Confidentiality and Attitudes Toward Data Sharing By Federal Agencies" in *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*, Pat Doyle, et al., (Amsterdam: Elsevier Science B.V., 2001). In some cases after controlling for other factors associated with response rates, such as respondents' age, the length of the form, race, and education, privacy concerns were not significant predictors of response rates.

[32]B.K. Atrostic et al., "Nonresponse in U.S. Government Household Surveys: Consistent Measures, Recent Trends, and New Insights," *Journal of Official Statistics*, vol. 17 no. 2, 2001, 209-226.

[33]We have discussed options for protecting privacy and confidentiality while conducting record linkage in U.S. General Accounting Office, *Record Linkage and Privacy: Issues in Creating New Federal Research and Statistical Information*, GAO-01-126SP (Washington, D.C.: April 2001).

[34]26 CFR Part 301, *Federal Register* vol. 68, no. 13, January 21, 2003, p. 2691.The IRS shares responsibility with SSA for protecting the confidentiality of Social Security earnings records compiled from W-2 forms submitted to the IRS.

GAO-03-337 Retirement Income Data

number of its research data centers where approved analysts working on approved projects can access confidential data.

Researchers continue to explore statistical techniques for providing more information from survey data sets while reducing the risks that confidential information will be compromised. One well-known technique for doing this, "top coding," involves reporting in data files only that an individual respondent's income, for example, exceeds the "top code" amount, not the actual value. In this way, so many individuals are included in the high-income group that their identities cannot be determined. More sophisticated techniques include the use of methods to substitute artificial records containing estimated values based on knowledge of the real data. These simulated subjects are assigned number values selected to ensure that relationships between important variables are preserved. This allows people to remain anonymous to the researchers using the data. However, pension experts state that this method can only accommodate certain kinds of variables and, therefore, complex relationships between variables may not be maintained. [35]

Recent legislation may enable the Census Bureau and BLS to link their data on businesses for statistical purposes. In December 2002, the Congress enacted the Confidential Information Protection and Statistical Efficiency Act of 2002 as part of the E-Government Act of 2002 (P.L. 107-347). This act permits designated statistical agencies to share information concerning businesses for statistical purposes, but not information concerning individuals or households. It authorizes three agencies—the Census Bureau, the Bureau of Economic Analysis, and BLS—to share data on businesses with one another for statistical purposes.

Conclusions

With the proportion of retirees to workers expected to increase dramatically over the next couple of decades, important decisions lie ahead. Access to retirement income data needed to inform those decisions has actually decreased in some respects, despite the recommendations of

---

[35]Critics of these efforts say that these techniques cannot preserve all the relationships between variables in a data set. Moreover, the techniques are not workable for variables, such as the age of retirement, that do not conform to a simple mathematical pattern. The frequency at which people retire, for example, is spiked at certain ages, such as 60, 62, and 65 years of age, which makes it difficult to summarize the data using a statistical formula. One of our panelists noted, however, that if only one variable has such characteristics, the actual data for that variable could be left as long as other variables were masked.

the 1997 National Research Council's Panel on Retirement Income Modeling. Although many sources of useful retirement income data remain, retirement analysts we consulted cited shortcomings. They noted, for example, that data shortcomings persist when no one federal agency is responsible for coordinating efforts to fill retirement income data needs.

Indeed, several such data needs could be met with information the federal government already possesses or to which it already has access. For example, Labor has the authority to obtain existing documents describing the features of private pension plans. However, though it has this authority, it gives employers in its National Compensation Survey a choice about whether to provide them to support Labor's BLS statistical studies of pension plans. To encourage voluntary participation in the survey, Labor does not make them available to other agencies, outside analysts or the general public. The Census Bureau gathers or collects information about some general features of private pension plans through surveys of households and individuals but has not yet had the opportunity to corroborate and supplement this data using information from respondents' employers available through Labor. While information on pension and individual retirement accounts is gathered through forms collected by the IRS, the data are not regularly tabulated or linked to survey data, and thus, are not available for the study of pensions.

While improvements in data are essential for more reliable forecasts of retirement income, protecting respondents' information and minimizing the burden data collection efforts impose on firms and individuals are also crucial. To sustain programs for compiling statistics about firms and individuals, respondents must be able to trust that their personal information will not be misused. Finding an appropriate balance between providing wider access to data to support policy analysis and keeping data secure is a persistent and evolving challenge that policymakers must continually address. In addition, federal agencies need to consider both the federal cost of these efforts and the financial and nonfinancial costs imposed on respondents in comparison with the benefits expected from improvements in data. While taking into account these cost considerations, respondent concerns about increased reporting burden, and agency concerns about maintaining confidentiality, the Congress, and the Departments of Labor and the Treasury could take the next steps to help fill some of the data needs.

Matter for Congressional Consideration

To facilitate plan participants', beneficiaries', and analysts' timely access to information about employer-provided pension plans, the Congress should consider directing the Secretary of Labor to obtain from plan administrators electronic filings of all SPDs and summaries of material modifications required by ERISA and make them publicly available in electronic form.

Recommendations

To help provide the data needed to inform important policy decisions concerning retirement programs, the Secretary of Labor should direct the Bureau of Labor Statistics to prepare a plan to improve data for analyzing retirement income and wealth in coordination with OMB, the Federal Reserve Board, IRS, and the agencies represented in the Federal Interagency Forum on Aging Related Statistics, including the Census Bureau, SSA, and the National Institute on Aging. The plan should include cost-effective strategies to

- make better use of existing statistics by linking survey and administrative data,
- improve access to linked data consistent with privacy and confidentiality legislation, and
- improve data collected from employers related to retirement income and wealth.

For plans in place before these new filing requirements go into effect, and where it is cost-effective, the Secretary of Labor should use existing authority to obtain copies of summary plan descriptions and summaries of material modifications in cases where analysts working on federally conducted or federally sponsored research seek SPDs for statistical purposes. This should assist analysts of retirement income in obtaining information about the features of employer-sponsored benefit plans that are not electronically available.

To help analysts improve analyses of retirement plan finances, the Internal Revenue Service should publish on an annual basis aggregate tabulations, such as those prepared in the Statistics of Income Bulletin, of information filed on IRS Forms 5498, 1099-R, and W-2.

Agency Comments

We provided a draft of this report to the Departments of Commerce, Labor, the Treasury, and the IRS and received technical comments from all four. In response we modified the draft as appropriate. We also provided a draft of this report to the 11 members of our expert panel and modified the

draft as appropriate in response to their comments. Commerce had no major comments on the report (see app. VI). Labor agreed on the need for access to accurate data but did not agree with our recommendations to the Secretary of Labor (see app. VII).

Regarding our recommendation to collect electronic copies of SPDs, Labor concluded that this is at odds with the Taxpayer Relief Act of 1997, which eliminated requirements that SPDs be regularly filed with Labor. Although Labor still has authority to request SPDs, it had indicated in final regulations concerning SPDs that it generally intended to limit the exercise of its authority to requesting SPDs on behalf of participants and beneficiaries. Labor also indicated that there was little public interest in the SPDs. Although there was not substantial public demand for paper copies of SPD's in Labor's Public Disclosure Room in Washington, D.C., in the last decade there has been a great increase in policymakers' demand for better data on retirement income in light of the rapid future increase in the retiree population. Although in both 1993 and 1995 we supported elimination of paper filing requirements, we also endorsed electronic access to SPDs. We continue to believe that it is time to phase in a requirement that SPDs be filed electronically. The costs involved should be considerably less than those incurred filing and storing paper copies of the documents. The burden imposed on plan administrators would not be unreasonable. Labor estimated in 1998, for example, that it would cost an average of $1.55 to provide SPDs for health benefit plans to each plan participant, of which $1.00 was the estimated cost of materials and postage.[36] Cognizant of the shortcomings of many SPDs, experts we consulted nonetheless indicated that better access to SPDs is a top priority for improving retirement income data. Accordingly, we changed our recommendation to a Matter for Consideration for the Congress to direct the Secretary of Labor to obtain from plan administrators electronic filings of all SPDs and summaries of material modifications required by ERISA and make them publicly available in electronic form.

Regarding our recommendation that BLS prepare a plan to improve data for analyzing retirement income and wealth, Labor indicated that past efforts to coordinate improvements in retirement income data have not been successful due to privacy concerns, and other issues. They stated that the type of planning and coordination we envisioned usually is the

---

[36]U.S. Department of Labor, "Proposed Amendments to Summary Plan Description Regulations," *Federal Register*, vol. 63, no. 174, September 9, 1998, p. 48384.

purview of OMB's Office of Statistical Policy, and that BLS could participate in that coordination. Furthermore, the extra demands placed on staff would not be negligible. In our view, the need for improvements in retirement income data warrants renewed efforts to address the priorities identified by the experts we consulted. The recommendations of these experts focused primarily on improved use of existing data to support policy analysis, and not on additional data collection. We recognize OMB's role overseeing information collection and developing policies to improve government statistics. However, this does not preclude efforts by other agencies to take the lead in developing plans for improvements focused on data within specific subject areas such as retirement income. Because OMB does not have the retirement income focus needed to coordinate in this way, we have retained our original recommendation. That recommendation specifically identifies OMB as one of the agencies that should be involved in the development of a plan, and the plan should be developed in a manner consistent with OMB's policy.

With respect to our recommendation regarding provision of SPDs before new electronic filing requirements go into effect, Labor stated that the need to protect the confidentiality of survey data may hamper wider access to SPDs. BLS, for example, assures respondents to its National Compensation Survey that their identities will be kept confidential. The Secretary's authority to request SPDs is delegated to EBSA. If BLS were to obtain SPDs from EBSA it would have to reveal the identity of its respondents and therefore would have to obtain their consent. We conclude that the need for improvements in retirement income data warrants Labor's use of its existing authority to obtain SPDs for analysts engaged in federally conducted or federally sponsored research. Arrangements could be developed through which BLS and other statistical agencies could both obtain SPDs and protect the identity of respondents. They could, for example, request SPDs from a larger number of employers without identifying which employers were being surveyed. This is the kind of improved access to data that we envisioned BLS could take the lead in identifying in coordination with other agencies.

We are sending copies of this report to the Secretary of Labor, the Secretary of the Treasury, the Secretary of Commerce, and the Commissioner of the Internal Revenue Service. We will also make copies available to others on request. In addition, the report will be available at no charge on GAO's Web site at http://www.gao.gov/.

If you have any questions concerning this report, please contact Barbara Bovbjerg at (202) 512-7215, Robert Parker at (202) 512-9750. See appendix VIII for other contacts and staff acknowledgments.

Sincerely yours,

Barbara D. Bovbjerg
Director, Education, Workforce
  and Income Security Issues

Robert P. Parker
Chief Statistician

Appendix I: Scope and Methodology

To identify information that experts say is a priority for improving data for the study of retirement income and wealth, we conducted a Web based survey of experts in the field and convened an 11 member panel of experts to discuss opportunities for improving these data. We used the Web based survey instrument to survey 276 experts in retirement income data.[1] Before implementing the survey, we contacted respondents via email and asked them to participate. Out of our initial list of 326 experts, 27 declined to participate, with the majority citing busy schedules or lack of sufficient expertise as their reasons. In addition, we concluded that we had inaccurate or out-of-date E-mail addresses for 23 of the experts.

We studied available research and interviewed experts in order to develop a questionnaire of options to improve retirement income. The questionnaire asked respondents to indicate the priority (from highest to lowest) they would place on 22 actions to improve retirement income data. Respondents were asked to rate each action independently, without making comparisons between the actions. We pretested the content and format of the questionnaire with 6 experts in the area of retirement income. The questionnaire was then refined and posted on our Web site and an E-mail message informed participants of its availability. This E-mail message also contained a unique user name and password that allowed each respondent to log on and fill out his or her own questionnaire.

As of December 12, 2002, 190 of the experts responded to the survey (a 69% response rate). Eighteen percent of respondents indicated that they were affiliated with federal agencies, about half were affiliated with colleges and universities, 24 percent were affiliated with other nonprofit organizations, and the remaining 9 percent were affiliated with for profit or other organizations.

Our preliminary results of the survey identified two areas for improvement that respondents most often cited as having the highest priority: (1) matching survey and administrative data and (2) employer data. We used these areas to serve as the principal topics at a 1-day expert panel meeting at our headquarters on September 10, 2002. The 11 panelists included 5 federal officials with responsibilities related to retirement income data, 3 university based analysts, and 3 from private not-for-profit agencies.

---

[1]We identified these people through literature searches on topics related to retirement income data and by asking members of our Retirement Advisory Panel for suggested names and in turn asking them for additional names.

Barbara Bovbjerg, Director, Education, Workforce, and Income Security Issues, and Robert Parker, Chief Statistician, moderated the discussion. (For a summary of the panel's discussion and a list of panelists, see app. IV.)

To identify factors limiting the availability of information needed for the study of retirement income and wealth, we reviewed documents obtained from and interviewed officials at the Department of Labor's (Labor) Employee Benefit Security Administration (EBSA), and Bureau of Labor Statistics (BLS), the Census Bureau, the Treasury Department's Office of Tax Analysis, the Internal Revenue Service's (IRS) Statistics of Income Division, the National Institute of Health's National Institute on Aging, and the Social Security Administration's (SSA) Office of Research, Evaluation, and Statistics. We also attended conferences related to retirement income analysis sponsored by the Retirement Research Consortium and the Society of Actuaries and interviewed analysts at the Urban Institute, The Brookings Institution, the Employee Benefit Research Institute, and the National Research Council.

The scope of our work did not include an evaluation of estimated costs and benefits of specific proposals for improving retirement income data. We did not independently verify the federal funding figures provided to us by longitudinal survey administrators or agencies sponsoring the surveys.

Appendix II: Status of Recommendations From the 1997 Report of the Panel on Retirement Income Modeling

Below are recommendations concerning retirement income data needs excerpted from the 1997 report of the National Research Council's Panel on Retirement Income Modeling followed by summaries on the status of each as of December 2002.[1]

1. Continue support of longitudinal studies

Recommendation: Existing panel surveys of middle-aged and older people should receive continued government support. Longitudinal data from these surveys are essential to analyze retirement and savings decisions and determine behavioral responses to changes in public and private sector policies. Such analyses in turn are essential to develop better models for forecasting the likely effects of alternative policy proposals on retirement income security. In particular, the Health and Retirement Study (HRS) and Asset and Home Dynamics Among the Oldest Old (AHEAD) surveys should receive continued support. These surveys should be refreshed periodically with new cohorts in order to offer insight into how behavior changes over time.

Status: As shown in table 2, the amounts of federal support for three major longitudinal surveys have been sustained. The HRS and AHEAD studies, which are now jointly funded, have increased after taking the effects of inflation in account.

Table 2: Federal Outlays for Selected Longitudinal Studies—Fiscal Years 1997-2001

Millions of constant fiscal year 2001 dollars[a]

	1997	1998	1999	2000	2001	2002
HRS and AHEAD	$6.8	$7.0	$7.4	$9.0	$9.4	$10.1
National Longitudinal Studies[b]	14.0	11.8	13.2	12.8	12.8	12.4
Panel Study of Income Dynamics[c]	2.5	2.5	3.7	3.7	3.6	2.8
Total	**$23.3**	**$21.3**	**$24.3**	**$25.5**	**$25.8**	**$25.3**

Source: GAO analysis of data from the National Institute of Aging and the University of Michigan's Institute for Social Research.

[a]These figures are adjusted for inflation using the Bureau of Economic Analysis's gross domestic product price index.

[b]NLS

[c]PSID

---

[1]Constance F. Citro and Eric A. Hanushek, eds., *Assessing Policies for Retirement Income: Needs for Data, Research and Models* (Washington, D.C.: National Academy Press, 1997).

The National Institute on Aging continues to fund the HRS and AHEAD studies. Supplemental funding also comes from SSA. Both the AHEAD and HRS studies have been refreshed with new cohorts. In 1998, interviews began for a cohort of people born from 1924 to 1930 and a cohort of people born from 1942 to 1947.

The National Science Foundation continues to fund the PSID. Additional support comes from the National Institute on Aging, the National Institute on Child Health and Human Development, and the Departments of Health and Human Services, Agriculture, Housing and Urban Development (HUD), and Labor.

Labor sponsors the National Longitudinal Survey (NLS) through BLS. The NLS is conducted under contract with researchers at Ohio State University, the University of Chicago, the Census Bureau and the University of Wisconsin. In addition to Labor funding, financial support has come from agencies including the National Institute of Child Health and Human Development, the National Institute on Aging, the National Institute on Drug Abuse, the National Institute on Alcohol Abuse and Alcoholism, and the Departments of Defense, Justice, and Education.

2. Measure expenditures

Recommendation: Panel surveys of middle-aged and older people should experiment with methods to develop measures of families' total expenditures and expenditures on housing and medical care. Such consumption measures are important for projections of economic well being in retirement.

Status: The HRS and AHEAD surveys provide on a longitudinal basis respondents' estimates of many categories of expenditures including housing expenditures, total out-of-pocket medical expenditures, and total expenditures, total expenditures relative to income, as well as information on savings preferences. Labor's Consumer Expenditure Survey provides cross-sectional rather than longitudinal data on many types of expenditures, including housing expenditures, and medical expenditures, and total expenditures and related income. The American Housing Survey, sponsored by HUD, provides detail on housing expenditures . The proposed American Community Survey also would provide limited detail on housing expenditures. The Medical Expenditure Panel Survey (MEPS) provides extensive data on medical expenditures.

3. Gather more data on younger people

Recommendation: Panel surveys of younger people, such as the National Longitudinal Survey of Youth (NLSY), should include detailed questionnaire modules on pension coverage, wealth, health status, and retirement- and savings-related expectations. Such information is needed to fully understand life-cycle behavior and to track the disparities in income and wealth that are evident by middle age.

Status: The NLSY asks about the total amount of retirement savings, amounts contributed, and amounts withdrawn, and pension coverage. It also provides information about assets and debt and limited information about health (height, weight, and general evaluation of health), but not retirement or savings expectations.

In 1995-99 interviews, the National Longitudinal Survey of Young Women (NLSYW) asked respondents the extent to which they agreed or disagree with statements such as "Work is the most meaningful part of life" and "People who don't retire when they can afford to are foolish." The NLSYW also asked respondents about the availability of a retirement or pension plans in 1978 and in each round from 1978, 1983, through 1999. The 1991 and 1995-99 rounds of the survey included more detailed questions on amounts in retirement plans, formula for calculating benefits for defined benefit plans provided by current and previous employers, vesting status, and expectations about retirement, such as expected amounts of benefits.

4. Collaborate to improve data quality and utility

Recommendation: Agencies and researchers involved in retirement income-related panel surveys of individuals, and other surveys as appropriate (such as the Survey of Consumer Finances (SCF) and the Survey of Income and Program Participation (SIPP)), should collaborate regularly in reviewing questionnaire content and data collection practices to identify ways to improve data quality and utility. For example, the bracketing technique used in HRS and AHEAD that has been demonstrated to reduce nonresponse to important items should be adopted in other surveys. Also, such surveys might include a common core of questions on specific topics. The National Institute on Aging should facilitate such collaborative efforts.

Status: Federal agencies collaborate through entities such as (1) the Federal Interagency Forum on Aging-Related Statistics, originally established by the National Institute on Aging, National Center for Health

GAO-03-337 Retirement Income Data

Statistics, and Census Bureau and (2) the Inter-Departmental Committee on Employment-Related Health Insurance Surveys, headed by the BLS, National Center for Health Statistics, and the Agency for Healthcare Research and Quality. The Federal Committee on Statistical Methodology and the Interagency Committee on Confidentiality and Data Access have coordinated research on techniques such as the bracketing technique are now widely used and researchers do collaborate on questionnaire design. In addition, the Census Bureau and other statistical agencies seek comments from a wide range of researchers on the content of questions before fielding survey instruments. For example, the Interagency Committee on the SIPP, which consists of representatives of interested federal agencies, and the American Statistical Association's Survey Research Methods Section SIPP Working Group, which consists of federal and non-federal analysts, both advise the Census Bureau on that survey.

5. Establish interagency task force on employer data

Recommendation: Labor should establish an interagency task force on employer data to specify an integrated plan for collecting retirement income-related information. The plan should specify short-term and long-term goals that consider user needs, resource constraints, and the problems of obtaining information from employers due to such factors as low response rates, locating the appropriate respondents, and confidentiality concerns. The task force should involve researchers, private benefit consultants, and representatives of public and private employers in its work.

Status: According to officials at the Department's EBSA, Labor has not made this a priority because its resources are limited.

6. Gather more data on benefit plan offerings

Recommendation: The employer data collection plan should include short-term and long-term goals for obtaining improved information on the distribution across employers of all benefit plan offerings (including pensions, health insurance, disability insurance, retiree health insurance, life insurance). Comprehensive baseline information is a priority need, along with a plan for regular updating. Needed data elements include benefit plan characteristics and costs, employer characteristics (e.g., number of employees, financial characteristics, wage structure), and workforce characteristics (e.g., age structure) for public and private employers and the self-employed.

Status: Labor's BLS has increased the amount of information gathered through its National Compensation Survey, which includes components that had been gathered through the Employee Benefits Survey, the Employment Cost Index survey, and the Occupational Compensation Survey. The Employee Benefits Survey did not provide tabulations of the estimated percentage of establishments providing specified types of benefits, but the National Compensation Survey does. The survey includes components for state and local government employers, medium and large private employers, and small private employers, but not federal employers or the self-employed. The survey provides data on number of employees, employer and employee costs, wage structure, and the characteristics of retirement benefit plans. It does provide data on the employer's cost of providing defined benefit and defined contribution retirement programs. It does not provide information on financial characteristics of plans (such as assets and liabilities) and does not provide information on the age structure of the workforce.

In addition, the MEPS, sponsored by the Agency for Healthcare Research and Quality, provides additional data on health insurance plans.

7. Redesign and enhance employee benefits survey.

Recommendation: The employer data collection task force should give priority to redesigning and enhancing existing data collection systems on employer benefit offerings and related topics. Such systems include the Employee Benefits Survey, which currently provides information for broad categories of employe[rs] but not for employe[es], and the Form 5500 data series, which serves regulatory purposes and currently has limited research use. Consideration should be given to improving the Form 5500 series, including:

- making the data more timely and accessible (e.g., on-line);
- linking records over time to provide panel data;
- merging the Form 5500 benefit plan information with the kind of employer financial characteristics found in the Compustat database;
- working to standardize the reporting for health care and disability plans, so that they can be added to the Form 5500 database; and
- finding ways to add information about benefit plan features to the database, perhaps by abstracting analytically useful information from the narrative plan descriptions that are filed with the Form 5500.

Status: The Employee Benefits Survey (now part of the National Compensation Survey) continues to provide data for categories of employers, not categories of employees. It provides data by employer

industry group, employer employment size group, employee union status, employee occupational group, and employer geographic location. BLS has begun providing aggregate estimates for all private employers, rather than only providing data separately for small private employers, and large private employers, as it did in the past. However, the survey does not cover federal, military, agricultural, fishing, forestry, or private household employers.

Data from IRS Form 5500 continues to become available well after the end of the reporting year. (Data for 1998 or the 1998-99 reporting year became available in 2002.) In part, this is because the IRS deadline for submitting the forms is 7 months after the end of the reporting year. For example, for a firm with a 1998 reporting year beginning December 15, 1998, and ending December 14, 1999, the Form 5500 was due July 31, 2000. In addition, Labor takes several months to review and edit the returns before making them publicly available. Labor makes Form 5500 data files available to researchers and policy analysts, but its Web site does not include links to the data. A private firm provides a Web site with images of the completed forms, but not compiled data sets. Labor does not link Form 5500 reports by firm to facilitate longitudinal analysis. Linking these consolidated reports of publicly held companies with Form 5500 data is difficult because these reports can cover only parts of a company, more than one company, or privately held companies. Some researchers have linked these data for selected firms.

Labor no longer requires that it receives the summary plan descriptions regularly, and as a result the public no longer has access rights to new or revised versions. Labor does, however, incorporate some data from these plans in its National Compensation Survey data. This includes, for example, information on normal retirement ages and formulas for calculating employer contributions to pension plans.

8. Gather more data on labor demand for older workers

Recommendation: The employer data collection plan should include short-term and long-term goals for obtaining information on labor demand for older workers and the factors that may affect that demand. Needed data elements include employment patterns of older workers, compensation and benefit costs by age, and worker productivity by age. Very little information on these topics is currently available, and some raise difficult measurement issues. A reasonable short-term goal is to sponsor case studies of employers that can help identify important variables and feasible means of collecting them on a larger scale.

Status: The Health and Retirement Study and other surveys provide much
data on employment patterns and salary and wages of older workers. Little
has been done, however, concerning compensation and benefits costs or
productivity by age.

9. Conduct longitudinal survey of employers and their workers

Recommendation: The employer data collection task force should
consider the feasibility and cost-effectiveness of a panel survey, which is
periodically refreshed that collects detailed information on employers and
their workers. Such a survey should cover the full universe, including
private for profit, nonprofit, and government employers, and the self-
employed. Longitudinal data from an employer-based survey are needed to
analyze the factors that affect employer decisions about recruitment and
retention of older workers and benefit plan offerings and how these
decisions, in turn, affect workers.

Status: Such a task force has not been formed and no survey like the one
the panel recommended has been undertaken. BLS has developed a
longitudinal database of business establishments, the "LDB", based on
data from states' unemployment insurance programs, and the Census
Bureau continues to maintain a longitudinal establishment database
covering all private nonagricultural establishments. Neither of these
databases includes data on pensions or other retirement plans. The BLS
database provides data on employees' hours and wages. The Census
Bureau database also provides data on employment and wages, and
periodically includes data on employer contributions to pension plans and
health insurance. To some extent the HRS links data from individuals and
their employers on a longitudinal basis, but it tracks the individuals, not
the firms over time. The *E-Government Act of 2002* (P.L. 107-347) may
facilitate collaboration between the Census Bureau and BLS and could
provide for linking National Compensation Survey (NCS) data on
retirement plans to the Census Bureau's data on business establishments.

10. Gather more data from employers of HRS/AHEAD
sample members

Recommendation: HRS and AHEAD should develop and implement a
plan for obtaining information on a continuing basis on the pension and
health insurance offerings of the employers of the HRS/AHEAD sample
members.

Status: Data from HRS sample members' employers is available. Similar efforts have been undertaken with other surveys, such as the National Longitudinal Study of Mature Women.

11. Match panel survey responses and key administrative records

Recommendation: Matched files of panel survey responses and key administrative records should be regularly produced for retirement-income-related policy analysis and projection purposes. Examples include exact matches of survey records and Social Security earnings histories and benefit records, Medicare and Medicaid records, and the National Death Index. The added information in matched files is obtainable at low marginal cost and is essential for analysis of retirement and savings decisions and the effect of medical care use and expenditures on retirement security.

Status: Some matches between administrative data and panel survey data have been achieved. HRS investigators have completed matches between HRS and Social Security earnings and benefit record data, National Death Index data, and Medicare records. These are available on a restricted basis to selected researchers and policy analysts.

The Census Bureau has matched SIPP data with Social Security earnings records and extracts from individuals' IRS tax return data. The Census Bureau's Longitudinal Employer Household Dynamics (LEHD) project is exploring options for more extensive matches between Form 5500 data from employers and Census Bureau survey data on establishments.[2] Recently finalized Treasury regulations give the Census Bureau access to

---

[2]The LEHD program, which was started in 1998, is designed to evaluate and improve the quality of data collected in the Census Bureau's demographic and economic censuses and surveys through longitudinal analysis. The program combines longitudinal micro data from federal and state administrative data on employers and employees with these census and survey data. In addition to its use to improve its census and surveys, the Census Bureau conducts policy-relevant research on labor force and employment issues and creates new data products. Currently, the LEHD program provides quarterly workforce indicators for a number of participating states. The linkage of Form 5500 data and Census Bureau establishment data is described in Julia Lane, et al., "New Uses of Health and Pension Information: The 5500 file at the Census Bureau, LEHD Technical Paper No. TP-2002-03" (Washington, D.C.: U.S. Bureau of the Census, January 2002).

specified IRS tax return records to support SIPP and LEHD data linkage efforts.[3]

Also, researchers at the Employee Benefit Research Institute (EBRI) linked survey data to state Medicaid records in order to develop state specific projections of Medicaid expenditures.

12. Increase researchers' access to exact match files safeguarding confidentiality

Recommendation: Agencies should collaborate on the development and oversight of matched data sets for individuals and employers, with input from researchers on content. They should also vigorously explore creative solutions for providing research access to exact match files that safeguard the confidentiality of individual responses. Possible solutions include: (1) developing public use files that contain summary variables derived from the administrative records portion of the matched file (2) requiring researchers to sign nondisclosure agreements with significant penalties for violations; and (3) providing researchers with access to matched files on site at secure data centers.

Status: Access to linked data sets remains quite limited. Access to HRS linked data for example, is typically made available via a rigorous application process resulting in a data use agreement with the University of Michigan. To date, none of the linked data sets are available in encrypted public use files. However, according to HRS researchers, "The HRS, in conjunction with several other funded projects, has established a secure data facility to broaden access to the restricted datasets. We are exploring ways to eventually implement a system for encrypted online delivery of sensitive data files, as well as extend access to our restricted data." (One such method is the further use of data centers, which provide access to restricted information, including linked data sets, for approved researchers working on data sets. For more information on data centers, see page 19).

---

[3]Final regulation, "Disclosure of Return Information to the Bureau of the Census, Department of the Treasury, Internal Revenue Service, 26 CFR Part 301," *Federal Register* vol. 68, no. 13, January 21, 2003, p. 2691.

13. Fund regular evaluation of data quality

Recommendation: Budgets for retirement income-related surveys should include sufficient resources for regular evaluation of data quality. Evaluation methods include reinterviewing sub-samples of respondents to measure consistency of reporting; experimentation with alternative question wording to identify possible reporting problems; and comparing survey estimates with administrative records to determine the completeness and accuracy of survey reporting, taking care to adjust for differences in definitions and other aspects of the two sources.

Status: Several studies using the recommended methods have been conducted, with mixed results. The Census Bureau conducted a study comparing estimates of various types of 1990-96 incomes in SIPP and CPS to benchmark data estimated by the Census Bureau from the personal income estimates in the National Income and Product Accounts. The study found disparities between the survey based Census estimates and the administrative record based personal income estimates that could not be explained by differences in definitions.[4] For 1996, the aggregate wages and salary estimate based on CPS survey data was 102 percent of the benchmark based on administrative records and the CPS social security and railroad retirement benefit payments data were 92 percent of the benchmark. In contrast, the study found more substantial disparities for several other types of income. The aggregate CPS data for property income was 71 percent of the benchmark, and CPS data for pension income was 77 percent of the benchmark. The study is being replicated with data for 1999. Other studies have noted definitional and quality differences between estimates of personal savings from the Flow of Funds Accounts and National Income and Product Accounts.[5] Another study compared estimates of wealth from the SCF, PSID, and SIPP.[6]

---

[4]Marc I. Roemer, *Assessing the Quality of the March Current Population Survey and the Survey of Income and Program Participation Income Estimates, 1990-1996*, Income Surveys Branch Housing and Household Economics Statistics Division, U.S. Census Bureau (Washington, D.C.: June 16, 2000).

[5]U.S. Congressional Budget Office, *The Budget and Economic Outlook: Fiscal Years 2004-13*, (Washington, D.C., January 2003), p. 33.

[6]See for example, R. Curtin, F.T. Juster and J. Morgan, "Survey Estimates of Wealth: An Assessment of Quality" *in The Measurement of Saving*, *Investment and Wealth*, R.E. Lipsey and H.S. Tice, editors, National Bureau of Economic Research, Studies in Income and Wealth, 52 (Chicago: University of Chicago Press, 1989).

The Department of Treasury's Office of Tax Analysis has compared estimates of pension plan participation and contributions with estimates from the Census Bureau survey data. The Office of Tax Analysis linked W-2 data with Statement of Income (SOI) data from Form 1040. The results from this data set matched results from Census Bureau surveys, except for low-income households. The tax records obviously don't include nonfilers, but in addition, by design, the SOI sample includes relatively few low-income filers (in the order of 1 in 5,000 filers), but all filers in the highest income brackets. The Census Bureau surveys such as the SIPP over sample low income households and have much less data for the highest income households. For most income brackets, however, the data match quite well, according to Treasury officials.

The Census Bureau periodically assesses the quality of CPS data by reinterviewing a subsample of respondents, but none of the reinterviewing has covered questions on income for at least the last 4 years, according to the Census Bureau quality assurance staff.

The Census Bureau has also studied the accuracy of respondent data by matching income data in the March CPS with selected income detail on individual IRS income tax returns and SSA earnings and benefit records. A similar effort is underway using 1999 data.

One of the components of the BLS's National Compensation Survey is a program of re-interviews of a sub-sample of respondents to verify and clarify survey data, including data on retirement plans.

HRS investigators have compared employee responses about retirement income to employer provided data and Social Security records and found wide discrepancies. It is unclear to what extent these result from respondents' limited knowledge of their pensions or data errors.

Appendix III: GAO Survey on Retirement Income Data Needs and List of Respondents

Questionnaire

Survey on Retirement Income Data Needs

U.S. General Accounting Office

The U.S. General Accounting Office (GAO), an agency of the U.S. Congress, is conducting this survey as part of a study of data needed to increase accuracy in:
--the measurement of current retirement plan assets, income, and wealth,
--the projection of future retirement plan assets, income, and wealth for each worker cohort, and
--the estimation of the effects of potential retirement policy changes.

We will use the results of this survey to identify the data areas that researchers view as having the highest priority for improvement.

We are conducting this study at the request of Representative Rob Andrews, Ranking Minority Member of the Employer-Employee Relations Subcommittee of the House Committee on Education and the Workforce.

About This Questionnaire

This questionnaire has 8 questions, divided into two parts:

Part I: Suggested Actions for Improving Retirement Income and Wealth Data

Part II: General Information

It should take only 10-15 minutes to complete.

Participants in this survey will be listed in our report, however names will not be associated with individual responses. You will be notified when our report is issued.

Navigating, Exiting, Changing Your Answers, and Printing the Questionnaire

To read to the bottom of a screen: Use the scroll bar on the right hand side of the screen.

To move from screen to screen: Use the buttons at the bottom of each screen.

To exit: Click on the "Exit" button at the bottom of the screen. You may

log in again to re-open your questionnaire. The questionnaire will restart at the last response you entered.

To change your answers: To change an answer marked with a "button" (circle), click on another answer. To "uncheck" a checked box, click on the box again (this will "uncheck" it), then check the box(es) you wish to check. Note: You can change your answers, even after logging off, by logging on again (see above).

To print your responses: Click on the "Print" button at the end of each section. You cannot print the entire questionnaire at one time.

Indicating You Have Completed the Questionnaire

When you have completed the questionnaire, please be sure you go to the end of the questionnaire, to the section called "When The Questionnaire Is Complete," and click on "Completed." Your answers will not be used unless you have done this.

Questions?

If you have any questions, please contact Ben Pfeiffer, in our Seattle office, at pfeifferb@gao.gov or (206) 287-4832, or Tim Fairbanks, in our Washington office, at fairbankst@gao.gov or (202) 512-6916.

Part I: Suggested Actions for Improving Retirement Income and Wealth Data

This section asks about a number of actions, some prompted by the recommendations in a 1997 National Research Council report, to improve retirement income and wealth data. Efforts to improve data can involve increasing the amount of available data, improving the quality of data and/or increasing the timeliness of data.

A number of different types of data are needed to accurately measure current, and project future, retirement income and wealth. We are considering data on all types of retirement income and wealth, including pensions, individual retirement accounts, wages and salaries, Social Security and other government program benefits, housing equity, and other financial and non-financial assets.

This section asks for your assessment of the _priority_ to be given to various actions for improving retirement income and wealth data. Although you may regard many of the suggested actions as important, please keep in mind that because of limited resources some may need to be lower priority than others.

Data on Households and Individuals

1. Please indicate the priority you would place on taking the following actions to improve retirement income data. Please keep in mind that because of limited resources some actions may need to be lower priority than others. *(Click one button in each row.)*

	Highest priority	High priority	Medium priority	Low priority	Lowest priority	No opinion	No Response
Increase support for longitudinal studies of individuals over 50 years of age, such as the Health and Retirement Study (HRS).	○	○	○	○	○	○	○
Expand longitudinal studies of retirement savings of younger individuals (age 50 or below).	○	○	○	○	○	○	○
Increase support for other studies of households and individuals' retirement and wealth, such as the Survey of Income and Program Participation (SIPP), and the Survey of Consumer Finances (SCF).	○	○	○	○	○	○	○
Improve measurement of family and household consumption expenditures in surveys such as the Consumer Expenditure Survey (CEX) and in panel surveys such as HRS.	○	○	○	○	○	○	○

Data on Employers and Employee Benefits

2. Please indicate the priority you would place on taking the following actions to improve retirement income data. Please keep in mind that because of limited resources some actions may need to be lower priority than others. *(Click one button in each row.)*

	Highest priority	High priority	Medium priority	Low priority	Lowest priority	No opinion	No Response
Improve the design and reporting of retirement income-related data from employers, such as mandatory pension plan disclosures and surveys on benefit plan offerings.	○	○	○	○	○	○	○
Conduct research on labor demand for older workers and the factors that may affect that demand.	○	○	○	○	○	○	○
Conduct a longitudinal survey of employers and their workers.	○	○	○	○	○	○	○

Improve data from the <u>employers</u> of respondents in surveys such as the Health and Retirement Study (HRS), and the Asset and Health Dynamics Among the Oldest Old Study (AHEAD).

Other Areas for Improvement

3. **Please indicate the priority you would place on taking the following actions to improve retirement income data. Please keep in mind that because of limited resources some actions may need to be lower priority than others.** (*Click one button in each row.*)

	Highest priority	High priority	Medium priority	Low priority	Lowest priority	No opinion	No Response
Improve national data on aggregate retirement and non-retirement assets such as Flow of Funds, and National Income and Product Accounts.	○	○	○	○	○	○	○
Improve matching of survey respondents with key administrative records, while protecting confidentiality.	○	○	○	○	○	○	○
Increase researchers' access to datasets on individuals or households matched with administrative datasets.	○	○	○	○	○	○	○
Improve collaboration between agencies and researchers to improve questionnaires and data collection and dissemination practices.	○	○	○	○	○	○	○

Crosscutting Actions

4. **Please indicate the priority you would place on taking the following actions to improve retirement income data. Please keep in mind that because of limited resources some actions may need to be lower priority than others.** (*Click one button in each row.*)

	Highest priority	High priority	Medium priority	Low priority	Lowest priority	No opinion	No Response
Collect additional data in existing surveys	○	○	○	○	○	○	○
Begin new surveys	○	○	○	○	○	○	○
Improve the quality of existing surveys	○	○	○	○	○	○	○
Improve the timeliness of existing data	○	○	○	○	○	○	○
Improve researchers' access to existing administrative data, such as Social Security earnings records or Medicare records	○	○	○	○	○	○	○
Improve techniques for protecting the privacy of respondents' survey and administrative data	○	○	○	○	○	○	○

Fund research on retirement income and wealth	○	○	○	○	○	○	○
Fund expanded data collection efforts	○	○	○	○	○	○	○
Fund increased or improved matching of data	○	○	○	○	○	○	○
Fund new or improved modeling efforts	○	○	○	○	○	○	○

5. **Are there other types of actions that are important for improving the availability, timeliness, or quality of retirement income and wealth data?**
   (*Please enter your suggestions in the text box below. The textbox will expand to accommodate your response.*)

Part II: General Information

6. **Which of the following describe the way that you work with retirement income and wealth data?**
   (*CHECK ALL THAT APPLY*)
   1. ☐ Data collection and/or data management
   2. ☐ Primary data analysis
   3. ☐ Secondary data analysis
   4. ☐ Policy analysis or development
   5. ☐ Other

   If you checked "other" (above), please specify the way you work with retirement income data in the text box below.
   (*The textbox will expand to accommodate your response.*)

7. **Which of the following best describes your affiliation?**
   (*CHECK ONLY ONE ANSWER*)
   1. ○ Federal government
   2. ○ State or local government

3. ○ University or college

4. ○ Other not-for-profit organization

5. ○ Private for-profit organization

6. ○ Other

7. ○ No response

If you checked "other" (above), please specify your affiliation in the text box below.
(*The textbox will expand to accommodate your response.*)

8. **Which of the following data sets have you used in your work during the past 5 years?**
   (*CHECK ALL THAT APPLY*)

   1. ☐ Asset and Health Dynamics Among the Oldest Old (AHEAD)

   2. ☐ Consumer Expenditure Survey (CEX)

   3. ☐ Current Population Survey (CPS)

   4. ☐ Flow of Funds data

   5. ☐ Health and Retirement Study (HRS)

   6. ☐ IRS Form 5500 data

   7. ☐ National Compensation Survey (NCS) (incorporating the Employee Benefits Survey)

   8. ☐ National Income and Product Accounts data

   9. ☐ Panel Study of Income Dynamics (PSID)

   10. ☐ Social Security Administration administrative files

   11. ☐ Survey of Consumer Finances (SCF)

   12. ☐ Survey of Income and Program Participation (SIPP)

   13. ☐ Other

If you checked "other" (above), please specify the data sets in the text box below. (*The textbox will expand to accommodate your response.*)

Additional Comments

If you would like to make additional comments concerning any topic covered in this questionnaire, please feel free to enter them in the textbox below, or, if you would prefer, send an email message to Ben Pfeiffer at pfeifferb@gao.gov. (*The textbox will expand to accommodate your response.*)

When The Questionnaire Is Complete

When you have completed this questionnaire, please check the "Completed" box below.
(CHECK ONLY ONE ANSWER)

1. ○ Completed
2. ◉ Not completed

Thank You

Thank you for your participation.

Click on the **Exit** button below to exit the questionnaire, then close the browser windows associated with this questionnaire by clicking on the small "X" in the upper right hand corner of your screen.

| Print |
| Submit |
| Cancel |

Questionnaire Programming Language - Version 5.0
U.S. General Accounting Office

Survey Results

1. Data on households and individuals: Please indicate the priority you would place on taking the following actions to improve retirement income data.

	Highest priority (percent)	High priority (percent)	Medium priority (percent)	Low priority (percent)	Lowest priority (percent)	No opinion (percent)	Number of cases
Increase support for longitudinal studies of individuals over 50 years of age, such as the HRS	27.6	40.0	23.8	5.9	2.2	0.5	185
Expand longitudinal studies of retirement savings of younger individuals (age 50 or below)	26.9	39.2	24.2	7.5	1.6	0.5	186
Increase support for other studies of households and individuals' retirement and wealth, such as the SIPP, and the SCF	18.2	28.3	40.6	10.7	1.1	1.1	187
Improve measurement of family and household consumption expenditures in surveys such as the CEX and in panel surveys such as HRS	15.2	33.2	32.1	12.0	6.0	1.6	184

2. Data on employers and employee benefits: Please indicate the priority you would place on taking the following actions to improve retirement income data.

	Highest priority (percent)	High priority (percent)	Medium priority (percent)	Low priority (percent)	Lowest priority (percent)	No opinion (percent)	Number of cases
Improve the design and reporting of retirement income-related data from employers, such as mandatory pension plan disclosures and surveys on benefit plan offerings	33.0	31.4	25.0	8.0	1.6	1.1	188
Conduct research on labor demand for older workers and the factors that may affect that demand	13.8	28.2	31.9	19.1	6.4	0.5	188
Conduct a longitudinal survey of employers and their workers	14.0	25.3	33.9	21.5	3.2	2.2	186
Improve data from the employers of respondents in surveys such as the HRS, and the AHEAD	30.1	35.5	21.9	7.1	2.2	3.3	183

3. Other areas for improvement: Please indicate the priority you would place on taking the following actions to improve retirement income data.

	Highest priority (percent)	High priority (percent)	Medium priority (percent)	Low priority (percent)	Lowest priority (percent)	No opinion (percent)	Number of cases
Improve national data on aggregate retirement and non retirement assets such as Flow of Funds, and National Income and Product Accounts	9.2	20.0	30.3	28.6	8.1	3.8	185
Improve matching of survey respondents with key administrative records, while protecting confidentiality	45.2	30.1	14.5	5.4	2.2	2.7	186
Increase researchers' access to datasets on individuals or households matched with administrative data sets	56.8	25.1	9.8	4.4	2.7	1.1	183
Improve collaboration between agencies and researchers to improve questionnaires and data collection and dissemination practices	24.1	40.1	27.3	7.0	0.5	1.1	187

4. Crosscutting actions: Please indicate the priority you would place on taking the following actions to improve retirement income data.

	Highest priority (percent)	High priority (percent)	Medium priority (percent)	Low priority (percent)	Lowest priority (percent)	No opinion (percent)	Number of cases
Collect additional data in existing surveys	23.0	36.1	31.1	6.6	0.5	2.7	183
Begin new surveys	4.0	14.7	29.9	35.0	13.6	2.8	177
Improve the quality of existing surveys	25.7	44.3	22.4	2.7	2.2	2.7	183
Improve the timeliness of existing data	21.6	33.0	24.9	14.1	4.3	2.2	185
Improve researchers' access to existing administrative data, such as Social Security earnings records or Medicare records	50.8	28.6	13.5	3.8	2.2	1.1	185
Improve techniques for protecting the privacy of respondents' survey and administrative data	20.5	22.2	34.1	16.2	3.2	3.8	185
Fund research on retirement income and wealth	28.1	38.9	22.7	7.6	1.6	1.1	185
Fund expanded data collection efforts	18.4	45.3	27.4	5.0	1.7	2.2	179
Fund increased or improved matching of data	31.1	37.3	22.0	5.6	1.7	2.3	177
Fund new or improved modeling efforts	11.7	18.4	40.8	20.7	6.1	2.2	179

5. Are there other types of actions that are important for improving the availability, timeliness, or quality of retirement income and wealth data?

Percent writing comments	Number of cases
43.2	190

6. Which of the following describe the way that you work with retirement income and wealth data? (Check all that apply.)

	Percent checking	Number of cases
Data collection and/or data management	26.7	187
Primary data analysis	53.5	187
Secondary data analysis	58.8	187
Policy analysis or development	74.3	187
Other	9.1	187

	Percent explaining other	Number of cases
6a. If you checked 'other' (above), please specify the way you work with retirement income data in the text box below.	100	17

7. Which of the following describes your affiliation?

	Percent checking	Number of cases
Federal government	18.0	189
State or local government	0.0	189
University or college	48.7	189
Other not-for-profit organization	24.3	189
Private for profit organization	7.4	189
Other	1.6	189

	Percent explaining other	Number of cases
7a. If you checked "other" (above), please specify your affiliation in the text box below.	66.7	3

8. Which of the following data sets have you used in your work during the past 5 years? (Check all that apply.)

	Percent checking	Number of cases
AHEAD	20.1	184
CEX	35.9	184
CPS	76.6	184
Flow of funds data	30.4	184
HRS	63.6	184
IRS Form 5500 data	38.0	184
NCS (incorporating the Employee Benefits Survey	17.9	184
National Income and Product Accounts data	35.9	184
PSID	35.9	184
SSA administrative files	36.4	184
SCF	47.3	184
SIPP	51.1	184
Other	21.7	184

	Percent explaining other	Number of cases
8a. If you checked "other" (above), please specify the data sets in the text box below.	100	40

Additional comments: If you would like to make additional comments concerning any topic covered in this questionnaire, please enter them in the textbox below.

Percent writing comments	Number of cases
14.7	190

List of Respondents to the Survey

Henry J. Aaron
The Brookings Institution

Julie Agnew
College of William and Mary

John Ameriks
TIAA-CREF Institute

Joseph M. Anderson
Capital Research Associates

Kenneth Apfel
University of Texas at Austin

Katherine Baicker
Dartmouth College

Vickie Bajtelsmit
Colorado State University

Dean Baker
Center for Economic and Policy Research

Laurel Beedon
Public Policy Institute, AARP

Dan Beller
Employee Benefits Security Administration

Department of Labor

Keith A. Bender
University of Wisconsin-Milwaukee

Mark C. Berger
University of Kentucky

B. Douglas Bernheim
Stanford University

Merton C. Bernstein
Washington University

Joydeep Bhattacharya
Iowa State University

Andrew Biggs
Cato Institute

Emily Blank
Howard University

Robert B. Blancato
Matz, Blancato & Associates, Inc.

Henning Bohn
University of California - Santa Barbara

Christopher M. Bone
Actuarial Science Associates

Barry P. Bosworth
The Brookings Institution

Linda Smith Brothers
University of Wisconsin - Madison

Charlie Brown
University of Michigan

Jeffrey Brown
Harvard University

Richard V. Burkhauser
Cornell University

Gary Burtless
The Brookings Institution

John Y. Campbell
Harvard University

William J. Carrington
Welch Consulting

Yung-Ping Chen
University of Massachusetts Boston

Constance F. Citro
National Research Council

Robert L. Clark
North Carolina State University

Denise M. Clark
Feder Semo Clark & Bard, P.C.

Joao Cocco
London Business School

Lee Cohen
Social Security Administration

Steven B. Cohen
Agency for Healthcare Research and Quality|
Department of Health and Human Services

Courtney C. Coile
Wellesley College

Craig Copeland
 EBRI

Christopher Cornwell
University of Georgia

Julia Lynn Coronado
Federal Reserve Board

David Cutler
Harvard University

Kimberly Darling
SAG Corporation

Angus Deaton
Princeton University

Jeff Dominitz
Carnegie Mellon University

Stuart Dorsey
Baker University

Karen Dynan
Federal Reserve Board

Ryan D. Edwards
Stanford University

Douglas W. Elmendorf
Federal Reserve Board

Gary V. Engelhardt
Syracuse University

Eric M. Engen
American Enterprise Institute

William E. Even
Miami University

Jeff Faux
Economic Policy Institute

Melissa Favreault
Urban Institute

Karen W. Ferguson
Pension Rights Center

Douglas Fore
TIAA-CREF Institute

Jonathan Barry Forman
University of Oklahoma

Leora Friedberg
University of Virginia

Robert B. Friedland
Georgetown University

Don Fullerton
University of Texas at Austin

William G. Gale
The Brookings Institution

Ron Gebhardtsbauer
American Academy of Actuaries

Thomas Glass
Glass & Co. CPAs

Jagdeesh Gokhale
Federal Reserve Bank of Cleveland

Carol Gold
Internal Revenue Service

Nancy M. Gordon
Census Bureau, Department of Commerce

Pierre-Olivier Gourinchas
Princeton University

Brian Graff
American Capital Strategies

Matthew Greenwald
Matthew Greenwald & Associates, Inc.

Lijia Guo
University of Central Florida

Alan L. Gustman
Dartmouth College

Steven Haider
RAND

Eric Hanushek
Stanford University

Brian Headd
Small Business Administration

Joni Hersch
Harvard University

Roger Hickey
Institute for America's Future

Catherine Hill
National Academy of Social Insurance

Richard Hinz
World Bank (formerly, Department of Labor)

Lorrie L. Hoffman
University of Central Florida

Karen C. Holden
University of Wisconsin-Madison

Sarah Holden
Investment Company Institute

Kevin M. Hollenbeck
W.E. Upjohn Institute

Martin Holmer
Policy Simulation Group

Marjorie Honig
City University of New York

M. Cindy Hounsell
Women's Institute for a Secure Retirement

Warren Hrung
Department of the Treasury

Michael D. Hurd
RAND

Edwin C. Hustead
The Hay Group

Robert M. Hutchens
Cornell University

Howard M. Iams
Social Security Administration

Estelle James
Consultant

David C. John
Heritage Foundation

Richard W. Johnson
Urban Institute

David Joulfaian
Department of the Treasury

F. Thomas Juster
University of Michigan

Arthur B. Kennickell
Federal Reserve Board

Surachai Khitatrakun
University of Wisconsin-Madison

Kilolo Kijakazi
Center on Budget and Policy Priorities

Geoffrey Kollmann
Congressional Research Service

Sophie M. Korczyk
Analytical Services

Marvin H. Kosters
American Enterprise Institute

Douglas L. Kruse
Rutgers University

Julia Lane
Urban Institute

Annamaria Lasardi
Dartmouth College

Ronald Lee
University of California, Berkeley

Jules Lichtenstein
AARP

Jeffrey B. Liebman
Harvard University

Dennis Logue
Price College

Robin Lumsdaine
Brown University

David A. Macpherson
Florida State University

Brigitte C. Madrian
University of Chicago

Joyce Manchester
Social Security Administration

Charles F. Manski
Northwestern University

Ann A. McDermed
North Carolina State University

Andrew Metrick
University of Pennsylvania

Daniel J. Mitchell
Heritage Foundation

Olivia S. Mitchell
University of Pennsylvania

H. Fred Mittelstaedt
University of Notre Dame

Catherine Phillips Montalto
Ohio State University

James H. Moore
Social Security Administration

Kathryn L. Moore
University of Kentucky

Brent R. Moulton
Bureau of Economic Analysis, Department of Commerce

James Musumeci
Southern Illinois University at Carbondale

Randall J. Olsen
Ohio State University

Van Doorn Ooms
Committee for Economic Development

Peter R. Orszag
The Brookings Institution

Beverly J. Orth
William M. Mercer, Inc

Michael Packard
Pension Benefit Guaranty Corporation

Michael G. Palumbo
Federal Reserve Board

Constantijn W. A. Panis
RAND

Jonathan A. Parker
Princeton University

Donald O. Parsons
George Washington University

Christina Paxson
Princeton University

Cynthia D. Perry
Massachusetts Institute of Technology

Pamela J. Perun
Urban Institute

Joseph S. Piacentini
Employee Benefits Security Administration

Department of Labor

Christopher Polk
Northwestern University

James Poterba
Massachusetts Institute of Technology

Elizabeth T. Powers
University of Illinois at Urbana-Champaign

Patrick J. Purcell
Congressional Research Service

John W. R. Phillips
Social Security Administration

Anna M. Rappaport
William M. Mercer, Inc.

Robert R. Reed III
University of Kentucky

Cordelia W. Reimers
Hunter College

John C. Rother
AARP

John E. Sabelhaus
Congressional Budget Office

Dallas L. Salisbury
EBRI

Andrew A. Samwick
Dartmouth College

Thomas R. Saving
Texas A & M University

Patricia L. Scahill
Ernst & Young LLP

Sylvester J. Schieber
Watson Wyatt Worldwide

Robert F. Schoeni
University of Michigan

James H. Schultz
Brandeis University (retired)

John C. Scott
American Benefits Council

Lois B. Shaw
Institute for Women's Policy Research

Stuart A. Sirkin
Pension Benefit Guaranty Corporation

Jonathan S. Skinner
Dartmouth College

Timothy A. Smeeding
Syracuse University

Kent Smetters
Wharton School

Karen E. Smith
Urban Institute

James P. Smith
RAND

Paul Smith
Department of the Treasury

Ralph Smith
Congressional Budget Office

Frank P. Stafford
University of Michigan

Norman Stein
University of Alabama

Thomas L. Steinmeier
Texas Tech University

C. Eugene Steuerle
Urban Institute

Ann Huff Stevens
Yale University

Annika E. Sundén
Boston College

Richard M. Suzman
National Institute on Aging

Kenn B. Tacchino
Widener University

Albert Teplin
Federal Reserve Board (Retired)

Lawrence H. Thompson
Urban Institute

Shripad Tuljapurkar
Mountain View Research

Cori E. Uccello
Urban Institute

Jack L. VanDerhei
Temple University

Steven F. Venti
Dartmouth College

Alice Wade
Social Security Administration

Daniel Weinberg
Census Bureau, Dept. of Commerce

David R. Weir
University of Michigan

Joseph White
Case Western Reserve University

William J. Wiatrowski
Bureau of Labor Statistics, Department of Labor

Joshua L. Wiener
Oklahoma State University

Ernest Wilcox
Bureau of Economic Analysis, Department of Commerce

Samuel H. Williamson
Miami University

Robert J. Willis
University of Michigan

Doug A. Wolf
Syracuse University

Aliya Wong
Thelen Reid & Priest, LLP

Jing Jian Xiao
University of Rhode Island

Paul J. Yakoboski
American Council of Life Insurers

Sheila R. Zedlewski
Urban Institute

Stephen P. Zeldes
Columbia University

Appendix IV: Views of GAO's Expert Panel on Retirement Income Data Needs

This appendix provides the summary of discussion by members of an expert panel that we convened on retirement income data needs on September 10, 2002. The panel consisted of 11 nationally recognized experts who, during a day-long meeting, discussed the issues the federal government should address in order to improve the quality of retirement income data statistics. All the ideas presented in this appendix may not represent the view of every member of the panel. Moreover, these ideas should not be considered to be our views.

Members of Our Expert Panel

The following individuals were members of our expert panel on retirement income data:

- Eric Engen, Resident Scholar, The American Enterprise Institute
- Melissa Favreault, Research Associate, The Urban Institute
- Nancy Gordon, Associate Director for Demographic Programs, US Census Bureau
- Alan Gustman, Professor of Economics, Dartmouth College
- Richard Hinz, Chief Economist and Director of Policy and Research, EBSA, Department of Labor (now at the World Bank)
- Howard Iams, Director, Division of Policy Evaluation, Office of Policy, Social Security Administration
- John Sabelhaus, Unit Chief, Long-Term Modeling Group, Congressional Budget Office
- Dallas Salisbury, President, Employee Benefit Research Institute
- Jack VanDerhei, Associate Professor, Department of Risk, Insurance & Healthcare Management, Fox School of Business and Management, Temple University
- William Wiatrowski, Chief, Compensation Data Analysis and Planning Division, Bureau of Labor Statistics, Department of Labor
- Robert Willis, Professor of Economics, Institute for Social Research, University of Michigan

Views of the Panelists

The panel members discussed a number of issues the federal government needs to address in order to improve the quality of retirement income statistics. Specifically, panelists discussed actions and strategies the federal government could undertake related to (1) improving matching of survey and administrative data, (2) improving access to administrative data, and (3) improving the quality of employer data.

Need for Better Matched Data

Panelists said that a significant amount of the needed information about American workers and pension behavior is already being collected in government and private surveys and government administrative reports. While no one single survey or report collects all the pension information panels expressed interest in, they said that some household surveys could be linked with administrative report data from employers. Specifically, panelists discussed the following:

- The Survey of Income and Program Participation (SIPP) a household survey conducted by the Bureau of the Census, has information on demographic characteristics, labor force participation, and detailed information on income, including some basic pension information. The SIPP does not gather detailed information about the characteristics of these pensions.[1]
- Summary Plan Descriptions (SPD), prepared by employers as required under ERISA, provide detailed information about the characteristics of the pension plans that they provide to their employees.
- In the past, the Health and Retirement Study (HRS), which is primarily a household survey conducted by the University of Michigan, has matched survey data from households who permitted the Social Security Administration to provide Social Security earnings records and benefit records, the National Death Index, and to Medicare records for those individuals who are Medicare eligible. HRS records also have been matched to SPDs obtained either from Labor or from employers.
- The panelists were interested in the information available from the 1979 National Longitudinal Survey, which is sponsored by the Bureau of Labor Statistics and gathers a wide range of information over a long time period. Panelists felt that this information could be potentially be useful because the participants, who initially participated as youth, are now approaching retirement age, which would give researchers access to a lifetime of data.
- In addition to matching currently being done, some panelists said that they were interested in matching existing pension information to other sources of employer information and public financial data, such as information from reports filed with the SEC and available, for example, from Compustat.

Limitations on Access to Data

Members of our expert panel felt that there are several sources of administrative data that could give researchers valuable information, especially those that could be linked, but legal and logistic restrictions and

---

[1]While some employee information is gathered from other surveys, the panel felt that the SIPP survey is extremely important for matching to administrative records because it gives demographic and labor force characteristics not available in most administrative record files.

limitations prevent access to some of this pension data. Specifically, panelists discussed different types of limitations to access.

Legal Restrictions

- The Federal Code Title 13, Section 9 sets very specific limitations on the access to any data identifying individuals gathered by the Census Bureau and any data that are linked to Census data has the Title 13 limitation "attached" to it as a result.[2]
- Individual records from the IRS have some access restrictions similar to those of the Census Bureau except that legislation allows specified agencies access to certain tax return records.[3]
- As a result of amendments to ERISA legislation SPDs are no longer routinely collected by Labor.

Fragmented Retirement Data Responsibilities Contributes to Data Shortcomings

- The responsibility for gathering and analyzing pension data is fragmented between different government agencies. There is no central agency responsible for retirement data – it is fragmented between the Pension Benefit Guaranty Corporation, Labor and IRS. As a result, individual agencies do not have the incentive to provide access to information or to collect statistics for research purposes.
- While Labor's regulations specify that employers must supply SPDs if requested, it does not specify that employers must have a copy of the SPDs. As a result, researchers who request SPDs from employers are frequently told that the Plan Administrator has the documents. The Plan Administrators in turn tell researchers that they have no authority for providing them to anyone except plan participants.
- Since 1980, OMB's Office of Information and Regulatory Affairs has provided government wide leadership and oversight of efforts to reduce the burden on respondents to government information collection, including statistical surveys". This "reduction" effort has meant that some research data previously collected in administrative reports has been eliminated.

Other Suggestions for Improving Access to Data

The expert panelists made many suggestions for improving access to pension data. More specifically, panelists discussed the following:

---

[2]The newly enacted *E-Government Act of 2002* (P.L. 107-347) extended Census type confidentiality restrictions to all data collected in a federal statistical survey.

[3]See 26 U.S.C. 6103.

- It was suggested that agencies with access to data, such as Labor, take advantage of improvements in technology to require and provide electronic copies of Form 5500s and SPDs.
- Panel experts suggested that some sort of license be issued for research vehicles (such as the SCF, the SIPP and the HRS) to have legal access to employer pension information, provided that they take measures to ensure confidentiality.[4]
- Panelists suggested creating more research data centers to match otherwise restricted data, including Census Bureau data. They suggested changing legislation so that the Census Bureau's data would not make everything subject to Title 13. The Census Bureau, however, is concerned about possible reidentification issues.[5]
- Panelists suggested studying more closely the Census Bureau's results in its experimentation with the development of "synthetic data" in the LEHD program, a technique in which many relationships between variables are maintained in a data set, but in a manner that makes it impossible to identify specific individuals.[6] Panelists also cautioned, however, that in many cases these techniques are not workable and researchers will need access to the original survey data.

Restricted Access Sites Have Provided Some Increased Opportunities for Access to Survey Records and Matched Files

Experts discussed research data centers, operated by several Federal agencies and private survey organizations, as an effective technique to make data not available because of confidentiality restrictions more available to researchers, but with certain restrictions. While the data centers provide opportunities for conducting research with survey records or matching records between surveys or with administrative records, experts said that there are limitations to the data centers, which make data

---

[4]Unlike the Census Bureau, some agencies, such as the National Center for Education Statistics, have authority to use a data license procedure. Under these licenses, licensees must submit detailed research plans, sign disclosure protection agreements, and agree to restrictions similar to those by the Bureau of the Census at their research data centers. For more detail, see Paul B. Massell and Laura Zayatz, "Data Licensing Agreements at U.S. Government agencies and Research Organizations," presented at the International Conference on Establishment Surveys-II, Buffalo, N.Y., June 17-21, 2000.

[5]Implementation of this suggestion would require changing not only Title 13 but other laws as well–like the new *E-Government Act*.

[6] U.S. General Accounting Office, *Record Linkage and Privacy: Issues in Creating New Federal Research and Statistical Information*, GAO-01-126SP (Washington, D.C.: April 2001) p. 88 and Federal Committee on Statistical Methodology, *Report On Statistical Disclosure Limitation Methodology*, *Statistical Policy Working Paper*, *22* (Washington, D.C.: May 1994).

much more difficult to access. Specifically, panelists discussed the
following:

- Restricted access sites are a useful means of allowing confidential information
  to be accessed by researchers, subject to certain restrictions.
- A federal storage data center where a number of federal data sets could be
  brought together could allow agencies to share some otherwise inaccessible
  information.
- Data centers could aid the work that researchers are doing by storing research
  already conducted in the data center.
- The Census Bureau, which operates Research Data Centers in eight locations
  throughout the country, allows researchers with pre-approved projects to use
  confidential economic and demographic survey data, such as SIPP, for which
  either no public-use version is available, or the public-use version does not
  contain the information required by the researcher. While researchers can
  access these data at the centers, they are not allowed to remove individual
  records from the data center. This restriction prohibits researchers from
  matching Census Bureau data with data sets available to researchers, unless
  those data are imported into the data center.
- BLS has a similar data center located in Washington, D.C. in which data
  extracted from SPDs collected as part of the National Compensation Survey
  are stored. As with the Census Bureau centers, there are limitations on who
  can access this information and restrictions on removing data from the data
  center.
- Because confidential data cannot be removed from either the BLS or the
  Census Bureau data centers, it is currently not possible for researchers to
  match data sets from the two agencies.

More Employer Information Needed on the Value and Provisions of Employer-Provided Pensions

There was wide agreement on the panel that greater access to employer
information was needed to accurately capture the value and provisions of
employer-provided pensions. More specifically, panelists discussed the
following:

- Employee-provided information about pensions is not a viable source because
  individuals often do not have a good understanding of the value or
  characteristics of pensions. In addition, agencies expressed concern about the
  impact of additional questions on nonresponse, and it may be difficult to
  obtain OMB approval for adding questions to statistical surveys or to
  administrative reports.
- Valuable information on the value and characteristics of employer pension
  information is already collected by the Department of Labor on the Form
  5500s.
- Through the LEHD program, the Census Bureau has been trying to link
  together employer information with data from surveys.
- Agencies have concerns about maintaining the privacy and confidentiality of
  data. For example, there was concern that linking SIPP information about

individuals with Form 5500 files or summary plan descriptions could facilitate the reidentification of individual data reported to the Census Bureau. As a result, those linked data would be available only within the data centers.

Appendix V: Characteristics of Selected Surveys for Analysis of Retirement Income and Wealth

Below is a table highlighting some of the features of selected surveys pertinent to the analysis of retirement income and projection of future retirees' income.[1]

Table 3: Summary Table of Selected Survey Data Sources

	Asset and Health Dynamics Among the Oldest Old and Health and Retirement Study	Consumer Expenditure Survey	Current Population Survey	Employee Benefits Survey [a]	Panel Study of Income Dynamics	Survey of Consumer Finances
Unit of observation	AHEAD: Men and women aged 70 and above in 1993-94 and aged 69 –75 in 1998-99 and spouses HRS: Men and women aged 51-61 in 1992 and aged 51 to 56 in 1998-99 and spouses[a]	Consumer units/ household members	Households	Employer	Head of household	Household/ primary economic unit
Cross sectional or longitudinal	Longitudinal	Cross-sectional	Cross-sectional	Cross-sectional	Longitudinal	Cross-sectional
Cohorts or years covered	Began in 1992. Planned to include a new cohort of people aged 51 to 56 every 5 years	Civilian non-institutionalized population[b]	Civilian non-institutionalized population	In even years state & local governments & beginning in 1999, private employers surveyed annually	All	All
Panel period	AHEAD: 1993-present HRS: 1992-present	Current ongoing survey: 1980-present	1942 to present	1997 to present	1968 - present	1983-present

[1]AHEAD, HRS, and PSID are conducted by university research entities with support from federal agencies. The other surveys are conducted by federal government agencies.

Survey of Income and Program Participation	National Longitudinal Survey of Mature Women	National Longitudinal Survey of Older Men	National Longitudinal Survey of Young Women	National Longitudinal Survey of Young Men	National Longitudinal Survey of Youth
Household members	Women aged 30-44 in 1967 and spouses	Men aged 45-60 in 1966 and spouses	Women aged 14-24 in 1968 and spouses	Men aged 14-24 in 1966 and spouses	Men and women aged 14-21 in 1979 and spouses
Longitudinal	Longitudinal	Longitudinal	Longitudinal	Longitudinal	Longitudinal
All	Women born 1922 to 1937	Men born 1906 to 1921	Women born 1943 to 1953	Men born 1941 to 1952	Men & women born 1957 to 1964
1981 - present New panels approx. every 3 years	1967 to present	1966 to 1990	1968 to present	1966 to 1981	1979 to present

	Asset and Health Dynamics Among the Oldest Old and Health and Retirement Study	Consumer Expenditure Survey	Current Population Survey	Employee Benefits Survey	Panel Study of Income Dynamics	Survey of Consumer Finances
Frequency of interviews	Biennial	The interview survey has a rotating panel design; consumer units are interviewed every 3 months for 5 periods. The diary survey also has a rotating panel design - consumer units keep weekly diaries.	Monthly for a total of 8 months; households are interviewed 4 months in a row, rest for 8 months, and are interviewed for another 4 months (4-8-4 sample rotation)	Once in a survey period	Annual/ biennial	Triennial
Over samples	Blacks, Hispanics, and Florida residents	No	Hispanics in the March data collection, (includes the March supplement.)	N/A	Low income	Wealthy households
Most recent sample size:	22,000 individuals	Interview: 30,600 Diary: 15,300	55,000 households interviewed	4,500 establishments	7,406 families	4,500
Administrative record linkages	Medicare files, Nat'l Death Index, SSA earnings and projected benefits files, Employer Pension Plan SPDs	None	SSA and IRS (restricted)[c]	N/A	None	None

Survey of Income and Program Participation	National Longitudinal Survey of Mature Women	National Longitudinal Survey of Older Men	National Longitudinal Survey of Young Women	National Longitudinal Survey of Young Men	National Longitudinal Survey of Youth
Every 4 months	Every 1 to 3 years	Every 1 to 2 years until 1983; then 1990	Every 1 to 3 years	Every 1 to 2 years until 1981	Every year until 1994 then every 2 years beginning in 1996
Low-income households for 1990, and 1996 panels, and thereafter	Blacks	Blacks	Blacks	Blacks	Blacks, Hispanics, and economically disadvantaged
35,000 households interviewed	5,000	5,000	5,000	5,000	9,000
SSA & IRS[d] (restricted)	Employer pension plan SPDs for 1989	None	None	None	None

[a]The Employee Benefits Survey is now a component of the National Compensation Survey (NCS). It excludes federal, military, agricultural, fishing, forestry, or private household employers.

[b]The total civilian non institutional population of the United States, as well as that portion of the institutional population living in the following group quarters are included in the CEX: Boarding houses; housing facilities for students and workers; staff units in hospitals and homes for the aged, infirm, or needy; permanent living quarters in hotels and motels; and mobile home parks. Excluded are military personnel living on military bases and nursing home residents.

[c]Restricted refers to those persons at the SSA, BLS, and Census Bureau with authority to access these data.

[d]Links with Social Security earnings records and extracts from individuals' IRS tax returns.

THE SECRETARY OF COMMERCE
Washington, D.C. 20230

MAR - 4 2003

Ms. Barbara D. Bovbjerg, Director
Education, Workforce, and Income
  Security Issues
U.S. General Accounting Office
Washington, DC 20548

Dear Ms. Bovbjerg:

The U.S. Department of Commerce appreciates the opportunity to comment on the General
Accounting Office's draft report entitled *Retirement Income Data: Improvements Could Better
Support Analysis of Future Retirees' Prospects.* The Department of Commerce has no major
comments to the report. However, we have enclosed suggestions for clarifying the report.

Sincerely

Donald L. Evans

Enclosure

U.S. Department of Labor   200 Constitution Avenue, NW
Washington D.C. 20210

March 12, 2003

Ms. Barbara D. Bovbjerg
Director
Education, Workforce, and
  Income Security Issues
United States General Accounting Office
441 G Street, N.W., Room 5930
Washington, D.C. 20548

Dear Ms. Bovbjerg:

We have reviewed the draft report prepared by the General Accounting Office (GAO) entitled, Retirement Income Data: Improvements Could Better Support Analysis of Future Retirees' Prospects (GAO-03-337). Based upon our review of the report, we are providing you with the following comments. Technical comments have already been provided directly to your staff.

GAO Recommendation Number 1: "To help provide the data needed to inform important policy decisions concerning retirement programs, the Secretary of Labor should direct the Bureau of Labor Statistics (BLS) to prepare a plan to improve data for analyzing retirement income and wealth in coordination with . . . [other agencies]."

Experience indicates that there are a number of concerns that must be recognized in attempting the types of coordination called for in this recommendation. Much of the coordination that retirement income researchers have suggested involves linking data that typically are collected from individuals (such as demographics and wealth accumulation) with data typically collected from employers (such as details of retirement benefit plans and the costs associated with those plans). Past efforts at coordination of these types of data from employers and employees by different agencies have had only limited success, due to privacy concerns and other issues.

Coordination activities such as those recommended that have taken place in the past have often concluded that different surveys exist because they have – quite appropriately -different missions. For example, surveys of employer-provided retirement income benefits conducted by BLS are part of the Bureau's mission to report on all types of compensation provided to persons in the labor force. Surveys conducted by other agencies may be directed toward total wealth accumulation by individuals or toward all sources of retirement income, regardless of whether the income comes from an employer-sponsored plan. The operational, statistical, and data confidentiality environments of these different data-gathering programs make coordination of surveys a daunting task.

Further, the type of planning and coordination activities envisioned in the recommendation usually are the purview of the Office of Management and Budget's Office of Statistical Policy. BLS and other statistical agencies routinely participate in such activities as envisioned in this GAO recommendation under OMB direction. BLS could participate in the coordination effort as described in the recommendation, but the extra demands placed on staff will not be negligible.

GAO Recommendation Number 2: "To facilitate plan participants', beneficiaries', and analysts' timely access to information about employer-provided pension plans, the Secretary of Labor should obtain from plan administrators electronic filings of all SPDs and summaries of material modifications required by ERISA and make them publicly available in electronic form."

Although the recommendation that the Employee Benefits Security Administration (EBSA)[1] resume the collection of SPDs in electronic form could help BLS in its attempt to obtain plan provision data, especially if there were a standard electronic format for such documents, as discussed below there are various problems associated with implementing such a requirement.

The recommendation that EBSA resume the collection of summary plan descriptions is at odds with amendments to ERISA enacted just six years ago as part of the Taxpayer Relief Act of 1997 (TRA'97). As noted in GAO's draft report, TRA'97 amended ERISA to eliminate the requirement to file copies of summary plan descriptions and summaries of plan changes (collectively SPDs) with EBSA. Congress also eliminated the related requirement that EBSA make those documents available to the public in its public disclosure room. Although TRA'97 preserved EBSA's authority to make individual requests for copies of SPDs from plan administrators, that authority would not support re-imposition of a general SPD filing requirement.[2] Furthermore, EBSA issued final regulations in 2002 implementing its authority to request SPDs, and reiterated in the preamble that EBSA generally intends to limit the exercise of its authority to requesting SPDs on behalf of participants and beneficiaries.

EBSA's experience with pre-TRA'97 SPD filings indicated that there was insufficient public interest in obtaining copies of SPDs from EBSA's public disclosure room to justify the compliance costs and burdens placed on plans. While EBSA maintained approximately 2 million SPDs and received over 150,000 SPDs and related documents annually before the filing requirement was eliminated by TRA'97, we received only approximately 1,000 requests for SPDs in an average year.[3] The requesters included participants and beneficiaries, other governmental agencies, congressional offices, media representatives, and others.

---

[1] EBSA was formerly the Pension and Welfare Benefits Administration.
[2] The GAO electronic filing recommendation appears to repeat a comment from a 1993 GAO "Report to Congressional Requesters on Management Reform" entitled "GAO's Comments on the National Performance Review's Recommendations," GAO/OCG-94-1. One of the NPR recommendations was to eliminate the SPD filing and storage requirements. GAO stated in connection with that recommendation "[o]ur work indicates that DOL could reduce costs by reducing or eliminating storage of Summary Plan Description (SPD) hard copies while continuing to provide electronic access to SPDs." As noted above, however, after GAO's 1993 report, Congress, in 1997 eliminated the SPD filing requirement without any provision for continuing an electronic collection program.
[3] The GAO report focuses on data shortcomings related to pension plans and retirement income. The SPD filing requirement applied more generally to ERISA–covered pension and welfare plans. Accordingly, the number of requests for pension plan SPDs would be a smaller subset of the total annual requests.

2

Accordingly, based on EBSA's pre-1997 experience, there did not appear to be broad-based interest in SPDs on file and analysts generally did not appear to treat the collection as a valuable source of research data. This situation was due in large part to shortcomings in the type of detail, consistency, and timeliness of SPDs that would be required for it to be a useful analytic tool. For example, BLS' experience with the collection of SPDs and other data on employer-sponsored retirement plans indicates that those representatives of employers who are interviewed by BLS staff often do not have sufficient knowledge of the plan to be able to provide the desired information. Comparisons of information contained in SPDs with information provided by survey respondents often yield discrepancies because of the wide variety of plan designs and descriptions adopted by plan sponsors.

Although EBSA establishes content as well as readability and similar format requirements for SPDs, EBSA is statutorily precluded from requiring SPDs be submitted on prescribed forms. An underlying rationale was that administrators needed flexibility to present plan information in a manner best suited to the particular workplace and workforce involved. Due to the resulting diversity that exists among SPDs, having copies of SPDs on file, even in electronic form, might not provide a usable research database.

Another shortcoming with the SPD filing requirement recommendation is that EBSA could not give requesters any assurances that the information on file was up to date due to ERISA's provisions which allow summaries of pension plan changes to be furnished up to 210 days after the end of the plan year in which a plan amendment or change is adopted. An electronic filing requirement would suffer from the same shortcoming.

The GAO report suggests an electronic filing requirement would facilitate timely access to information by plan participants and beneficiaries, however, EBSA expects that a very small number of participants and beneficiaries will ask for SPDs because ERISA already requires plan administrators to automatically furnish SPDs to participants and beneficiaries within specified time limits set by statute in ERISA. Administrators are also required to provide additional copies of SPDs to participants and beneficiaries on request. Further, if the participant or beneficiary wanted an up to date copy of the SPD, EBSA would still be required to communicate with the plan administrator to determine whether the copy we had on file was current because of the above noted time lag allowed for distributing information about pension plan changes.

Implementing GAO's recommendations would impose costs and burdens on employee benefit plans to maintain and file electronic SPDs as well as EBSA to establish a system capable of receiving, storing, and disclosing the electronically filed documents. Under current law, there is no requirement that plan administrators prepare SPDs in electronic form. Although EBSA published regulations in April 2002 designed to encourage the use of new technologies for disclosing information electronically to plan participants, we did not require that plan's maintain electronic versions of their SPDs. Such a mandate would have to be imposed to implement GAO's recommendation. Also, to accommodate such an electronic filing requirement, EBSA would have to develop and operate a system to receive, process, store and retrieve the electronic SPDs. The GAO report does not include estimates of the costs and burdens an electronic filing requirement would impose on plan administrators and EBSA. At the time the SPD filing requirement was eliminated, we estimated the change would reduce the aggregate reporting burden on plans, many of which were maintained by small employers, by approximately 150,000 burden hours per year.

3

<u>GAO Recommendation Number 3:</u> "For plans in place before these new filing requirements go into effect, and where it is cost-effective, the Secretary of Labor should use existing authority to obtain copies of summary plan descriptions and summaries of material modifications in cases where analysts working on federally conducted or federally-sponsored research seeks SPDs for statistical purposes. This should assist analysts of retirement income in obtaining information about the features of employer-sponsored benefit plans that are not electronically available."

As noted above, EBSA issued final regulations in 2002 implementing its authority to request SPDs, and reiterated in the preamble that EBSA generally intends to limit the exercise of its authority to requesting SPDs on behalf of participants and beneficiaries. Use of this authority for collection of SPDs as part of a Federal government survey (such as those conducted by BLS) may be hampered by confidentiality restrictions placed on the identity of survey respondents.

In closing, the Department supports exploring less burdensome and more effective means of addressing the shortcomings in available data identified by GAO's Expert Panel. We agree that Congress, federal agencies, researchers, and others in the private sector need accurate data to assess employee benefit, tax, and economic trends and policies. Rather than reinstating an electronic variant of an SPD filing requirement, the Department would support an effort to identify alternative sources of such employee benefit plan information and ways to make that information available to researchers.

Thank you for providing us with an opportunity to comment on your draft report.

Sincerely,

Ann L. Combs
Assistant Secretary
Employee Benefits Security Administration

Kathleen P. Utgoff
Commissioner
Bureau of Labor Statistics

4

Appendix VIII: GAO Contacts and Staff Acknowledgments

GAO Contacts	Alicia Puente Cackley (202) 512-7022 Benjamin Pfeiffer (206) 287-4832
Staff Acknowledgments	Timothy Fairbanks, Nicholas Larson, Lynn Musser, Emily Pickrell, and Roger Thomas also contributed to this report.