CENTER for RETIREMENT RESEARCH at BOSTON COLLEGE

A GUIDE TO MEASURING WEALTH, INCOME, AND REPLACEMENT RATES IN THE HEALTH AND RETIREMENT STUDY

Nilufer Gok, Anqi Chen, and Laura D. Quinby

June 2024

Center for Retirement Research at Boston College
Haley House
140 Commonwealth Avenue
Chestnut Hill, MA 02467
Tel: 617-552-1762 Fax: 617-552-0191
https://crr.bc.edu

All of the authors are with the Center for Retirement Research at Boston College; Nilufer Gok is a research manager and Anqi Chen and Laura D. Quinby are senior research economists. The research reported herein was pursuant to a grant from the U.S. Social Security Administration (SSA) funded as part of the Retirement and Disability Research Consortium. The findings and conclusions expressed are solely those of the authors and do not represent the views of SSA, any agency of the federal government, or Boston College.

About the Center for Retirement Research

The Center for Retirement Research at Boston College, part of a consortium that includes parallel centers at the National Bureau of Economic Research, the University of Michigan, and the University of Wisconsin-Madison, was established in 1998 through a grant from the Social Security Administration. The Center's mission is to produce first-class research and forge a strong link between the academic community and decision makers in the public and private sectors around an issue of critical importance to the nation's future. To achieve this mission, the Center conducts a wide variety of research projects, transmits new findings to a broad audience, trains new scholars, and broadens access to valuable data sources.

Center for Retirement Research at Boston College
Haley House
140 Commonwealth Avenue
Chestnut Hill, MA 02467
phone: 617-552-1762 Fax: 617-552-0191
https://crr.bc.edu

Affiliated Institutions:
The Brookings Institution
Mathematica – Center for Studying Disability Policy
Syracuse University
Urban Institute

Introduction

The *Health and Retirement Study* (HRS) is a primary source of information on retirement wealth, income, and replacement rates, but calculating these measures requires a host of methodological choices that affect the results. Since researchers have not yet established clear best practices for dealing with the survey's complex structure, studies using the HRS are often inconsistent and difficult to replicate. Additionally, the steep learning curve is daunting for young scholars interested in exploring retirement issues. The CRR aims to make the HRS more accessible by providing: 1) a methodological guide that identifies the key conceptual and technical choices that must be made when analyzing a household's financial resources in the HRS and 2) clean, well-documented code that builds on RAND's efforts to calculate retirement wealth, income, and replacement rates.

The discussion proceeds as follows. The next section provides a brief overview of the HRS. The third section discusses our methodology for calculating household wealth, which includes housing, financial, retirement plan pension, and Social Security wealth. The fourth section discusses income. The final section covers replacement rates, including different ways to define average pre-retirement income.

HRS and RAND HRS Data Files

The HRS is a panel survey, conducted biennially since 1992, that interviews a nationally representative sample of about 20,000 people ages 50+ and their spouses. Over the decades, the number and types of questions in the survey have evolved and expanded, making it a very rich source of information but also very complex.[1] The latest core interview, for example, covers over 20 topics, including assets, income, health, disability, employment, family structure, and expectations.

Fortunately, the RAND Corporation has improved the usability of the survey by publishing the RAND HRS, which harmonizes variables over time and identifies when they are

[1] The HRS survey consists of two main sections: the core and the exit interview. Information on income, wealth, and replacement rates comes from the core interview. The exit interview surveys surviving spouses or proxy respondents after a respondent passes away and includes information on finances, health, and bequests, which is not relevant for calculating income, wealth, or replacement rates.

at the household or individual level (see Box 1 on defining household head).[2] It also creates summary variables that might be useful for researchers, such as net worth, total income, or Social Security wealth, and imputes missing data.

Box 1. *Defining the Head of Household*

Evaluating household wealth, income, or replacement rates often requires defining a household head. We follow the RAND HRS definition for consistency. The household head is the respondent in single households or the financial respondent in coupled households. If no financial respondent is defined, then the head is defined as whoever is not the family respondent. If no family respondent is defined, the respondent with the lowest person number is defined as the head.[3]

The two main downsides to the RAND HRS are: 1) it comes out about 1.5 years after the HRS is released; and 2) it does not include all the variables.[4] We build on the RAND HRS to provide an easy-to-use code for researchers who want to use the most recent data to investigate questions related to retirement wealth, income, and replacement rates. This code requires researchers to have access to the restricted summary earnings files from the U.S. Social Security Administration (SSA).

Household Wealth

Household wealth can be grouped into four broad categories: 1) financial wealth, excluding defined contribution (DC) assets; 2) housing wealth; 3) employer-sponsored retirement plans; and 4) Social Security. The following sections outline how each of the four wealth components is measured and how our code builds on or differs from the RAND HRS.

Non-DC Financial and Housing Wealth

Measuring non-DC financial wealth and housing wealth requires the fewest assumptions because the value of the assets can be measured directly. Since methodologies for measuring the

[2] The HRS harmonizes variables across years to help researchers who want to examine trends over time. Along with the longitudinal file, RAND also publishes RAND HRS fat files. These combine all HRS sections into one

[3] For detailed variable names on how to define a household head, see the RAND HRS codebook.

[4] While many of the key wealth and income measures are included, the RAND HRS is more limited in other topics, such as health.

two sources of wealth are similar, they are discussed together. In our code, non-DC financial wealth includes the net value of stocks, mutual funds, bonds and bond funds; the value of checking, savings, and money market accounts, certificates of deposit; and government savings bonds, excluding any of these assets held in DC plans such as 401(k)s and IRAs. It also includes the value of businesses and vehicles. Debt in this category includes revolving credit card balances, student loans, medical debt, personal loans, and other debt. Our measure of total non-DC financial wealth is allowed to be negative for households where debt exceeds wealth. Housing wealth is the net value of residences, which is calculated as the gross value of the primary and secondary residences, less any relevant mortgages and home loans. Other real estate is also included. For households where debt exceeds the value of the house, total housing wealth is also allowed to be negative.

The HRS asks a sequence of questions to determine the value of each asset and debt type. First, households are asked if they own certain types of assets or debt. If the answer is yes, they are asked about the dollar value of the asset or debt. If the respondent does not know the value or is unable to provide an exact dollar amount, they are asked a series of unfolding bracket questions to narrow down the dollar value range. Table 1 shows the share of households that does *not* provide an exact value for their various assets and debts. For example, 65 percent of households sampled (unweighted) have checking, savings, or money market accounts in 2020, and of those, 18 percent do not report a value.

Table 1. *Percentage of Households with Missing Asset/Debt Values, by Type, 2020*

Financial wealth	Owns asset/ debt type	Missing values, among owners
Net value of vehicles	67%	14%
Net value of businesses	6	31
Net value of IRA, Keogh accounts*	27	26
Net value of stocks, mutual funds, and investment trusts	16	28
Value of checking, savings, or money market accounts	65	18
Value of CD, government savings bonds, and T-bills	8	24
Net value of bonds and bond funds	3	40
Net value of all other savings	11	20
Value of other debt	30	2
Housing wealth		
Value of primary residence	56%	12%
Value of all mortgages/land contracts (primary residence)	23	11
Value of other home loans (primary residence)	5	6
Value of secondary residence	10	2
Value of all mortgages/land contracts (secondary residence)	1	1
Net value of real estate	10	18

* We categorized IRA and Keogh accounts as employer-sponsored plans. However, the imputation method for missing values is the same as other non-DC financial wealth and housing wealth.
Source: University of Michigan, *Health and Retirement Study* (HRS) (2020).

To preserve the sample size and representativeness of the survey, RAND imputes values for households that fail to provide the exact values of their various assets and debts using the nearest neighbor approach, which finds the most similar household and imputes their asset and debt values.[5] In order to reduce the discrepancies between the latest data release, which does not include imputations, and the RAND HRS, our code also imputes missing wealth values using a simpler nearest-neighbor approach (see Box 2 for details on our imputation methodology).

[5] Their latest imputation methodology is outlined in Bugliari et al. (2023).

Box 2. *Imputation Methodology for Missing Values*

We impute missing values for each type of asset and debt using a nearest-neighbor approach. The majority of respondents provide a range even if they are not able to provide an exact value. For these respondents, we find a "neighbor" donor who has similar demographic characteristics (based on their gender, race, education, birth year, and marital status) and reported an exact value within the range reported by the respondent with the missing wealth. [6] If no donors fall within a range, we assign the midpoint of the range. We impute from the full sample for respondents who do not provide a range.[7] This approach allows us to preserve the distribution of each variable. The imputations are conducted for each asset and debt type separately.

For a very small percentage of married households, the spouse is not a respondent in the HRS, and we have almost no information about them or their financial resources, such as the spouse's earnings history, Social Security, and employer-sponsored retirement plans. The respondent's spouse only answers questions about themselves even though they are a coupled household. In order to account for the financial resources of the non-respondent spouse, we adjust each type of household-level resource using a nearest-neighbor approach where we use the respondent spouse's characteristics to find the donor.[8]

Our imputation method differs from the RAND HRS along two dimensions. First, RAND's nearest-neighbor approach incorporates more factors (employment status, income, etc.) when finding an appropriate donor. Second, RAND imputes ownership even if the respondent is unsure whether they own a particular type of asset or debt. In contrast, we assign respondents a value of zero if they are unsure. Despite these differences, our imputations provide similar values to the RAND HRS (see Box Table).

[6] The "mi impute pmm" Stata package is used to conduct the imputations. For more details, see https://www.stata.com/manuals13/mimiimputepmm.pdf.

[7] On average, across all years, about 11 percent of households provided a range for the value of their primary house, 20 percent for stocks, 13 percent for checking accounts, and 10 percent for IRAs. This is compared to 2 percent, 11 percent, 9 percent, and 7 percent of respondents who did not provide a range.

[8] We only adjust at the household level, non-respondent spouses' resources can't be used in individual level analysis.

Box Table. *Distribution of IRA Balances in 2020*

	Full HRS, including respondents with missing values		Only respondents with missing values	
	CRR imputed	RAND imputed	CRR imputed	RAND imputed
Mean	$106,358	$107,214	$232,166	$222,162
25th	0	0	1,500	0
50th	0	0	70,000	60,000
75th	38,750	40,000	269,000	250,000
Max	8,250,000	8,250,000	4,100,000	4,000,000

Notes: Those with more than $10 million in their IRA accounts are dropped.
Source: Authors' calculations using HRS raw and RAND HRS files (2020).

Employer-Sponsored Retirement Plans

Wealth in both defined benefit (DB) pensions and defined contribution (DC) plans is an important resource for many households. However, measuring wealth from these plans is not straightforward for several reasons. First, while the HRS asks meticulous questions about each retirement plan from current and prior employers, keeping track of the status of each plan over time is challenging (see Box 3 on retirement plan questions in the HRS).[9] Second, until recent cohorts, many workers had access to DB plans, which provide benefits as a stream of income. Converting that stream of income into wealth that can be compared with DC wealth or other wealth components requires several calculations and assumptions. The RAND HRS, which is the primary source for many other wealth and income variables, only includes the balance of DC plans from the respondent's current job. IRA balances are captured in the RAND HRS in the financial wealth section. Our code incorporates DC plans from current and prior jobs and includes IRA balances in employer-sponsored plans because IRAs are, to a large extent, rollovers from DC plans.

[9] The HRS provides data files named *Employer-Sponsored Pension Wealth from Current Jobs* every year a new cohort enters, but these come with a 6-year lag and don't capture total pension wealth as they ignore dormant plans from previous employers. The HRS also provides researchers with the *Pension Estimation Program*, which calculates pension wealth using pension formulas and other details collected from employers of HRS respondents in *Summary Plan Descriptions*. This program is only available to researchers with restricted HRS access, and is updated every 6 years.

Box 3. *The Evolution of HRS Retirement Plan Questions and Data Structure*

The HRS refers to both DB and DC plans as pension plans. Prior to 2012, pension questions were included in the employment section of the HRS, and pension information was collected at the individual respondent level. When respondents first joined the HRS, they were asked about the pensions from their current and prior jobs. In subsequent interviews, respondents were asked to provide status updates on plans from which they still expect to receive benefits or have a positive balance. If, however, a respondent is currently receiving income from a DB pension plan, that respondent is queried about the plan is only in the income section.[10]

A new pension section, which includes both DB pensions and DC retirement accounts, was introduced in 2012, which dramatically changed the structure of the data. Questions were now designed to follow individual pensions over time rather than follow the respondents who owned the pensions.[11] Many respondents have more than one plan, so it is important to match respondents with all of their plans from current and prior employers. First-time respondents, however, were still only asked about their pensions from prior jobs in the employment section, which is tracked at the individual level. The survey also asked about the characteristics of DB pensions that were already paying out benefits.

The pension section was revised again in 2016 and has stayed largely the same since then. Now, all pensions, from both current and prior employers are tracked at the pension level.

Fortunately, Gustman, Steinmeier, and Tabatabai (2014) (GST) constructed retirement plan wealth variables for the years 1992-2010 using publicly available HRS data. These wealth calculations are widely used by retirement researchers.[12] Following GST's approach, our code produces pension wealth variables for the years 2012 forward.

The first step in producing updated pension wealth estimates is to determine the type of plan(s), but many respondents provide inconsistent information on their retirement plan type.

[10] Prior to 2012, only the benefit payout was recorded in the income section, with no details about the plan.
[11] The analysis of employer-sponsored retirement accounts is always done at the individual or household level. However, the change in data structure from individual level to plan level makes linking across time more challenging mechanically and requires extra care.
[12] Details about their methodology are found in their book.

Venti (2011) estimates that around 30 percent of respondents revise their plan type in later interviews, even though neither their job nor their plan has changed. GST revised prior year estimates based on the latest responses regarding expected benefits or account balances. Our code, on the other hand, only uses the information provided at the current interview and does not revise prior year estimates. Users of the code should be aware of this difference when comparing our retirement plan wealth estimates after 2012 with those of GST in 2010 and prior years. The following sections detail our methodology for estimating DB and DC wealth and compare our wealth estimates with GST estimates.

DC + IRA Wealth. Determining how much households have in their current or prior job DC plans is straightforward in theory, as it equals the sum of all account balances at a given time. In reality, it is difficult to track all the different DC accounts that are left with prior employers. Here we refer to vested accounts that are left at prior employers as "dormant accounts." As discussed in Box 3 above, the structure of pension questions has evolved over the years. Since 2016, respondents have been asked about all plan(s) from current and past employers each time they are interviewed. For subsequent interviews, respondents confirm whether they still own the plan(s) mentioned in prior interviews or, if not, what happened to the plan.[13]

GST assumes that these dormant accounts grow by a nominal 5.8 percent per year (3 percent real interest rate and 2.8 percent inflation). However, we assume that if respondents do not mention the account in the current interview, they no longer have the account. The rationale is that, once all the plans from current and prior employers are discussed, respondents are asked if they have any other pensions. This question should capture any plans that were missed. Additionally, the average balance of dormant undiscussed plans is $150,000 (in 2010). While respondents may forget small balances from prior employers, it is unlikely they would forget about an account worth $150,000. The good news is that only 8 percent of respondents have

[13] If respondents say they no longer have a plan, the survey asks what happened to the plan. The options are: "rolled into IRA," "withdrew all money," "combined with another plan," "transferred to new employer," "lost all benefits," "withdrew some money," "plan frozen or lost," "converted to annuity," "receiving regular benefits," and "left money in the account." They can choose more than one option and based on their answer, they are asked further questions.

dormant plans, and despite the difference in how to treat undiscussed pensions, our estimates are similar to those of GST (see Table 1).[14]

Table 1. *Percentage of Respondents with Dormant DC Accounts in 2010*

	55-64	All
GST	9%	8%
CRR	5	4

Note: The shares represent that of an unweighted sample.
Sources: GST (2014) and authors' calculations using HRS (2010).

The second way that our methodology differs from GST is our imputation methodology for respondents who do not provide an exact amount for their plan balances. As discussed in Box 2, we use nearest-neighbor to impute balances based on the ranges provided by the respondent. GST imputes account balances alongside other pension-related variables, such as pension coverage and number of plans, using a mixed regression and hot deck approach that controls for industry, firm size, and personal characteristics.[15] Despite the two methodological differences, our estimates of DC balances in 2010 are similar to those of GST (see Table 2).[16]

Table 2. *Average Household DC Wealth, by Wealth Quintile, 2010*

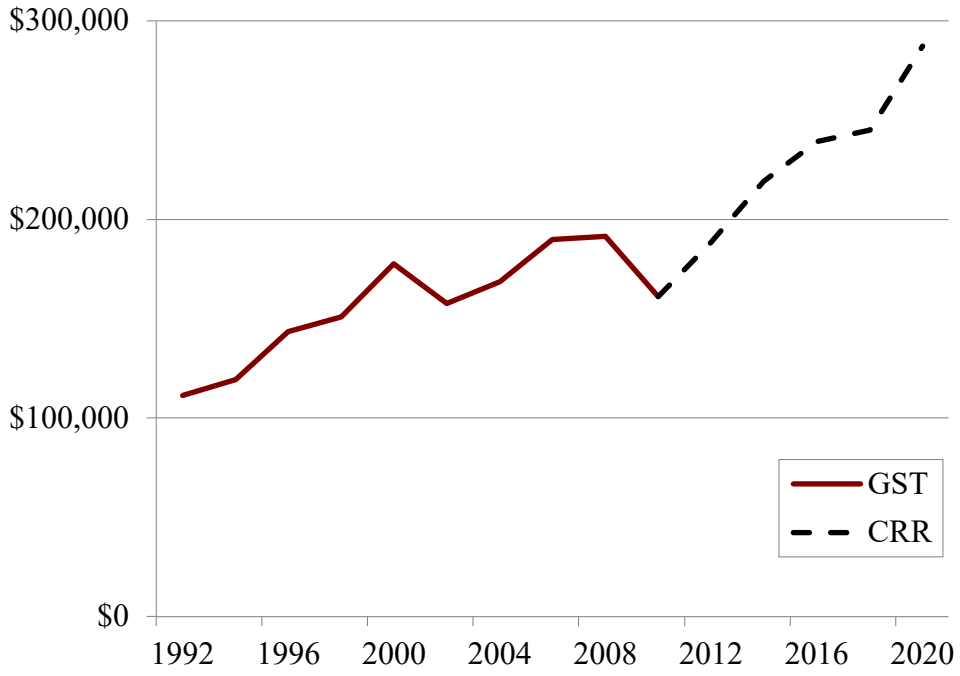
Quintile	GST	CRR
Bottom	$11,197	$11,228
Second	26,202	21,049
Middle	62,559	56,906
Fourth	91,828	75,443
Top	140,951	125,839

Note: Wealth quintiles are based on total household wealth excluding pension, IRA, and Social Security wealth.
Sources: GST (2014) and authors' calculations using HRS (2010).

[14] The share of individuals with dormant accounts has been increasing over time. In 2018, 16 percent of HRS respondents ages 55-64 had at least one dormant DC account. Capitalize (2021), using the Form 5500, found that 25 percent of plan participants were dormant (inactive vested accounts). However, 5500 data are at the participant level and participants do not translate to individuals. For example, a worker who has left a plan each at two of his prior employers will be counted as two participants in the 5500.
[15] For more details, see Gustman, Steinmeier, and Tabatabai (2012) – admin supplement.
[16] The difference between the two methods is larger for higher wealth groups.

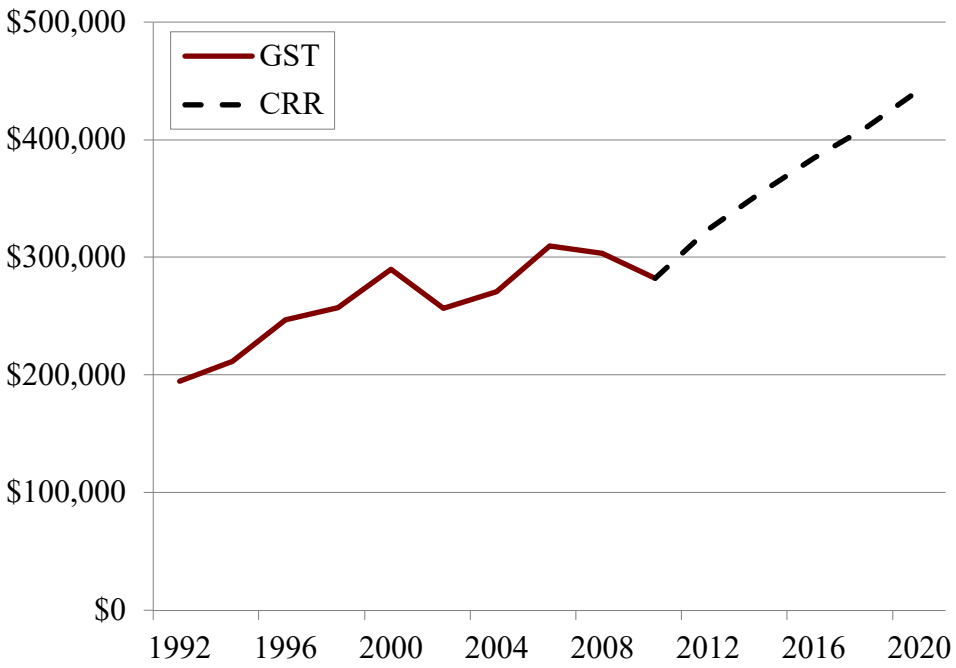
Figures 1 and 2 track average DC and IRA balances over time, combining GST with our method, for all households and those with a DC or IRA account, respectively. At first glance, DC wealth appears to break in 2012, the year we transition from GST estimates to our estimates. However, given how similar the two methods are for previous years, as shown above, two other factors are likely contributing to the break in trend. First, the stock market began to rebound after the Great Recession, resulting in a surge in DC and IRA wealth. Second, 2012 was also the year when the HRS introduced a dedicated pension section and included more questions that tracked plans from prior jobs. The dedicated pension section may capture a fuller picture of household DC and IRA wealth.

Figure 1. *Average DC and IRA Wealth ($2020) for All Households Ages 55-64, 1992-2020*



Sources: HRS (2012-2020); GST (2014); and RAND longitudinal file 1992-2020v1.

Figure 2. *Average DC and IRA Wealth ($2020) for Households Ages 55-64 with DC Assets, 1992-2020*

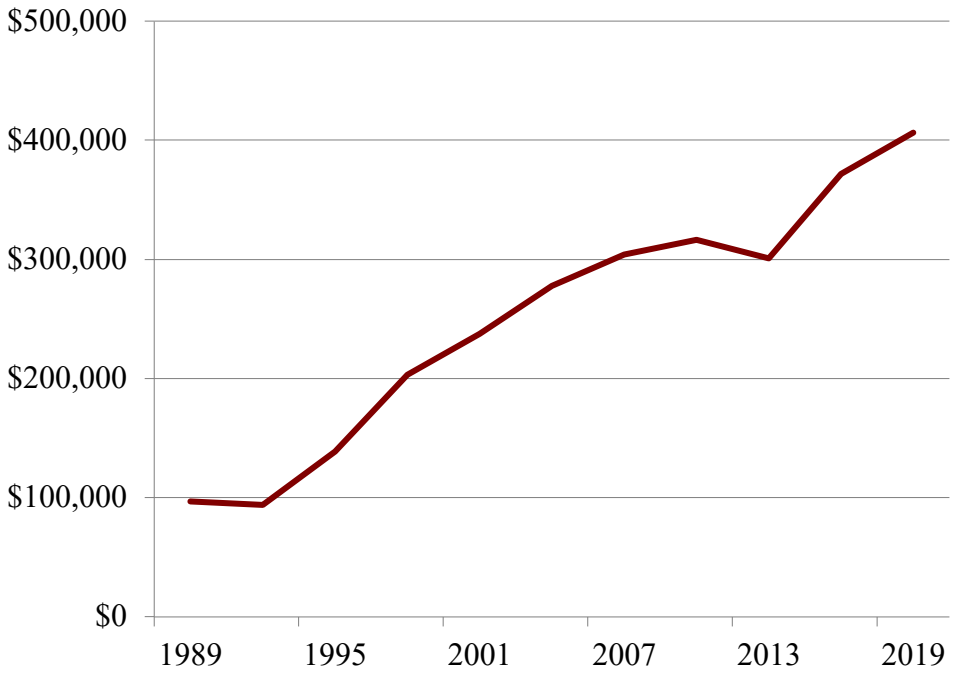


Sources: HRS (2012-2020); GST (2014); and RAND longitudinal file 1992-2020v1.

Comparing balances in the HRS with those in the SCF suggests that much of the growth is due to stock market recovery after the Great Recession. Average values are similar in both datasets and follow a similar trend (see Figure 3).[17]

[17] Interestingly, the dip DC assets was documented in 2013 instead of in 2010 for the SCF. In email correspondence economists at the Federal Reserve Board hypothesize that respondents may been in denial about how fare their accounts had fallen in 2010.

Figure 3. *Average DC and IRA Wealth ($2020) for Households Ages 55-64 with DC Assets, 1992-2020*



Source: U.S. Board of Governors of the Federal Reserve System, *Survey of Consumer Finances* (SCF) (1989-2019).

DB Wealth. The first step in calculating households' DB wealth is to determine their DB income stream. Respondents are asked whether they expect to receive or are currently receiving benefits, the amount of benefits, and the ages at which benefits started or are expected to start.[18] They can report expected benefits in the form of an amount per month/year, as a percentage of pay, or as a lump sum. Similar to other financial variables, some respondents do not provide an exact value for their current or expected DB income. Respondents already receiving DB pension income who do not provide an exact value are not asked to provide a range for their benefit amount. So, we impute the annual benefit from those who provide an amount. Respondents who are not yet receiving DB benefits but expect to receive benefits in the future are asked questions using unfolding brackets to determine the range of their future benefits. For these respondents, we use the impute using the nearest neighbor approach, as discussed in Box 2.

Once we have benefit streams for everyone currently receiving or expecting to receive DB income, we calculate the present value of the DB plan at the age when respondents expect to

[18] A small number of respondents have DB pensions that do not get paid out for their entire lives. Our code does not incorporate end dates for DB pensions.

start receiving benefits and prorate this value to the interview year. As for the discount rate, our calculations default to the long-run interest rate from the corresponding year's *Social Security Trustees Report*, while GST uses a discount rate of 5.8 for all years they calculate. For consistency over time, we convert GST wealth calculations using a multiplier: for every percentage-point decrease in the discount rate, the present value of wealth increases by 8%.[19] Survival probabilities are based on cohort mortality from the most recent *Social Security Trustees Report*, which is currently in 2023. We ignore cost-of-living adjustments that pensions might have in order to be consistent with GST.

So far, we have relied on just responses in the pension section to determine respondent's DB wealth. Respondents are asked again in the Assets and Income section, if they receive a pension and the amount. Theoretically, the responses from the two sections should be consistent, however, they are not. Some respondents report receiving pension income but do not report having a pension in earlier sections. To see the magnitude of this issue, we compared percent of people with positive DB wealth with a similar measure in the *Survey of Consumer Finances* across HRS years. While in earlier years, HRS calculations without the income section were similar or higher than SCF values, they start to underestimate especially at older ages after 2014. A comparison for 2010 and 2018 is shown in Tables 3 and 4 respectively. The second column only uses the Pensions section, and the third uses both sections.

Table 3. *Percent with DB Holdings, SCF 2010 vs HRS 2010*

Age group	SCF	HRS with just pension section	HRS with pension + income section
50-60	19%	27%	29%
60-70	27	34	40
70-80	39	34	46

Sources: SCF (2010); GST (2014); RAND longitudinal 1992-2020v1.

[19] The long-run interest rate from the *Social Security Trustees Report* is very similar to GST's 5.8% value. If researchers wish to use 5.8% for all the years, they can adjust the code accordingly.

Table 4. *Percent with DB Holdings, SCF 2019 vs HRS 2018*

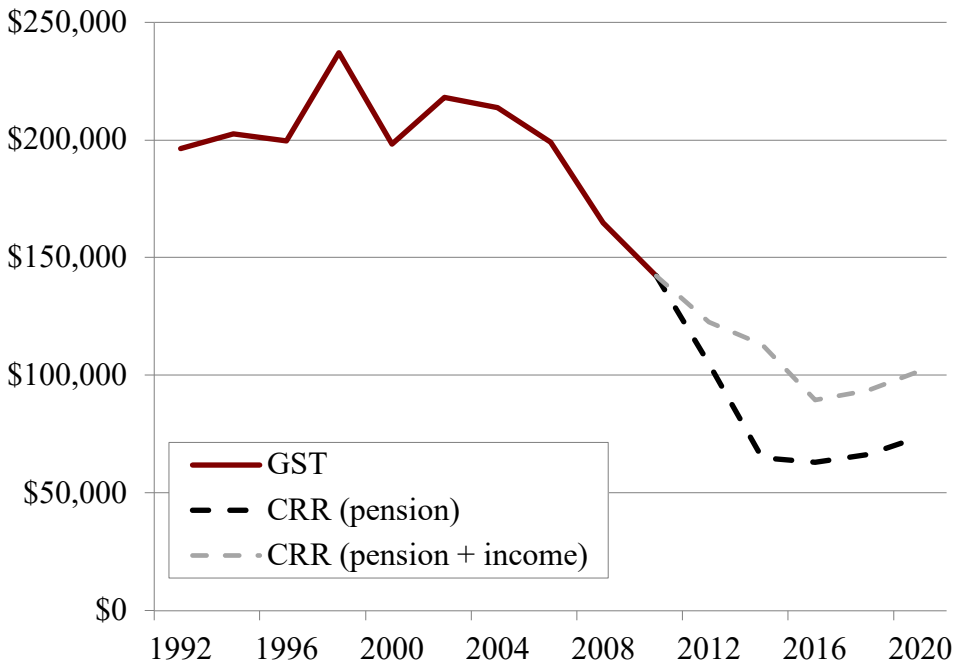
Age group	SCF	HRS with just pension section	HRS with pension + income section
50-60	16%	18%	20%
60-70	25	14	28
70-80	37	11	40

Sources: SCF (2019); GST (2014); RAND longitudinal 1992-2020v1.

While just using respondents' answers about pensions undercount the percentage of households with a DB pension, combining the pension and the income sections seems to overcorrect. Because of this, we include both versions of DB wealth. For the second version, we add the present discounted value of total pension income reported in the Assets and Income section to respondents who didn't mention any current payments in the Pensions section.

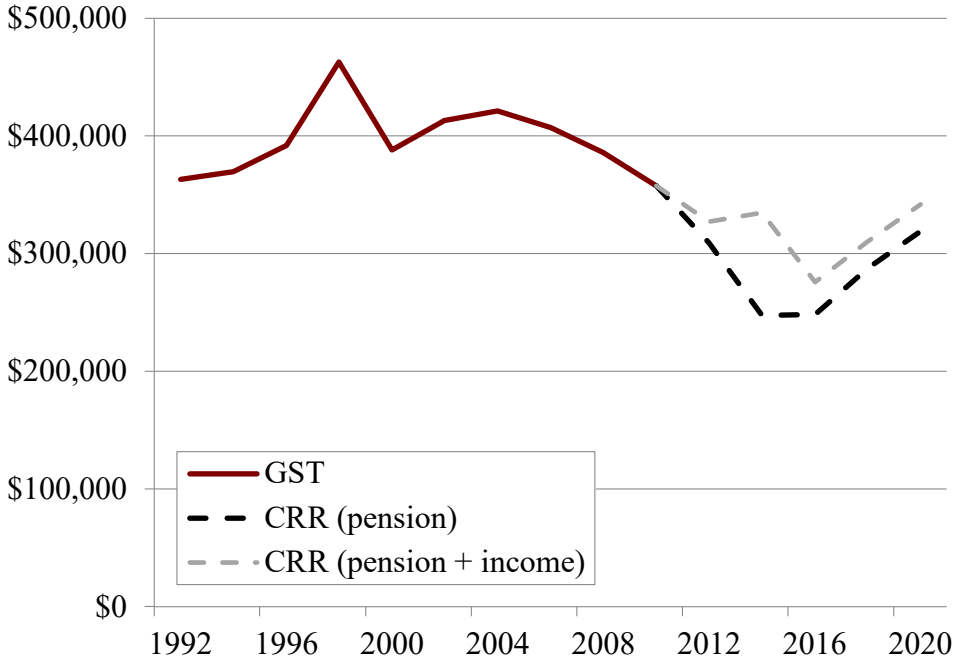
Figures 4 and 5 illustrate the average DB wealth for all households and those with DB wealth, respectively. The trend is as expected – average DB wealth is declining rapidly since fewer and fewer employers offer DB pensions.

Figure 4. *Average DB Wealth ($2020) for All Households Ages 55-64, 1992-2020*



Sources: HRS (2012-2020); GST (2014); and RAND longitudinal file 1992-2020v1.

Figure 5. *Average DB Wealth ($2020) for Households Ages 55-64 with DB Wealth, 1992-2020*



Sources: HRS (2012-2020); GST (2014); and RAND longitudinal file 1992-2020v1.

Social Security

The final, and often most important, source of wealth for retired households is Social Security, but converting the stream of income into wealth and incorporating the value of spousal and survival benefits can be quite challenging.

The RAND longitudinal files calculate Social Security wealth at ages 62, the FRA, and age 70 from linked restricted earnings data from SSA.[20] While these calculations are convenient and a valuable foundation, they have several drawbacks. First, Social Security wealth is not calculated for respondents who have already claimed Social Security benefits and for respondents who are older than each claim age.[21] As a result, the calculations understate Social Security wealth for households where one spouse has already claimed, and this wealth measure is missing entirely for households where both spouses have claimed. Second, RAND only

[20] The Cross-Wave Prospective Social Security Wealth Measures of Pre-Retirees provides Social Security wealth calculations for survey years 1992, 1998, and 2004 (waves 1, 4, 7). The Cross-Wave Prospective Social Security Wealth Measures of Pre-Retirees Waves 10 and 13 provides calculations for survey years 2006 and 2012.
[21] RAND does not calculate Social Security wealth at age 62 for anyone who is age 63 or older, wealth at the FRA for anyone older than the FRA, and wealth at age 70 for anyone older than age 70.

provides wealth estimates for these households every six years, when new cohorts are introduced to the sample.[22]

Our code calculates Social Security wealth for all respondents and households in all survey years. Specifically, we calculate *counterfactual* Social Security wealth at ages 62, 65, FRA, and 70 – that is, what respondents' Social Security wealth would have been had they claimed at the ages above.[23] Our code also includes a variable for when respondents actually claim Social Security.[24] Researchers who wish to evaluate actual Social Security wealth can select the counterfactual wealth calculation that most closely aligns with the respondent's actual claiming age and that of the spouse to create their own household wealth variable.[25]

Primary Insurance Amount. The first step in calculating Social Security wealth is to determine each respondent's Primary Insurance Amount (PIA).[26] This calculation is straightforward if we know the respondent's earnings history. Fortunately, many respondents (around 74%) have given permission to link their SSA earnings record to the HRS. The linked data includes detailed earnings information from 1951 to the permission year or 6 years after the permission year, depending on the agreement. Thus, we can construct earnings histories for respondents from when they started working until their 50s (when respondents joined the HRS).[27]

Earnings from when the linked records end to the four claiming ages (62, 65, FRA, and 70), are assumed to equal an average of the last five observed years, indexed by the Average Wage Index (AWI).[28] We also assume that respondents work until they claim.[29] In other words, for our age-62 wealth calculation, we assume that respondents work until age 62, claim Social

[22] While most respondents join the HRS when new cohorts are brought in every six years, some respondents join on off years. Marriages and new couplings would also introduce new people into the sample.
[23] Counterfactual Social Security wealth is useful for two main reasons. First, for data privacy reasons, we do not want to reveal exact benefit amounts and exact claim ages. Second, when respondents claim is highly correlated with sociodemographic and health factors. Researchers may want to evaluate wealth absent of the claiming behavior which is endogenous.
[24] We cannot identify actual claiming ages for respondents born before 1917 and those in the AHEAD cohort.
[25] Actual Social Security wealth is only for respondents receiving benefits off of their own record. Our code does not provide accurate measures of wealth from spousal benefits because we do not incorporate different spousal claiming ages.
[26] Formulas and examples of how to calculate Social Security benefits can be found at https://www.ssa.gov/oact/ProgData/retirebenefit1.html and are outlined in the code files.
[27] We have earnings records until age 54 for around 95 percent of the sample with linked records.
[28] This methodology follows Kapinos et al. (2022).
[29] Friedberg (2000) found that the earnings test substantially reduced the probability that retirees would work after they claim, if they claimed before the FRA. While there is some evidence of those with higher earnings working after they claim, additional earnings after their claiming age would not impact their PIA.

Security, and stop working. Similarly, for our age-70 wealth calculation, we assume respondents work until age 70 and claim then too.[30]

However, not all respondents give permission to link their earnings records, and while the majority of respondents eventually do give permission, it is often not when they first join the survey (see Table 3). Restricting the analysis to respondents who can be linked to their earnings records would thus exclude a large share of respondents from younger cohorts.

Table 3. *Percentage of Respondents with Linked SSA Earnings Records, by Birth Cohort*

Cohort	Percentage with linked earnings data	Percentage linked at first interview	Percentage linked within first three interviews
CODA (1924-1930)	71%	25%	25%
HRS (1931-1941)	80	26	26
War Baby (1942-1947)	75	13	13
Early Baby Boomers (1948-1953)	67	12	17
Middle Baby Boomers (1954-1959)	57	24	35
Late Baby Boomers (1960-1965)	44	24	44

Source: HRS Respondent Cross-Year Summary Earnings Data Description (2022).

Respondents who cannot be linked to SSA earnings data fall into two groups – those who have claimed Social Security and those who have not claimed. For respondents who have already claimed Social Security, we use their self-reported benefit amount and claiming age to determine their PIA.[31] For respondents who have not claimed, we impute their PIA using the nearest-neighbor approach based on gender, race, birth year, education, industry of longest job, and HRS cohort. This approach is similar to that used by RAND. To account for changes in the PIA from counterfactual early or delayed claiming, we adjust the PIA of these respondents based on respondents with linked data. For example, we calculate the mean PIA ratio between working

[30] The rationale is because workers often respond by working longer when the full retirement age increases (Behaghel and Blau, 2012). So, we assume that if workers had to claim at higher ages, they would work until those ages. The exception is if the respondent stopped working before the early eligibility age of 62. In that scenario, their earnings are zero between when they stopped working and the four prospective claiming ages. But workers who retire before the early claiming age likely have other sources of income, such as disability insurance, and are unlikely to adjust their labor decisions.

[31] While the HRS does not distinguish between income from Social Security's disability insurance and income from its old-age and survivors insurance, disability income converts into retirement income at the FRA. Anyone receiving Social Security income prior to 62 is designated as receiving disability insurance.

and claiming at age 70 relative to working and claiming at age 65 for respondents with linked data and apply it to respondents who cannot be linked.[32]

Present Discounted Value of Benefits. Once PIAs are calculated, they can be converted to benefit amounts for the four claiming ages, depending on the respondent's birth year.[33] The EPV of lifetime worker benefits is calculated using the following:

$$EPV_a = \sum_{t=a}^{119} P_t \; SSB_t \; (1 + r_t)^{(a-t)} \tag{1}$$

where *a* represents one of the four claiming ages (62, 65, FRA, and 70), P_t is the probability of surviving to age t, and r_t is the interest rate. Benefit, denoted as SSB_t, is projected out to age 119 using the COLA adjustment and long-run inflation assumptions from the *Social Security Trustees Report* for the corresponding year. Following Fang (2022), survival probabilities, P_t, for each age and birth cohort are from the Social Security actuaries for the most recent year and r_t is the long-run nominal interest rate from the *Social Security Trustees Report* for the corresponding year.

Spousal and Survivor Benefits. Some respondents are able to receive a higher benefit based on the earnings history of their spouse.[34] While the spouse is still alive, the respondent's benefit is equal to the higher of their own retired worker benefit or 50 percent of their spouse's benefit, both calculated from equation (1). After their spouse dies, the respondent's benefit is equal to the higher of their own benefit or 100 percent of their spouse's worker benefit. The EPV of lifetime benefits *in addition* to what respondents can receive based on their own work history, weighted for the two states of the world (if the spouse is alive and if the spouse is deceased), is calculated using the following:

$$\begin{aligned} EPVS_a = & \sum_{t=a}^{120} \max\left(0, (0.5 * SSB_{s,t} - SSB_{r,t})\right) P_{s,t} \, P_{r,t} (1 + r_t)^{(a-t)} + \\ & \sum_{t=a}^{120} \max\left(0, (SSB_{s,t} - SSB_{r,t})\right)(1 - P_{s,t}) \, P_{r,t} (1 + r_t)^{(a-t)} \end{aligned} \tag{2}$$

[32] Social Security benefits for respondents with linked data are before Medicare Part B deductions while benefits for respondents without linked data are after Medicare Part B deductions. To make the two sets of estimates more comparable, we add the standard Part B premium for that year to the benefits of respondents who are not linked to earnings data.

[33] For a table on how benefits as a share of the PIA vary at different claiming ages and for different birth cohorts, see: https://www.ssa.gov/oact/ProgData/ar_drc.html

[34] Respondents are only eligible for spousal/survivor benefits if they have been married for 10 years or more.

where *r* represents the respondent, and *s* the spouse. Survival probabilities are the respondent's and spouse's individual birth year. We assume that spouses claim at the same age. In other words, the age-62 wealth calculation assumes both the respondent and spouse claim at age 62.[35]

The EPV of each respondent's total Social Security wealth is the sum of equation (1) and equation (2). Household Social Security wealth is therefore the sum of the respondent and spouse's total Social Security wealth.[36]

Current Income

Aside from wealth, researchers may also want to understand the income for retirees and workers near retirement. Fortunately, most components of current income can be directly measured.[37] In our code, income includes earnings (wage, salary, self-employed, contractor, bonuses, tips); DB pensions and annuities; Social Security income; interest and dividends; business and rental income; government transfers (e.g., disability insurance, TANF, SNAP, unemployment benefits): alimony; and other sources. Household income for couples is simply the sum of income from both spouses.

Similar to financial and housing wealth components, the HRS asks a sequence of questions to determine the value of each income component. If a respondent receives a particular source of income but is unable to provide an exact dollar amount, they are asked a series of questions using unfolding brackets to narrow down the dollar value range. Table 4 shows the share of households that does *not* provide an exact value for their various sources of income. For example, 40 percent of the HRS sample households (unweighted) have earnings in 2020, and of those, 24 percent do not report a value. Once again, our code also imputes missing wealth values using a simpler hot deck approach (see Box 2 for details on our imputation methodology).[38]

[35] This methodology follows Kapinos et al. (2022).
[36] For a very small percentage of households, the spouse is not a respondent in the HRS and we have almost no information on their earnings history or claiming decision. For these households, we adjust household Social Security wealth using the nearest neighbor method.
[37] Information on current income is mostly contained in the Income and Assets section of the HRS. For some respondents, income from DB pensions or annuities comes from the Pension section.
[38] Future versions of the code will switch to using the nearest neighbor approach.

Table 4. *Percentage of Households with Missing Income Values, by Type, 2020*

Source	Receives income type	Missing values, among those with income type
Earnings	40%	24%
DB pension income	27	19
Annuities	6	50
Disability insurance	14	33
Social Security retirement income	62	23
Unemployment or worker's compensation	4	44
Government transfers	21	20
Capital income	49	47
Other income	7	35

Note: A household is categorized as owning an asset if either the respondent or the spouse owns that type of asset. Similarly, a household is categorized as having missing values if either the respondent or spouse has a missing value.
Source: HRS (2020).

DC Withdrawals

The key issue, for retirees, is whether to include withdrawals from DC and IRA accounts as income. These withdrawals are categorized as dissaving in the RAND data. However, withdrawals are considered income for tax purposes and are a resource available in retirement. Our code defines DC and IRA withdrawals in separate variables so researchers can decide whether they want to include these withdrawals in the definition of income. Withdrawals include both regular and irregular withdrawals. [39]

Required Minimum Distribution (RMD). Once retirees reach a certain age, they are required to make minimum withdrawals from their DC and IRA accounts.[40] Many respondents say they did not make any withdrawals even when they are subject to the RMD. This inconsistent response is perhaps because they may not consider RMD withdrawals the same as other DC/IRA withdrawals. Our code calculates the additional withdrawals that respondents will need to make in order to follow the RMD, if respondents' combined DC and IRA withdrawals are less than the RMD for their age in a given year. Once again, our code defines RMDs in

[39] Due to changes to the pensions section – as described in Box 3 - researchers should be careful comparing DC withdrawals across years as variations in the questions and wording may lead to differences.
[40] Prior to 2020, RMD distributions started at age 70½. From 2020-2022, RMDs started at age 72. In 2023, the RMD increased to 73, and it will further rise to 75 by 2035.

separate variables so that researchers can decide whether to include them in the definition of income.

Replacement Rates

Once we calculate total wealth and income in retirement, researchers often want to evaluate if these resources are adequate for retirement. A common measure to assess retirement adequacy is replacement rates – the ratio of the retirement income that could be generated by the household's resources divided by its pre-retirement income. A lively debate exists over how to define both retirement income and pre-retirement income, and over what replacement rate is considered adequate.[41] Our code will present options for calculating potential retirement and pre-retirement income and leave the assessment of what is considered adequate to researchers.

A second consideration is defining when "pre-retirement" ends and retirement begins. The challenge is that it is often impossible to precisely define when work ends and retirement begins. The growth of bridge jobs means that many workers may be working part-time or in less demanding jobs but have already claimed Social Security or started drawing down assets. And higher-income workers may continue to work full-time well beyond when they claim Social Security and, thus, may have substantial earnings at older ages. For this reason, our code will calculate replacement rates at four different ages (62, 65, FRA, and 70), the same as for calculating counterfactual Social Security wealth, so researchers can decide which age best fits their needs.

Potential Retirement Income

The numerator of the replacement rate ratio is the retirement income that could be generated from households' retirement resources. This calculation is done by converting Social Security wealth, employer-sponsored retirement wealth, and other financial wealth into annual income flows using an actuarially fair single-life immediate annuity. Although few households voluntarily annuitize wealth, annuities are a proxy for a sustainable withdrawal rate. Our code provides two options for annuitizing wealth. One includes the net value of the house as a source of potential retirement income and the other does not.

[41] Biggs and Springstead (2008), Goss et al. (2014), and Munnell and Soto (2005).

The annuity factor is a combination of the discount rate and the population-wide, cohort-specific mortality rate denoted in the following equation:

$$Annuity\ factor = \frac{1}{\sum_{t=a}^{119} P_t\ (1 + r_t)^{(a-t)}} \qquad (3)$$

where a is one of four retirement ages (62,65, FRA, 70), P_t is the cumulative probability of surviving to age t, and r_t is the discount rate. To be consistent with the DB wealth calculations, the default discount rate is 5.8 percent. This value is similar to the long-run interest rate assumption from the latest *Social Security Trustees Report*, which has been fairly stable over time. The code is flexible so researchers can adopt a discount rate that is best suited for their analysis.[42]

Average Pre-retirement Income

The denominator for replacement rates – pre-retirement income – can be defined in a variety of ways.[43] A common definition is the average of the highest 35 years of wage-indexed earnings. This measure is used to determine Social Security benefits and includes lower earning years early on in a respondent's career as well as higher earning years in their 40s and 50s. But living standards typically increase as you earn more, and households may wish to maintain the lifestyle they are accustomed to close to retirement. So, another frequently used definition of pre-retirement income is the highest 5 years of earnings between age 45 and the retirement age. Our code will include both definitions of pre-retirement income.

Pre-retirement income using the highest 35 years of wage-indexed earnings has already been estimated, as it is a component of the Social Security wealth calculations. Pre-retirement income based on the highest 5 years between age 45 and retirement can be estimated using a similar methodology as the Social Security PIA. For respondents with linked earnings data, this measure is easily calculated.[44] For respondents without linked earnings records, we impute the last five years of earnings based on their PIA, gender, race, birth year, education, industry of

[42] See footnote 19 for how to convert historical GST data into DB wealth calculations.
[43] For the vast majority of households, earnings are equivalent to pre-retirement income. Very lower earners may also receive government transfers while very high earners may have capital income. Unfortunately, we do not have a good way of measuring non-earned income across the lifecycle so our measure will overstate retirement adequacy for both very high and very low earners.
[44] As when calculating PIA, earnings from when the linked records end to the four claiming ages (62, 65, FRA and 70) are assumed to equal the average of the last five years observed, indexed by the AWI.

longest job, and HRS cohort using the nearest-neighbor approach.[45] Tables 5 shows Social Security replacement rates by claim age, and they are consistent with replacement rates calculated by SSA's actuaries.[46]

Table 5. *Median Social Security Replacement Rate, by SS Claim Age*

	Social Security claim age			
	62	65	FR	70
Highest 5 years	26%	31%	34%	38%
AIME	36	44	48	59

Sources: HRS (1992-2020); GST (2014); HRS Restricted Summary Earnings file (2020) and RAND longitudinal file 1992_2020v1.

Median total and non-housing replacement rates are shown in Tables 6 and 7, respectively.

Table 6. *Median Total Replacement Rate, by SS Claim Age*

	Social Security claim age			
	62	65	FR	70
Highest 5 years	56%	62%	66%	71%
AIME	71	84	89	105

Sources: HRS (1992-2020); GST (2014); HRS Restricted Summary Earnings file (2020) and RAND longitudinal file 1992_2020v1.

Table 7. *Median Non-Housing Replacement Rate, by SS Claim Age*

	Social Security claim age			
	62	65	FR	70
Highest 5 years	45%	51%	54%	58%
AIME	59	71	76	87

Sources: HRS (1992-2020); GST (2014); HRS Restricted Summary Earnings file (2020) and RAND longitudinal file 1992_2020v1.

[45] All the components of the retirement income are already adjusted for non-respondent spouses while the pre-retirement income isn't adjusted. In our current simple methodology, we won't be picking the same neighbor for the pre-retirement income as we do for retirement income. Therefore, for replacement rate calculations, we don't impute for households with non-respondent spouses and drop them out of any analysis.

[46] The latest version can be found here: https://www.ssa.gov/oact/NOTES/ran9/an2024-9.pdf

Concluding Comments

This guide and accompanying code create an easy to use and flexible methodology for calculating retirement wealth, income, and replacement rate variables in the HRS. It builds on existing work from the RAND Institute and Gustman, Steinmeier, and Tabatabai and establishes a first pass at standardizing best practices for calculating the wealth and income components needed when analyzing a household's financial resources in the HRS.

References

Behaghel, Luc and David M. Blau. 2012. "Framing Social Security Reform: Behavioral Responses to Changes in the Full Retirement Age." *American Economic Journal: Economic Policy* 4(4): 41-67.

Biggs, Andrew G. and Glenn R. Springstead. 2008. "Alternate Measures of Replacement Rates for Social Security Benefits and Retirement Income." *Social Security Bulletin* 68(2): 1-19.

Fang, Chichun. 2022. "Health and Retirement Study Prospective Social Security Wealth Measures of Pre-Retirees Waves 10 and 13." Version 3.0. Ann Arbor, MI: University of Michigan, Institute for Social Research.

Friedberg, Leora. 2000. "The Labor Supply Effects of the Social Security Earnings Test." *Review of Economics and Statistics* 82(1): 48-63.

Health and Retirement Study. 2022. "RAND HRS Longitudinal File 2020 V1 Public Use Dataset." Produced and Distributed by the University of Michigan with funding from the National Institute on Aging (Grant Number NIA U01AG009740). Ann Arbor, MI.

Goss, Stephen, Michael Clingman, Alice Wade, and Karen Glenn. 2014. "Replacement Rates for Retirees: What Makes Sense for Planning and Evaluation?" Actuarial Note Number 155. Baltimore, MD: U.S. Social Security Administration.

Gustman, Alan L., Thomas L. Steinmeier, and Nahid Tabatabai. 2010. *Pensions in the Health and Retirement Study*. Cambridge, MA: Harvard University Press.

Kapinos, Kandice, Charlie Brown, Michael Nolte, Helena Stolyarova, and David Weir. 2022. "Prospective Social Security Wealth Measures of Pre-Retirees." Ann Arbor, MI: University of Michigan, Institute for Social Research.

Munnell, Alicia H. and Mauricio Soto. 2005. "What Replacement Rates Do Households Actually Experience in Retirement?" Working Paper 2005-10. Chestnut Hill, MA: Center for Retirement Research at Boston College.

Quinby, Laura D., and Geoffrey T. Sanzenbacher. 2020. "Do Pensions Matter for Recruiting State and Local Workers?" *State and Local Government Review* 52(1): 6-17.