What's New in Version P of the RAND HRS?

Version P incorporates the Early Release for 2014, which includes the Mid Baby Boomer cohort and the most recent versions of the cross-wave Tracker and Region and Mobility files. It contains 37,495 observations or rows. It is a Respondent-level file so each row represents a unique Respondent. It also adds new variables and makes adjustments and corrections. The current versions of the core and cross wave data used in Version P are:

- 1992 Final V1.01
- 1993 Final V2.1
- 1994 Final V1.0
- 1995 Final V2.0
- 1996 Final V4.0
- 1998 Final V2.3
- 2000 Final V1.0
- 2002 Final V2.0
- 2004 Final V1.0 (October 2006)
- 2006 Final Release V2.0 (September 2010)
- 2008 Final Release V2.0 (October 2012)
- 2010 Final Release V4.0 (March 2014)
- 2012 Final Release V1.0 (December 2014)
- 2014 Early Release V1.0 (January 2016)
- Tracker 2012 Final V1.0 (March 2015)
- Cross-Wave Region and Mobility File V4.0 (May 2013)
- Master ID File V5 (December 2009)
- Cross-Wave Imputation of Cognitive Functioning Measures 1992-2014

We have added the following to the file:

– **For questions relating to Alzheimer's and Dementia, beginning in Wave 12, we have added the following variables: RwALZHEQ, RwALZHEE, RwALZHEF, and RwALZHES and RwDEMENQ, RwDEMENE, RwDEMENF, and RwDEMENS.** RwALZHEQ and RwDEMENQ indicate the question wording used, based on preload variables, and whether the preloaded information is correct. RwALZHEE and RwDEMENE indicate whether or not a doctor has ever told the Respondent he/she had this condition in this or any previous wave. RwALZHEF and RwDEMENF indicate whether the current wave conflicts with information given at a prior wave. RwALZHES and RwDEMENS are documented in the "Change in Health" section and indicate whether Alzheimer's or dementia is new in a particular wave, i.e, reported as new since the last interview.

– **We have added a second set of variables to estimate out-of-pocket medical expenditures, which are available beginning in Wave 10.** RwOOPMDO contains the sum of estimates of out-of-pocket spending for nine categories of medical services: (1) hospital; (2) nursing home; (3) doctor visits; (4) dental visits; (5) outpatient surgery; (6) prescription drugs; (7) home health care; (8) special facilities; and (9) any other services not covered in categories 1-8. RwOOPMDOF indicates whether any components of RwOOPMDO have been imputed.

– **Interview wave in which retirement date was reported (RwRETDTWV)**: We have added a new variable to capture when the Respondent reported retirement month and year. When the Respondent reports that they consider themselves retired, an HRS branchpoint variable indicates where the retirement date may be found. Data is often available from a previous interview or it may be in the current wave employment section or simply following the self-report of retirement status. This variable is available for all 12 waves of the RANDHRS.

– **Probability Respondent will bequest $500,000 or more (RwBEQ500)**: We have added this larger bequest variable which is available beginning in Wave 6.

– **We have added an additional weight variable (RwWTCRNH)**: It is the person-level weight and the nursing home

weight combined. For those not living in a nursing home, RwWTCRNH equals RwWTRESP and for those living in a nursing home, RwWTCRNH equals RwWTR_NH.

We have made the following adjustments, improvements, and corrections to the data and documentation:

Retirement

- Retirement Month and Year (RwRETMON RwRETYR): We filled in some previously missing retirement dates by using data reported in interviews earlier than the one immediately preceding. We also ensured that retirement variable values missing due to proxy interview were consistently coded as .S.
- For indicator variables RwLIV75F, RwLIV85F, RwLIV8XF, we converted missing values from 9 to .M.
- Data on the Respondent's financial planning horizon (RwFINPLN) was not collected in Waves 9 and 10, but the question is asked again in Waves 11 and 12. R11FINPLN data was inadvertently excluded from the previous version of RANDHRS.

Pension

- To improve the organization of derived variables, we have moved the documentation of two variables, Respondent belongs to a union (RwUNION) and Firm Size (RwFSIZE), from the Pension section to the Employment section.
- In Waves 11 and 12, we integrated Section J2 information into variables where the Respondents report receiving income from pensions (RwPENINC, RwPENI_N, RwPENIC1, RwPENIC2).

Employment

- We fixed a programming error that improves the accuracy of the Total years worked from self report (RwJYEARS). The following lists the number of Respondents affected in each Wave: Wave 1=6, Wave 2=7, Wave 3=8, Wave 4=12, Wave 5=15, Wave 6=19, Wave 7=29, Wave 8=29, Wave 9=42, Wave 10=134 and Wave 11=225 .

Income and Wealth Imputations

- For Version P, we ran the cross-sectional income and wealth imputations for Waves 11 and 12, and the cross-wave wealth imputations for Wave 10 forward (to pick up changes made to the Wave 11 cross-sectional imputations). During this process, we made some minor corrections to annuities (Wave 11), improved the method for imputing ownership of primary residences, including any associated mortgages or other home loans (Waves 11 and 12), updated our assumptions about the number of months food stamps were received in the last calendar year when limited information is available (Waves 11 and 12), adjusted the criteria in the cross-wave programs for selecting cases that should not undergo bracket imputation (Waves 10 and 11), and incorporated updates of the demographic and occupational data released by HRS.
- For Wave 10, re-running the cross-wave wealth imputations had a small impact overall. Most notably, compared to Version O, the average value of primary residences was less than 0.4% lower, and the average value of mortgages on primary residences was less than 1.1% lower. For Wave 11, the changes in imputed values were much larger, which is likely associated with some corrections we made to pension income. The corrections are described further below. Overall, average wealth changed by less than 1.0%, with the exception of the (cross-sectional) net value of non-housing financial wealth (H11WTOTF), which was roughly 2.6% lower. Average income, on the other hand, was roughly 6.0% higher. Please note that imputed values for individual households or Respondents can be different from the Version O imputations.
- An examination of the pension income (RwIPEN) variables in Version O showed markedly lower average pension income in Wave 11 compared to Wave 10. After a thorough analysis, we have concluded that this is largely due to the introduction of a new Pension Section (J2) in the 2012 questionnaire, which precedes and therefore affects some answers in the Assets and Income Section (Q). Because the question about pension income in the Assets and Income Section uses the wording "Not including Social Security or other retirement income you've already told me

about," some Financial Respondents did not report pension income they had reported earlier in the Pension Section. This interaction between different sections of the surveys was not intended by the survey designers. To remedy this problem in the data, beginning in Version P, we incorporate logic in Waves 11 and 12 to pull information about pension income of Financial Respondents from both Section Q and Section J2. We did not change the derivation of pension income for earlier waves, because those did not yet have the Pension Section (J2) that caused the interaction with the pension income questions in Section Q. Pension income for earlier waves continues to be calculated using only information from Section Q. See the "RAND Income and Wealth Imputation File Codebook" available on the HRS website http://hrsonline.isr.umich.edu, specifically Appendix C: Combining Pension Income from Section Q (Assets & Income) and Section J2 (Pension), for a detailed description of the methodology that we developed for combining information from the Pension Section (J2) and the Income Section (Q). After implementing this change for Waves 11 and 12, Financial Respondents' average pension income is roughly $3,200 higher in Version P compared to Version O, and the fraction reporting any pension income receipt is approximately 15.2% higher.

– Beginning in Wave 10, Respondents who report receiving income from an annuity, but say they received zero dollars last month, are asked whether they received any income last year. Those who say "yes" are asked to provide the total amount received. No unfolding bracket questions are asked. For purposes of the imputation, we divide this amount by the number of months they report receiving it, and treat the resulting amount as the "amount received last month." In previous versions, we have left this amount in the "monthly" annuity amount variables we provide in the Income and Wealth Imputation File (i.e., RwMPEN1, RwMPEN2, RwMPEN3 and SwMPEN1, SwMPEN2, SwMPEN3). We now set these amounts back to "0" to reflect the original response to the "amount received last month" question.

Poverty Threshold Measures

– We updated the poverty threshold variables to accommodate improvements made to the income and wealth imputations for Wave 11, including the cross-wave imputations for Wave 10. These improvements, in particular, affect total household income (HwITOT) and total household assets (HwATOTA), which are used as covariates in the imputation of household-member income (See the section entitled "Poverty Threshold Definitions and HRS Measures" for more information regarding how household-member income is used).

– The cumulative effect of all these changes was very small in Wave 10, where average household income (HwPOVHHI, HwPVHHIA) and the average percentage of cases classified as above or below the poverty threshold (HwINPOV, HwINPOVA) changed by less than 0.1% between Versions O and P. In Wave 11, however, the effects were much more pronounced. Specifically, average household income was nearly 6% higher in Version P, and the average percentage of cases classified as above or below the poverty threshold was nearly 8% lower. This is likely due to some corrections we made to pension income in Wave 11, which resulted in average pension income that was roughly $3,200 higher, and a rate of pension income receipt that was 15.2% percentage points higher in Version P.

Out-of-Pocket Medical Expenditures

– We re-ran the out-of-pocket medical expenditures (RwOOPMD) imputations for Waves 10 and 11. During this process, we corrected the calculation of the standard deviation for the "tobit" imputations (see the "Warning to Users" section above for details), updated the bracket imputation procedure for incomplete brackets when no donors are available, modified the hotdeck imputation to more fully randomize the order of value donors, corrected the race (i.e., "non-white") covariate used in the imputation models, and incorporated updates of the demographic and occupational data released by HRS. These improvements had a small impact overall. Specifically, average total out-of-pocket medical expenditures was essentially the same for Wave 10, and was roughly 0.5% higher in Version P compared to Version O for Wave 11. Please note that imputed values for individual Respondents can be different from the Version O imputations.

The SAS Format Library

Many of the derived variables on this file have been assigned SAS formats, or value labels. There are also SAS formats that are used to look up values using the SAS PUT function in data steps. For instance, the format library includes

yearly CPI-U values in this form.

We are no longer distributing SAS formats as a format catalog. Instead we are providing SAS code to allow the researcher to create their own formats catalog on any computing platform.

To create a formats catalogue, assuming sasfmts.sas7bdat is in `C:\randhrs\sasdata`, simly run the following code:

```
libname library "C:\randhrs\sasdata";
proc format library=library cntlin=sasfmts;
run;
```

This SAS code will create a file called `C:\randhrs\sasdata\formats.sas7bcat`

The formats can be found in text format in one of the `Fmt\*`.fmt files and are all included in a SAS dataset (`sasfmts.sas7bdat`).

The RAND/HRS data project is committed to producing high quality data for analysis. To this end, we have employed many innovative programming and quality assurance techniques including paired peer programming, standardized macros, and independent review. If you do, however, notice any undocumented discrepancies or apparent problems with the data, please let us know by e-mailing us (randhrshelp@rand.org).

Though we have attempted to derive measures that are consistent across waves, the underlying HRS data do not always allow this. Some of the native inconsistencies are present in our derived measures, but should be documented in detail in this codebook. Before using any measure comparatively across interview years, please be sure to read the variable description in this codebook carefully, particularly the sections on "How Constructed" and "Cross-Wave Differences in the Original HRS Data" that are included for each variable. If there are cross-wave differences that we have not documented, please let us know (randhrshelp@rand.org).