HRS | HEALTH AND RETIREMENT STUDY
A Longitudinal Study of Health, Retirement, and Aging
Sponsored by the National Institute on Aging

An Overview of Linked Social Security Administration Data in the Health and Retirement Study

User Guide

Joelle Abramowitz
Chichun Fang
Jody Schimmel Hyde

Survey Research Center
Institute for Social Research
University of Michigan
Ann Arbor, Michigan

May 24, 2024

Funding
The Health and Retirement Study is funded by a grant from the National Institute on Aging (U01 AG009740) with supplemental support from the Social Security Administration. HRS is conducted by the University of Michigan.

Suggested Citation
Abramowitz, J., Fang, C., & Hyde, J. S. (2024). An Overview of Linked Social Security Administration Data in the Health and Retirement Study. University of Michigan. https://hrs.isr.umich.edu/publications/biblio/14167

Table of Contents

In this user guide, we discuss the linkage of the Health and Retirement Study (HRS) to respondents' earnings and benefits records from the Social Security Administration (SSA). The SSA administers three programs that support many older adults in the United States: the Old-Age and Survivors Insurance (OASI) program, the Social Security Disability Insurance (SSDI) program, and the Supplemental Security Income (SSI) program. Taken together, these programs are a critical part of the social safety net. Benefit eligibility and payment amounts are connected to earnings—of an individual or others in their household—and as a result, the SSA also collects detailed earnings information derived from income tax records.

Linkages between survey and administrative data have the potential to increase the research value of both kinds of data. For benefits programs, administrative data can provide the universe of program participants as well as details about their participation. For earnings, administrative data can provide amounts of earnings of different types as they are reported to tax authorities. However, administrative data lack information outside of their domains and in the case of earnings data, may be sensitive to changes over time in reporting rules. Surveys complement administrative data because they ask respondents about a breadth of topics, for example, their demographics and household structures, as well as more complicated and evolving topics such as their retirement expectations, time use, and functional limitations.

Not all HRS survey respondents consent to having their records linked, and not all of those who consent are ultimately matched to the administrative files. But the information available for those who do is rich. These administrative files can be linked to the HRS after receiving approval from the University of Michigan to conduct research on a stated objective. As we describe below, there are also several derivative files based on the administrative data that

can be used without approval to access restricted data, though the bulk of SSA files require permission for use.

This guide provides information about the available SSA data products and the process for matching those records. In Section 1, we provide background about the programs that SSA administers and information collected about those programs and related measures in the HRS. We then provide an overview of the administrative files as well as files derived from them in Section 2. In Section 3, we outline the process by which HRS respondents consent to having their records linked to SSA files as well as the steps to successfully match consented records. In Section 4, we describe weights that account for the fact that not all HRS respondents consent to or have their records successfully linked to the SSA administrative files. These non-response weights can be used by researchers in place of those that the HRS produces for the main survey to produce population-representative estimates.

1 Background

1.1 Programs Administered by SSA

SSA administers programs to support older adults, their dependents, and people who are blind or have significant disabilities. The SSI program is a means-tested safety net program for people of older ages or with disabilities who have limited income and assets. Old Age, Survivors, and Disability Insurance (OASDI) is an insurance program funded by payroll taxes with two components: OASI, commonly referred to as Social Security retirement, and SSDI. As the name indicates, this insurance program is intended to provide income support to workers who cannot work because of age or disability, as well as their dependents in certain instances.

Each of the programs has complicated rules around eligibility and benefit amounts that we do not attempt to document here. Primers including those by SSA (2005/2006) and Moula-

Ali (2013) have more detailed information about SSDI and SSI; Tatem and Huston (2023) provides a more detailed view of the OASI and SSDI programs. The rules surrounding SSDI and SSI benefit eligibility are notoriously complex, and become even more complicated near retirement. Some salient considerations for users interested in the SSA administrative data include:

- OASDI eligibility is based on being "insured" for the program based on quarters of coverage, which are tied to an individual's own earnings, or that of a spouse or parent in certain circumstances. OASI eligibility is based on lifetime quarters of coverage; the DI program requires that a share of the quarters be accumulated in recent years. In both programs, monthly benefit amounts are tied to average indexed monthly earnings (AIME) of the person whose record entitles an individual to benefits. Thus, SSA tracks individuals' lifetime earnings for purposes of setting DI and OASI benefit amounts.
- Unlike OASDI, SSI benefits are set to the federal benefit rate (FBR), though some states offer additional supplements to augment that amount.
- Both SSDI and SSI benefits are reduced if the beneficiary has earnings, though the programs vary dramatically in how earnings affect benefits (Levere et al., 2018).
- For those who are insured, the SSDI program ends at the full-retirement age, or FRA, which is age 67 for those born after 1959. At that point, SSDI benefits convert to OASI. OASI can begin as early as age 62, with monthly benefits reduced to actuarially account for the additional years of benefit receipt before FRA. Between the ages of 62 and FRA, an individual can be covered by DI (if they meet the disability program rules) or OASI, but not both.

- SSI eligibility spans all ages, though eligibility based on disability ends at age 65, at which point eligibility is based on old age. A person can concurrently receive SSI and DI or SSI and OASI. Without sufficient quarters of coverage, a person might not be eligible for DI or OASI but be able to receive SSI if they have income and assets under the program's thresholds.
- After a 24-month waiting period, DI beneficiaries receive Medicare; SSA coordinates with the Centers for Medicare & Medicaid Services on payment of premiums and subsidies for Part B and Part D of the Medicare program. SSI beneficiaries are generally eligible for Medicaid without a wait, but the process to apply for Medicaid varies by state.

Though the SSDI and SSI programs are distinct, the application process for SSDI and SSI are handled jointly. To make a disability determination, SSA uses a five-step sequential determination process; Wixon and Strand (2013) provides a nice overview of the disability determination process. In both programs, applicants must meet the program's financial eligibility criteria (sufficient quarters of coverage in the SSDI program and income and assets below the specified thresholds for SSI). Applicants must also be determined to have a significant medical impairment that is expected to last at least twelve months or result in death and that makes the applicant unable to engage substantially in the labor force (based on a measure called substantial gainful activity, or SGA). If claimants seek OASI before SSA's full retirement age, SSA will consider whether they meet the criteria to be considered for SSDI and/or SSI.

The process to receive an allowance for disability benefits can take months or years. Initial disability applications are processed by state disability determination services (DDS), which first verify that an applicant meets the financial requirements of the program; applications

that do not meet that standard receive a "technical denial." After making that determination, the DDS then makes an initial allowance or denial of benefits and notifies applicants. In some cases, these initial decisions can be reconsidered by the DDS. The DDS decision usually takes months before being shared with the applicant. After that point, applicants have additional options to appeal a denial, first to an administrative law judge (ALJ), then to an Appeals Council, and finally, to federal court. The appeals process can take years.

1.2 The Health and Retirement Study

The HRS is a longitudinal study of a representative sample of approximately 20,000 Americans over age 50 and their spouses. Since 1992 it has been conducted every two years with new cohorts added every six years. The HRS asks questions on a breadth of topics including work history, current employment, disability, retirement plans, net worth, income, health insurance and health status. See Sonnega et al. (2014) for more detail on the design of the study. Here, we focus on features of the HRS that are important for understanding the value of the linked SSA data.

Waves and Cohort Structure: The HRS started in 1992 with a nationally representative sample of the U.S. population aged 51-61 (born between 1931 and 1941). Another study, Asset and Health Dynamics among the Oldest Old (AHEAD), started in 1993 with a nationally representative sample of the U.S. population aged 70 and above (born before 1924). Each newly added sample is called a "cohort." Both HRS and AHEAD were conducted every other year, and each iteration of the longitudinal study is called a "wave." For example, wave 1992 refers to the 1992 interview. The two studies merged in 1998 and retained the HRS name; two cohorts, one born between 1924 and 1930 (Children of the Depression, CODA), and another born between 1942 and 1947 (War Babies, WB), were also added in order to keep the HRS representative of

the U.S. population over aged 50. As the HRS sample ages, new cohorts are added every six years to keep the same age representativeness of the HRS. Early Baby Boomers (EBB, born 1948-53) were added in 2004, Mid Baby Boomers (MBB, born 1954-59) were added in 2010, Late Baby Boomers (LBB, born 1960-65) were added in 2016, and Early Generation X (EGenX, born 1966-71) were added in 2022. Figure 1 illustrates when each cohort was added to the study, how each cohort aged over time, and the representativeness of the HRS in each wave. The vertical axis of Figure 1 is truncated at 90-year-olds only for simplification; the HRS intends to follow respondents until their death, and respondents do not age out of the study.

Survey Mode: When respondents enter the HRS for the first time, their first interviews (i.e., their "baseline interviews") are mostly conducted face-to-face (FTF). Prior to 2004, follow-up interviews could be telephone-based, or FTF for respondents aged 90 or older (Sonnega et al., 2014). Starting in 2006, the HRS re-designed the follow-up interviews so that a randomly chosen half of the sample receives an FTF interview with physical and biological measures taken (Enhanced Face-to-Face (EFTF) interviews). The rotation is built into the HRS sample design, so each respondent will have an EFTF interview every two waves (four years). In non-EFTF waves, respondents are interviewed via telephone. Starting in 2018, in non-EFTF waves, respondents could choose to take web-based interviews. In 2020, pandemic-induced disruptions to in-person interviewing meant that nearly all interviews were conducted by phone or internet. The normal mode split (half in-person, half by phone/internet) was resumed for 2022.

Survey Reports of Earnings and Income: Respondents report about benefits and income in different parts of the survey, using different look-back periods. Each respondent and spouse reports their earnings from their current main job in Section J of the survey; these can be reported in the interval most appropriate to the respondent's job. Section Q collects additional

information about annual income amounts for the respondent and spouse, including earnings and other forms of earned income. These amounts are collected in survey year T for reference year T-1, so that in the HRS survey data there is a sequence of observations from every other year. This income information is generally provided by one household member, the designated household financial respondent, for both the respondent and their spouse.

Survey Reports of Participation in Programs Administered by SSA: As we describe in more detail in the next section, SSA administers federal disability and benefit programs. At each HRS interview, respondents are asked whether they receive any income from Social Security in Section J3.[1] Additionally, respondents are asked about work-limiting health conditions in Section M and provide information about application to and/or receipt of benefits from Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI), for which eligibility is tied to one's ability to work. For individuals who say they have or are currently receiving benefits, respondents report the amount of their last monthly check in Section M.

2 SSA-Linked Data Products

The HRS receives three types of SSA administrative record extracts that are made available to HRS users as restricted data: (1) earnings records that SSA receives from the Internal Revenue Service and uses to determine benefit amounts; (2) a file documenting applications to SSDI and SSI, and (3) data files documenting the receipt of benefits from programs administered by SSA including OASI, SSDI, and SSI.[2] In addition to the data files provided by SSA, HRS produces derivative data products to support their use. Here, we provide

[1] Questions on the eligibility for and receipt of Social Security retirement benefits previously followed the survey questions on employment. They have been included in Section J3 since the 2016 wave.
[2] Given the connection between SSDI, OASI, and Medicare, the benefits files for those programs also include premium amounts for Medicare Parts B, C, and D.

a high-level overview of each of the available files and the information they do (and do not) contain. Each of these files has specific documentation available on the HRS website, including a data description and codebook with information on the file layout and data elements. The files that SSA provides and HRS derivative data products are updated with each HRS survey wave. Table 1 summarizes these data products.

2.1 Administrative Earnings Extracts

Summary Earnings Records (SER)

The SER file is derived from SSA's Master Earnings File (MEF) and contains respondents' annual earnings in jobs covered by Social Security, including covered self-employed and agriculture jobs, back to 1951. All respondents have records back to 1951, even if they were not working or alive in the year; years with no recorded covered earnings have a value of 0. Data are reported at the respondent level; that is, the SER is a "wide" file with separate variables for earnings in each year. The annual earnings reported in this file are top-coded at the OASDI taxable maximum, which increased from $3,600 in 1951 to $168,600 in 2024 (U.S. Social Security Administration, 2023).

Detail Earnings Records (DER)

Like the SER, the DER file is derived from SSA's MEF. The DER is a "long" file containing annual earnings at the respondent-employer-year level, regardless of whether the job is covered by Social Security, back to 1978. Earnings reported in this file are not top-coded, although numbers larger than $250,000 are masked with a range indicator ($250K-$300K, $300K-$500K, or $500K+) to preserve confidentiality. Starting in 2004, records also contain tax-

deferred contributions into employer-sponsored employee benefits plans such as 401(k), 403(b), 457, or health savings accounts.

2.2 File documenting applications to SSDI and SSI

Form 831 records contain information on disability determinations made at the initial and reconsideration levels for SSDI and SSI applications filed from 1988 onward (and adjudicated in January 1989 onward). Form 831 records are structured as one row per application, though the version available to HRS researchers has the actual filing date omitted for disclosure reasons and replaced by a sequence variable that can be used to identify duplicate records. Within the Form 831 records, a person who applied concurrently to SSDI and SSI would have one row for DI and one for SSI, as each constitutes a separate application. A measure in the Form 831 file, CCF, indicates that the application was filed concurrently.

It is necessary to understand that Form 831 records are a subset of all applications to SSDI and SSI, and that they do not include:

- Applications that have been filed, but have not yet received an initial determination from the DDS. The average time from application to an initial or reconsideration decision has increased steadily over time and was about 6 months in 2023 (SSA, 2023).
- Applications that did not receive a medical review because the applicant was determined by the DDS to not meet the financial eligibility criteria for the program. Applications denied at this point are often called "technical denials." In the case of DI, this could mean not having sufficient quarters of coverage to be DI-insured. About one-third of DI applications receive a technical denial (SSA, 2021; SSA, 2022a). For SSI, this could mean income or assets that are above the eligibility threshold (SSA, 2009). In these

instances, HRS respondents might report having applied for benefits, but there would be no record in the Form 831 file.

- Applications that are initially denied but subsequently appealed. In this case, an initial application is observed in the Form 831 file, but subsequent appeals (and allowances based on those appeals would not be). In these cases, the respondent might report a pending application that would not be apparent in the Form 831 file. Additionally, applications that appear to be denied in the Form 831 file may ultimately lead to the participant receiving disability benefits, either as reported by beneficiaries or in other SSA administrative files. About half of DI applications and more than half of SSI applications that receive a medical review are rejected at the initial or reconsideration levels (SSA, 2022a; SSA, 2022b), with many ultimately receiving an award.

2.3 Files documenting receipt of benefits administered by SSA

Respondent Cross-Year Benefits File (CYBF)

The CYBF is a set of files that is derived from SSA's Master Beneficiary Record (MBR) and Payment History Update System (PHUS), with information available on benefits in the MBR from 1962 and on payments in the PHUS starting in 1984. The CYBF contains information on the receipt of SSDI and OASI—both on one's own earnings record and as an auxiliary beneficiary on another's record.

Reflecting that the CYBF merges two SSA administrative sources, it is sectioned into several components. Data from the MBR are contained in Sections A through F; Section G contains information from the PHUS. Section A is organized as one row per beneficiary and contains information on OASDI benefit spells. If respondents received SSDI at multiple points or moved from SSDI to OASI, variables in this section can be used to determine the reason that

benefits were awarded and the date that benefits began. This section also contains information on the disabling condition for which an individual was determined eligible for SSDI benefits.

Other parts of the CYBF (Sections B, C, D, E, and G) provide monthly information about benefit amounts, with one row per person-year. The Monthly Benefit Amount (MBA), Monthly Benefit Credited (MBC), and Monthly Benefit Payable (MBP) measure different concepts of benefits that users should consult to align with their needs. The CYBF users' guide distinguishes the specifics of each of these concepts. These variables should be used in conjunction with the Ledger Account Status (LAF) variable (Section F), which indicates a beneficiary's benefit eligibility in a particular month. When the LAF starts with a C, it indicates that the beneficiary was eligible for a benefit payment in that month. Eligibility alone does not determine that a beneficiary received a cash benefit, it must be used in conjunction with MBP.

The PHUS payment values in Section G offer an alternative to the monthly benefit amounts in other sections. MBR-based payment amounts reflect SSA's best current information and can be overwritten as the agency learns more about the beneficiary over time. In contrast, the PHUS values in Section G are snapshots of what occurred in the month. Specifically, the amounts reflect the check the beneficiary received that month (and a separate value indicates the amount that Social Security paid in Medicare premiums for Parts B, C, and D); the sum of the values is the total amount Social Security paid for the beneficiary in a month). Thus, the PHUS may be most closely aligned to what beneficiaries would report as income from OASI or SSDI benefits.

SSI File

The SSI file contains information on SSI benefits for disabled and aged recipients, and is derived from SSA's Supplemental Security Record (SSR). It can be thought of as analogous to

the CYBF, but for SSI. The SSI file is structured with two sections. Section A contains eligibility information and contains one row for each period of eligibility.[3] Section B contains monthly data about benefits received starting in January 1974 (when the SSI program began).

Similar to the CYBF, the amount actually paid to an SSI recipient in a given month requires considering two variables in the file. The Payment Status Code (PSTAT) variable indicates a beneficiary's SSI entitlement status in a particular month and can be considered analogous to the LAF for SSDI. When the recorded PSTAT is C01, the beneficiary was eligible for a benefit payment in that month. This variable can be used alone to determine a beneficiary's eligibility for payment in a month. The amounts that were due to the beneficiary are recorded in FEDAMT and STATAMT; these values can be overwritten by SSA as additional information about monthly income affecting benefits is received by the agency.

SSI monthly amounts are affected by income and earnings using a retrospective monthly accounting process, meaning that a beneficiary can be in C01 status, but not receive a cash payment in the month based on income from an earlier month. The Federal Money Paid Amount (FEDPMT) indicates the amount that SSA paid to the beneficiary from the federal disability program. Another variable, STATPMT, records any federally-administered state supplement in addition to FEDPMT that was paid. The total amount received by the beneficiary in C01 in the month can be determined by adding FEDPMT and STATPMT together.

Disability Analysis File (DAF)

For users interested only in SSDI and/or SSI receipt on the basis of disability (e.g. not OASI nor SSI on the basis of age), the DAF offers a streamlined product that can be used in

[3] There are instances where new records are created for logistical reasons and therefore do not represent a new period of eligibility.

place of the CYBF and SSI file. SSA produces the DAF to facilitate research based on administrative records, cutting down on the steps necessary to process information from across different programs and files with different structures. A key advantage of the DAF is that it contains information from both SSDI and SSI in a single source, rather than separately consulting source files from each program. Another benefit is that it has extensive documentation about the file contents, structure, and variable elements (see U.S. Social Security Administration (2021)).

The DAF contains records on all disability beneficiaries with at least one month of benefits since 1996; each SSI beneficiary from age 18 through FRA, and each SSDI beneficiary from 18 through FRA. The file contains one-time information recorded at application as well as monthly information on benefits due, received, and monthly measures as early as 1994. While the DAF time span is shorter than the CYBF and SSI File, its contents may be useful to many HRS researchers given the close alignment to the start of the survey. In addition, the DAF contains information on benefit suspense and termination for work, work-related program participation (such as Ticket to Work), and other topics related to beneficiary employment that is not readily available in the CYBF and SSI file.

The HRS maintains two versions of the DAF—one without geographic identifiers and another with those identifiers; the latter requires an approved research plan that requires geography to access the file. Each beneficiary in the respective DAF has a single record, with information on SSDI and SSI populated as relevant to each person's benefits history. Limited information collected at the time of the initial application is available for each beneficiary, and the file does not include information about applications that did not result in benefits.

2.4 HRS Data Products Derived from SSA-Linked Data

HRS has produced data products derived from the earnings information in the SSA administrative records extracts that estimate respondents' retirement benefits or lifetime earnings. The purpose of these files is to facilitate use of the information in the earnings files, constructed in a way to support common research interests.

Cross-Wave Prospective Social Security Benefit Wealth Measures of Pre-Retirees

The motivation behind this public-use data product is to provide estimates of respondents' expected (hence "prospective") OASI benefits upon retirement in interview waves that are "pre-retirement." "Pre-retirement" waves are defined at the respondent level; for each respondent, these are the waves that the respondent has not yet claimed (according to either their linked administrative records[4] or survey self-reports) OASI benefits.[5] The file contains projections at each wave; earlier iterations only projected benefit wealth at the baseline interview. Because the file only includes information for pre-retirement waves, it excludes AHEAD and CODA cohorts, who entered the study at older ages.

For respondents with matched SSA earnings records, we project their covered earnings through various future claiming ages (62, the respondent's FRA based on their birthdate, and 70). Benefit wealth is the present discounted value (PDV) of all future benefits, accounting for longevity. We use benefit wealth to calculate the AIME, Primary Insurance Amount (PIA), and monthly benefits amounts at those hypothetical claiming ages. For respondents without any

[4] Respondents do not appear in the MBR until they apply for OASDI benefits. Here, "respondents who have not claimed according to their linked administrative records" actually means we count a respondent as "not claimed" if this respondent has linked earnings records but no records of old-age, spousal, or survivors benefits.
[5] The amount of benefits paid for those who have claimed can be obtained directly from the survey or from the MBR (for those whose linked administrative records are available.)

matched SSA earnings records, we impute their PIA from which we calculate their monthly benefit payments and PDVs accordingly.

The PDVs reported in these data products include benefits from respondents' own earnings, spousal benefits, and survivors benefits. The data product only provides PDVs for pre-retirement waves, and only for ages that a respondent has not attained in a particular wave. For example, if a respondent entered the HRS at age 55 in 2004 and claimed the OASI benefits at age 66 in 2015, we only provide wealth estimates for waves between 2004 and 2014 for this respondent. Furthermore, since the respondent reached age 62 in 2011, the age-62 wealth estimates in the 2012 and 2014 waves became irrelevant (this respondent could no longer claim at age 62 in 2011, based on information available in 2012/2014) and hence not provided in the data product. The same concept is also applicable for wealth estimates at FRA and age 70.

Imputed Lifetime Earnings

This data product provides estimates of respondents' total earnings throughout their lifetime. This file is made available as an HRS restricted data product with permission required for use.

For respondents with matched SSA earnings records, the lifetime earnings include the summation of available earnings data as well as projections. Take a respondent who only has matched earnings records through age 62 as an example. We use their actual earnings records to calculate lifetime earnings through age 50, 55, and 60; we also project the respondent's earnings through age 70 using the same algorithm we use to estimate Social Security benefits wealth in order to calculate lifetime earnings through age 65 and 70.

For respondents without any matched SSA earnings records, we impute their cumulative lifetime earnings at HRS entry and then their annual earnings after HRS entry. Imputed lifetime earnings through various ages are calculated accordingly.

Cross-Wave Permissions Consent History

As described in more detail in Section 3, HRS respondents are asked to consent to the SSA administrative linkage at various points in time, reflecting their selection into the sample, the last time they consented, and the SSA rules that apply to consents at the time. Because the timing of consent varies for each respondent, this file exists to provide users with details about whether, in a given wave, a respondent was asked for a consent, whether the respondent granted permission, whether the permission form was sent to SSA (after a check for having complete information), and whether the respondent was ultimately matched to the SSA administrative earnings and/or benefits files. Based on the year the consent was provided, the file can be used to determine other years for which the consent remains valid.[6]

3 Consenting to Linkage

3.1 The Consent Process

The primary determinant of whether respondents have matched records in the administrative file is whether they consent to the study linking their data. The HRS has asked respondents to consent to link their survey responses with their administrative earnings and benefits records since its inception in 1992. SSA requires that respondents consent to the linkage

[6] One important caveat is that the consent permissions history file indicates status as of the year the consent was provided; it does not necessarily reflect all of the years the consent remains valid. This can be important as benefits status changes. For example, someone who consented in 2016 might not have any benefits records available in that year, and the permissions consent history would correctly indicate that. If, however, the respondent applied for SSDI (or began receiving SSDI, SSI, or OASI) while the consent period was still active, the respondent would have benefits records in later years, even though no records were found in the year consent was provided.

by providing a valid Social Security number, their date of birth, and their full first and last name. Additionally, a "wet" signature (a physical signature on pen and paper) and current date is required for the consent to be valid; as a result, the HRS only asks for consent during FTF interviews.[7]

The period for which respondents consent to have their data linked to administrative records has changed over the course of the HRS. Before the 2006 wave, consents were only valid through a date specified on the consent form, meaning that all information prior to the consent was available for the respondent, but nothing beyond that time. Starting in (and including) the 2006 wave, consent forms became prospective, meaning that the respondents consent to their data being linked through the current year and into the future. Currently, the consent for earnings records is valid for 6 years; for example, the 2020 match includes respondents who consented in the 2014, 2016, 2018, and 2020 waves.

The consent forms also changed over time: in the early years of the study, respondents consented to the linkage of both earnings and benefits records on the same form (with a single signature); starting in the 2006 wave, respondents separately consent to the earnings and benefits linkages on two separate forms. The consent permissions history file distinguishes whether respondents consented to only the earnings file, only the benefits file, or both.

Respondents who consent at one point are asked to consent again when their previous consent expires. If consents for both earnings and benefits records have expired, the respondent is asked to re-consent for both records. If the consent for earnings records expired but the consent for benefits records is still valid, HRS only asks for re-consent if the respondent is working for pay as of the current interview or was working for pay in the previous interview.

[7] An exception is the 2020 wave. Face-to-face interviews were suspended during the COVID-19 pandemic, and consent forms were mailed to the respondents, who then could sign paper forms and return them by mail.

There is no difference in the process and forms used for consent versus re-consent. Table 3 shows which respondents were asked for consents in each wave. Table 4 shows how long the consents remain valid for each of the data products.

Consent forms are provided during the interview, then completed and returned by respondents. Forms that are returned but not completed properly are not sent to SSA; these respondents are counted as consenters, but they will not have a matched record. The permissions consent history file tracks consent separately from sent and matched records. The HRS does not verify that the information provided on the form is accurate before sending, only that it is complete.

Properly completed consent forms are sent to SSA, which first validates the information provided by respondents to its Numident file (to verify that the individual provided the correct information confirming that the SSN is for the right person),[8] then matches to earnings or benefits files based on the permission that the respondent provided.[9] This match process is completed separately for earnings and benefits, where matching to earnings means a record in the SER or DER, and matching to benefits means their presence in the CYBF file (see footnote 6). SSA always sends HRS the "full" records of matched respondents that it has. For example, each batch of linked summary records include earnings information extending back to 1951, not just back to the time of the previous match.

Because SSA returns the latest records at the time of match, the availability of SSA records for each respondent only depends on their most recent match. In other words, provided a

[8] The Numident, or SSA's Numerical Identification System, is the electronic record of information reported when individuals applied for a Social Security Card. It contains name, date of birth, and other information on applicants for all Social Security numbers back to 1936.

[9] SSA does not send the match results with Numident back to the HRS. Hence, for a respondent who is not matched, we do not know whether the respondent was not matched with Numident (indicating invalid information on the consent form), or was not matched with the earnings/benefits records (indicating the lack of records in SSA's files).

new consent/match, the consent/match history prior to that becomes irrelevant. In the case where a respondent consented before but never re-consented (or the consent has expired), HRS retains the matched records that were previously received. For example, if a respondent consented in 1992 but never re-consented, the matched summary earnings records for this respondent cover only 1951 through 1991. Without a valid consent, the respondent will not receive updates in subsequent matches, and the earnings in 1992 and beyond are set to missing in the linked administrative records data product. However, the 1951-1991 records for this respondent remain available in the HRS distribution of the linked administrative record data products, so the records in the data products are "cumulative."

3.2 Implications of Consent for the Number of Respondents with Matched Records

Over the 1992-2020 waves, 75.2 percent of HRS respondents gave permission to SSA to provide earnings or benefits records, and 68.8 percent of HRS respondents were matched to their earnings or benefits records. Because of changes over time in the consent process and general patterns of the willingness to consent to administrative linkages, each HRS cohort has a different likelihood of having a matched record. Figure 2 shows how consent and match rates vary by cohort; the consent rate is the share of respondents that consented to have their responses linked to their earnings or benefits records of those who were asked while the match rate is the share of respondents that were matched to their earnings or benefits records of those who were asked. The figure includes one record for each respondent. The consent rate is higher than the match rate because not every record sent to the SSA was matched. We see the highest consent and match rates for the HRS cohort, with lower consent and match rates for more recent cohorts. These patterns reflect both higher baseline consent rates as well as additional opportunities to consent for respondents who have been in the sample longer.

Table 5 shows the number of respondents by cohort who ever have matched records in each of the SSA administrative records data products. Because each of the files covers a different population (earners, beneficiaries, applicants), the maximum potential size of each file varies for reasons unrelated to the HRS consent process. Having a record in the file at any point does not mean that the respondent has a record for every year covered by the data product, because a respondent's records are only updated with a valid consent. For example, a respondent who last consented in 1992 will only have earnings records between 1951 and 1991 in the summary earnings records (that is, earnings records in 1992 and beyond will be missing), while a respondent who still has a valid consent as of 2020 will have earnings records in all years between 1951 and 2020.

Figure 3 shows how consent and match rates for earnings records vary by wave. The figure includes one record for each respondent for each wave they are in the survey. The number of respondents asked to consent in each wave varies substantially, with typically many more consent requests made during years that new cohorts are added to the sample. In the early waves of the HRS, prior to 2004, we see higher consent and match rates in new cohort waves, which decline in intervening waves. In 2004 and later waves, we see higher sent and match rates initially, but which decrease over time. There are several factors that affect the dynamics in these trends. First, the consents made between 1992 and 1996 only apply to the 1992 match, and those made in 1998 and 2000 only apply to the 1998 match. For example, in 1994 through 1996, only those who did not consent in 1992 and 1993 were asked for consents[10], hence the lower sent/match rates in 1994 through 1996. The same goes for 2000. Second, consents became prospective in 2006, so those who consented in 2004 were re-asked in their next face-to-face

[10] Recall that the original HRS cohort respondents were interviewed in 1992, 1994, and 1996, and the AHEAD cohort respondents were interviewed in 1993 and 1995. These two cohorts were merged in 1998.

interview. This explains the higher rates in 2004 and waves immediately after. Respondents who consented in 2006 or later are only re-asked when their consents expire, so the dynamics are also different from the earlier periods. Finally, the declining rates in recent waves reflect cohort effects, as the younger cohorts are less likely to respond to surveys, and also less likely to provide consent when asked. This declining pattern is common in population surveys in the U.S. and is not a HRS-specific issue.

4 Social Security Weights

As we described, not every HRS respondent who consents to the linkage ultimately has a matched record in the SSA administrative files, either because they provide incomplete or inaccurate information. Additionally, respondents can consent to link their earnings records but not their benefits records, or vice versa.

For these reasons, users need to be aware that the sample who matches to the SSA administrative records are not nationally representative when using the usual HRS wave-specific survey weights. Previous research determined that the likelihood of having a matched record is not random; matched and unmatched respondents vary on several observable dimensions including gender, race, education, geographic location, labor force participation, health status, income, wealth and retirement expectations (Olson, 1999; Haider and Solon, 2000; Michaud, Kapteyn, Smith and van Soest, 2011). Others found that inverse probability weights (Wooldridge, 2010) that adjust only for race, gender, household composition and birth cohort do not completely remove selection on observable biases (Michaud et al. 2011).

To account more fully for the nonrandom likelihood of a matched SSA record, HRS staff construct “Social Security weights” to be used to produce population-representative estimates accounting for the likelihood of having a matched record. Unlike other SSA administrative

records files available to HRS users, researchers do not need to select this file; it is provided by default for all users who are approved to access the earnings or benefits files. Respondents in the "matched sample" have information in either the administrative earnings *or* benefits records.[11] The Social Security weights are inverse probability weights that result from a logit model predicting the likelihood of being in the matched sample at each wave, accounting for demographic, health, work, and financial characteristics.[12] This logit model predicts a "likelihood of match," and the Social Security weights are calculated as the usual wave-specific HRS respondent-level weights (available in the HRS tracker file) adjusted for the likelihood of match. There is hence no need to use the HRS respondent-level weights in the tracker file in addition to the Social Security weights when weighting the variables in the linked HRS-SSA data. However, users working with the application and benefits files may want to consider whether additional weighting is needed to account for having a record in those files (and not the earnings files)

Social Security weights are constructed for each wave.[13] The wave-specific values take into account the most recent information that the HRS has to facilitate a match; as we describe below, the process to match records allows for users to opt-in to matching at various points, and the rules governing the match process have changed over time as well. We provide two sets of

[11] An important nuance here is that the matched sample means that the respondent has records in the earnings (SER or DER) *or* the benefits (CYBF) files. Weights are based on having a record in any of the following files: SER, DER, or CYBF. The SER includes records for all SSNs, even if the person does not have earnings (and therefore has values of 0 in all years). Thus, a respondent who consented to the earnings linkage and provided valid information would have a SER record. But, the CYBF contains records for OASDI beneficiaries, so not all consenters to the benefits linkage will have information in that file. Additionally, even those with a CYBF record may not have information in all of the benefits files; a respondent who did not apply for SSDI or SSI will be missing from the 831 file, a respondent who did not receive SSI will be missing from the SSI file, and a respondent who did not receive SSI or SSDI for disability will be missing from the DAF.

[12] See the appendix in Fang (2024) for the details regarding how the respondent-level Social Security weights are calculated.

[13] Earlier versions of the Social Security weights were previously only available for baseline waves. HRS has since "filled in" the gaps of data availability for non-baseline waves.

Social Security weights: “currently matched” (a respondent has matched records as of a particular wave) and “ever matched” (a respondent has ever been matched as of a particular wave). To the extent that consents for earnings and benefits have different expiration dates, we further distinguish “currently matched for earnings records” and “currently matched for benefits records” in waves when such distinction is needed.

This data product also incorporates newly-available data linkages. For example, say a respondent who entered the HRS in 2016 did not consent (and was not matched by SSA) until 2020. Weights calculated based on the data linkage available as of 2016 would count this respondent as “unmatched,” because no linked data were available when the calculation was done in 2016. However, in 2020, the respondent’s linked data became available, and the HRS now has this respondent’s earnings records not just for 2020, but for all years between 1951 and 2020. To utilize the newly available information between 2016 and 2020, this respondent is moved to the “matched sample” in all earlier waves, so that a non-zero Social Security weight can be assigned for all waves that this respondent has a non-zero respondent-level weight.

To evaluate the extent to which the Social Security weights address differences in characteristics of the matched sample as compared to the full sample, Table 2 compares select characteristics, as examined by Haider and Solon (2000), using HRS sample weights (the full sample) to those using Social Security weights (the matched sample) by wave for new cohort waves. Across waves, results show statistically significant differences for the share male, the share non-white, and the share that never worked, though the estimates are not meaningfully different. These results suggest that the Social Security Weights produce sample characteristics that replicate those of the full sample using HRS sample weights. Hence, although the matched

sample is not a random subsample of the entire HRS, a properly weighted matched sample is still population representative.

5 Conclusion

The HRS-SSA data linkage can be incredibly powerful to support analyses of lifetime earnings and of interactions with the retirement and disability programs administered by the SSA. This guide, along with detailed documentation of each of the data products available on the HRS website, is designed to support users in understanding important details about working with those files.

6 References

Fang, Chichun. 2024. “Health and Retirement Study: Social Security Weights. Data Description and Usage.” Retrieved from https://hrs.isr.umich.edu/data-products/restricted-data/available-products/9696.

Haider, Steven, and Gary Solon. 2000. “Nonrandom Selection in the HRS Social Security Earnings Sample.” Working Paper. RAND Labor and Population Program.

Michaud, Pierre-Carl, Arie Kapteyn, James P Smith, and Arthur van Soest. 2011. “Temporary and permanent unit non-response in follow-up interviews of the Health and Retirement Study.” *Longitudinal and Life Course Studies* 2 (2): 145-69.

Olson, Janice A. 1999. “Linkages with Data from Social Security Administrative Records in Health and Retirement Study.*” Social Security Bulletin*. 62 (2): 73–85.

Sonnega, Amanda, Jessica D. Faul, Mary Beth Ofstedal, Kenneth M. Langa, John WR Phillips, and David R. Weir. 2014. "Cohort profile: the health and retirement study (HRS)." *International Journal of Epidemiology* 43 (2): 576-585.

U.S. Social Security Administration, Office of Retirement and Disability Policy, Office of Research, Demonstration, and Employment Support. (2021). “Disability Analysis File 2021 (DAF21) Documentation: Data from January 1994 through December 2021.” Washington, DC: Mathematica, April 2023. [Volume 1: Getting Started with the DAF21]. Retrieved from https://www.ssa.gov/disabilityresearch/daf.html.

U.S. Social Security Administration. 2023. Contribution and Benefit Base. Retrieved from https://www.ssa.gov/OACT/COLA/cbb.html.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: The MIT Press.

Figures

Figure 1: Cohorts in the Health and Retirement Study

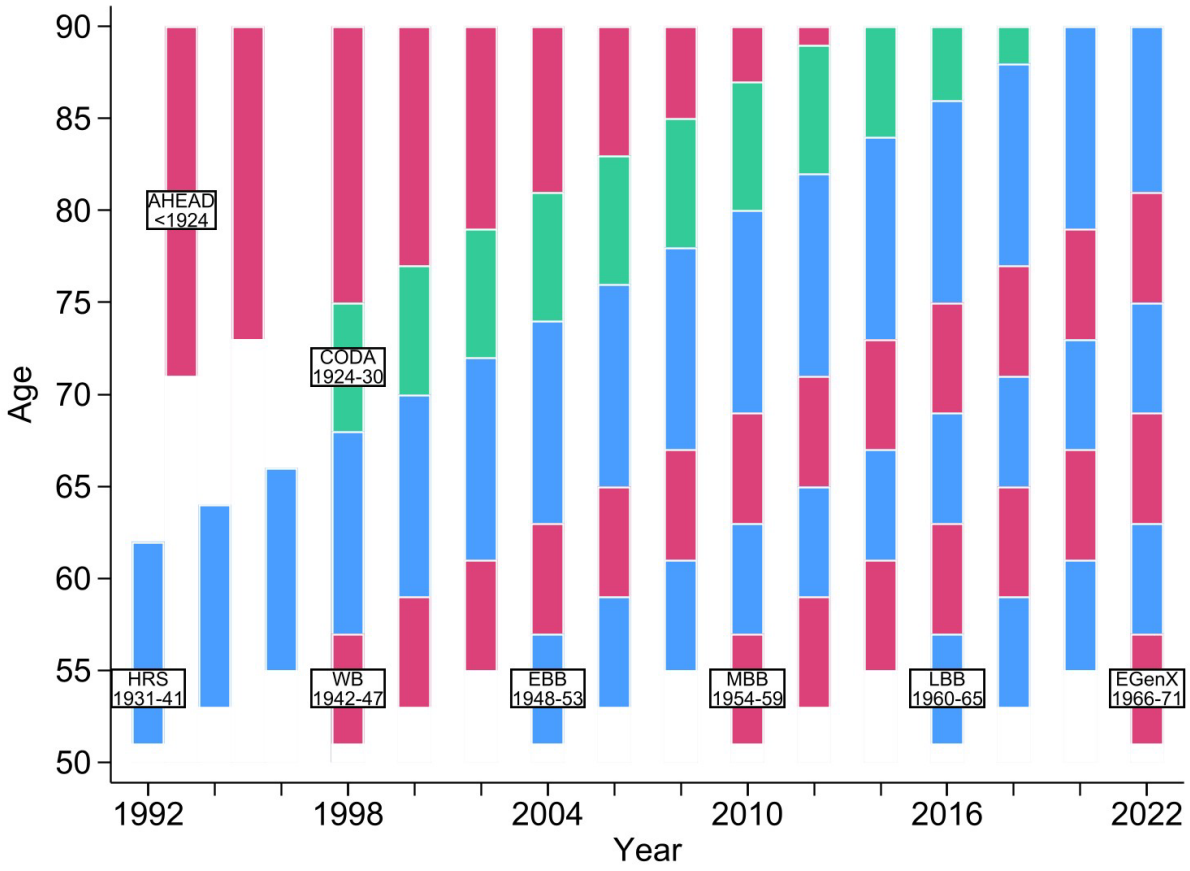
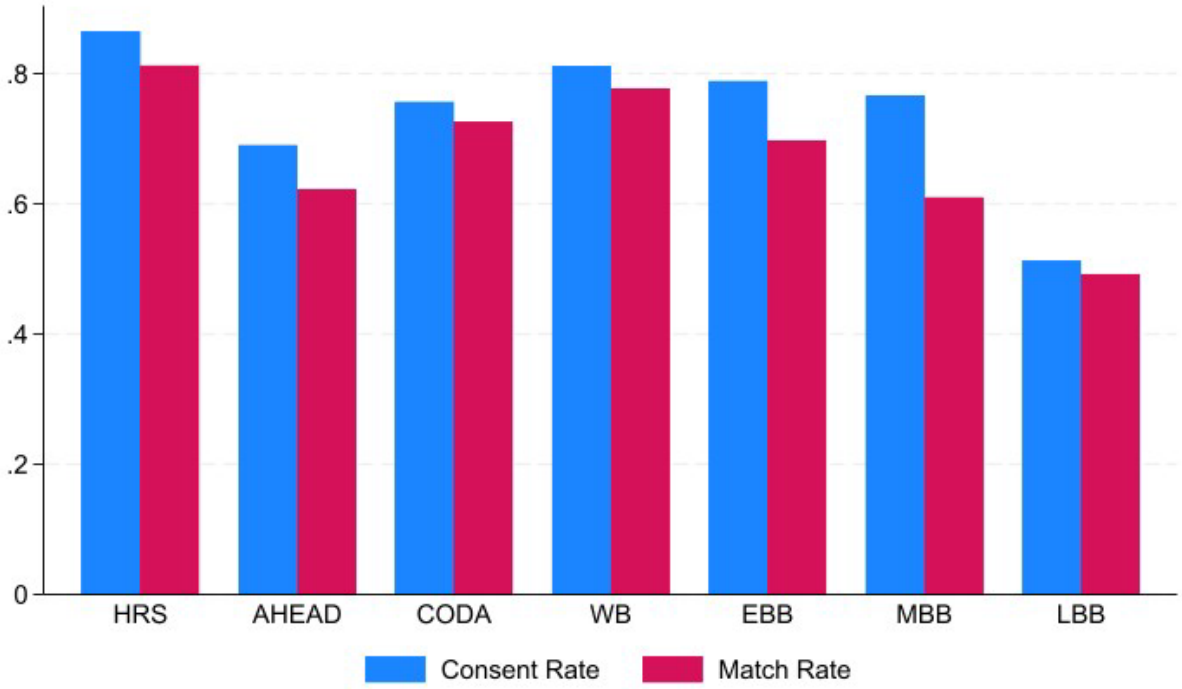
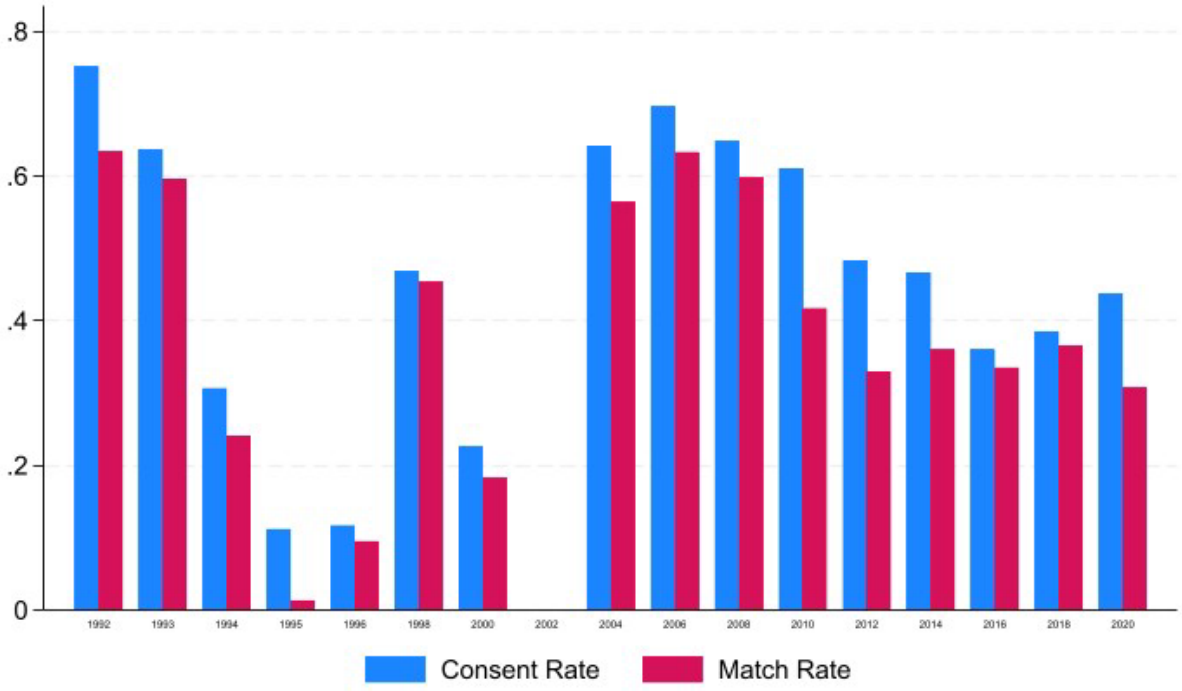


Figure 2: Ever Consent and Match Status by Cohort



Source: 2020 SSA Permissions Consent History file.
Notes: The figure shows the share of respondents in each cohort that ever consented to linkage to earnings or benefits records (consent rate) and were ever matched to their earnings or benefits records (match rate).

Figure 3: Consent and Match Status by Wave



Source: 2020 SSA Permissions Consent History file.
Notes: The figure shows the share of respondents that consented to linkage to earnings or benefits records (consent rate) and were matched to their earnings or benefits records (match rate) of those who were asked in each wave.

Tables

Table 1: Overview of SSA-HRS Linked Data Products

Data product (restricted use unless otherwise noted)	File inclusion criteria	Unit of observation	Earliest data in file
Administrative earnings records			
Summary Earnings Records (SER)		Respondent	1951
Detailed Earnings Records (DER)		Respondent-Employer-Year	1978
Applications to SSA disability programs			
Form 831	All initial applicants to SSI or SSDI who received an medical review	Respondent-Application	1989[14]
Receipt of benefits administered by SSA			
Cross-year Benefits File (CYBF); a mix of the Master Beneficiary Record (MBR) and Payment History Update System PHUS)	Beneficiaries with at least one month of SSDI or OASI receipt since 1962	Respondent; Respondent-Year[15]	1962; PHUS only back to 1984
Supplemental Security Income (SSI) file	Beneficiaries with at least one month of SSI since 1974	Respondent; Respondent-Year[16]	1974
Disability Analysis File (DAF)	Beneficiaries with at least one month of SSI or SSDI between birth and full retirement age since 1996	Respondent; Respondent-Year	1994
Data products derived from SSA linked data			
Cross-wave prospective Social Security wealth measures of pre-retirees (two versions; publicly available)		Respondent	1992
Imputations of lifetime earnings records		Respondent	See Footnote[17]
Files to support analysis of SSA linked files			
Cross-wave Social Security weights		Respondent	1992
Cross-wave permissions consent history (two versions; publicly available)		Respondent	1992

[14] Indicating the date of adjudication on the initial application, not the date of the initial application. For example, an application filed in September 1988 and adjudicated in January 1989 is in the data, but an application filed in October 1988 and adjudicated in December 1988 is not.

[15] There are multiple files in this data product. One is at the respondent level, while the rest are at the respondent-year level.

[16] There are two files in this data product. One is at the respondent level, and the other is at the respondent-year level.

[17] Respondent-level lifetime earnings measures include earnings from both covered and uncovered jobs starting in 1978, and earnings from covered jobs starting in 1951.

Table 2: Means for Select Variables using HRS Weights and Social Security Weights in New Cohort Waves

	New cohort survey wave									
	1992		1998		2004		2010		2016	
	HRS Weights	SS Weights	HRS Weights	SS Weights	HRS Weights	SS Weights	HRS Weights	SS Weights	HRS Weights	SS Weights
% Male	48.14	47.62***	44.59	44.73***	45.56	45.45***	46.16	45.83***	47.07	46.38***
% HRS Cohort	100	100	37.7	40.48	31.85	33.92	23.17	25.57	16.18	18.58
% AHEAD Cohort	0	0	27.93	25.8	14.78	14.04	5.89	6.04	2.3	2.55
% CODA Cohort	0	0	12.28	11.63	8.07	7.91	5.07	5.25	2.21	2.44
% WB Cohort	0	0	22.09	22.1	22.1	22.51	17.18	18.27	13.43	14.94
% EBB Cohort	0	0	0	0	23.2	21.63	22.88	22.34	19.35	20.05
% MBB Cohort	0	0	0	0	0	0	25.8	22.52	24.36	23.34
% LBB Cohort	0	0	0	0	0	0	0	0	22.16	18.09
% U.S. Born	90.9	91.14	91.92	92.06	90.9	91.36	89.48	90.39	87.66	89.2
% Non-white	17.47	16.89***	16.01	15.56***	18.1	17.56***	21.59	20.15***	26.14	23.98***
% HS Education or Less	38.68	39.08	35.66	35.78	34.36	34.58	32.15	32.55	30.48	30.8
% Some College	19.76	19.59	20.12	20.41	23.29	23.3	25.8	25.67	27.59	27.54
% Completed College	18.51	18.49	18.8	19.06	23.5	23.26	27.44	27.32	29.67	29.66
% Retired	13.54	13.32	42.99	42	40.32	40.82	41.21	43.49	41.12	44.74
% Never worked	3.51	3.34**	4.32	3.78**	3.4	3.17**	3.51	3.28**	3.07	2.95**
% Homeowner	81.08	81.25	80.96	81.28	80.47	80.37	78.95	79.06	75.11	75.39
% Reporting Good Health	79.32	79.86	70.32	71.11	72.6	72.53	74.34	74.33	74.47	74.1
% with a Job-limiting Disability	20.82	20.67	30.06	29.45	12.22	12.16	29.41	30.66	32.1	34.1
Household Earnings ($10K)	4.99	5	5.35	5.36	6.81	6.7	7.68	7.57	9.67	9.47
Household Wealth ($100K)	2.43	2.41	3.25	3.26	4.47	4.4	4.8	4.75	5.88	5.96
Household Wealth ($100K) Squared	30.47	30.05	124.04	107.7	140.13	145.16	149.57	144.65	265.47	261.42

Source: 2020 RAND Longitudinal File, 2020 HRS Tracker file, 2020 Social Security Weights file.

Notes: The table shows means for select variables using HRS sample weights (the full sample) and Social Security weights (the matched sample) in new cohort waves. *** p<0.01, ** p< 0.05, * p < 0.10 for t-test of difference in means using Social Security weights compared to HRS weights.

Cohort abbreviations: HRS: Health and Retirement Study; AHEAD: Asset and Health Dynamics Among the Oldest Old; CODA: Children of the Depression; WB: War Babies; EBB: Early Baby Boomers; MBB: Mid Baby Boomers; LBB: Late Baby Boomers.

Table 3: HRS Respondents Asked for Consents

Cohort	1992/94/96	1993/95	1998/2000	2004	2006: Asked If Face-to-face and…	2008: Asked if Enhanced Face-to-face and…	2010: Asked if Enhanced Face-to-face and…	2012: Asked if Enhanced Face-to-face and…	2014: Asked if Enhanced Face-to-face and…	2016: Asked if Enhanced Face-to-face and…	2018: Asked if Enhanced Face-to-face and…	2020: Asked if Enhanced Face-to-face and…
New spouses or never-interviewed R in any cohort	Asked all	Asked all	Asked all	Asked all	Asked all	Asked all	Asked all	Asked all	Asked all	Asked all	Asked all	Asked all
AHEAD (Born <=1923)	NO	Asked all	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
CODA (1924-1930)	NA	NA	Asked all	NO	R was not asked in 2004	R refused, did not provide valid permission information, was not asked, or was non-response in 2004 or 2006	R refused, did not provide valid permission information, was not asked, or was non-response in 2006 or 2008	R refused, did not provide valid permission information, was not asked, or was non-response in 2006, 2008 or 2010	R refused, did not provide valid permission information, was not asked, or was non-response in 2006, 2008, 2010 or 2012	R refused, did not provide valid permission information, was not asked, or was non-response in 2006, 2008, 2010, 2012 or 2014	R refused, did not provide valid permission information, was not asked, or was non-response in 2006, 2008, 2010, 2012, 2014, or 2016; or consent expired and re-consent required.	R refused, did not provide valid permission information, was not asked, or was non-response in 2006, 2008, 2010, 2012, 2014, 2016, or 2018; or consent expired and re-consent required.
HRS (1931-1941)	Asked all	NO	NO	Asked all	R was not asked in 2004							
WAR BABY (1942-1947)	NA	NA	Asked all	Only asked refusers	R was not asked in 2004							
EARLY BOOMERS (1948-1953)	NA	NA	NA	Asked all	R refused or was non-response in 2004							
MID BOOMERS (1954-1959)	NA	NA	NA	NA	NA	NA	Asked all					
LATE BOOMERS (1960-1965)	NA	NA	NA	NA	NA	NA	NA	NA	NA	Asked all		

Table 4: Consent Wave and Availability of Data Products

Most Recent Consent/Match	Summary Earnings Records (SER)	Detailed Earnings Records (DER)	Benefits (Cross-year Benefits File; CYBF)	Form 831 and Supplemental Security Income (SSI) file
1992/1994/1996	Yes, through 1991	Yes, through 1991	Partial, through 1991[18]	No
1993/1995	Yes, through 1992	No[19]	Partial, through 1992[20]	No
1998/2000	Yes, through 1999	Yes, through 1999	Yes, through 1999	No
2004	Yes, through 2003	Yes, through 2003	Yes, through 2003	No
2006	Yes, through 2013	Yes, through 2013	Yes, through 2018	Yes, through 2018
2008	Yes, through 2016	Yes, through 2016	Yes, through 2020	Yes, through 2020
2010	Yes, through 2016	Yes, through 2016	Yes, through 2020	Yes, through 2020
2012	Yes, through 2018	Yes, through 2018	Yes, through 2020	Yes, through 2020
2014	Yes, through 2020	Yes, through 2020	Yes, through 2020	Yes, through 2020
2016	Yes, through 2020	Yes, through 2020	Yes, through 2020	Yes, through 2020
2018	Yes, through 2020	Yes, through 2020	Yes, through 2020	Yes, through 2020
2020	Yes, through 2020	Yes, through 2020	Yes, through 2020	Yes, through 2020

Note: "Yes" indicates records in this data products are available if consents/matches occurred in the certain year, and the records are updated through the year specified. For example, in the first row, if respondents were matched in 1992, 1994, or 1996 waves, their earnings in the Summary Earnings Records are updated through 1991. Form 831 and SSI records are only available for consents/matches in 2006 and later waves. As long as a respondent is ever matched for a particular data product, their records are available in that data product and updated through what their most recent consent/match year would indicate.

[18] Some information in the MBR is not available for this consent wave. PHUS is not available for this consent wave.

[19] These records come from the AHEAD cohort. The DER for this cohort were not available to the HRS per the agreement between HRS and SSA.

[20] PHUS is not available for this consent wave.

Table 5: Number of Linked Respondents by Cohort and Data Product

	Cohort						
	HRS	AHEAD	CODA	WB	EBB	MBB	LBB
Total HRS respondents	13,652	8,445	2,430	2,825	5,072	5,365	4,617
In SER	11,133	5,240	1,751	2,163	3,472	3,164	2,218
In DER	10,483	851	1,566	2,106	3,391	3,136	2,241
In CYBF	7637	4899	1657	1692	2899	2185	928
In Form 831	911	18	38	365	930	995	667
In SSI	1,224	119	131	312	908	894	612
In DAF	764	9	8	311	730	673	331

Source: 2020 Summary Earnings Record File, 2020 Detail Earnings Record File, 2020 Cross-Year Benefits File, 2020 Form 831 File, 2020 Supplemental Security Income File, 2020 Disability Analysis File, 2020 RAND Longitudinal File.
Notes: The table shows the number of total HRS respondents and the number of respondents that ever appear in each file by cohort.
Cohort abbreviations: HRS: Health and Retirement Study; AHEAD: Asset and Health Dynamics Among the Oldest Old; CODA: Children of the Depression; WB: War Babies; EBB: Early Baby Boomers; MBB: Mid Baby Boomers; LBB: Late Baby Boomers.

Table 6: Number of Linked Records by Survey Wave

	Survey wave														
	1992	1994	1996	1998	2000	2002	2004	2006	2008	2010	2012	2014	2016	2018	2020
Total HRS respondents	12,652	19,642	17,991	21,384	19,578	18,165	20,129	18,469	17,217	22,034	20,554	18,747	20,912	17,146	15,723
Respondents ever matched															
In SER	10,581	14,993	14,075	16,858	15,767	14,916	16,582	15,457	14,586	17,230	16,157	14,800	15,333	12,908	11,648
In DER	9,893	10,032	9,745	12,977	12,499	12,282	14,388	13,702	13,222	16,190	15,321	14,156	14,874	12,584	11,412
In CYBF	7,134	11,645	11,033	13,614	12,928	12,438	13,996	13,220	12,568	14,416	13,561	12,438	11,988	10,111	9,022
In Form 831	770	762	770	1,138	1,141	1,198	1,770	1,784	1,810	3,012	2,882	2,707	3,121	2,669	2,350
In SSI	1,082	1,128	1,138	1,558	1,569	1,622	2,201	2,112	2,053	3,058	2,891	2,667	3,008	2,535	2,199
In DAF	653	640	648	951	994	995	1,464	1,478	1,500	2,339	2,214	2,051	2,135	1,765	1,536
Respondents matched in survey wave															
In SER	6,928	6,657	6,694	11,411	10,425	10,674	10,517	10,581	10,754	14,511	14,022	12,768	13,991	6,823	5,709
In DER	5,090	4,700	4,338	5,752	5,336	4,752	5,104	4,594	4,209	6,295	5,602	4,803	5,368	3,403	2,706

Source: 2020 Summary Earnings Record File, 2020 Detail Earnings Record File, Cross-Year Benefits File, 2020 Form 831 File, 2020 Supplemental Security Income File, 2020 Disability Analysis File, 2020 RAND Longitudinal File.

Notes: The table shows the number of total HRS respondents, the number of respondents that ever appear in each file by wave (i.e., who had a valid consent in the wave and had at least one record in the linked file; these values do not indicate that the earnings/benefits information aligns to the year shown), and for the SER and DER, the number of HRS respondents matched to their earnings record in each wave.