RAND Center for the Study of Aging
A RAND LABOR AND POPULATION CENTER

RAND Income and Wealth Imputation File (v.P)

The RAND Income and Wealth Imputation File contains the component and ownership variables that were used in RAND HRS income and wealth summary measures. These imputations are distributed as companion files to the RAND Enhanced Fat Files and are incorporated into the RAND HRS file.

The method used to develop these imputations is described in this document. Each original income and wealth component is imputed separately. The original components can vary from wave to wave and are combined when appropriate into the main components of income and wealth to provide consistent categories. The main component and total variables are the same as those included on the RAND HRS file.

For income, the reference period for totals and main components is the last calendar year. When needed, another income measure is added to estimate income received in the last calendar year. For example, if a measure is given as income last month, we impute the monthly amount and examine other information, such as a start date, to determine how many months the income was received in the last calendar year to produce an estimate for that year.

The Imputation file is a respondent-level file with household-level data merged to both individuals in a couple household. The file is sorted by respondent ID (HHID and PN). Prior to 2004, there are two files for each year, one for income and one for wealth. Starting in 2006, the income and wealth variables are combined into one file.

For more detailed information about the variable naming conventions, please see the 2014 RAND Income and Wealth Imputation codebook. You can find more about the data products on the RAND Center of Aging web site: http://www.rand.org/labor/aging/dataprod.

Please contact RAND HRS Help (RANDHRSHelp@rand.org) with any questions.

1. Introduction and Overview

This document describes the variables on the RAND Income and Wealth Imputation File, Version P. This is a cleaned, processed, streamlined, and where necessary imputed, collection of income and wealth variables derived from the Health and Retirement Study (HRS). The HRS is a national panel survey of individuals over age 50 and their spouses. Its main goal is to provide reliable data on the decisions, choices, and behaviors of people as they age and respond to changes in public policy, the economy, and health. The survey elicits information about demographics, income, assets, health, cognition, family structure and connections, health care utilization and costs, housing, job status and history, expectations, and insurance. Over the years the HRS has expanded its multi-diciplinary approach to additional subject areas, including the collection of biomarkers, and added to the external data sources that can be linked to the HRS data.

The HRS is sponsored by the Social Security Administration and National Institute of Aging (NIA), and data collection is conducted by the Institute for Social Research (ISR) at the University of Michigan. It consists of six cohorts:

- Initial HRS cohort, born 1931 to 1941. This cohort was first interviewed in 1992 and subsequently every two years.
- AHEAD cohort, born before 1924, initially a separate study (The Study of Assets and Health Dynamics Among the Oldest Old or AHEAD). This cohort was first interviewed in 1993 and subsequently in 1995, 1998, and every two years thereafter.
- Children of the Depression Age (CODA) cohort, born 1924 to 1930. This cohort was first interviewed in 1998 and then subsequently every two years.
- War Baby (WB) cohort, born 1942 to 1947. This cohort was also first interviewed in 1998 and subsequently every two years.
- Early Baby Boomer (EBB) cohort, born 1948 to 1953. This cohort was first interviewed in 2004 and subsequently every two years.
- Middle Baby Boomer (MBB) cohort, born 1954 to 1959. This cohort was first interviewed in 2010.

In addition to respondents from eligible birth years, the survey interviewed the spouses of married respondents or the partner (if in the same household) of a respondent, regardless of age. Some spouses of the initial HRS entry cohort respondents were age 70 or older and were subsequently included in the AHEAD study instead. These so-called HRS/AHEAD overlap cases may thus have been interviewed in 1992, 1993, 1995, and from 1998 forward. From 1998 onward, households sampled for the new cohorts exclude households in which the spouse or partner would be age-eligible for an earlier cohort. For example, a couple in which one spouse was born in 1935 (HRS age cohort) and the other spouse was born in 1945 (WB age cohort) could have been sampled as part of the original HRS cohort sample but not as part of the WB cohort sample.

The RAND Income and Wealth Imputations are distributed as a companion file to the RAND Enhanced HRS Fat Files[1] and are incorporated into the RAND HRS Data File.[2] The RAND Income and Wealth Imputation File (Version P) contains all six cohorts. This document refers to the entire survey as the HRS and the 1931-1941 cohort that was first interviewed in 1992 is labeled the "initial" or "original" HRS entry cohort.

The method used to develop the imputations is described in Chapter 2. Each original income and wealth component is imputed separately. Some original components lend themselves to combination, such as summing up to three IRAs to provide a total IRA value. Some original income components are combined into logical main components, such as summing income from wages, tips and bonuses, professional practice or trade, and a second job into an earned income component. In addition, total wealth and income variables are derived.

For income, the reference period for totals and main components is the last calendar year. When needed, another income measure is added to estimate income received in the last calendar year. For example, if a measure is given as income last month, we impute the monthly amount and examine other information, such as a start date, to determine how many months the income was received in the last calendar year to produce an estimate for that year.

If the original components are combined for any of these reasons, these original components are referred to as "partial components" in this document, while the combinations or uncombined original components are called "main components" of wealth and income. Analysts may wish to recombine components or to see the individual parts of a total amount. All individually imputed parts are provided in this file. It is important to understand how the parts and the combinations overlap, so that parts of income and wealth are not double-counted.

The HRS contains several auxiliary files. The RAND Income and Wealth Imputation File only incorporates the core interviews. It does **not** include exit interviews or any restricted data, but does use information from the current Tracker, Region and Mobility, and Master ID files.

The data include any individual interviewed at least once. This includes individuals who were age-eligible (born in eligible years) at the time of their first interview, spouses that were not age-eligible at baseline, and spouses who married an age-eligible respondent between survey waves.

The HRS over-samples Hispanics, Blacks, and residents of Florida, and provides weighting variables to make it representative of the community-based population.

The data described in this document are based on the final data releases of 1992, 1993, 1994, 1995, 1996, 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012 and the early data release of 2014.

---

[1] The RAND-enhanced Fat Files are interview year-specific files containing most raw HRS variables merged to the respondent level to facilitate merging longitudinally and with the RAND HRS Data file. These are available on the HRS web site http://hrsonline.isr.umich.edu/. Please see www.rand.org/labor/aging/dataprod for more information.

[2] The RAND HRS Data file is a comprehensive but not complete collection of derived variables based on HRS public release data. It contains all waves of data, with consistent variable names across waves. It is available for download from the HRS web site. Please see www.rand.org/labor/aging/dataprod for more information and sample programs for its use with the Fat Files.

The file is based on data from the latest release of each wave's HRS or AHEAD core data, and Version 1.0 of Tracker 2014.

1.1. Confidentiality and Access Restrictions

The data described in this document are based on HRS public release files. Before using the data, you must have obtained permission from HRS by registering with them for downloading the public release files. The HRS website contains information on the process to register for access to HRS public release data (https://ssl.isr.umich.edu/hrs).

By registering with HRS you agree to the "Conditions of Use" governing access to the data. There is NO RESTRICTED DATA on the RAND Income and Wealth Imputation File.

1.2. Data File Structure

The RAND Income and Wealth Imputation data are distributed in SAS, Stata, and SPSS formats, as a single file that includes all waves of the HRS. The data contain respondents within the HRS, AHEAD, CODA, WB, EBB, and MBB entry cohorts. Table 1 lists the source year of data for each of the entry cohorts, by wave. The 1993 data are treated as Wave 2 data and the 1995 data are treated as Wave 3 data for the AHEAD entry cohort. The 1994 data are treated as Wave 2 data and the 1996 data are treated as Wave 3 data for the HRS entry cohort. The AHEAD and HRS survey instruments in these years differed significantly. This documentation distinguishes between the instruments by using Wave 2A and Wave 3A to refer to the 1993 and 1995 data for the AHEAD entry cohort , and Wave 2H and Wave 3H to refer to the 1994 and 1996 data for the HRS entry cohort.

Table 1. Source of Data for Entry Cohorts by Wave.

<table>
<tr><th rowspan="2">Wave</th><th colspan="6">Entry Cohort</th></tr>
<tr><th>HRS
HACOHORT=3</th><th>AHEAD
HACOHORT=0,1</th><th>CODA
HACOHORT=2</th><th>WB
HACOHORT=4</th><th>EBB
HACOHORT=5</th><th>MBB
HACOHORT=6</th></tr>
<tr><td>1</td><td>1992</td><td>1992
(HRS/AHEAD
overlaps only)</td><td colspan="4" rowspan="3">Not available</td></tr>
<tr><td>2</td><td>1994
(Wave 2H)</td><td>1993
(Wave 2A)</td></tr>
<tr><td>3</td><td>1996
(Wave 3H)</td><td>1995
(Wave 3A)</td></tr>
<tr><td>4</td><td>1998</td><td>1998</td><td>1998</td><td>1998</td><td colspan="2" rowspan="3">Not available</td></tr>
<tr><td>5</td><td>2000</td><td>2000</td><td>2000</td><td>2000</td></tr>
<tr><td>6</td><td>2002</td><td>2002</td><td>2002</td><td>2002</td></tr>
<tr><td>7</td><td>2004</td><td>2004</td><td>2004</td><td>2004</td><td>2004</td><td rowspan="3">Not available</td></tr>
<tr><td>8</td><td>2006</td><td>2006</td><td>2006</td><td>2006</td><td>2006</td></tr>
<tr><td>9</td><td>2008</td><td>2008</td><td>2008</td><td>2008</td><td>2008</td></tr>
<tr><td>10</td><td>2010</td><td>2010</td><td>2010</td><td>2010</td><td>2010</td><td>2010</td></tr>
<tr><td>11</td><td>2012</td><td>2012</td><td>2012</td><td>2012</td><td>2012</td><td>2012</td></tr>
<tr><td>12</td><td>2014</td><td>2014</td><td>2014</td><td>2014</td><td>2014</td><td>2014</td></tr>
</table>

The unit of observation is an individual. Each individual is uniquely identified by a household ID (HHID) and a person number (PN). We combined these variables into a single ID variable, HHIDPN (HHIDPN: HHold ID + Person Number /Num), where HHIDPN = 1000 * HHID + PN.[3] This file may be merged with other HRS data by HHIDPN.

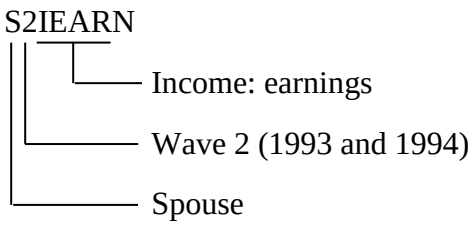
The file includes a flag (HwPICKHH where "w" is the corresponding Wave 1-12) to allow easy reduction to a household level file. A household level file may be extracted by selecting observations where the flag has a value of one (HwPICKHH=1). This will select all single-R households, and the record of the Financial Respondent in couple households where both individuals responded. Note that this only works on a per-wave basis, because household composition may change across waves through divorce and other reasons.

Section A of the data codebook describes identifiers and flags pertaining to the household, respondent, and spouse. Section B provides descriptive statistics and frequencies for the income variables, and Sections C and D provide the same for the cross-wave and cross-sectional wealth variables, respectively. Appendix A provides additional details for any changes made from Version M forward, Appendix B lists the primary residence/mobile home inconsistencies due to Asset Verification, and Appendix C describes methodology that we developed for combining information from the Pension Section (J2) and the Income Section (Q).

[3] This HHIDPN variable is numeric. Also available is RAHHIDPN (RAHHIDPN: HHold ID + Person Num /9-Char), its 9-character string equivalent.

1.3. Variable Naming Conventions

With few exceptions, variable names in the RAND Income and Wealth data follow a consistent pattern. The first character indicates whether the variable refers to the reference person ("R"), spouse ("S"), or the household ("H").[4] The second character indicates the wave to which the variable pertains: "1", "2", "3", "4", "5", "6", "7", "8", "9", "10", "11", "12", or "A". The "A" indicates "all," i.e., the variable is not specific to any single wave. In the Income and Wealth file, this is only used for RAHHIDPN, the string version of the unique identifier of the respondent, but in the RAND HRS it is used for many more variables. The remaining characters describe the concept that the variable captures. For example:



Variable S2IEARN captures the earnings of the spouse. The "I" refers to income, more specifically, income received in the previous calendar year ("M" is used for income received in the previous month).

The name of the variable does not indicate who provided the information. For example, the spouse's earnings may have been reported by the spouse himself or herself, or they may have been reported by a proxy. The HRS obtains many variables, particularly on financial and family matters, by proxy.

In the text below, we may refer to variables such as SwIEARN for example, without specifying the wave. This reference points at the group of variables S1IEARN, S2IEARN, S3IEARN, S4IEARN, S5IEARN, S6IEARN, S7IEARN, S8IEARN, S9IEARN, S10IEARN , S11IEARN and S12IEARN.

Variable labels also follow a consistent pattern. The first characters denote the name of the variable, followed by a colon. Then follows the wave to which the variable pertains (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or 12). The remainder of the label describes the concept that the variable captures. For example, the variable label of S2IEARN is:

S2IEARN:W2 Income:Sp Earnings

It may seem duplicative to include the name of the variable and the wave in the variable label. However, SAS and Stata often suppress the variable name and instead use its label in the presentation of results.

[4] The reference person need not be the person who responded to the question. It is the person whose information is recorded in the variable.

In previous versions of the income and wealth data, the names of some detailed component variables followed a different convention. As of Version M, naming has been made consistent for all income and wealth variables. For a complete list of which variables have been renamed from previous versions to Version M of this codebook, see Appendix A.

Wealth variables that use the cross-wave imputations (see Chapter 2) have an "A" after the wave number, whereas the corresponding wealth variables that use the cross-sectional imputations have a "W". For example, H9ASTCK is the value of stocks and mutual funds of the household in Wave 9, using cross-wave imputations if imputation is necessary, and H9WSTCK is the value of stocks and mutual funds of the household in Wave 9, using cross-sectional imputations if imputation is necessary.

For both income and wealth variables, a subsequent "O" is used for variables that indicate ownership (i.e., whether the household owns the asset or the household or individual receives the type of income), whereas "F" is used for variables that indicate the type of imputation necessary (imputation flags; see Chapter 3).

These conventions are summarized in Tables 2 and 3 for the income and wealth variables, respectively. The remainder of the variable name, called the *stem*, indicates the type of income or asset. For example, "EARN" denotes earnings and "STCK" denotes stocks and mutual funds.

For example, for Wave 10, R10IFEARN indicates whether the respondent's earned income is imputed. R10IOEARN indicates whether the respondent reports (or is imputed to) having any earned income. R10IEARN is the reported or imputed amount the respondent earned in the last calendar year, set to zero for those who do not receive any. For the spouse, SwIFEARN, SwIOFEARN, and SwIEARN are analogous to RwIFEARN, RwIOEARN, and RwIEARN, respectively. If the respondent does not have a spouse, the value for the spousal variable is set to .U, and ownership for the spouse is also set to .U. The imputation flag for the spouse variable then has the special value 8 (see Chapter 3).

In previous versions of the RAND Income and Wealth Imputation File, partial component values were set to missing if the partial component was not owned. The current version harmonizes the definition of partial component variables with main component variables and totals, and thus sets the value to zero if the component is not owned.

Table 2. Variable Name Prefixes for Income Variables

Prefix	***Description***
	Household Level
HwIF	Household annual income imputation flag (previous calendar year)
HwIO	Whether household received income from this source in the previous calendar year
HwI	Household annual income value, set to zero for non-owners (previous calendar year)
HwM	Household partial income component value – monthly value, where original value was reported as a monthly amount. If total amount received last year is available and amount received last month is not, then total annual amount is used to estimate a monthly amount.
HwN	Household partial income component – number of months received in last calendar year, where the original value was reported as a monthly amount. This is computed from month and year began receiving income. If month began is unknown and year began is the last calendar year, then number of months received is assigned a value of 6 months. If year began is unknown, then number of months received is assigned a value of 12 months and a flag is set to 1.
HwNF	Household partial income component – flags cases where there is no information on the number of months received. When the year began receiving income is unknown, the number of months received in the last calendar year is assigned a value of 12 and the flag is set to 1.
	Respondent Level
RwIF	Respondent income imputation flag (previous calendar year)
RwIO	Whether respondent received any income from this source in the previous calendar year
RwI	Respondent income value, set to zero for non-owners (previous calendar year)
RwM	Respondent monthly income; apart from respondent vs. household, analogous to HwM
RwN	Respondent income – number of months received in last calendar year; apart from respondent vs. household, analogous to HwN
RwNF	Respondent income – flags cases where there is no information on the number of months received; apart from respondent vs. household analogous to HwNF
	Spouse Level
SwIF	Equivalent of RwIF for the spouse
SwIO	Equivalent of RwIO for the spouse
SwI	Equivalent of RwI for the spouse
SwM	Equivalent of RwM for the spouse
SwN	Equivalent of RwN for the spouse
SwNF	Equivalent of RwNF for the spouse

Table 3. Variable Name Prefixes for Wealth Variables

Prefix	*Description*
HwAF	Household asset cross-wave imputation flag for a single-component asset, e.g., HwAFIRA1 to summarize imputation status of HwAIRA1.
HwWF	Household asset cross-sectional imputation flag for a single-component asset, e.g., HwWFIRA1 to summarize imputation status of HwWIRA1. (These are equal to the corresponding HwAF variables and only included for convenience.)
HwAX	Household asset cross-wave imputation summary flag for a combined asset, e.g., HwAXIRA to summarize imputation status of HwAIRA, which is the sum of HwAIRA1, HwAIRA2, and HwAIRA3.
HwWX	Household asset cross-sectional imputation summary flag for a combined asset, e.g., HwWXIRA to summarize imputation status of HwWIRA. (These are again equal to the corresponding HwAX variables.)
HwAO	Whether household owns asset, using cross-wave imputations
HwWO	Whether household owns asset, using cross-sectional imputations
HwA	Household asset value, set to zero for non-owners, using cross-wave imputations
HwW	Household asset value, set to zero for non-owners, using cross-sectional imputations

1.4. Missing Values

Variables may contain missing values for several reasons. SAS and Stata offer the capability to distinguish multiple types of missing values, and we have attempted to record as much information as possible. Generally, the codes adhere to the classification in Table 4.

Table 4. Missing Codes

Code	Reason for missing
.	Reference person did not respond to this wave
.D	Don't know
.R	Refused
.X	Does not apply (specifics depend on variable)
.Q	Data not available because of HRS and AHEAD survey instrument differences in Wave 2 or 3
.U	Reference person is not married (for spousal variables)
.V	Spouse did not respond this wave (for spousal variables)
.S	Information not available due to skip patterns, typically because the interview is by proxy respondent
.M	Other missing

The coding scheme varies across variables. Not all missing codes are used in the RAND Income and Wealth Imputation File. Consult the Data Codebook section of this document for details on individual variables. The RAND Income and Wealth Imputation File in Stata format is for use with Version 11 or later.

2. Imputation Methodology

2.1. Background

The RAND Income and Wealth Imputation data contain a number of wealth and income variables. Where missing, we imputed their values. In this chapter, we give an overview of the question sequences and imputation methods. A detailed technical description can be found in Hurd, Meijer, Moldoff, and Rohwedder (2016).[5]

Most HRS and AHEAD questions on wealth and income follow the same pattern. Consider holdings of stocks and mutual funds as an example. First, the interviewer asks whether the respondent (or his/her spouse or partner) owns any shares of stock or stock mutual funds. If affirmative, the interviewer asks the value of these stock holdings. If the respondent is unable or unwilling to provide an exact amount, the interviewer asks whether it is more than $25,000. If the answer is "more than $25,000," the interviewer asks whether it is more than $125,000, whereas if the answer is "less than $25,000," the interviewer asks whether it is more than $2,500. Depending on the responses, the ranges are narrowed down to $0-2,500; $2,500-25,000; $25,000-125,000; $125,000-400,000; $400,000 or more. These ranges are known as "brackets," and the sequence of probes into increasingly narrow ranges are known as "unfolding brackets" questions. The brackets vary by asset and income category, and the cut-off values (also known as thresholds), though generally stable, can change between waves. For example, the cut-off values for dividend and interest income in HRS 1994 (Wave 2H) are $200, $500, $2,500, and $10,000, while in HRS 1996 (Wave 3H) they are $1,000, $5,000, and $25,000.

The respondent may opt out of the question sequence at any time. As a result, the raw data contain valid zero-value responses, exact amounts, complete bracket responses, incomplete bracket responses, claim of ownership without any information about the value, and unknown ownership. (We also use the term "ownership" to indicate whether a household or individual receives a certain kind of income or holds a certain kind of debt.) An incomplete bracket results if the respondent provided some information about the value, but was unable or unwilling to respond through the last unfolding bracket probe. For example, if the respondent indicated that the stock holdings amount to more than $25,000, but did not say whether they are worth more than $125,000. In that case, the range is an open-ended $25,000 or more. A claim of ownership without value results if the respondent indicated that stock is owned, but revealed neither the exact amount nor a range. A claim of ownership without value is a special case of an incomplete bracket, namely an open-ended bracket of greater than zero dollars.

As an illustration, Table 5 shows the frequency distribution of response types on the ownership and value of stock holdings in HRS 1994 (Wave 2H). The stock holdings question is asked from the so-called "financial respondent" in the household, and the unit of observation in the table is a household. The majority of respondents, 68.1 percent, report not owning any stocks (other than

[5] Hurd, M. D., E. Meijer, M. Moldoff, and S. Rohwedder (2016). *Improved wealth measures in the Health and Retirement Study: Asset reconciliation and cross-wave imputation*. Santa Monica, CA: RAND Corporation, Center for the Study of Aging (http://www.rand.org/pubs/working_papers/WR1150.html).

in retirement plans, which are not covered by this question). About one-fifth, 20.3 percent, own stocks and provide an exact value. All other categories require imputation.

Table 5. Distribution of Response Types on Stock Holdings (HRS 1994)

	Frequency	Percent
Continuous value	1,431	20.3
Complete bracket	487	6.9
Incomplete bracket	34	0.5
Owns, no value/bracket	111	1.6
No asset	4,803	68.1
Don't know ownership	66	0.9
No financial respondent	119	1.7
Total	7,051	100.0

Note the last category in Table 5 "No financial respondent." These are cases in which the HRS, for whatever reason, did not interview a financial respondent.[6] For those cases, virtually nothing is known about financial issues. The RAND Income and Wealth Imputation data contain imputed values for these households, but the user should be aware that these imputations are subject to potentially large errors. They may be identified through flag variables. For example, Table 5 is derived from variable H2AFSTCK ("H2AFSTCK:W2 Asst Flag:Stocks").

In summary, the data contain continuous responses and several types of responses that require imputations. In decreasing order of informational content:

- Case 1: We may know a "complete" range of values;
- Case 2: We may know that the household owns the asset (or has the income type), but have no information on its value, or only coarse information in the form of incomplete brackets; or
- Case 3: We may not even know whether the household owns an asset, much less its value.

2.1.1. Alternative Question Sequences

While the majority of income and asset questions follow the pattern described above, there are deviations.

Some questions, particularly income questions in HRS 1992 (Wave 1), do not probe for brackets if the respondent is unable or unwilling to provide an exact amount. In the data, we classify missing responses in this case as if there were unfolding bracket questions, but the respondent refused to provide any range information, i.e., as Case 2 above.

[6] In some households, a financial respondent was designated, but in fact provided no financial data, perhaps due to a partial interview that ended before the main financial section of the survey was conducted. For some of these cases, house and mortgage information may have been provided but no other wealth or income data was collected in the main income and wealth module which follows the housing module in the survey instrument. For imputation of income and non-housing wealth, cases missing the entire module of financial data are treated as though there is no financial respondent.

Some interviewers in Wave 1 used so-called "range cards" instead of the sequence of unfolding bracket questions. This was especially prevalent for questions on the value of primary residence, mortgages, home loans, equity lines of credit, and debt. The range cards contain a list of ranges. For example, a card may have shown $0-100; $100-500; $500-1,000; $1,000-5,000; $5,000-10,000; $10,000 or more. The cards were intended for other purposes, but were sometimes used inadvertently when respondents were unable or unwilling to provide an exact amount.[7] Respondents who were presented a range card had instant knowledge of all cut-off values, as opposed to gradual access in unfolding bracket questions. The cut-off values on the range cards were typically different from those in the appropriate unfolding bracket sequence. We treat responses from range cards in the same way as complete brackets, i.e., Case 1 above. Naturally, we account for the cut-off values on the range cards, even if they are different from those in the unfolding bracket sequence.

Starting in HRS 1998 (Wave 4), the "entry point" of the unfolding bracket sequence was randomized in questions about assets. In other words, respondents who were unable or unwilling to provide an exact amount were asked whether the value was more than a certain value, where that value varied across respondents. The underlying idea was to reduce at the population level any response bias that might arise from the value of the entry point ("acquiescence bias"). The fact that the entry point varied across respondents does not enter our imputations, and we process the resulting information in the same way as in previous waves.

Also starting in Wave 4, interviewers were able to record a new response. For example, suppose the interviewer asked, "Is the amount greater than $5,000, less than $5,000, or what?" The potential answers now are "less than $5,000," "about $5,000," and "more than $5,000." In earlier waves, the second response was combined with the first or third response. The additional option is present at each subsequent branch. Where the respondent indicated that the amount was "about" equal to a certain value, we took that value as an exact response and did not impute anything. However, this is still classified as a complete bracket in the imputation flag variable and not as a continuous amount. We treated range responses in the same way as those in earlier waves.

Finally, some cut-off values of specific asset and income questions changed between waves, as did the entry point. In addition, the way unfolding bracket information is presented in the raw data changes over time, from variables representing the "yes/no" questions through Wave 5, to variables summarizing them as a "minimum/maximum" of the range in Wave 6 forward. This did not affect the resulting response types.

[7] The Wave 1 interview was face-to-face; subsequent interviews were conducted mostly by telephone. For all subsequent waves, interviews with respondents 80 years of age or older, as well as initial interviews (for all cohorts and new cohorts), are also conducted face-to-face. Starting with the HRS 2006 wave, about half of the sample is interviewed face-to-face (to facilitate the collection of additional content), and half is interviewed by telephone.

2.2. Imputation Process

The HRS public release files provide imputations for many asset and income types in earlier waves, but the imputation method is not consistent across all waves, and there are no imputations in more recent waves. The RAND Income and Wealth Imputation data contain imputations of all asset and income types using a consistent method for all waves. Beginning with HRS 2006, RAND has provided the income and asset imputations for the HRS. The RAND HRS data file contains summary measures of income and assets. The RAND Income and Wealth Imputation File is a more complete and detailed file, containing all individual component imputations in addition to the summary measures.

As defined previously, there are three types of missing values that require separate types of imputation. Correspondingly, we developed three progressive imputation steps: to impute an exact amount, given that a range is known; to impute a range, given that ownership or only incomplete range is known; and to impute ownership, in case nothing is known. Table 6 illustrates the type of imputation necessary for each type of missing value.

The imputation process is progressive in the sense that we first impute ownership for those for whom nothing is known. Given ownership, we impute brackets. Given brackets, we impute exact amounts. We always use all available information. In particular, where incomplete brackets are known, we impute complete brackets in the given range.

Table 6. Response Types and Required Imputations

Reported Information	Required Imputation
Continuous value	None
Complete bracket	Amount
Incomplete bracket	Bracket, Amount
Owns, no value/bracket	Bracket, Amount
No asset	None
Don't know ownership	Ownership, Bracket, Amount
No financial respondent	Ownership, Bracket, Amount

2.2.1. Ownership Imputation

To impute ownership, we first estimate a logistic regression model of ownership based on the sample of respondents with nonmissing ownership information for the asset or income type at issue. The covariates are discussed in Section 2.2.4. Next, we calculate the predicted probability of ownership for households with missing ownership information. Finally, we draw a random number from a uniform distribution between zero and one. We impute ownership if the predicted probability exceeds the random number, and non-ownership otherwise.

The estimation sample varies by asset and income type. For example, a household that reports whether they own their primary residence but does not report whether they own stocks or mutual funds enters the estimation sample for ownership of the former, but not the estimation sample for ownership of the latter.

In some waves and for some asset and income types, ownership is rare, and the logistic regression model fits the data poorly. Instead of imputing on the basis of a logistic model, we randomly assign ownership with a probability of ownership found in the nonmissing sample. This is equivalent to fitting a logistic model without covariates. We apply this procedure for assets in waves with fewer than 50 households reporting ownership.

2.2.2. Bracket Imputation

We impute brackets for asset owners and income recipients (imputed and reported) who do not report a continuous value, and do not fully complete the questionnaire bracketing sequence on asset or income value. First, we estimate an ordered logit model based on the sample of households who do not report a continuous value but do complete the bracketing sequence. The covariates are discussed in Section 2.2.4. Next, we calculate the predicted probabilities of being in each bracket for respondents with missing or incomplete bracket information. For those who partially complete the bracketing sequence, we calculate conditional probabilities based on the range of possible values from their answers. Finally, we draw a random number from a uniform distribution between zero and one, and assign a bracket based on a comparison of the random number with the cumulative distribution of range probabilities.

For some asset and income types, notably Wave 1 incomes, no bracket questions were asked. For these items, this step is skipped and we treat the strictly positive dollar range as a single large open-ended bracket.

Beginning in Wave 4, households that do not report a continuous value at the first opportunity but do give an "about" response during the unfolding bracket sequence of questions are included in the logit model. Probabilities are then estimated for being in each bracket as well as each cutpoint value. Households imputed to one of the cutpoint values need no amount imputation for that particular component.

For some asset and income types, fewer than 50 households completed the bracket sequence in a certain wave. If this is the case, we follow the same strategy as with ownership in such situations. That is, we impute the bracket using only the marginal probabilities and no covariates. In some rare cases, there is not even enough information in the marginal probabilities, and we skip the bracket imputation and impute amounts directly. This happens when only one complete bracket range is reported (so this would otherwise get probability 1), or when the incomplete bracket range reported is completely outside the range of the reported complete bracket ranges.

2.2.3. Amount Imputation

We impute exact amounts for all cases with (reported or imputed) bracket information. The procedure is different for cases in closed vs. open-ended brackets.[8] For closed brackets, we use a "nearest neighbor" approach; for open-ended brackets, a tobit-based approach. The following discusses the two approaches in turn.

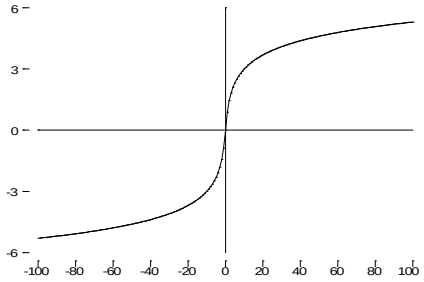
In the "nearest neighbor" approach for closed brackets, we first estimate a linear regression model based on the sample of households who report an exact continuous value. The covariates are discussed in Section 2.2.4. The distribution of asset and income amounts tends to be roughly lognormal, so we would like to apply a logarithmic transformation to the outcome (asset, income) variable. However, some outcomes, such as business income, may be negative. The frequency with which this occurs is very low—too low to allow for a fully flexible model specification. Instead, we therefore apply the inverse hyperbolic sine transformation[9] and use this as the dependent variable. Next, we compute predicted values for all cases, both with and without exact amounts. For each missing observation in a closed bracket, we find the household that is closest in *predicted value*, among the households that report an exact amount within that bracket. This is the nearest neighbor. We then impute the actual value reported by the nearest neighbor.

For some asset and income types, fewer than 50 households reported a continuous value in a given wave. In this case, instead of a nearest neighbor from a regression model, a donor household is found using a conditional hotdeck procedure. This can be viewed as a nearest neighbor method without covariates, but because there are multiple households that are equally near (i.e., they all gave continuous answers that fall in the given bracket), one is chosen at random.

Another exception is that sometimes there are not enough donors in the given bracket, where we define "not enough" as being less than two observations. In this case, a value within the bracket

[8] A closed bracket has finite cut-off values; an open-ended bracket is, for example, $500,000 or more, so the upper limit is infinite.

[9] The inverse hyperbolic sine transformation is given by $y = \ln(Y + \sqrt{Y^2 + 1})$. For positive values of outcome Y, not close to zero, this transformation closely mimics the logarithmic transformation. Only for small amounts, on the order of between -$10 and +$10 is the transformation appreciably different from the logarithmic transformation. The transformation is point-symmetric around zero. It may be graphed as:



If the logarithmic transformation were added to this graph, it would be indistinguishably close on the north-east quadrant (up to a scale factor).

is imputed from a lognormal regression model. This is similar to the tobit procedure discussed below, but includes all positive values in the estimation.

In exploratory work, we applied the nearest neighbor method to missing amounts in both closed and open-ended brackets. However, we found that the resulting imputations generated implausible distributions at the top of the distribution. The data contain some outliers, which the nearest neighbor approach selects with too high a frequency to be plausible. We therefore developed an alternative approach for open-ended brackets.

For missing observations in open-ended brackets, we estimate a separate model. In essence, we estimate a lognormal regression model. However, as indicated above, the observed distributions differ from the lognormal distribution because there occasionally are negative values, and in general the smaller observed values do not fit the lognormal distribution as well. But we only use this model for imputing values in the upper bracket, so the goal is to approximate the right tail of the distribution closely. Therefore, we first censor observations in the bottom 25 percent of the outcome variable. That is, we keep the information that they are less than the 25th percentile, but not the actual values. This ensures that the values used in the estimation are all positive and that their distribution is close to (censored) lognormal. Based on this sample of nonmissing observations, we estimate a censored regression (tobit) model. The covariates are discussed in Section 2.2.4. The dependent variable in this model is a logarithmic transformation of the actual amount, and the residuals on this scale are assumed to be normally distributed and homoskedastic. We then compute predicted values of the log-amount for missing observations. To preserve the spread of the distribution of outcomes in the imputations, we add a draw from the residual distribution to this log-amount, and then untransform (exponentiate) the result. The residual distribution is assumed to be normal (with a zero mean and a standard deviation equal to the estimated standard deviation of the residual in the tobit regression model), but truncated from the left to ensure the resulting value is in the top bracket. Note that this truncation point is higher than the censoring point used in the estimation.

For some asset and income types, fewer than 100[10] households reported a continuous value in certain waves. If this is the case, we follow an analogous strategy as with ownership and bracket in such a situation. That is, we impute the value using only the marginal distribution and no covariates, where the marginal distribution is a truncated lognormal distribution. We implemented this by estimating the aforementioned tobit model, but omitting the covariates, and imputing the value accordingly.

2.2.4. Covariates

The number of model specifications in the imputations is large. There were four waves of data when these methods were initially developed, with well over a dozen asset and income types, and each requires four equations. We experimented extensively with model specifications. On the one hand, we would like to select covariates that predict the outcome variables best. On the other hand, we would like the specifications to be parsimonious and consistent across asset and

[10] We use 100 here rather than the 50 used for the other models, because the estimates of the tobit models with covariates were still unstable with sample sizes between 50 and 100.

income types. Consistency across asset and income types caused problems with asset and income types where only small samples were available. In the end, we opted for the same set of covariates in all asset model specifications (ownership, bracket, and amount for all asset types), and another set in the income specifications. The sets are formed by principal components of approximately 30 underlying covariates. For income imputations, the underlying covariates include (transformations of) husband and wife's employment status, education, health status, age, race, marital status, occupation class[11], cognition, and bequest expectations. For wealth imputations, the same set applies; however, they exclude employment status and include a number of income amounts (imputed, when necessary) and indicators of pension or government benefit receipt. We found that the first ten principal components resulted in model fits that in most cases were very close to the fit from the larger set of covariates, with very few sample size issues. (The previous sections explained what we do in cases where there are sample size issues.) The set of regressors varies slightly across waves, and the principal component loadings are computed separately for each wave. For example, Table 7 shows summary statistics and loadings of the covariates that make up the principal components of the Wave 2 income imputations.

When there is no financial respondent, the income variables are not available as covariates for the assets, and therefore a smaller set of underlying covariates is used. To impute these cases, we compute principal components from this more limited set of covariates, and estimate the corresponding imputation models. Thus, this doubles the number of imputation models used.

The imputations using this set of covariates should be satisfactory for many purposes. Comparisons of the resulting distributions of income and assets with distributions from external sources (e.g., asset distributions from the Survey of Consumer Finances) indicate that the income and asset distributions match well. If they differ substantially from distributions in other surveys, the distributions in the HRS often seem to be more plausible. However, the imputations are sometimes less satisfactory for analyses of longitudinal patterns, especially asset changes between waves. In particular, when a household reports a "no value/bracket" or "don't know

[11] Beginning in Wave 8 (2006), HRS changed the coding of occupation from the 1980 U.S. Census Occupation Codes to those for 2000 (a supplemental file has since been released for Wave 7 (2004) that uses the 2000 U.S. Census Occupation Codes as well). One of the covariates we use in the imputation process is an indicator for whether or not a person's occupation is defined as "professional" (See Table 7, M_PROF and F_PROF). The 1980 codes were collapsed into 17 categories, and determining who was professional was relatively straightforward (See JMW201AM in the HRS 2004 codebook for a list of these categories). Specifically, we defined professionals as anyone in the first two categories (i.e., "Managerial specialty operation," or "Professional specialty operation and technical support").

The 2000 U.S. Census Occupation Codes were used for Wave 7 (2004) to Wave 10 (2010) (a supplemental file has since been released for Wave 10 that uses the 2010 Census Occupational Classification System as well). The 2000 codes were collapsed into 25 categories, which were quite different from the 17 categories produced using the 1980 codes. We determined that the first 11 categories could appropriately be considered professional occupations (see KMW201AM in the HRS 2006 codebook for a list of these categories).

The 2010 Census Occupational Classification System was used for Waves 10, 11 and 12 (2010, 2012 and 2014). The 2010 codes were collapsed into 23 categories, and are not that different from the 25 categories produced using the 2000 codes. Therefore, the first 10 categories were considered professional occupations (see NMW201AM in the HRS 2012 codebook for a list of these categories). In the end, we evaluate all of the information described above regarding occupation when deriving the "professional" covariate.

ownership," this method may lead to large changes between waves, more than is to be expected in the population, because the method does not take the correlation (or persistence) across time into account. Therefore, starting with version M of the RAND Income and Wealth Imputation File, we provide *cross-wave* imputations of asset variables, which take information from adjacent waves into account. These cross-wave wealth imputations are in addition to the cross-sectional wealth imputations, which are computed without using adjacent wave information. The main RAND HRS file only includes the cross-wave imputed variables, but the RAND Income and Wealth Imputation File includes both the cross-wave and the cross-sectional imputations.

Table 7. Loadings on the First Two Principal Components

(Wave 2 Income Imputations)

			Loadings on		
Variable	Mean	Std. Dev.	First comp't	Second comp't	Description
BEQ10	61.54	42.43	0.3163	-0.0648	Probability Bequest $10,000+
BEQ10M	0.0527	0.2234	-0.1277	0.1219	Bequest $10,000 missing
BEQ100	31.80	40.39	0.3008	-0.0415	Probability Bequest $100,000+
BEQ100M	0.0568	0.2315	-0.1208	0.1140	Bequest $100,000 missing
M_COLLEG	0.1533	0.3603	0.2299	-0.0211	Male: College Graduate
F_COLLEG	0.1245	0.3302	0.1500	-0.1255	Female: College Graduate
M_HSGED	0.3966	0.4892	0.1476	0.1601	Male: HS Diploma or GED
F_HSGED	0.5304	0.4991	0.1070	0.0217	Female: HS Diploma or GED
M_EXHLTH	0.3529	0.4779	0.2874	0.0348	Male: Excellent/Very Good Health
F_EXHLTH	0.4283	0.4949	0.2436	-0.0867	Female: Excellent/Very Good Health
M_PRHLTH	0.1747	0.3797	-0.1119	0.2326	Male: Fair/Poor Health
F_PRHLTH	0.2008	0.4006	-0.2311	0.0722	Female: Fair/Poor Health
M_PROF	0.2162	0.4117	0.2557	0.0348	Male: Professional Work
F_PROF	0.2077	0.4057	0.1667	-0.1149	Female: Professional Work
M_WORK	0.4577	0.4982	-0.2311	0.0722	Male: Currently Working
F_WORK	0.4883	0.4999	0.2557	0.0348	Female: Currently Working
M_UNEMP	0.0202	0.1407	0.1667	-0.1149	Male: Unemployed
F_UNEMP	0.0214	0.1446	0.2506	-0.0675	Female: Unemployed
M_DISAB	0.0923	0.2895	0.1314	-0.2130	Male: Disabled/Temp Laid Off
F_DISAB	0.9665	0.2955	-0.0131	-0.0057	Female: Disabled/Temp Laid Off
M_RETIR	0.2014	0.4011	-0.0316	-0.0436	Male: Retired
F_RETIR	0.0978	0.2971	-0.1175	0.1382	Female: Retired
AGE	58.68	4.85	-0.2017	0.0259	Age of Oldest Partner
AGESQ	3466.35	580.17	0.0737	0.3712	Squared Age of Oldest Partner
SINGLFEM	0.2308	0.4214	0.0301	0.1841	Single Female Financial Respondent
MARRIED	0.6722	0.4694	0.0163	0.4601	Married Couple
NONWHITE	0.2914	0.4544	0.0161	0.4618	Non-White Financial Respondent
MISSCOGN	0.0906	0.2871	-0.2655	-0.3032	Missing Cognition Score
LOWCOGN	0.2804	0.4492	0.2819	0.2625	Low Cognition Score

The cross-wave imputations add a few covariates to the models: in addition to the 10 principal components, we include (the inverse hyperbolic sine of) the value of the asset in the previous wave and (the inverse hyperbolic sine of) the value of the asset in the next wave. If there is a strong persistence in asset values over time, including these past and future values, it should capture this and lead to imputations that show smoother patterns across time. If in the adjacent wave, the asset is not owned, we include zero as its value, and we include dummies for whether

the household owns the asset in the previous and next wave. Changes in marital status (where we treat cohabitation the same as marriage, as is usual in the HRS) have potentially large effects on wealth, and therefore, in this case, wealth changes should be less smooth. Therefore, we also include a set of dummies for such changes: divorce or separation, death of the spouse/partner, or remarriage/new partnership. There is one set for changes between the last wave and the current wave and one set for changes between the current wave and the next wave. See Hurd et al. (2016)[12] for details of the definitions.

If the household did not report a continuous value (or a no asset response) in an adjacent wave, we have a missing covariate, and thus we cannot impute the value for the current wave. To solve this problem, we first compute cross-sectional imputations (i.e., without the cross-wave information) for each household, and then use the cross-sectional imputation of the adjacent wave as the covariate in the current wave. However, for households for which we have no adjacent wave information, this still does not solve the issue. This happens when a household enters the sample (the first wave a cohort is entered); in the latest wave (in Wave 12); or when a household did not participate in a wave. For imputing these cases, we use imputation models that include the cross-wave information from only the previous wave or only the next wave, whichever is applicable. If neither previous nor next wave information is available for a household, we use the cross-sectional imputation.

In principle, all imputation models are wave-specific. That is, principal components are computed for each wave separately, and coefficients of the covariates in the imputation models are computed for each wave separately. However, the number of marital status changes is typically too small to estimate the coefficients of these dummies with enough precision in each wave separately. Therefore, we use all waves jointly to estimate these coefficients. That is, we estimate wave-specific coefficients for all covariates except the indicator variables for the change in marital status.

2.3. Asset Verification

After a successful experiment in 2001, described in Hill (2006),[13] the HRS added the *asset verification* section (section U; also called *asset reconciliation* section) to the questionnaire in 2002. Whenever there is a large discrepancy, defined as more than $50,000, between an asset value in the previous wave and the value of the same asset in the current interview, the respondent is asked to verify, or correct when necessary, the asset values in the previous and current wave. The idea behind this is that large changes in asset values are rare, and therefore, if we see them in the data, there may have been a reporting or data entry error, which we want to correct. The respondent is only asked these verification questions if the same (financial) respondent reported the asset in both waves and total wealth differs by more than $150,000. The

[12] Hurd, M. D., E. Meijer, M. Moldoff, and S. Rohwedder (2016). *Improved wealth measures in the Health and Retirement Study: Asset reconciliation and cross-wave imputation*. Santa Monica, CA: RAND Corporation, Center for the Study of Aging (http://www.rand.org/pubs/working_papers/WR1150.html).

[13] Hill, D. H. (2006). Wealth dynamics: Reducing noise in panel data. *Journal of Applied Econometrics*, *21*, 845-860.

former is for disclosure reasons and the latter is to avoid flagging portfolio rebalancing (e.g., selling stocks and buying bonds instead).

In HRS 2002 and HRS 2004 (Waves 6 and 7), respondents were asked at most about three assets in the asset verification section. From 2006 onward, respondents could potentially be asked about all assets that were checked in this section, although it does not happen often that a respondent is asked to verify more than a few assets. There are other changes between waves, for example in the preloaded information from the previous wave and how it is used. See Hurd et al. (2016)[14] for the details.

Starting with version M, the RAND wealth measures take corrections from the asset verification section into account. These corrections precede the imputations. First, corrections for the current wave are implemented. Then, corrections of the previous wave's value are evaluated. If the asset was not asked about in the previous wave's asset verification section, then we implement the correction of the previous wave's value. If the asset was also asked about in the previous wave's asset verification section, we generally use the result from the previous wave's asset verification section (the contemporaneous one). However, we inspect all cases where the results from the previous wave's asset verification section and the current wave's asset verification section differ by more than a factor of 9 to catch gross errors, in particular accidental errors in the number of zeros.

The corrections from the asset verification section occasionally lead to some complications, such as inconsistent data that cannot occur in the main questionnaire because of the main questionnaire's skip patterns, and complications with the implementation of the imputation methods. An example of the former is where the main questionnaire has a branch of questions in case the primary residence is a mobile home, and another branch of questions if it is another type of dwelling. In some cases, the corrections from the asset verification section lead to the household having a number in both branches. We check all such cases and determine what is the most reasonable way to interpret the data and then make it consistent with the logic of the (main) questionnaire. A list of these primary residence/mobile home inconsistencies is provided in Appendix B. An example of a complication with the imputations is where unfolding bracket thresholds are not always the same. See Hurd et al. (2016)[15] for the details and how we dealt with these issues.

[14] Hurd, M. D., E. Meijer, M. Moldoff, and S. Rohwedder (2016). *Improved wealth measures in the Health and Retirement Study: Asset reconciliation and cross-wave imputation*. Santa Monica, CA: RAND Corporation, Center for the Study of Aging (http://www.rand.org/pubs/working_papers/WR1150.html).

[15] Ibid.

2.4. Substantive Differences Across Waves

In addition to survey-technical changes between waves, and the introduction of the asset verification section, there have been a few changes that may affect the comparability of asset values across waves. We discuss the most important changes here.

Notable Differences between Waves HRS 1992 (Wave 1) and HRS 1994 (Wave 2)

Net value of vehicles: Wave 1 includes a measure of the value of a recreational vehicle or motor home in the Housing Section, and a measure of the value of other vehicles in the Asset Section. We separately imputed these values and summed them. For Wave 2, the two components are incorporated into one measure of the net value of vehicles in the Asset Section.

Notable Differences between HRS 1994 (Wave 2) and HRS 1996 (Wave 3)

Asset income: In Waves 1 and 2, asset ownership and value were asked in the Asset Section, whereas income from assets was asked separately in the Income Section. Starting in Wave 3, income from asset questions were incorporated in the Asset Section. For example, if the respondent indicated owning stocks, the interviewer followed up with a question about dividends. This increased the response rate for asset income. Assuming that this increases the quality of responses, it also improves the accuracy of our asset imputations, because asset income is an explanatory covariate of our imputation model.

Net value of IRA/Keogh accounts: In Waves 1 and 2, respondents were asked to report the total value of all Individual Retirement Account (IRA) and Keogh accounts. In Wave 3, separate questions were asked about the largest, second largest, and all other accounts.

Notable Differences between HRS and AHEAD (Waves 2 & 3)

HRS 1994 and AHEAD 1993 (Wave 2)

AHEAD 1993 has a very different structure than HRS 1994. With the exception of Social Security benefits, SSI, and food stamps, respondents are expected to specify the types of income received. For example, they are asked if they have "any regular income," and if so, they are asked to identify the source. So there are no specific questions such as, "Do you receive any income from pensions?" However, they can describe up to 4 regular incomes per partner, and up to 3 household investment incomes. The result is many separate components for those combined in other waves. For example, Wave 2 asks about several types of stock income, and later waves ask only about total stock income. Specific question wording differences are described in the "Cross Wave Differences in Original HRS Data" subsections in the codebook.

HRS 1996 and AHEAD 1995 (Wave 3)

The structures of HRS 1996 and AHEAD 1995 are very similar. However, there are several notable differences. These are outlined under the "Cross Wave Differences in Original HRS Data" subsections in the codebook.

Notable Differences between HRS 1998 (Wave 4), HRS 2000 (Wave 5), and HRS 2002 (Wave 6)

The structures of Waves 4, 5, and 6 are very similar. Any important differences are specified under "Cross Wave Differences in Original HRS Data" subsections in the codebook.

Notable Differences between HRS 1998 (Wave 4) through HRS 2002 (Wave 6) and HRS 2004 (Wave 7) and later

The structures of questions from Wave 7 forward are very similar to those in Waves 4 to 6. However, Waves 7 and later no longer ask for income from trusts and alimony specifically. One can assume that these types of income would now be reported with non-specific other income. We have looked at this for cases that previously did report alimony or trust income, and found that other income did not increase as one might expect.

Notable Differences between HRS 2000 (Wave 5) and HRS 2002 (Wave 6) and later

In Wave 5, respondents who are 65 years of age or older, and report not working for pay in the last calendar year, skip the questions about income from unemployment and worker's compensation. Respondents who are younger than 65 years of age, on the other hand, are asked both sets of questions, even if they are not working. From Wave 6 onward, the same pattern is true for respondents who are 65 years of age or older. However, those who are younger than 65 years of age, and report not working for pay in the last calendar year, are asked the questions about income from unemployment, but skip those related to worker's compensation.

Notable Differences between HRS 2006 (Wave 8) and HRS 2008 (Wave 9) and later

Business assets reported earlier in the interview: Beginning in Wave 6, after the value of business or farm assets is collected, a question (Q492) asks whether these assets were reported previously in the interview. Beginning in Wave 9, a follow-up question was added which asks respondents to indicate what percentage was previously reported (Q523). For example, in Wave 9, about 30% of business owners indicate that they had reported their business wealth as either primary residence (which could be a farm or ranch), secondary residence, or other real estate earlier in the interview (Q492 = yes), and of these, most say that all of the business asset was previously reported (Q523 = 100%). These two variables are available in the RAND Income and Wealth Imputation File for the convenience of the analyst, and can be used for adjusting total wealth to reflect the amount of wealth that is twice reported. We do not use these two variables in the imputations or derivations of any of the variations on total wealth.

Housing loan question wording: In the housing section, there is a series of questions about whether the respondent has "…a mortgage, land contract, second mortgage, or any other loan that uses the property as collateral." In Wave 9, the qualifier "Do not include reverse mortgages" was added.

Notable Differences between HRS 2008 (Wave 9) and HRS 2010 (Wave 10) and later

Food stamps: The questions asking about number of months receiving food stamps changed in Wave 10. Specifically, questions are only asked for up to four "eligible years". For example, a person interviewed in 2011 and 2008 is asked questions about food stamp receipt in 2008, 2009, 2010, and 2011. Prior to Wave 10, respondents can report on food stamp receipt for several years prior to the current interview year.

Income from annuities: Beginning in Wave 10, respondents who report receiving income from annuities, but a value of "0" is indicated for the amount received last month (which indicates that payments have not started yet, or the respondent did not receive a payment last month) are asked a new sequence of questions about any income they may have received from those annuities last year. Similar questions are asked of spouses/partners. No unfolding bracket questions are asked.