*Linkages With Data From Social Security Administrative Records in the Health and Retirement Study**

The Health and Retirement Study (HRS) is a major longitudinal study designed for scientific and policy researchers for study of the economics, health, and demography of retirement and aging. The primary HRS sponsor is the National Institute of Aging, and the project is being conducted by the Survey Research Center of the Institute for Social Research (ISR) at the University of Michigan. Several agencies, including the Social Security Administration (SSA), are supporting the project. This is the second note describing SSA's data support for the HRS. The conditions under which SSA has been able to provide data and the respondent consent procedures developed for the release of SSA data are discussed in Olson (1996). This note describes the data from SSA records that have been released for linking to HRS data, linkage rates resulting from the consent process, and subgroup patterns in linkage rates.

HRS Sample and Design Features

The original HRS is a panel survey of a nationally representative sample of households containing at least one person born in the 1931-41 period. The name "Health and Retirement Study" has come to mean two things: (1) both the original cohort, which began in 1992 and (2) the larger project, which includes other survey cohorts that began in 1993, 1998, or are planned for future years. The several cohorts in the full study are summarized in chart 1. This note is generally restricted to the original cohort first interviewed in 1992, although SSA has agreed to provide administrative data for all four cohorts described in the chart.

Members of the cohorts born in the 1931-41 period were aged 51-61 in 1992, and in HRS terms, they are "age-eligible persons." They constituted 9,824 persons or 78 percent of the 12,652 persons in the initial 1992 interview. Spouses or partners were also interviewed, regardless of their year of birth, but spouses or partners born in years other than 1931-41 are not representative of other age groups, and their HRS weights were set to zero.[1]

*Janice A. Olson, Division of Economic Research, Office of Research, Evaluation and Statistics, Office of Policy, Social Security Administration.

Acknowledgments: The author wishes to thank Keith Bender, Benjamin Bridges, Barry Bye, Sharmila Choudhury, and Olivia Mitchell for their helpful comments on earlier drafts of this note.

Record Linkages

From the outset, the HRS design recognized the benefits to be gained from augmenting survey information with data from SSA's administrative files and from other sources. Examples from non-SSA sources include detailed pension plan data used to estimate pension wealth that were developed from summary pension plan documents from employers of survey respondents.[2] Data from the National Death Index of the National Center for Health Statistics on the fact and date of death have been added, as relevant, and information is being developed from Medicare records about diagnoses, procedures, hospital stays, and medical use and costs.

Data derived from SSA records and currently available for HRS research are of three types: earnings histories, Social Security benefit histories, and Supplemental Security Income (SSI) payment histories.[3] First and of primary interest to most HRS restricted data users are data on earnings histories. They are primarily used to develop estimates of Social Security benefits and wealth and pension wealth, which are used to understand retirement behavior, preparedness for retirement, and economic well-being and in related studies.

Two restricted files containing earnings data are currently available. One includes annual taxable earnings up to the Social Security taxable maximum and annual quarters of coverage (or coverage credits) for the 1951-91 period. For the early years, a summary taxable earnings amount for the 1937-50 period and total quarters of coverage for years 1947-50 are also available. An earnings record can exist for persons with no covered earnings over their working lives, although the overwhelming majority of persons in the HRS cohort have some earnings. A second file—the Summary Earnings and Projected Benefits File (SEPBF)—was created from the annual covered earnings data (Mitchell, Olson, and Steinmeier forthcoming). It includes summary earnings and employment measures, a Disability Insurance (DI) eligibility status indicator, and projections of Social Security benefits and wealth assuming retirement at age 62 and at normal retirement age. The SEPBF includes measures developed to facilitate user updates to summary estimates and projections using earnings as reported in the HRS surveys.

A third earnings file, released as a restricted file by the ISR, includes wages and self-employment income in noncovered jobs as well as in jobs covered by Social Security

and wages above the taxable maximum. For each consenting respondent, the file consists of one record per employer per year from 1980 through 1991. Records include the employer identification number (EIN), which has been scrambled to preserve confidentiality. The scrambled EIN will allow study of individual job changes, including the return of a worker to a prior job. Wage information on the file is drawn from W-2 reports, while the self-employment income information is drawn from the Internal Revenue Service Form 1040 Schedule SE.

The second type of data derived from SSA records includes benefit amount information from the Retirement, Survivors, and Disability Insurance (RSDI) programs. Monthly benefit amounts (MBA) received from 1962 through 1991 and monthly benefits paid (essentially the MBA minus any withholdings for Medicare requested by the beneficiary) are available. Whether benefits were received as retired or disabled workers or as secondary beneficiaries is indicated in the benefit data. A benefit record is available for persons who received any Social Security benefits in the 1962-91 period, even if it was for only 1 month in the period.

The third type of data is from the SSI program—SSA's program for the low-income aged, blind, and disabled. Payment amounts received under the federally administered SSI program are provided for 1974 (the program's initial year) through 1991. A payment record is available for persons who received any SSI payments in the 1974-91 period, even if it was for only 1 month in the period.

Consent Agreements

For SSA to release identifiable administrative record data to ISR, it was necessary to have the signed agreement of respondents, as well as their cooperation in providing Social Security numbers (SSNs). As described in Olson (1996), the agreements directed SSA to release earnings and benefit data to the University of Michigan for the HRS and assured respondents that the University would protect the privacy and confidentiality of the data. HRS interviewers made three attempts (in 1992, 1994, and 1996) to obtain respondent consents, but even after these attempts, some respondents failed to give their consent. In the 1998 interviewing cycle, the HRS project did not make a fourth attempt to persuade reluctant respondents.

The absence of a record derived from SSA data does not always indicate a refusal, and data could be missing for several reasons. First, in the start-up 1992 interview, procedures for handling the respondent consent agreements were not always well understood, and SSA could not process some signed forms because the forms were received after a 60-day deadline.[4] Second, respondents had to personally sign the forms, and signatures from proxies, even spouses or those with powers of attorney, were not acceptable. Third, some respondents may have been willing to cooperate but have been unable to provide an SSN that SSA could validate, and a very small number of persons do not have SSNs. Among the HRS-aged cohort, such respondents may be concentrated among recent immigrants.

Linkage Rates

To explore the success of the consent procedures, an indicator of the presence or absence of a record from any one of the earnings, RSDI, or SSI benefit files was created.[5] The indicator was used in conjunction with personal characteristics from the 1992 HRS survey data file. Table 1 shows the number and percent of 1992 HRS respondents with data derived from SSA administrative records. For the age-eligible group, percents based on unweighted and weighted data are shown for those interested in any differences. (As noted, respondents who are not age eligible have no weights.) Table 1 shows 75 percent of all respondents and 75 percent of the age-eligible respondents have linked data. Weighted and unweighted results are essentially the same.

Virtually everyone with any data derived from SSA records also has linked earnings data, but only 10 percent of all respondents and less than 7 percent of age-eligibles have linkages with data derived from RSDI records. The fact that the all-respondent group includes more than 1,500 respondents aged 62 or older and potentially eligible for retirement benefits probably accounts for the higher rate in that group. Less than 2 percent have linkages with data derived from SSI records.

The linkage rate of only 75 percent raises concerns about the degree to which the subsample with linked data derived from Social Security records reflects the larger sample population. Tables 2 and 3 investigate that concern and show the proportion of HRS respondents with linked data by demographic, health and employment, financial, and interview characteristics. Because Social Security policy interest is often directed toward subgroups, fairly detailed categories are shown in some cases. Categories for race or ethnic status and marital status, for example, reflect that desire for detail. Information is shown first for age-eligible respondents (table 2) and then for all respondents (table 3). The discussion focuses on results for the age-eligible group.

The left sides of the tables show the number of HRS respondents and the proportion with linked data. Proportions based on unweighted and weighted data for the age-eligible group are shown. For example, in table 2 the proportion of men with linked data is .737 when unweighted data are used. Also shown are p-values for chi square statistics under the null hypothesis that all the proportions are the same for the categories of each variable.[6] For example, the p-value for the null hypothesis that the proportion of men with linked data is the same as that for women is .0016, and most would reject the null hypothesis in that case and conclude that there is a difference in the proportions of men and women with linked data. In contrast, the p-value of .3852 for marital status suggests that the variation in the proportions for marital status groups might be due to chance.

As seen on the left side of table 2, significant differences (using $p < .05$) in simple proportions with linked data occur for more than a dozen characteristics—by sex, race or ethnic status, education, birthplace and age of immigration, health status, emotional health status, Disability Insurance program

status, work status, earnings, household income, net worth, interview mode, region, and language of interview. In most cases, however, the variation is not large and particularly low or high values tend to occur among relatively small groups. Low proportions among blacks (.704), Hispanics (.681), Asians (.711), American Indians (.711), those with little education (.720), and those with low household incomes (.728) may cause particular concern because of the interest in retirement analyses of these groups.[7] Respondents with the highest earnings (.705), those with the highest assets (.712), and those who reported that they had never worked (.684) also have below-average proportions with linked data. The low proportion for the latter group may reflect the inclusion of a relatively high fraction of respondents without SSNs.

The right side of table 2 (and table 3) addresses the question of whether differences in proportions persist when other factors are held constant. It shows the results of a logit regression with variables entered additively, that is, without interaction terms. The dependent variable equals 1 for those with linked data, and zero otherwise. As noted, some variables are omitted from the regression because of multicollinearity concerns. Shown are the individual logit coefficients, chi square statistics and their associated p-values, and an overall p-value for each variable. Logit coefficients were constrained to sum to zero across the categories of each of the predictor variables.[8] When comparing coefficients for two categories of a particular variable, respondents in the category with the larger coefficient have a higher probability of having a linked record than those with a smaller coefficient, other factors held constant. For each variable, the overall p-value in the right-most column refers to the null hypothesis that the coefficients for all categories of the variable are zero.

For example, for race or ethnic status, the overall p-value of < .00005 suggests that the significant variation by race seen on the left persists when other factors are controlled. The large positive coefficient for whites (p=.0001) shows they have significantly higher linkage rates than other race or ethnic groups, but the relationships for the other groups are not clear. For marital status, while the results on the left side suggest no variation, the logit results show that, controlling for other factors, married respondents are more likely to have linked data (p=.0191) than the nonmarried groups. There appears to be little variation across the nonmarried categories shown.

A comparison of results on the left and right sides shows that most variables with significant relationships in the tabulations when nothing is held constant continue to exhibit significant variation when other factors are held constant. Only sex, health status, and work status drop out, although the coefficient for those who have never worked remained significantly lower than other categories. As noted, those who had never worked would be more likely than others not to have an SSN to report. Controlling for other factors also clarifies that those with higher net worth are less likely to have linked data. The almost monotonic relationship of the coefficients for net worth persists even when household income is omitted and net worth is the only financial variable in the logit (result not shown).

The very strong role played by the 1992 interview characteristics is worth some attention. It is not surprising that interviews taken with the 476 proxy respondents resulted in low linkage rates.[9] As previously noted, the data consent agreements needed the personal signature of the respondent, and only about 30 percent of cases with 1992 proxy interviews had linked data after interviewing attempts in 1994 and 1996 as well as in 1992. Results for telephone interviews with respondents themselves are similarly discouraging: only half of the relatively small number of 1992 personal interviews completed by telephone eventually resulted in successful linkages.[10]

Linkage patterns are also presented by region, using classifications of the Bureau of the Census. Differences in the characteristics of regional populations may account for regional differences in linkage rates. However, regional differences might also reflect variations in field procedures. Since significant differences by region remain after controlling for many demographic factors, it may be that linkage rates could be improved by reviewing methods or techniques used by interviewers in the more successful regions for ideas that might lead to greater success with the consent forms in the future.[11]

Finally, below-average proportions (.691) of respondents with interviews conducted in Spanish have linked data. The large majority of these respondents are immigrants, and it may be that they are less likely to have SSNs.[12] Other factors may also be important. The ISR is considering translating the consent agreements into Spanish for future interviews with new HRS cohorts; insofar as language is a factor in the lower linkage rate, such a change might encourage greater participation by Spanish speakers.

Patterns in linkage rates among all respondents are shown in table 3. They are generally similar to those for the age-eligible group, with work status as something of an exception. In simple tabulations done on either sample, the proportion with linked data is lower among those who have never worked and somewhat higher among the unemployed. When other factors are controlled, however, the effects of work status remain significant only in the full sample.

Discussion

The HRS constitutes SSA's first use of consent forms to release identifiable data from SSA records for outside research and statistical purposes,[13] and the overall results are encouraging. The effort achieved linkage rates that are comparable to related rates in other surveys. For example, SSN reporting has been in the 70-74 percent range in the early 1990s in the Survey of Income and Program Participation (SIPP) conducted by the Bureau of the Census. Age breakdowns are available for SSN reporting rates in the 1996 SIPP when the overall rate was 74 percent. Reporting rates show a low of 63 percent for children under age 18, with rates of 78 percent for those aged 18-64 and 82 percent for those aged 65 or older.[14] The overall HRS 75-percent linkage rate is quite comparable, despite several differences in the nature of the estimates. In particular, the SIPP reporting task is a simpler (and perhaps less imposing)

one than the review and signature needed for the consent agreement in the HRS. Other factors that might lead to higher SIPP rates include long experience of SIPP interviewers with SSN collection, special efforts in 1996 to improve reporting, and the fact that the SIPP rates include a small proportion of cases that will prove incorrect in the subsequent validation process. On the other hand, the SIPP rates result from a single attempt rather that the three potential attempts in the HRS.

Comparisons with SSN reporting in the Current Population Survey (CPS) are more favorable to the HRS. In March 1998, the CPS reporting rate for respondents aged 15 or older was only 56 percent.[15] CPS interviews can be based on a single respondent for the household and that may be a primary reason for the lower rates.

Still, linked data from SSA records are not available for 25 percent of HRS respondents, and proportions with linked data differ among respondent subgroups. Two general strategies are available to address biases that might result. First, HRS has developed adjusted person weights for the 1992 HRS earnings data by cross-classifying respondents based on sex, race or ethnic status (Hispanic, black, all others), marital status, and age (whether older than 55). Documentation that accompanies the HRS Covered Earnings File contains a full description of the construction of the adjusted weights. Users seeking population estimates, for example, should use those weights instead of the person weights available for the full sample.

Second, the HRS has considerable survey information on work and benefit history to aid the development of comparable estimates for the missing quarter sample. The baseline 1992 HRS interview obtained information on tenure and wages in the current job or, for those not working, in the most recent job. Similar information on previous jobs lasting 5 years or more was also obtained as were periods of employment with federal, state, county, or local governments. Further job history questions, designed in part to aid the estimation of covered earnings for those without the linked earnings data described here, were asked in the 1996 interview. On the benefit side, the 1992 HRS has data on periods of previous and current receipt of both Social Security and SSI disability benefits. Amounts of SSI and Social Security income received and the type of Social Security benefit (for example, retirement, disability, and survivors) were also obtained. For both earnings and benefits, post-1991 receipt and amounts are obtained in each successive HRS interview.

Research Using the Linked Data

By the end of 1998, several dozen projects were underway or planned among researchers who had obtained HRS approval for access to the restricted data.[16] Several of these projects are looking at retirement behavior and the timing of retirement, often with a focus on the effects of health insurance coverage and incentives in pension plans. Dynamic models of retirement are being developed to estimate the impacts of potential Social Security reforms on employment and savings behavior. Several projects also look at total wealth among older workers and its relation to pensions, Social Security, and earnings histories. Sources of the heterogeneity in wealth across households are also being examined. Pension wealth and subgroup differences in pension wealth are under study.

At least two projects anticipate widowhood, one of which assesses the adequacy of savings and life insurance. The other estimates effects on poverty among widows of changes in the Social Security survivors insurance benefit formula. Some work also focuses on differences between immigrants and the native born in their participation in Social Security retirement and disability programs, work histories, and family patterns. Levels of eligibility for Social Security retirement benefits among immigrants are also being examined as part of a larger project on the economic status of older immigrants.

In addition to these projects, new applications for restricted data access continue to be received and processed by ISR. With the full public release of the 1996 HRS data expected this year, three sets of HRS interviews will be available and include health, work, and retirement reports while respondents are aged 55-65. The greater longitudinality and older sample are likely to expand the number of users conducting research on Social Security policy and retirement income security issues. The availability of earnings and benefit histories from SSA records will expand the range of that research and increase the accuracy of the results.

Chart 1.—Cohorts of the Health and Retirement Study

Survey and start:	*Health and Retirement Survey (HRS), 1992*
Sample (age at start):	1931-41 cohorts (aged 51-61) and spouses
Size at start:	12,652 including spouses of other birth cohorts
Primary mode:	1992—face-to-face; 1994, 1996, and 1998—mostly telephone followups
Survey and start:	*Asset and Health Dynamics of the Oldest Old (AHEAD), 1993*
Sample (age at start):	1923 and earlier birth cohorts (aged 70 or older) and spouses
Size at start:	8,224 including spouses of other birth cohorts
Primary mode:	1993—telephone if aged 70-79; face-to-face if aged 80 or older
Survey and start:	*War Babies (WB), 1998*
Sample (age at start):	1942-47 cohorts (aged 51-56) and spouses
Estimated size at start:	2,511 plus spouses from HRS, AHEAD, and CODA
Primary mode:	1998—face-to-face
Survey and start:	*Children of the Depression (CODA), 1998*
Sample (age at start):	1924-30 cohorts (aged 68-74) and spouses
Estimated size at start:	2,355 plus spouses from HRS, AHEAD, and WB
Primary mode:	1998—face-to-face
Proposed:	*Early Boomers* (born 1948-53), starting *2004*. A long-term plan would add a new 6-year cohort aged 51-56 every 6 years, keeping the study sample representative of the population aged 51 or older.

Table 1.—Number and percent of 1992 Health and Retirement Study (HRS) respondents with data derived from Social Security administrative records,[1] by type of data and HRS age eligibility status

Type of data	Total		Age eligible (1931-41 birth cohort)		
	Number (unweighted)	Percent (unweighted)	Number (unweighted)	Percent (unweighted)	Percent (weighted)
Total respondents	12,652	100.0	9,824	100.0	100.0
With Social Security data	9,478	74.9	7,373	75.1	75.4
Earnings data	9,472	74.9	7,370	75.0	75.4
RSDI benefit data	1,237	9.8	667	6.8	6.4
SSI benefit data	228	1.8	186	1.9	1.6

[1] Shows the number of 1992 interviewed cases with valid SSNs for whom an administrative record was successfully found following the 1992, 1994, or 1996 HRS interviewing cycle. Data are from HRS's Version 3 Earnings History; Retirement and Survivors Disability Insurance (RSDI); and Supplemental Security Income (SSI) data files.

Table 2.—Characteristics of 1992 age-eligible Health and Retirement Study (HRS) respondents: Proportion with data derived from Social Security administrative records and logistic regression results

Characteristic	Tabulation results				Main effects regression results[2]			
	Total	Proportion (unweighted)	p-value[1]	Proportion (weighted)	Logit coefficient	Chi square	p-value	Variable p-value
Total	9,824	0.751		0.754	(3)	(3)	(3)	(3)
Part I. Demographic Characteristics								
Sex:								
Men	4,602	.737		.742	-0.0403	2.27	0.1322	
Women	5,222	.763	(0.0016)	.767	.0403	2.27	.1322	(0.1322)
Race or ethnic status:								
White	7,024	.772		.768	.3047	15.89	.0001	
Black	1,685	.704		.699	-.1471	3.00	.0833	
Hispanic	918	.681		.684	-.1276	1.75	.1853	
Asian	114	.711		.716	.0532	.08	.7751	
American Indian	83	.711	(.0000)	.726	-.0832	.16	.6892	(.0000)
1992 marital status:								
Married	7,226	.755		.762	.1347	5.50	.0191	
Partner	248	.730		.743	.0463	.12	.7280	
Separated	310	.719		.708	-.0476	.17	.6845	
Divorced	1,065	.746		.744	-.0669	.79	.3741	
Widowed	617	.737		.737	-.0410	.20	.6514	
Never married	358	.732	(.3852)	.717	-.0254	.05	.8204	(.1626)
1992 age group:								
51-54	3,347	.752		.760	.0248	.45	.5007	
55-59	4,564	.749		.749	-.0033	.01	.9220	
60-61	1,913	.752	(.9229)	.760	-.0215	.26	.6107	(.7932)
Education:								
Unknown	59	.559		.513	-.3353	1.78	.1826	
Less than 8 years	1,120	.720		.722	.0676	.69	.4060	
8-12 without diploma	1,477	.756		.769	.1878	6.36	.0117	
H.S. diploma/G.E.D.	3,681	.771		.772	.1589	6.68	.0098	
Some college/A.A./A.S.	1,873	.736		.737	-.0752	1.19	.2759	
Bachelor's degree	913	.735		.737	-.0332	.15	.7018	
More than B.A./B.S.	701	.756	(.0001)	.767	-.0752	1.19	.2759	(.0093)
Birthplace and age upon U.S. entry:								
Born in United States	8,832	.756		.759	(2)	(2)	(2)	
Entry age unknown	28	.536		.574	(2)	(2)	(2)	
Before age 20	167	.766		.762	(2)	(2)	(2)	
20-29	319	.715		.721	(2)	(2)	(2)	
30-39	258	.733		.736	(2)	(2)	(2)	
40 or older	220	.632	(.0000)	.645	(2)	(2)	(2)	(2)

See footnotes at end of table.

Table 2.—Characteristics of 1992 age-eligible Health and Retirement Study (HRS) respondents: Proportion with data derived from Social Security administrative records and logistic regression results—*Continued*

Characteristic	Tabulation results				Main effects regression results[2]			
	Total	Proportion (unweighted)	p-value[1]	Proportion (weighted)	Logit coefficient	Chi square	p-value	Variable p-value
Part II. 1992 Health, Disability, and Work Characteristics								
Health:								
Excellent	2,130	0.761		0.763	0.0342	0.30	0.5852	
Very good	2,736	.763		.770	-.0002	.00	.9977	
Good	2,731	.747		.750	-.0509	1.03	.3092	
Fair	1,415	.722		.722	-.0659	1.19	.2751	
Poor	812	.743	(0.0368)	.742	.0828	.74	.3890	(0.5819)
Health compared with last year:								
Much better	534	.800		.795	.2466	6.02	.0141	
Somewhat better	977	.747		.759	-.0828	1.21	.2716	
Same	7,067	.749		.754	-.0385	.50	.4791	
Somewhat worse	974	.746		.751	-.0117	.02	.8765	
Much worse	272	.713	(.0585)	.695	-.1136	.73	.3918	(.1461)
Emotional health:								
Excellent	1,900	.753		.753	.0253	.17	.6780	
Very good	2,852	.776		.780	.1676	9.94	.0016	
Good	3,247	.745		.751	.0386	.63	.4266	
Fair	1,389	.712		.719	-.1609	7.07	.0079	
Poor	436	.734	(.0001)	.730	-.0705	.45	.5038	(.0050)
Work limitation status:								
Not limited	7,785	.751		.756	-.0346	.24	.6255	
Somewhat limited	711	.782		.777	.1508	2.66	.1031	
Part time/occasional	148	.743		.719	-.0744	.24	.6230	
Cannot work	1,180	.731	(.0821)	.734	-.0418	.22	.6427	(.3356)
Disability Insurance program:								
Not disabled	7,782	.750		.756	(2)	(2)	(2)	
Never applied	960	.763		.764	(2)	(2)	(2)	
Awarded/in process—								
on first application	453	.746		.745	(2)	(2)	(2)	
on appeal/reapplication	260	.746		.748	(2)	(2)	(2)	
Denied/DK status—								
on first application	207	.672		.661	(2)	(2)	(2)	
on appeal/reapplication	162	.821	(.0397)	.809	(2)	(2)	(2)	(2)
HRS-reported work status:								
Never worked	374	.684		.663	-.2059	4.11	.0427	
Out of the labor force	2,418	.755		.762	.0460	.52	.4695	
Unemployed	326	.779		.778	.1343	1.46	.2271	
Working	6,706	.751	(.0122)	.756	.0255	.18	.6676	(.2286)

See footnotes at end of table.

Table 2.—Characteristics of 1992 age-eligible Health and Retirement Study (HRS) respondents: Proportion with data derived from Social Security administrative records and logistic regression results—*Continued*

Characteristic	Tabulation results				Main effects regression results[2]			
	Total	Proportion (unweighted)	p-value[1]	Proportion (weighted)	Logit coefficient	Chi square	p-value	Variable p-value
Part III. 1991 Financial Characteristics								
HRS-reported earnings:								
Unknown	71	0.535		0.498	(2)	(2)	(2)	
Zero	2,719	.733		.735	(2)	(2)	(2)	
Negative or $1-$21,812	3,428	.772		.777	(2)	(2)	(2)	
$21,812-$53,399	2,885	.751		.759	(2)	(2)	(2)	
$53,400-$99,999	582	.737		.741	(2)	(2)	(2)	
$100,000 or more	139	.705	(0.0055)	.719	(2)	(2)	(2)	(2)
Household income:								
Unknown	71	.535		.498	(2)	(2)	(2)	
Negative or zero	81	.642		.677	-0.5155	6.07	0.0138	
$1-$20,000	2,140	.728		.735	-.1644	3.78	.0518	
$20,000-$49,999	3,821	.756		.755	.0022	.00	.9766	
$50,000-$99,999	2,885	.769		.775	.1788	5.40	.0202	
$100,000-$249,999	742	.749		.750	.1872	3.38	.0661	
$250,000 or more	84	.762	(.0051)	.782	.3118	1.74	.1873	(0.0009)
Nonhousing equity:								
Unknown	71	.535		.498	(2)	(2)	(2)	
Negative or zero	1,131	.751		.756	(2)	(2)	(2)	
$1-$9,999	1,743	.759		.771	(2)	(2)	(2)	
$10,000-$49,999	2,451	.754		.759	(2)	(2)	(2)	
$50,000-$99,999	1,358	.752		.758	(2)	(2)	(2)	
$100,000-$199,999	1,247	.774		.774	(2)	(2)	(2)	
$200,000-$499,999	1,072	.735		.740	(2)	(2)	(2)	
$500,000 or more	751	.720	(.1534)	.716	(2)	(2)	(2)	(2)
Total net worth:								
Unknown	71	.535		.498	(2)	(2)	(2)	
Negative or zero	810	.743		.748	.3114	12.81	.0003	
$1-$9,999	798	.752		.772	.2244	7.51	.0061	
$10,000-$49,999	1,666	.786		.796	.3136	26.00	.0000	
$50,000-$99,999	1,704	.748		.759	-.0626	1.18	.2781	
$100,000-$199,999	1,984	.756		.762	-.1344	5.70	.0170	
$200,000-$499,999	1,828	.747		.745	-.2320	14.86	.0001	
$500,000 or more	963	.712	(.0040)	.708	-.4204	27.96	.0000	(.0000)
Part IV. 1992 Interview Characteristics								
Interview mode/type:								
Face-to-face; own	8,869	.786		.791	1.3819	444.47	.0000	
Face-to-face; proxy	344	.334		.319	-.6380	37.86	.0000	
Telephone; own	479	.511		.510	.1350	2.16	.1416	
Telephone; proxy	132	.288	(.0000)	.300	-.8788	33.68	.0000	(.0000)
Region:								
New England	377	.743		.751	-.0003	.00	.9982	
Middle Atlantic	1,406	.724		.725	-.1421	4.94	.0263	
East North Central	1,528	.776		.781	.0784	1.47	.2247	
West North Central	816	.819		.820	.2733	9.35	.0022	
South Atlantic	2,518	.749		.756	.0331	.39	.5328	
East South Central	624	.760		.747	.0102	.01	.9120	
West South Central	992	.708		.691	-.1296	3.07	.0797	
Mountain	403	.737		.771	-.1489	1.86	.1726	
Pacific	1,160	.743	(.0000)	.757	.0258	.13	.7170	(.0132)
Language of interview:								
English	9,397	.753		.757	(2)	(2)	(2)	
Spanish	427	.691	(.0022)	.681	(2)	(2)	(2)	(2)

[1] Chi square tests were done excluding the 71 age-eligible cases with 1991 financial status unknown.

[2] Model excludes the 71 age-eligible cases with 1991 financial status unknown. Birthplace, DI Program status, HRS earnings, nonhousing equity, and language of interview are excluded because of multicollinearity concerns.

[3] Not applicable.

Table 3.—Characteristics of 1992 Health and Retirement Study (HRS) respondents: Proportion with data derived from Social Security administrative records and logistic regression results

Characteristic	Tabulation results			Main effects regression results[2]			
	Total	Proportion (unweighted)	p-value[1]	Logit coefficient	Chi square	p-value	Variable p-value
Total	12,652	0.749		(3)	(3)	(3)	(3)
Part I. Demographic Characteristics							
Sex:							
Men	5,867	.738		-0.0301	1.47	0.2260	
Women	6,785	.759	(0.0038)	.0301	1.47	.2260	(0.2260)
Race or ethnic status:							
White	9,131	.771		.3112	22.13	.0000	
Black	2,072	.699		-.1733	5.48	.0192	
Hispanic	1,180	.680		-.1294	2.38	.1228	
Asian	155	.677		-.0578	.14	.7120	
American Indian	114	.719	(.0000)	.0492	.07	.7869	(.0000)
1992 marital status:							
Married	9,894	.752		.1173	4.87	.0273	
Partner	385	.738		.0962	.76	.3833	
Separated	312	.718		-.0594	.27	.6058	
Divorced	1,079	.749		-.0624	.73	.3943	
Widowed	620	.737		-.0444	.25	.6174	
Never married	362	.729	(.4868)	-.0473	.19	.6668	(.2063)
1992 age group:							
Under age 51	1,229	.745		.0757	1.33	.2483	
51-54	3,347	.752		.0412	.97	.3246	
55-59	4,564	.749		.0131	.12	.7278	
60-61	1,913	.752		-.0106	.04	.8343	
62 or older	1,599	.744	(.9744)	-.1194	4.21	.0402	(.3031)
Education:							
Unknown	83	.614		-.1554	.52	.4727	
Less than 8 years	1,493	.720		.0631	.79	.3734	
8-12 without diploma	1,858	.747		.1207	3.43	.0639	
H.S. diploma/G.E.D	4,686	.771		.1374	6.53	.0106	
Some college/A.A./A.S	2,442	.739		-.0782	1.68	.1948	
Bachelor's degree	1,200	.726		-.1138	2.29	.1302	
More than B.A./B.S	890	.760	(.0000)	-.0782	1.68	.1948	(.0052)
Birthplace and age upon U.S. entry:							
Born in United States	11,358	.755		(2)	(2)	(2)	
Entry age unknown	41	.512		(2)	(2)	(2)	
Before age 20	235	.770		(2)	(2)	(2)	
20-29	431	.705		(2)	(2)	(2)	
30-39	308	.718		(2)	(2)	(2)	
40 or older	279	.631	(.0000)	(2)	(2)	(2)	(2)

See footnotes at end of table.

Table 3.—Characteristics of 1992 Health and Retirement Study (HRS) respondents: Proportion with data derived from Social Security administrative records and logistic regression results—*Continued*

Characteristic	Tabulation results			Main effects regression results[2]			
	Total	Proportion (unweighted)	p-value[1]	Logit coefficient	Chi square	p-value	Variable p-value
Part II. 1992 Health, Disability, and Work Characteristics							
Health:							
Excellent	2,807	0.763		0.0945	2.95	0.0859	
Very good	3,481	.758		.0307	.41	.5195	
Good	3,544	.747		-.0108	.06	.8056	
Fair	1,807	.725		-.0720	1.81	.1780	
Poor	1,013	.731	(0.0147)	-.0425	.25	.6185	(0.4291)
Health compared with last year:							
Much better	714	.800		.2056	5.54	.0186	
Somewhat better	1,277	.749		-.0776	1.36	.2434	
Same	9,072	.746		-.0833	2.96	.0852	
Somewhat worse	1,248	.748		-.0162	.06	.8080	
Much worse	341	.721	(.0250)	-.0284	.06	.8133	(.0717)
Emotional health:							
Excellent	2,455	.754		-.0146	.07	.7858	
Very good	3,692	.766		.0794	2.90	.0885	
Good	4,189	.746		.0266	.38	.5365	
Fair	1,772	.717		-.1204	4.97	.0257	
Poor	544	.737	(.0021)	.0290	.09	.7601	(.0994)
Work limitation status:							
Not limited	10,040	.749		-.0627	1.05	.3047	
Somewhat limited	919	.778		.0803	1.00	.3172	
Part time/occasional	209	.766		-.0085	.00	.9482	
Cannot work	1,484	.732	(.0632)	-.0091	.01	.9083	(.4603)
Disability Insurance program status:							
Not disabled	10,037	.748		(2)	(2)	(2)	
Never applied	1,274	.761		(2)	(2)	(2)	
Awarded/in process—							
on first application	581	.738		(2)	(2)	(2)	
on appeal/reapplication	309	.748		(2)	(2)	(2)	
Denied/DK status—							
on first application	256	.688		(2)	(2)	(2)	
on appeal/reapplication	195	.826	(.0291)	(2)	(2)	(2)	(2)
HRS-reported work status:							
Never worked	474	.671		-.2528	7.94	.0048	
Out of the labor force	3,398	.753		.0326	.34	.5614	
Unemployed	399	.787		.1854	3.37	.0663	
Working	8,381	.750	(.0002)	.0349	.44	.5059	(.0451)

See footnotes at end of table.

Table 3.—Characteristics of 1992 Health and Retirement Study (HRS) respondents: Proportion with data derived from Social Security administrative records and logistic regression results—*Continued*

Characteristic	Tabulation results			Main effects regression results[2]			
	Total	Proportion (unweighted)	p-value[1]	Logit coefficient	Chi square	p-value	Variable p-value
Part III. 1991 Financial Characteristics							
HRS-reported earnings:							
Unknown	95	0.516		(2)	(2)	(2)	
Zero	3,698	.734		(2)	(2)	(2)	
Negative or $1-$21,812	4,477	.768		(2)	(2)	(2)	
$21,812-$53,399	3,519	.751		(2)	(2)	(2)	
$53,400-$99,999	703	.735		(2)	(2)	(2)	
$100,000 or over	160	.731	(0.0091)	(2)	(2)	(2)	(2)
Household income:							
Unknown	95	.516		(2)	(2)	(2)	
Negative or zero	93	.656		-0.4519	5.31	0.0212	
$1-$19,999	2,601	.724		-.1815	5.70	.0170	
$20,000-$49,999	4,973	.756		-.0085	.02	.8988	
$50,000-$99,999	3,801	.767		.1606	5.47	.0194	
$100,000-$249,999	979	.740		.1378	2.37	.1237	
$250,000 or over	110	.773	(.0006)	.3435	2.73	.0987	(0.0002)
Nonhousing equity:							
Unknown	95	.516		(2)	(2)	(2)	
Negative or zero	1,365	.745		(2)	(2)	(2)	
$1-$9,999	2,216	.759		(2)	(2)	(2)	
$10,000-$49,999	3,222	.757		(2)	(2)	(2)	
$50,000-$99,999	1,758	.749		(2)	(2)	(2)	
$100,000-$199,999	1,625	.765		(2)	(2)	(2)	
$200,000-$499,999	1,407	.737		(2)	(2)	(2)	
$500,000 or over	964	.720	(.1340)	(2)	(2)	(2)	(2)
Total net worth:							
Unknown	95	.516		(2)	(2)	(2)	
Negative or zero	952	.738		.2632	11.20	.0008	
$1-$9,999	997	.748		.1930	7.02	.0081	
$10,000-$49,999	2,149	.784		.2998	30.55	.0000	
$50,000-$99,999	2,228	.750		-.0328	.42	.5184	
$100,000-$199,999	2,594	.755		-.1117	5.14	.0234	
$200,000-$499,999	2,392	.745		-.2069	15.38	.0001	
$500,000 or over	1,245	.710	(.0003)	-.4045	33.84	.0000	(.0000)
Part IV. 1992 Interview Characteristics							
Interview mode/type:							
Face-to-face; own	11,374	.787		1.3932	611.16	.0000	
Face-to-face; proxy	465	.316		-.7136	63.19	.0000	
Telephone; own	630	.521		.1762	4.92	.0266	
Telephone; proxy	183	.290	(.0000)	-.8559	44.26	.0000	(.0000)
Region:							
New England	468	.731		-.0470	.21	.6466	
Middle Atlantic	1,774	.722		-.1403	6.11	.0134	
East North Central	1,982	.767		.0398	.50	.4795	
West North Central	1,057	.814		.2282	8.64	.0033	
South Atlantic	3,260	.754		.0786	2.80	.0941	
East South Central	817	.763		.0432	.28	.5973	
West South Central	1,273	.711		-.0855	1.70	.1918	
Mountain	506	.737		-.1253	1.66	.1982	
Pacific	1,515	.737	(.0000)	.0084	.02	.8927	(.0087)
Language of interview:							
English	12,090	.752		(2)	(2)	(2)	
Spanish	562	.692	(.0006)	(2)	(2)	(2)	(2)

[1] Chi square tests were done excluding the 95 total cases with financial status unknown.

[2] Model excludes the 95 cases with 1991 financial status unknown. Birthplace, DI program status, HRS earnings, nonhousing equity, and language of interview are excluded because of multicollinearity concerns.

[3] Not applicable.

Notes

[1] The initial HRS response rate was 82 percent. HRS data were weighted to adjust for nonresponse and design features (for example, blacks, Hispanics, and residents of the state of Florida were oversampled). In the larger HRS, most of these spouses or partners are or will be incorporated into other cohort-based HRS surveys.

[2] The detailed pension plan data that were collected following the 1992 interviews are available from HRS/Michigan only under restricted conditions. A software package is also available that allows researchers to generate pension benefits for workers with the pension plan data. Gustman et al. (forthcoming) provide information on pension data linkages. In 1999, HRS is collecting new pension plan data from employers identified in the 1998 interviewing cycle to identify changes in plan content or changes resulting from job shifts.

[3] All HRS data derived from SSA records are available for research purposes only under restricted conditions from the HRS staff at the ISR, University of Michigan (see *www.umich.edu/~hrswww)*. At this time, the files may be linked with the 1992 and 1994 HRS public use files. ISR specifies permissible linkages with other restricted files when access to the restricted data is approved.

[4] Forms were processed according to Internal Revenue Service requirements, which included a 60-day deadline for SSA's receipt of the form. Some respondents, who may have consented in 1992, may not have been sufficiently committed to do so again in 1994 or 1996 when a higher level of effort was required. Those interviews were generally by telephone and relied heavily on mailout-mailback procedures for the consent forms.

[5] The version 3 HRS files used here include data based on consents given in 1992, 1994, or 1996.

[6] For a discussion of p-values, see, for example, Wonnacott and Wonnacott, Section 9.2, 1977.

[7] The HRS experienced considerably higher attrition among Hispanics by the 1994 interview and, to a lesser extent, higher attrition among blacks, according to preliminary work by Hill and Willis (1997). Insofar as attriters in 1994 (or 1996) did not have a validated SSN from a consent form signed in 1992, they will be less likely to have linked data after the 1994 (or 1996) interview.

[8] For a statistical reference, see Chapter 2 of Bishop, Fienberg, and Holland (1975). The logits were estimated with SAS using PROC CATMOD. (See SAS Institute, Inc., 1989, especially pp. 452-453). Unweighted data were used. Unweighted data were used. The coefficient, chi square, and p-values for the dropped category have been computed to provide more complete information.

[9] If an interview could not be taken because of physical or cognitive limitations, a proxy *interview* could be done. For a small number of other cases, proxies were used when the sample person was unavailable, could not speak or understand sufficient English or Spanish, or was unwilling to be personally interviewed but did not object to another person providing information for him or her (Survey Design: Study Detail, revised December 3, 1998, *www.umich.edu/~hrswww)*.

[10] Procedures for handling consent forms in telephone interview situations were newly developed for the 1992 start. As noted, mailout-mailback procedures require greater respondent cooperation, which probably contributed to the generally lower rates.

[11] Although regional office boundaries differ in the SIPP and the HRS, some general comparisons of SSN reporting rates in the SIPP (discussed subsequently) with linkage rates in HRS are possible for six regions. The West North Central region shows quite high levels of cooperation in both surveys and the Middle Atlantic shows low ones. No particular similarities appear for the New England, East North Central, Mountain, or Pacific regions.

[12] For income tax purposes, a nonworking alien or spouse would use an individual taxpayer identification number issued by the Internal Revenue Service rather than the SSN.

[13] As noted, University of Michigan guarantees the privacy and confidentiality of the data.

[14] "Missing Social Security Numbers in Wave 1 of the 1996 Panel." Unpublished memorandum for The SIPP Record 96-07 from Michael F. McMahon, Demographic Surveys Division, U.S. Bureau of the Census, September 25, 1996. Final linkage rates of approximately 90 percent are obtained in the SIPP because attempts are made to locate SSNs for respondents not reporting SSNs (except for those who refused) in the SSN validation process. Such attempts have not been made in the HRS because SSA's release of identifiable data to ISR requires the SSN and signed consent of respondents.

[15] Unpublished CPS estimates were obtained from Greg Weyland, Demographic Surveys Division, U.S. Bureau of the Census. As it does in the SIPP, the validation process generally greatly increases final linkage rates in the CPS.

[16] "Annual Report on Distribution of Restricted Data Files." Memorandum from Cathy Liebowitz to Jan Olson and Robert Willis, December 10, 1998.

References

Bishop, Yvonne M.; Stephen E. Fienberg; and Paul W. Holland. 1975. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: The MIT Press.

Gustman, Alan L.; Olivia S. Mitchell; Andrew Samwick; and Thomas L. Steinmeier. Forthcoming. "Pension and Social Security Wealth in the Health and Retirement Study." In *Wealth, Work and Health: Innovations in Measurement in the Social Sciences,* James P. Smith and Robert J. Willis (eds.). Ann Arbor, MI: University of Michigan Press.

Hill, Dan. 1991. "HRS Wave 3 Response Rates." Memorandum to HRS Researchers. Ann Arbor, MI: Survey Research Center (May 28).

Hill, Daniel and Robert J. Willis. 1997. "Minority Attrition in the Health and Retirement Study: Racial and Ethnic Differences in the Response Process." Ann Arbor, MI: Survey Research Center. Unpublished paper (February 3).

Liebowitz, Cathy. 1998. "Annual Report on Distribution of Restricted Data Files." Memorandum to Jan Olson and Robert Willis. Ann Arbor, MI: Survey Research Center (December 10).

McMahon, Michael F. 1996. "Missing Social Security Numbers in Wave 1 of the 1996 Panel." Memorandum for The SIPP Record 96-07. Demographic Surveys Division, U.S. Bureau of the Census (September 25).

Mitchell, Olivia S.; Jan Olson; and Thomas Steinmeier. Forthcoming. "Earnings and Benefits in the Health and Retirement Study. " In *Forcasting Retirement Needs and Retirement Wealth*, O. Mitchell, B. Hammond, and A. Rappaport (eds.). Philadelphia: University of Pennsylvania Press. An earlier version of the paper appeared as "Construction of the Earnings and Benefits File (EBF) for Use with the Health and Retirement Survey" as NBER Working Paper Number 5707 and as HRS/AHEAD Documentation Report Number DR-001.

Olson, Janice A. 1996. "The Health and Retirement Study: The New Retirement Survey." *Social Security Bulletin*, Vol. 59, No. 1 (Spring), pp. 85-88.

SAS Institute, Inc. 1989. *SAS/STAT User's Guide*, Version 6, Fourth Edition, Volume 1, Chapter 17. Cary, NC: SAS Institute, Inc.

Wonnacott, Thomas H. and Ronald J. Wonnacott. 1977. *Introductory Statistics*, Third Edition. New York: John Wiley & Sons, Inc.