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

Sampling Weights: Revised for Tracker 2.0 & Beyond

Documentation Report

HRS Staff

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

2019

Funding

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

Suggested Citation

HRS Staff. (2019). Sampling Weights: Revised for Tracker 2.0 & Beyond. University of Michigan. https://hrs.isr.umich.edu/publications/biblio/10018

Sampling Weights
Revised for Tracker 2.0 and Beyond

With the introduction of the new cohorts in HRS1998 the nature of the HRS sample changed dramatically and it became necessary to develop a revised set of sampling weights. Specifically the population of inference changed from American adults in the 1931-1941 or 1892 – 1923 birth cohorts to adults born prior to 1948. Furthermore we decided to revise weights for all past panel waves so as to use a consistent methodology. The new methodology involves post-stratifying each wave's weights to the March CPS on the basis of the birth cohorts of the respondent and spouse and on the basis of respondent gender and race/ethnicity. These factors capture all the major characteristics of the sample design and provide consistent adjustments for sample attrition and mortality. We will discuss the weight construction process in detail below. First, however, some notes on the appropriate use of weights in cross-sectional and longitudinal analysis are in order.

1. Appropriate Use of Weights

1.1 Household and Respondent Level Weights

There are two weights provided in the Tracker 2.0 file for each wave of the study. The first is the household weight xWGTHH where x ranges from a to f and corresponds to the wave in question (x=a HRS92, x=b AHD93,...,x=f HRS98). The household weight should be used in analyses of measures collected at the household level such as income, assets, debts and housing. The household weight in each year is designed to allow inferences to the U.S. population of households containing a not-initially-institutionalized adult born in the relevant birth cohort for that wave. The household weight is scaled so as to have the sum of the weights equal the number of households in the population as measured by the March CPS (with certain minor adjustments to be discussed in Section 2.2 below).

The second weight is the respondent weight xWGTR that should be used in analyses of measures collected at the respondent level such as health, labor supply and health care utilization. The respondent level weight is non-zero for living non-institutionalized respondents born in the appropriate years. It is zero for nonrespondents, deceased respondents and respondents residing in nursing homes. It is scaled so as to yield weight sums which correspond to the number of individuals in the U.S. population as measured by the March CPS for the year of data collection.

1.2 Weighting in Cross-sectional and Longitudinal Analyses

In cross-sectional analysis one should use either the respondent (xWGTR) or household (xWGTHH) weight for the year of data collection of interest depending on the level of measurement of the measure of interest. While all agree that weighting is necessary in univariate descriptive analyses, weighting in structural analyses is more

controversial and the user should probably conform to "best practice" for his or her discipline.

In longitudinal analyses the user needs to make choices about whether to weight and if so whether to use the base-year weight or the terminal-year weight. The terminal-year weight is appropriate for retrospective analysis where the goal is to model or describe the histories of individuals or households who were in the population as of the terminal year. An analysis of saving using the first difference in net worth as its measure is an example of such a retrospective analysis. When it is possible for new units to enter the sample at different waves, the user needs consider the implications of differential left-censoring. Terminal-year weighting does not correct for differential left censoring.

The base-year weight is more appropriate for prospective analyses where the goal is to model or describe the future experiences of the base-year population. Most event-history models are of this nature. In these models we take a sample of individuals or households "at risk" and follow them through time until either the event of interest occurs or observation ceases. We should note that in these analyses the impact of sample attrition needs to be considered. Base-year weighting does not correct for attrition and this attrition may result in serious biases if the attrition propensity is correlated with the propensity to experience the event of interest (see, e.g., Hill, 1997, or the special edition of the Journal of Human Resources on attrition in panel studies for methods of correcting for attrition bias).

1.3 Weighting the Dead and the Institutionalized

Because we are post-stratifying to the CPS and because it contains only living non-institutionalized individuals, a zero respondent-level weight is given for the dead and the institutionalized. This means that users wishing to analyze the dead or institutionalized will have to assign a positive weight to them. If death or institutionalization are among the outcomes of an event-history model then the base-year weighting option noted above should be used. If, on the other hand, one is using the dead and institutionalized in cross sectional analysis the respondent level weight for the wave prior to death or institutionalization can be used.

2. Weight Construction and Processing

Weight construction is a two-step process. The first step is to create a set of post-stratified household weights using the initial sampling probabilities and the birth years and race/ethnicity of the male and female household members. The second step is to use the newly created household weights to construct post-stratified respondent-level weights.

At both stages post-stratification is accomplished by constructing tables of the form given below for both the HRS and the CPS samples. Care must be taken that the sample definitions and treatment of cases is identical (or as close to identical as possible) in the two samples. Once the household grids are created the household weight is calculated according to:

$$W_{j\in k} = W^{1}_{h(j\in k)} \frac{\sum_{n\in k} W^{CPS}_{n}}{\sum_{j\in k} W^{1}_{h(j)}}$$

where $W^{1}_{h(I)}$ is the initial-wave post-stratified household weight of the $h(j)^{th}$ household and W^{CPS}_{n} is the household weight of CPS case n in cell k. For the War Babies cohort we used the average HRS household weight for the sample segment in which the War Baby resided. When this was missing (i.e. when there were no HRS1 households in the segment or when the case was in the CODA enrollment cohort) we used the overall mean HRS1 weight adjusted for the appropriate subsampling rate of the enrollment cohort.

Household Stratification Grid
(One for each race/ethnicity group)

	Male Cohort					
Female Cohort	No Male	1890 - 1923	1924 – 1930	1931 – 1941	1942 – 1947	1948 +
No Female	0	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$
1890 - 1923	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$
1924 – 1930	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$
1931 – 1941	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$
1942 – 1947	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$
1948 +	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	$\sum_{i\in c} w_i^{hh}$	0

2.2 Modifications of CPS Households

The basic sampling unit in the HRS sample is the age eligible economic unit that consists of at least one age-eligible individual and his/her spouse. When more than one economic unit was found in a household we randomly selected one and ignored the others. The basic sampling unit in the CPS is the household that can contain primary family members, subfamilies, primary individuals (unrelated persons) and secondary individuals. Any of these sub-household types could correspond to an HRS economic unit.

The first step in modifying the CPS households to correspond more closely to HRS economic units is to break the CPS households up into distinct sub-households. Primary families were separated from subfamilies and primary and secondary individuals were consolidated into new subhousehold units.

A second major difference in HRS and CPS is in the treatment of unmarried opposite gender couples. CPS imposes stricter rules in defining partners than does HRS.

In the CPS it is quite common for opposite gender individuals living in a partnership to be counted as a primary individual and an unrelated secondary individual. Before we estimated the weight-sums for the CPS we coupled opposite sex unrelated individuals whose ages were within 20 years of each other's. This rule resulted in a very close correspondence between HRS and CPS in the counts of unmarried opposite sex couples. In the HRS 5.16% of unmarried individuals are in (age-eligible) couples compared to 5.19% in the CPS.

2.3 Respondent Level Table Definitions

The respondent-level tables are similar to the household-level tables and are defined on the basis of respondent cohort, whether married, spouse cohort, gender and race. For each of the six gender-race groups we begin by defining a table of the form:

Unweighted HRS98 Case Counts
White Males (Cases as of Jan 14, 1999)

	Marital Status / Spouse Cohort					
Respondent Cohort	No Spouse	1890 – 1923	1924 - 1930	1931 - 1941	1942 – 1947	1948 - 1980
1890 - 1923	458	744	356	96	7	6
1924 - 1930	204	60	592	529	13	9
1931 - 1941	424	12	71	1482	677	224
1942 - 1947	136	2	1	100	348	315

One point that may not be completely obvious is that in the absence of inter-racial partnerships (and ignoring the unmarried column) the male and female tables for each race/ethnic group would be symmetrical around the main diagonal. There are, however, enough mixed race partners to destroy this symmetry. The practical implication of this is that the weighting scheme is sensitive not just to the respondent's own cohort, age, marital status and race/ethnicity but also to the age and race/ethnicity of the 'spouse'.

A final point that needs to be mentioned is that age-ineligible respondents do not fit in any of the six gender x race/ethnicity tables of the above form and are assigned an individual weight of zero.

2.3 Definition of Birth Cohort, Race and Ethnicity

For most individuals birth year is reported quite accurately. There is enough measurement error in birth year, however, that reports gathered in different waves vary. To prevent individuals from entering and leaving the age-eligible sample because of

measurement error in birth year reports, we decided to define the birth year for weighting purposes (WGTBYR) to be the first report of age given.

Because the purpose of breaking out the individual races and ethnicity is to account for the original oversample of Blacks and Hispanics we use a priority coding in which Hispanic ethnicity is given precedence over race. Thus, Hispanics of all races are grouped together and only non-Hispanic Blacks are considered Black. The White category also contains all non-Black, non-Hispanic individuals (e.g. Asian Americans).

Whenever we have missing data on the key variables we employ the assumption of heterosexual assortative mating—that is we assume individuals marry or partner with opposite sex individuals like themselves in terms of age, race and ethnicity. Thus, if age is missing for one partner we use the other partner's age.

2.4 Minimum Cell Size and Combining Cells

Since there are a total of 144 cells in the six gender/race tables of this form, it is not surprising that some cells will be empty or will contain only a few cases. The existence of these sparse cells can, because of small numbers, result in extreme weights. For this reason we decided to impose a minimum cell count of 20 cases. This required us to combine some cells with neighboring cells and since the respondent=s own cohort membership is more important in determining the effective selection probability than is his spouse=s we decided to collapse cells horizontally as depicted in the table. Thus for the 13 AHEAD age respondents married to War Baby or younger spouses we combined them with 96 HRS spouse cases. The corresponding combinations for the other respondent's cohorts are indicated by the shading in the table. It is important to note that the same pattern of cell combinations used in the HRS must be used in the CPS tables to avoid downward bias in the weights.

2.5 Refined Cohort Definition

Although the cohort definitions used in Table 1 do a good job of capturing differences in the initial selection probabilities, they leave something to be desired when it comes to controlling for differential mortality. The reason is that the age range of the original cohorts is quite broad, especially for the AHEAD subsample, and this means we are treating individuals with quite different mortality rates equivalently. For this reason, when sample size permitted, we refined the cohort definitions by breaking out AHEAD respondents born prior to 1914 from the younger ones and HRS respondents born prior to 1937 from those born during or after 1937. After careful examination we decided that cell counts were sufficiently large for Whites and for all unmarried respondents to support this refinement. The fact that the refined cohort definition could be used for all unmarried respondents is quite fortuitous since we are far less likely to learn of the death of these respondents than we are of those with a surviving spouse.

2.6 Initial Selection Probabilities and Weights

The initial selection probabilities and therefore the initial weight for the HRS tables are determined by the household=s situation at the time they were first interviewed. For individuals entering in 1992 (or via a 1992 respondent in 1994 or 1996) the appropriate weight is their household weight in 1992 (i.e. HRS Wave 1 Household Weight). Similarly for those entering in 1993 the AHEAD Wave 1 Household Weight is appropriate. For respondents entering in 1998, the inverse of the selection probability of the household should be used in constructing the HRS tables.