Longitudinal and Life Course Studies 2011 Volume 2 Issue 2 pp 145 - 169 SPECIAL SECTION ISSN 1757-9597

Temporary and permanent unit non-response in follow-up interviews of the Health and Retirement Study

Pierre-Carl Michaud
Université du Québec à Montréal and RAND
Arie Kapteyn, James P Smith
RAND
kapteyn@rand.org
Arthur van Soest
Netspar, Tilburg University

(Received September 2010 Revised March 2011)

Abstract

We study the effects of attrition and other unit non-response in the HRS on inferences about the distribution of socio-economic variables. A feature of the HRS is that efforts are made to bring non-respondents in a particular wave back in the next wave. We find that bringing back these temporary non-respondents substantially reduces the selection effects due to unit non-response. This applies to cross-section analyses but the same conclusion is obtained from our analysis of examples of panel data models, explaining changes in wealth, health, or labour force participation. This conclusion has important implications for users and designers of the HRS and other longitudinal socio-economic surveys with a similar design.

JEL codes: C33, C81, C44

Keywords: Selection bias, attrition, panel data, propensity scores

1. Introduction

Longitudinal surveys such as the Health and Retirement Study (HRS) provide a rich source of information to study the evolution of many socio-economic and health outcomes of a population of interest. The HRS, designed to be representative for the non-institutionalized U.S. population of ages 50 and over and their spouses, has become the most commonly used survey by economists for a variety of issues concerning the pre- and post-retirement years, with over 1100 published papers using the data, according to the HRS website.[1] European surveys like the English Longitudinal Study of Ageing (ELSA) and the Survey of Health, Ageing and Retirement in Europe (SHARE) with similar target populations have been modelled after the HRS and use similar longitudinal sample designs.

As in any socio-economic panel survey of individuals or households, an important potential weakness is that some respondents drop out over time, and when their characteristics are different from those in the retention sample, the sample may become less representative of the population of interest with every new wave. This may invalidate any inference drawn for the population of interest. Attention for this potential problem has been increasing over the past decade. See, for example, the special issues of *Journal of Human Resources* (Manski and Altonji 1998) and *Journal of the Royal Statistical Society* (Lynn 2006). Several studies analyze the nature of attrition in longitudinal studies targeted at the complete adult population in a given country, such as the Panel Study of Income Dynamics (PSID) in the US (Fitzgerald et al 1998) or the European Community Household Panel (ECHP; see Nicoletti and Peracchi 2005). To our knowledge, no such studies exist for a socio-economic survey targeted at the older part of the population, where non-response and attrition may play a specific role, due to health and cognition problems that increase with age, and due to mortality.

Existing studies tend to find that attrition, although often significantly correlated with socio-economic variables, often induces only a minor bias in the parameter estimates of econometric models of interest. See, for example, Fitzgerald et al (1998), and Lillard and Panis (1998), who consider earnings regressions, welfare participation, income dynamics, marriage formation and dissolution, and mortality risk in the PSID, Falaris (2003), who looks at equations explaining schooling attainment, labour force participation, self-employment, wages and fertility in several developing countries, Jones et al (2006), who consider dynamic models explaining self-assessed health in the BHPS (British Household Panel Survey) and the ECHP, or Behr (2006), and Behr et al (2005), who find that attrition in the ECHP does not bias estimates of earnings or income models. Whether this finding remains valid in different contexts and in the current era of reduced survey response rates is an open issue (Lynn 2006).

The original cohort entering the HRS in 1992 was composed of individuals born between 1931 and 1941 and their spouses (irrespective of their age). The sample drawn from this cohort was interviewed every two years. Other cohorts were added later (starting in 1993 with the study of Assets and Health Dynamics among the Oldest Old (AHEAD) cohort, born before 1923). In this study we will focus on the original HRS cohort, which was interviewed most often. The data we use cover the seven waves from 1992 until 2004. Every new wave has a substantial number of non-respondents, who may or may not come back in later waves. For analysis based upon this panel survey, it is important to know whether such unit non-response is selective and how potential selection effects can be tackled in order to draw unbiased inference for the US 50+ population of interest (US couples in 1992 with at least one partner born between 1931 and 1941, corresponding to how the original sample was drawn).

A specific feature of the HRS is that respondents who do not participate in a given wave, but do not explicitly state they refuse to participate in any future survey, are approached for an interview again for the next wave two years later (and again for later interviews, even if they miss several consecutive waves). This creates a distinction between attrition and temporary non-response. In order to investigate whether the effort to get people back into the survey is worthwhile, we will distinguish between these two groups. We will also distinguish attrition due to death from other attrition.

Other major American panels have also attempted to bring back non-respondents. For example, starting in 1992, the PSID has contacted all persons who dropped out in the prior wave and was successful in getting back 50% of them. The American NLSY (National Longitudinal Survey of Youth) rule is to try and interview essentially everyone from the original sample, regardless of how many times they were previously not interviewed. With the 1979 wave of the NLSY, for example, this policy resulted in a recapture of 46% of those who had ever dropped out by 2004.

Analyzing the value of bringing respondents back in is the main focus of this paper. Returning respondents have rarely been considered as a separate group. There are two exceptions. Olsen (2005) emphasizes the large number of returning respondents in the NLSY, stating that about half of respondents who missed one round will grant an interview for the next round. Hawkes and Plewis (2006) show that a substantial number of respondents in the NCDS (National Child Development Study, a UK cohort study following individuals from their birth in 1958) miss one wave but return in a later wave, and find that the characteristics of wave non-respondents differ from those of respondents who permanently leave the sample.

Reducing panel attrition is particularly desirable if the remaining respondents are a non-representative sample of the population. We are not aware of studies that have looked closely at how problems related to attrition affect the representativity of the HRS or other longitudinal surveys targeted at older population groups. Hill and Willis (2001) have considered the general problem of finding ways to increase response rates in the HRS but do not address the issue of whether a lower response rate leads to more selection bias. Hence, our contribution is twofold. First, we analyze how attrition affects the representativity of the HRS. Second, we aim at investigating whether re-contact efforts help to restore the representativity of the sample in following waves.

Obviously, unit non-response and attrition may have effects on some types of analyses and not on others. This will depend on the variables of interest and the type of analysis, for example: a cross-section analysis in a given year, a longitudinal analysis following respondents over time, the parameters of interest, and the specific model (such as, in particular, which conditioning variables are used). We consider some common examples - cross-section and panel data inference concerning wealth, home ownership and employment status.

The remainder of the paper is structured as follows. Section 2 presents data on interview participation and types of unit non-response for each wave. Section 3 analyzes the determinants of various types of unit non-response: attrition through death, other (permanent) attrition, and temporary unit non-response. Section 4 studies how these sources of unit non-response affect inference about the 2004 cross-sectional distribution of variables of interest, like wealth, health, or income. In section 5 we investigate the consequences of selective unit non-response for estimates of several examples of panel data models, considering wealth, home ownership and employment patterns. Section 6 concludes.

2. The HRS cohort born 1931–1941

The target population of the original HRS cohort consists of non-institutionalized households where at least one member was born between 1931 and 1941. The sample is drawn using a multi-stage area probability sample of households, and an interview is attempted with all age-eligible respondents and their spouses. Only non-institutionalized individuals are considered at baseline, but respondents entering nursing homes after the baseline interview are followed in later waves. The Institute for Social Research (ISR) in Michigan conducts the survey. For more technical details on the survey design, see Heeringa and Connor (1995).

The HRS over-samples respondents from three groups – African Americans, Hispanics, and residents of Florida. Of the 15,497 interviews attempted in 1992, 12,654 were realized, giving an individual unit response rate of 81.6% at baseline. The response rate is very similar for the African American (81.1%) and Floridian (82.2%) samples, but lower for the Hispanic supplement (77%).

We focus on the birth cohort 1931-1941 and drop spouses who are not in this cohort. This is because for a meaningful analysis at the individual level, the group of spouses not born in 1931-1941 is too small and specific. This leads to a sample of 10,089 respondents in 1992, aged 51 to 61 in 1992, and aged 63 to 73 in 2004. The population of interest, for our analysis of the data of a given wave, therefore consists of non-institutionalized individuals in the US born between 1931 and 1941 and alive in that wave. When using more waves, depending on the nature of the longitudinal analysis, it either consists of all individuals in this cohort alive in the first wave, or of all those still alive in the last wave used for the analysis.

We do not analyze unit non-response at baseline (which is inherently more difficult than follow-up non-response, since hardly any information is available for initial non-respondents). HRS provides sample weights based upon basic demographics, derived from a comparison with the much larger Current Population Survey (CPS); see Heeringa and Connor (1995, Section 5). We will maintain the assumption that these weights are sufficient to correct for non-response at baseline as well as for the over-sampling discussed above. [2]

Because the HRS is a study of an older population, it emphasizes tracking the vital status of respondents over waves. Deaths are reported by relatives contacted by an interviewer, or by a match with the National Death Index. Table 1 shows that the mortality rate grows from 1.7% between the first and second wave to 2.9% in 2004 as the cohort ages. The unweighted cumulative mortality rate over all waves is 14.4%. Weighting to correct for the over-sampling of African Americans, Hispanics and Floridians gives a cumulative mortality rate of 13.2%, which is close to what would be predicted from standard life-tables. If the respondent died, ISR attempted a so-called exit interview with a proxy respondent, usually the widow or widower, or a close relative of the deceased respondent – a short interview on the last period of the deceased respondent's life, cause of death, bequests, etc.

Table 1. Vital status in waves 1992-2004

Vital status	1992	1994	1996	1998	2000	2002	2004
alive	10,089	9,852	9,543	9,112	8,685	8,241	7,533
presumed alive	0	16	55	63	76	129	170
death reported in wave	0	167	211	213	272	343	246
mortality rate		1.7%	2.1%	2.2%	3.0%	3.9%	2.9%
vital status unknown	0	54	113	323	465	513	934

Notes. *A respondent is presumed alive if the interviewer cannot reach a respondent but has access to some information that the respondent might be alive. If no such information can be obtained, the respondent's vital status is classified as unknown.*

Table 2 presents interview status of all respondents who participated at least once. In 1992, 152 core interviews are missing – these are absent age-eligible spouses. Moreover, 187 respondents are not in the sample – these are future spouses of age-eligible HRS respondents. In later waves, numbers of missing interviews increase due to non-response. The response rate to core interviews (conditional upon participation in the first wave) is slightly falling over time (90.5% in 1994 versus 87.1% in 2004). The response rate to exit interviews is lower than to core interviews. Once an exit interview is completed, a respondent is classified as out-of sample. Respondents are also excluded from the sample if they explicitly request to be removed from the study. By 2004, 16.3% of the original respondents are out-of-sample.

Table 2. Interview status in waves 1992–2004

Interview status	1992	1994	1996	wave 1998	2000	2002	2004
core interview attempted							
core interview obtained	9,750	8,835	8,459	8,087	7,634	7,367	7,071
core interview missing	152	925	1,124	1,153	1,247	1,080	1,048
response rate	98.5%	90.5%	88.3%	87.5%	86.0%	87.2%	87.1%
exit interview attempted							
exit interview obtained		128	171	221	302	381	284
exit interview missing		39	41	49	76	79	45
response rate		76.6%	80.7%	81.9%	79.9%	82.8%	86.3%
other out of sample	187	162	294	579	830	1,182	1,641
% out of sample	1.9%	1.6%	2.9%	5.8%	8.2%	11.7%	16.3%
total	10,089	10,089	10,089	10,089	10,089	10,089	10,089

Notes.*: "Other out of sample" (other than respondents who are dead and for whom an exit interview was completed) includes non-eligible spouses that become eligible at a later wave and those who are permanently dropped from the sample (at their request or by HRS decision). For 1992, the response rate does not take account of the initial round of non-response as shown in Table 1.*

Interviewers re-contact every respondent who did not provide a core interview in the previous wave but is still classified as in-sample. Each participant normally gets $100 for a new interview and $60 for a panel interview on the phone.[3] As a result of re-contacts, there is a large variety of participation patterns. Figure 1 shows the various flows of entry and exit across years. For example, of the 9,750 respondents (5,156 women and 4,594 men) who provided core interviews in 1992, 167 (1.7%) were reported deceased the following wave, and 787 (8.2%) were missing because they could not be reached or refused to be interviewed. In 2004, of respondents providing a core interview in 2002, only 4.5% were missing.

Figure 1 also shows re-entry of previously interviewed respondents who skipped an interview. Starting in 1996, between 24.9% and 42.9% of respondents with missing interviews came back into the panel to provide a core interview in the next wave. This feature of the HRS helps to keep cumulative attrition down compared to a survey that does not attempt to re-contact respondents missing in a given wave. It implies that an analysis of attrition and non-response in the HRS should not consider non-response as an absorbing state.

Given that a fraction of respondents are not re-interviewed in later waves, one may ask if the remaining sample remains representative of the population of interest. If those leaving the panel have systematically different measured and unmeasured characteristics from those who stay in the panel, this will bias population inferences drawn from the HRS sample for variables of interest that are related to these measured or unmeasured characteristics.

3. Baseline determinants of non-response and attrition

In this section, we analyze how patterns of response behaviour between 1992 and 2004 are associated with respondent characteristics in 1992. We distinguish four types of participation sequences. First, 60.5% of the 1992 respondents provide core interviews in all six waves, from 1992 to 2004 (the *always in* group). Second, as seen in Figure 1, a sizeable fraction of respondents (9.4%) respond in both 1992 and 2004 but not in at least one intermediate wave. We refer to these as *temporarily out*. The last two groups are respondents who are not interviewed in 2004. These comprise 14.5% of respondents who die prior to the 2004 interview, and 15.6% of the 1992 respondents who are not interviewed in 2004 for other reasons than death. We refer to the latter as *attritors.* This term is not completely ideal here, since some of the respondents that we classify as *attritors* may come back into the survey in a later wave, after our observation window (2006 or later). Only a subset of the *attritors* has explicitly indicated to the HRS that they do not want to be contacted for future waves; these respondents definitely will not come back in after 2004. But the other respondents classified as *attritors* might still participate in waves later than 2004, outside our observation window.

Attrition due to mortality plays a special role, since in many cases the population of interest consists of survivors only. For example, if we want to analyze the wealth or income distribution at a given point in time, we will usually be interested in the distribution among survivors and not in the counterfactual distribution among survivors and deceased individuals. To be precise, for an analysis of the cross-section distribution of wealth or income in 2004, the population of interest are all non-institutionalized individuals in the US born between 1931 and 1941 and surviving until 2004.

On the other hand, particularly when looking at changes, the longitudinal analysis may be contaminated by selective mortality. See, for example, Attanasio and Hoynes (2000), who consider the age profile of wealth. Because of the well-known negative correlation between wealth and mortality, the part of an older birth cohort still alive at a given point in time is a relatively wealthy subset of the complete birth cohort. For some purposes, such as an analysis of wealth changes at the individual level, it may be desirable to correct for this. This makes it important to consider mortality as an explicit survey exit route in the analysis.

Figure 1. Exits and entry between 1992 and 2004

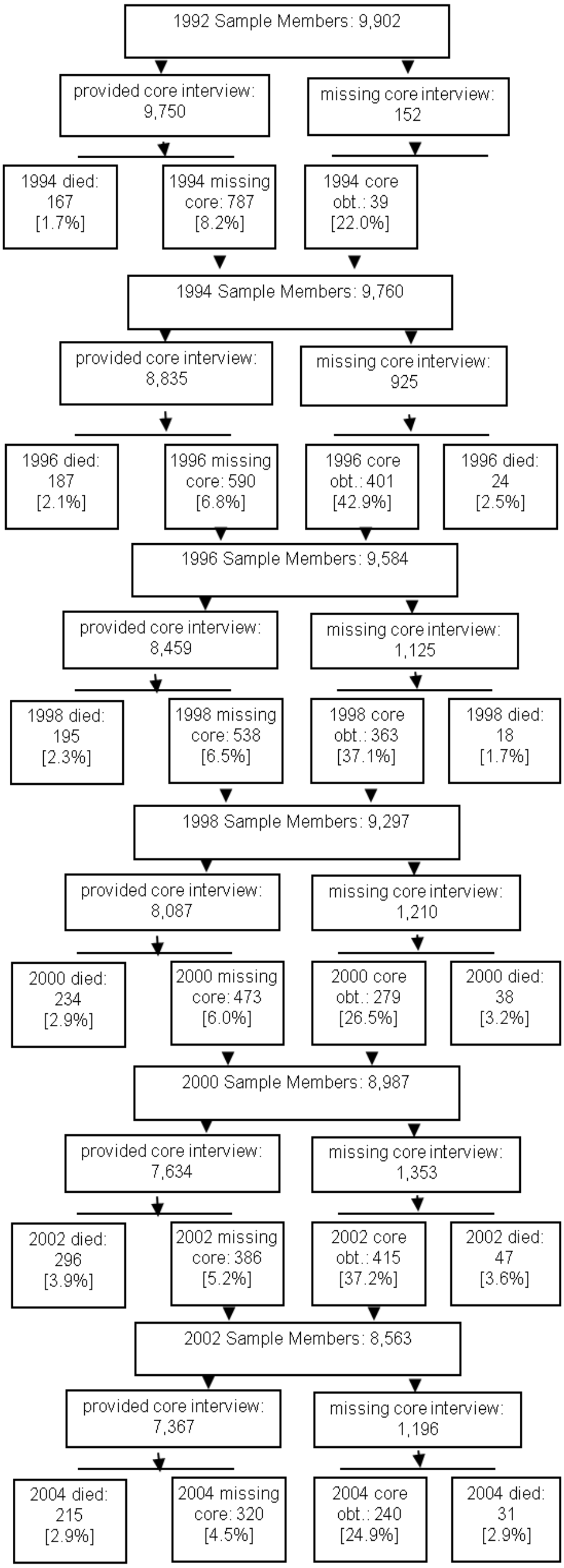

Table 3 summarizes baseline characteristics of respondents who gave an interview in 1992 by type of participation over the time period 1992–2004. Characteristics associated with mortality reflect the well-known positive association between health and socio-economic status (SES). Older individuals, African-Americans, unhealthy, and less educated respondents are more likely to die over the 12-years period. Compared to the *always in* group, the *attritors* group has an over-representation of individuals born outside the US and of Hispanics. This over-representation is even more pronounced among those *temporarily out*. In addition, African-Americans are also more likely to be in the *temporarily out* group. Several other demographics are similar for *temporarily out* and *attritors*. Both groups are more likely to have poor health and to be less educated than those *always in*. The *temporarily out* are significantly less likely to be home owners (73.4% in *temporarily out* compared to 83.9% for *always in,* and 83.5% for *attritors*), more likely to be working, and less likely to be retired (10.9% compared to 15.8% for *attritors*). A higher fraction of *temporarily out* respondents are divorced at baseline (17.4% compared to 12.1% for *always in* and 13.6% for *attritors*). Overall, the characteristics of the *temporarily out* group suggest that this group more often has an unstable life style, which makes them less likely to be reached by interviewers or to be available for an interview in a given wave.[4]

Table 3. Baseline characteristics by type of participation sequence 1992–2004 (weighted using baseline HRS weights)

	Status 2004				
Characteristics	***Always in***	***Temp.out***	**Died**	***Attritors***	**total**
Demographics in 1992					
age (yrs)	55.5	54.9	56.4	55.5	55.6
female (%)	55.2%	50.2%	40.6%	52.5%	52.3%
born outside U.S. (%)	8.8%	15.6%	7.0%	12.9%	9.8%
Black (%)	8.6%	15.0%	16.5%	9.0%	10.3%
Hispanic (%)	5.4%	14.1%	5.7%	7.2%	6.4%
married (%)	78.8%	73.2%	67.7%	78.2%	76.7%
widow(er) (%)	5.6%	5.7%	8.4%	4.7%	5.9%
divorced (%)	12.1%	17.4%	19.3%	13.6%	13.8%
ever divorced (%)	30.2%	36.3%	38.3%	30.6%	31.9%
single (%)	3.5%	3.8%	4.5%	3.5%	3.6%
household size (#)	2.62	2.76	2.51	2.56	2.61
Health Status in 1992					
health good (%)	25.7%	29.9%	26.3%	29.5%	26.7%
health fair/poor (%)	15.5%	20.8%	45.2%	16.1%	20.1%
ever had severe cond. (%)	15.9%	14.1%	41.8%	16.7%	19.4%
ever had mild cond. (%)	36.0%	40.6%	59.2%	38.0%	39.9%
at least one ADL (%)	3.4%	5.2%	12.7%	2.7%	4.7%
SES and Employment Status in 1992					
high school (%)	39.0%	36.8%	37.5%	40.0%	38.8%
some college (%)	20.4%	18.7%	17.3%	20.1%	19.8%
college and above (%)	20.7%	15.2%	12.1%	16.8%	18.5%
own house (%)	83.9%	73.4%	72.0%	83.5%	81.3%
working (%)	68.7%	69.6%	50.3%	68.8%	66.3%
retired or disabled (%)	15.2%	10.9%	28.6%	15.8%	16.7%
not in labour force (%)	13.5%	14.7%	11.7%	13.5%	13.3%
N	5,902	912	1,416	1,520	9,750
%	60.5%	9.4%	14.5%	15.6%	100.0%

Notes. *See Appendix for variable definitions. "Always in": respondents who provide core interviews in all 7 waves between 1992 and 2004. "Temp. out": respondents who provide core interviews in 1992 and 2004 but have skipped one or more interviews in intermediate waves. "Died": respondents in 1992 who died before 2004. "Attritors": respondents not in the HRS in 2004 and respondents with "vital status unknown". HRS 1992 weights used.*

Table 4 reports differences in the wealth, income and earnings distributions of the four groups.[5] Because the distribution of wealth is skewed, we do not only present mean values, but also several quantiles of the distribution. For those who died before 2004, the full extent of the socio-economic status (SES)-health gradient is revealed – they have lower wealth, household income and earnings than the other groups. Their median wealth in 1992 is about half that of those *always in* (about $82,600 versus about $150,600). Respondents *temporarily out* but present in 2004 are substantially different both from those *always in,* and also from *attritors*. For example, median wealth at baseline is about $98,500 for those *temporarily out,* compared to $150,600 for those *always in,* and $151,300 for *attritors*. Differences in wealth for *temporarily out* are partly explained by a lower home ownership rate (73.4% versus 83.9% for those *always in*). The second panel of Table 4, which presents the distribution of wealth excluding the value of the owned home, however, shows that this can only explain part of the difference: even if home ownership is ignored, most wealth quantiles of the *temporarily out* are substantially smaller than those of the *always in* group. In relative terms, differences are larger at the bottom of the distribution than at the top. Finally, differences in earnings (conditional on positive earnings) are much smaller than differences in household income or wealth.

Table 4. Baseline wealth, income and earnings distribution by type of response 1992–2004

Household wealth in 1992	**Mean**	**10th percentile**	**25th percentile**	**Median**	**75th percentile**	**90th percentile**
always in	275,997	7,062	55,959	150,558	319,768	606,227
temp. out but in for 2004	281,606	0	17,321	98,462	241,958	660,401
died prior to 2004	181,223	0	10,792	82,607	199,189	389,718
attritor	270,119	8,127	58,757	151,330	329,904	650,196
Total	262,711	2,665	45,301	134,036	299,783	591,291
Household wealth without housing (and mortgage) in 1992						
always in	176,662	0	11,991	57,292	172,542	445,011
temp. out but in for 2004	180,555	0	3,597	30,978	117,248	393,049
died prior to 2004	109,604	-799	1,332	19,919	89,269	233,831
attritor	160,403	0	10,659	50,209	177,205	442,586
Total	165,354	0	8,127	48,232	157,886	403,175
Household income in 1992						
always in	69,115	14,003	29,738	54,894	87,670	131,571
temp. out but in for 2004	68,851	10,659	23,903	47,965	79,942	117,248
died prior to 2004	48,536	7,275	15,892	33,309	62,621	97,396
attritor	68,585	13,803	30,106	53,295	80,209	128,707
Total	66,218	11,991	26,647	51,136	82,527	126,175
Earnings (conditional on positive earnings) in 1992						
always in	39,765	6,662	15,988	31,977	51,962	75,945
temp. out but in for 2004	45,950	7,994	17,321	30,644	50,497	74,613
died prior to 2004	34,561	5,329	13,324	26,647	46,633	66,618
attritor	43,072	7,994	17,321	33,043	50,630	74,613
Total	40,275	6,662	15,988	31,710	50,630	74,613

Notes. *All figures in 2004 US dollars. See the appendix for variable definitions. Weighted with HRS 1992 weights.*

To take account of the correlation among characteristics, we estimated a multinomial logit model explaining the type of response behaviour from baseline characteristics. We define indicators

$$s_{ij} \ \ (j = a, t, d, o)$$

denoting whether respondent *i* has been *always in*, *temporarily out*, has died, or was *o*ut in 2004 (*attritors*). The probabilities of the four outcomes, conditional on a vector of baseline characteristics x_{i0} are modelled as

$$P(s_{ij} = 1 | x_{i0}) = \frac{\exp(x_{i0}\beta_j)}{\sum_{j'} \exp(x_{i0}\beta_{j'})}$$

The parameter vectors to be estimated are β_j, $j = t, d, o$, with *always in* as the reference category, i.e. β_a=0. As explanatory variables, we include basic demographics, health indicators, and quintile dummies for wealth, household income, and earnings, allowing for non-linearities in the effects of these variables. In the appendix we give more details on the construction of the explanatory variables. We estimate the model for men and women separately, since pooling is strongly rejected. Since the main purpose of these estimates is the construction of weights based upon predicted probabilities, we prefer to keep a flexible model and do not aim at finding more parsimonious specifications.[6] Tables 5 and 6 present the results.

Table 5. 2004 Panel status explained from baseline characteristics – females

reference: *always in*	Parameter Estimates - Status 2004					
Covariates	***temp. out***		**died**		***attritor***	
age 50-55 spline	-0.010		0.053		-0.037	
age 56-60 spline	-0.071	**	0.081	**	-0.010	
born outside U.S.	0.281		-0.427	**	0.300	**
Black	0.234		0.071		0.065	
Hispanic	0.778	**	0.055		0.322	*
widow(er)	-0.197		0.227		-0.131	
divorced	-0.190		0.011		-0.172	
once divorced	0.125		0.082		-0.003	
single	-0.917	**	0.064		-0.251	
household size	0.018		0.000		-0.087	**
high school	-0.014		-0.179		-0.143	
some college	-0.073		-0.228		-0.370	**
college and above	-0.348		-0.204		-0.408	**
own house	-0.228		-0.179		-0.129	
retired	-0.223		0.330	**	-0.268	*
disabled	-0.172		0.484	**	-0.644	**
not in labour force	-0.074		0.003		-0.223	
1st wealth quintile	0.320		-0.175		-0.318	*
2nd wealth quintile	0.060		-0.347	**	-0.074	
4th wealth quintile	-0.032		-0.146		-0.043	
5th wealth quintile	0.136		-0.524	**	0.200	
1st earnings quintile	0.136		0.218		0.189	
2nd earnings quintile	-0.218		0.161		-0.131	
4th earnings quintile	0.339	**	0.026		0.085	
5th earnings quintile	0.546	**	0.091		0.001	
1st hld income quintile	0.257		0.113		-0.047	
2nd hld income quintile	0.158		0.162		-0.030	
4th hld income quintile	0.250		-0.316	*	-0.084	
5th hld income quintile	0.076		-0.142		-0.274	*
health good	0.338	**	0.357	**	0.139	
health fair/poor	0.351	**	0.781	**	0.059	
ever had severe cond.	-0.089		0.754	**	0.118	
ever had mild cond.	0.183	*	0.436	**	0.108	
at least one ADL	-0.274		0.322	**	-0.304	
constant	-1.669		-5.119	**	1.405	
Observations	5156		Chi-Sq. SES (df=12)			
LogLikelihood	-4988.89		Temp. Out Eq.		19.81	*
Pseudo-R2	0.071		Death Eq.		20.94	*
Chi-Sq. Equal Coeff.	80.33	**	*Attritors* Eq.		15.57	
			Chi-Sq. Region		29.30	**

Notes. *Multinomial logit point estimates. ** p-value<0.05, * p-value<0.10. The dependent variable is type of participation. Covariates refer to baseline characteristics of respondents in 1992. See Appendix for variable definitions. The reference category is always in; temp. out refers to respondents with core interviews in 1992 and 2004 but not in at least one wave between 1992 and 2002. Census division dummies are included in the estimation, but estimates are not reported. Chi-Sq. Region is a test on their joint significance. SES chi-square statistics test the null hypothesis of no SES effects (no wealth, earnings and income effects) in each equation. Chi-Sq. Equal Coeff. is a test for equal slope coefficients in the equations for attrition and temporarily out.*

Table 6. 2004 Panel status explained from baseline characteristics – males

reference: *always in*	**Parameter Estimates - Status 2004**					
covariates	***temp. out***		**died**		***attritor***	
age 50-55 spline	0.009		0.064	*	-0.014	
age 56-60 spline	-0.078	**	0.067	**	0.019	
born outside U.S.	0.213		-0.230		0.580	**
Black	0.729	**	0.338	**	0.141	
Hispanic	0.726	**	-0.080		0.041	
widow(er)	0.139		0.319		-0.601	
divorced	0.327	*	0.392	**	0.362	**
once divorced	0.226	*	0.228	**	-0.023	
single	0.262		0.224		0.257	
household size	0.043		-0.042		-0.039	
high school	0.068		0.029		-0.097	
some college	-0.038		0.092		0.107	
college and above	-0.256		-0.172		-0.239	
own house	-0.117		-0.074		-0.057	
retired	-0.438	**	0.149		0.205	
disabled	0.221		0.273		-0.135	
not in labour force	0.096		0.343	*	0.109	
1st wealth quintile	0.441	**	0.417	**	-0.219	
2nd wealth quintile	0.210		0.069		-0.019	
4th wealth quintile	-0.042		-0.018		-0.002	
5th wealth quintile	0.296		-0.097		0.084	
1st earnings quintile	-0.143		0.325	*	0.160	
2nd earnings quintile	-0.108		-0.021		-0.084	
4th earnings quintile	-0.129		-0.050		0.162	
5th earnings quintile	-0.160		-0.023		0.187	
1st hld income quintile	-0.110		-0.107		0.033	
2nd hld income quintile	-0.037		0.040		0.018	
4th hld income quintile	0.070		-0.173		0.013	
5th hld income quintile	-0.152		-0.207		-0.078	
health reported good	0.052		0.336	**	0.071	
health fair/poor	0.008		0.717	**	-0.034	
ever had severe cond.	-0.138		0.888	**	0.041	
ever had mild cond.	0.046		0.404	**	-0.002	
at least one ADL	0.396		0.430	**	-0.144	
Constant	-2.434		-5.559	**	-0.513	
Observations	4594		Chi-Sq. SES (df=12)			
LogLikelihood	-4828.69		Temp. Out Eq.		11.25	
Pseudo-R2	0.083		Death Eq.		21.15	**
Chi-Sq. Equality of Coeff.	115.79	**	*Attritors* Eq.		6.18	
			Chi-Sq. Region		11.63	

Notes. *See Table 5.*

Parameter estimates should be interpreted in comparison to *always in*, the benchmark outcome. First consider the demographic effects. Age effects are modelled as continuous piecewise linear, with a kink at 56 years (the mid-point in the age range in 1992). Older respondents are less likely to be *temporarily out* (and, as expected, more likely to die). Hispanic men and women and African American men are more likely to be *temporarily out*. Respondents not born in the United States are particularly likely to become *attritors*, possibly because of return migration. They are less likely to die while in the panel. Single women (mainly widows) are unlikely to be *temporarily out*.

Divorced men are more likely than married men to be in any of the three non-response categories. Highly educated women are less likely to become *attritors*, while no significant effect of education is found for men. Retired men seem to lead relatively stable lives and are less often *temporarily out*. Retired women and women on disability pensions are relatively likely to die.

Turning to the economic variables, male respondents in the lowest wealth quintile have a greater probability to be *temporarily out* or to die. The effect on *temporarily out* extends to the second lowest wealth quintile although it is not statistically significant.[7] For males, these wealth effects are the only significant SES link to non-response. For example, a likelihood ratio test does not reject the null hypothesis of no effect of income, wealth, and earnings on the odds of *attrition* versus *always in* at any conventional significance level, as indicated in the bottom part of the tables ("Chi-Sq. SES (df=12) *Attritors* Eq.").

Females with high earnings are more likely to drop out temporarily. Similarly, females in low wealth households (1st quintile) have a *lower* probability to be *attritors* in 2004. Joint likelihood ratio tests of no SES effects do not reject the null of no SES effects in the *attritor* equation, but do reject the null in the *temporarily out* equation.

The lack of a clear link between attrition and baseline wealth, income, and earnings is in line with results for the PSID reported in Fitzgerald et al (1998).[8] Overall, we do not find many significant effects of earnings, income or wealth on response behaviour. It seems that unconditional differences in, e.g. median wealth in Table 4, are largely due to other differences than in wealth itself, such as race and ethnicity, or, for women, education level. Some SES links are found for the *temporarily out* group but they work in opposite directions for females and males. Hence, it is unclear what effect this selection has on estimates of household wealth or income.

The link with income and wealth is much stronger for mortality, even conditional on our rich set of controls, including controls for baseline health. Joint tests looking at the null of no SES effects on mortality, reject this null hypothesis for both males and females.

4. Inference on univariate distributions in the 2004 cross-section

The common way to correct for unequal representation of population groups in the sample, when estimating the distribution of a variable of interest, is to use sample weights. Socio-economic surveys typically provide such weights with the data set, constructed on the basis of a number of key demographics like age, gender and race, and designed to make the weighted sample reproduce the population distributions of at least these key variables. In this section, we compare weighted distributions using standard weights and alternative weights that use more baseline information to analyze the consequences of attrition and temporary unit non-response, for inference on the distribution of a variable of interest *y* (such as wealth or health). A similar approach is used by Vandecasteele and Debels (2007) who analyze attrition in ECHP, but do not consider *temporarily out* respondents.

We consider two periods, the first and last available waves 1992 (*t=0*) and 2004 (*t=1*). The population of interest are all non-institutionalized individuals in the U.S. born between 1931 and 1941, surviving until time *t* (1992 or 2004). As a consequence, we do not correct for mortality – deceased persons are not in the population of interest.

The standard way to correct for over-sampling of minorities and initial unit non-response in each cross-section, is to use sample weights provided with the HRS dataset for each wave, the "HRS weights",[9] which use the ratio of the sample size in a given year for CPS (a cross-section) and HRS, in cells defined by gender, race and birth cohort of respondents and their spouses. Hence these weights are a function $w_t(q_{it}), t = 0,1,$ where q_{it} is a vector including gender, race, marital status and birth cohort for respondent *i* at time *t*.

Our maintained assumption in this section is that the weights $w_0(q_{i0})$ are sufficient to correct for stratified sampling and unit non-response in 1992. This relies on the *Missing at Random* assumption (MAR) (Little and Rubin 1987),[10] that initial non-response is independent of the variable of interest y_{i0} conditional on q_{i0}:[11]

$$MAR_0^q : y_{i0} \perp p_{i0} \mid q_{i0},$$

where p_{i0} is a dummy for participation in the interview at *t=0* and $\perp$ denotes conditional independence. MAR_0^q implies that a consistent estimator for the population mean of y_{i0} at *t=0* is given by $\sum_{i=1}^{n_0} w_0(q_{i0}) y_{i0} / \sum_{i=1}^{n_0} w_0(q_{i0})$, where n_0 is the size of the baseline sample; similarly, other statistics like quantiles can be estimated consistently using corresponding weighted sample statistics.

The HRS weights are adjusted each wave. The standard approach in applied work is to also use the HRS weights for 2004, estimating, for example, the mean y_{i1} at *t=1* with $\bar{y}_1^{w_1} = \sum_{i=1}^{n_1} w_1(q_{i1}) y_{i1} \sum_{i=1}^{n_1} w_1(q_{i1})$, where n_1 is the sample size at $t=1$. This is a consistent estimator under a similar *Missing at Random* assumption for 1992:

$$MAR_1^q : y_{i1} \perp p_{i1} \mid q_{i1},$$

Participation at *t=1* requires participation at *t=0* and retention. A sufficient condition for MAR_1^q is that both events are independent of y_{i1} given q_{i1}. We can say that MAR_1^q is stronger than MAR_0^q in the sense that MAR_1^q can be violated due to selective attrition or temporary non-response, even if initial unit non-response were completely random (so that MAR_0^q would certainly hold).

Comparing estimates of the distribution of y_{i1} using the HRS 1992 weights and the HRS 2004 weights gives insight in the role of selective follow-up non-response (attrition or temporary non-response) as far as this is related to the basic demographics q_{i0} and q_{i1}. Large differences between the two estimates may arise if, first, follow-up non-response is related to q_{i0} or q_{i1} and, second, y_{i1} is correlated with q_{i0} or q_{i1}. This comparison does not necessarily say much about the validity of MAR_1^q since if this is not satisfied, both estimates may well suffer from a bias in the same direction. For example, if, conditional on basic demographics, wealth is positively correlated with participation at *t=1*, both estimates will overestimate wealth statistics of the population at *t=1*.

To increase the likelihood that conditional independence is satisfied so that the weighted statistics indeed give consistent estimates of the population statistics, it is advisable to condition on as many variables that drive the participation probability as possible (Kalton and Brick 2000). In our case, an alternative weighting procedure can be based upon using a larger set of conditioning variables observed at baseline, stored in a vector x_{i0} (including q_{i0} but not q_{i1}). Using these weights relies on the assumption

$$MAR_1^x : y_{i1} \perp p_{i1} \mid x_{i0},$$

To construct the weights based upon MAR_1^x and the assumption that HRS 1992 weights correct for unit non-response at baseline, denote the retention probability (the probability that $p_{i1} = 1$, given participation in the baseline interview) conditional on x_{i0} by $p(x_{i0})$. This has the role of the propensity score in Little and Rubin (1987). If the baseline sample were a simple random sample and follow-up non-response were the only problem, MAR_1^x would imply that consistent estimates of means or other population statistics could be obtained using *inverse probability weights* $p(x_{i0})^{-1}$ (Horvitz and Thompson 1952; Horowitz and Manski 1998; Wooldridge 2002). Under MAR_1^x and the assumption that the 1992 HRS weights are sufficient to correct for baseline non-response, the 1992 HRS weights can be combined with $p(x_{i0})$ into new weights $\tilde{w}_1(x_{i0}) = w_0(q_{i0}) / p(x_{i0})$ that correct for stratified sampling and initial non-response, as well as for all forms of follow-up unit non-response.[12] For example, a consistent estimator of $E(y_{i1})$ is then given by the weighted sample average $\bar{y}_1^{\tilde{w}_1} = \sum_{i=1}^{n_1} \tilde{w}_1(x_{i0}) y_{i1} / \sum_{i=1}^{n_1} \tilde{w}_1(x_{i0})$

Our empirical strategy is to compare estimates of the mean and quantiles of the distribution of some variables of interest in 2004 using several sets of weights. First, we consider all participants in the 2004 survey (including those who were *temporarily out)* and compare the estimates of statistics of interest using no weights, the HRS 1992 weights, the HRS 2004 weights, and the inverse probability weights $\tilde{w}_1(x_{i0})$. For the latter, we construct retention probabilities from the estimates in Tables

5 and 6. The participation probability $p(x_i)$ for respondent *i* is the probability to be *always in* or *temporarily out*, conditional on being alive in 2004:[13]

$$p(x_{i0}) = \frac{p(s_{i,a} \mid x_{i0}) + p(s_{i,t} \mid x_{i0})}{1 - p(s_{i,d} \mid x_{i0})},$$

where *a* refers to "*always in*", *t* to "*temporarily out*" and *d* to "died" (cf. Section 3). The weights (after normalization so that their mean is 1) vary from 0.22 to 3.24 with a standard deviation of 0.418. Since there are no outliers, we did not consider stabilizing them to reduce variability.

Second, we repeat the same exercise, but now without the 2004 respondents in the *temporarily out* group who missed one or more intermediate waves (but returned in or before 2004), adjusting the inverse propensity scores and the weights $\tilde{w}_1(x_{i0})$ accordingly for the different selection process. In this case, the Inverse probability weights are the inverse of "participation probabilities"

$$p^a(x_{i0}) = \frac{p(s_{i,a} \mid x_{i0})}{1 - p(s_{i,d} \mid x_{i0})}$$

Comparing the results with the first set of estimates, including the *temporarily out* group, will show whether bringing respondents, who do not participate in one wave, back into the sample is worthwhile for reducing selection bias due to unit non-response in follow-up waves.

Results

We compared the distributions using the various weights of many variables of interest, referring to, for example, health, socio-economic status, and family composition. We often find substantial differences between weighted and unweighted statistics (mainly because of the oversampling of African Americans and Hispanics), but not between the statistics obtained using the three different weights. Details are available upon request. For most variables therefore, we do not find evidence of selective attrition, either including or not including the *temporarily out*. The exception is household wealth, which we describe in detail in Table 7.

The first panel of Table 7, including *temporarily out* respondents, presents unweighted statistics, and statistics using the three sets of weights discussed above. This leads to the same conclusion as for the other variables: if the *temporarily out* group is included, there is no evidence of selective non-response after the baseline interview.[14] For example, estimates of the median using inverse probability weights and HRS 2004 weights are virtually the same ($200,500 vs. $200,000). To be precise: the fact that HRS 1992 and HRS 2004 weights give virtually the same wealth quantiles suggests that unit non-response in 2004 is not related to the component of household wealth that can be explained by the basic demographics in *q,* and the fact that inverse probability weights give virtually the same results as HRS 2004 weights, suggests that unit non-response in 2004 is also not related to the components of household wealth, which is driven by the rich set of baseline characteristics in *x* (including baseline wealth).

This is different in the second panel of Table 7, where the *temporarily out* group is excluded, and only the 2004 observations that are in the balanced sample are considered. We then still find very similar results for the two sets of HRS weights, suggesting that temporary non-response is unrelated to the wealth component explained by the basic demographics, but we now obtain a much larger difference between quantiles using HRS weights and inverse probability weights. For example, the estimate of median total wealth, excluding *temporarily out,* is $213,500 using HRS 2004 weights, but only $203,400 if inverse probability weights are combined with HRS 1992 weights. This difference is statistically significant[15] and suggests that, conditional on basic demographics, wealthier families are more likely to be *always in*; not correcting for this leads to an overestimate of median total wealth in the population. If the *temporarily out* are included, the problem disappears, and all weights give about the same median total wealth (between $200,000 and $200,500), which is also rather close to the inverse probability adjusted median using the *always in* only. Thus the *temporarily out* are the group with relatively low wealth (given their demographic characteristics), and bringing them back into the sample is worthwhile to avoid selection problems. In other words, it is important to have (and use) the complete 2004 wave of the unbalanced panel sample, including those who missed one or more waves, rather than only those in the balanced sample. A qualitatively similar conclusion but with smaller selection effects is found for income; for other variables, no evidence of selective attrition is found, whether the *temporarily out* are included or not (results available upon request).

Table 7. Effects of weighting on household wealth: samples excluding and including temporarily out sequences

	10th	25th	Percentile Median	75th	90th
Household Wealth in 2004					
"*Always in*" and "*temporarily out*" sample (attrition weights correct for "*attritors*" only)					
Unweighted	2,000	48,775	166,550	430,000	864,000
HRS-92	5,000	61,500	200,100	487,000	967,500
Inverse probability weights (only *attritors*)	5,000	62,300	200,500	487,000	966,000
	(664.7)	(2358.6)	(5400.4)	(9042.5)	(31785.3)
HRS-04	5,000	62,000	200,000	487,000	969,200
	(660.4)	(2486.2)	(5350.1)	(8951.3)	(31817.3)
Test difference inverse probability weights-HRS04 (p-value)	0.088	0.159	0.371	0.479	0.254
Only "*always in*" (attrition weights correct for "*temporarily out*" and "*attritors*")					
Unweighted	3,800	55,000	179,000	448,000	875,000
HRS-92	7,350	69,000	213,200	500,000	969,200
Inverse probability weights (both *temp. ou*t and *attritors*)	5,598	64,000	203,400	488,000	951,200
	(894.8	(2314.7)	(5927.1)	(9479.2)	(31119.2)
HRS-04	7,300	$69,800	213,500	500,000	977,000
	(1010.8)	(2809.1)	(6193.5)	(10673.9)	(32722.3)
Test difference Inverse probability weights-HRS04 (p-value)	<0.001	<0.001	<0.001	<0.001	0.004
Test difference "*always in*" - "*always in* + *temp. out*" using HRS04 weights (p-value)	<0.001	<0.001	<0.001	0.002	0.287

Notes. *Amounts in 2004 USD. In the top panel, only "always in" respondents (interviews in all years from 1992 to 2002) are retained in the sample. Weights for attrition (includes "temporarily out" and "attritors") are constructed from the multinomial logit estimates in Table 5 and 6. In the bottom panel "always in" and "temporarily out" respondents are retained. IPW weights are derived again from the multinomial logit estimates and are the same as those used in Tables 7 and 8. Standard errors in parenthesis for IPW and HRS-04 calculations. Computed using 500 bootstrap replications. p-value for test of difference computed from normal distribution.*

5. Panel data models

In this section we analyze the consequences of selective non-response for panel data analysis. We consider three examples of static panel data models – a linear fixed effects model for log household wealth, and fixed effects logit models explaining home ownership and labour force participation. See below for details on these models. The regressors are age, indicators of health, and indicators of marital status. We include both current wave and previous wave values of these regressors to capture dynamic effects and to allow for differences in long run and short run effects (see Banks et al 2009).

Again, we focus on the value of the *temporarily out* sample for avoiding attrition bias. We do this by testing for attrition using the complete sample, including and excluding the *temporarily out* group after they have come back into the sample. If bringing back the *temporarily out* is essential for avoiding attrition bias, we expect an insignificant attrition bias in case they are included, and a significant attrition bias if they are excluded from the sample used for estimation.

The tests for attrition bias are Hausman tests, following Nijman and Verbeek (1996), who proposed to use a Hausman test for non-random attrition based upon comparing estimates using only the balanced sample of respondents participating in all waves, with estimates using the complete unbalanced sample, that includes those that participate in some waves and not in others. Under the null hypothesis of no selection on unobservables (or observables other than those included in the model), both estimators are consistent, and the one using all observations in the unbalanced panel is efficient. Hence, a test can be based upon the difference between the two sets of estimates. Let β be the k-vector of parameters. Denote the asymptotically efficient estimator under the null by $\hat{\beta}_e$ and the consistent but inefficient estimator under the null by $\hat{\beta}_c$. The Hausman test statistic is given by

$$D=\left(\hat{\beta}_c-\hat{\beta}_e\right)'Var(\hat{\beta}_c-\hat{\beta}_e)^{-1}\left(\hat{\beta}_c-\hat{\beta}_e\right)$$

where, as shown by Hausman (1978), $Var(\hat{\beta}_c-\hat{\beta}_e)$ simplifies to $Var(\hat{\beta}_c)-Var(\hat{\beta}_e)$. Under the null of no selective attrition, the test statistic asymptotically follows a chi squared distribution with *K* degrees of freedom. We perform this test for all parameters jointly and for subsets of the parameters. For one parameter, the test is equivalent to a simple t-test on significance of the difference in the two estimates.

Hausman tests are also used to choose between random effects and fixed effects models (see Cameron and Trivedi 2005) and to compare estimates based upon the unbalanced panel including all available observations, and upon the unbalanced panel excluding the observations on the *temporarily out* group after they have come back into the sample. In the latter case, under the null that non-response and attrition are random given the covariates included in the model, the estimator using all observations is efficient but the estimator dropping the observations on the *temporarily out* respondents after they have come back is consistent but not efficient, justifying the use of a standard Hausman test; this test has power if the observations not used in the latter case are different (in terms of unobservables driving the variable of interest) from the other observations.

Finally, we will also use Hausman tests to compare estimates that do and do not include observations on respondents who die later on (and are registered as deceased at a later survey wave). This can show whether any selective attrition that we find can be due to mortality.

Household wealth

We use a static linear panel data model with fixed effects to explain log household wealth y_{it} :[16]

$$y_{it}=x_{it}\beta+\alpha_i+\varepsilon_{it}$$

$\varepsilon_{it};t=1,...,T$, independent of each other and of $x_{it};t=1,...,T$

Here x_{it} is the vector of observed regressors (assumed to be strictly exogenous) and α_i the unobserved individual effect. Note that this model makes no assumptions on the α_i, in contrast to a random effects model in which α_i would be assumed to be independent of $x_{it};t=1,...,T$ and $\varepsilon_{it};t=1,...,T$

The results are presented in Table 8. All these estimates are obtained using standard within-group estimators for the static linear fixed effects panel data model, using Stata (xtreg with the option fe), automatically accounting for incomplete observations in an unbalanced panel, under the assumption that the error terms in the model are independent of non-response (see, for example, Cameron and Trivedi 2005).[17]

We also estimated random effects (RE) models (with varying intercepts only, not with varying slope coefficients) with the same samples and explanatory variables. The Hausman tests of the RE against the FE model always clearly reject the RE model. This is why we do not discuss the RE estimates in detail.

Table 8. Fixed effect regressions for log wealth

	Balanced		Unbalanced			Excluding returns		
	Estimate	t-value	Estimate	t-value	z-diff	Estimate	t-value	z-diff
age	3.082	3.48	2.854	3.35	0.92	3.073	3.56	0.05
age squared	-0.221	-3.10	-0.204	-2.98	-0.85	-0.222	-3.20	0.07
current wave								
ever had severe health condition	-0.039	-0.40	-0.005	-0.06	-0.87	-0.015	-0.17	-0.66
ever had mild health condition	-0.212	-2.25	-0.164	-1.85	-1.48	-0.172	-1.92	-1.34
health good	-0.167	-3.02	-0.176	-3.38	0.52	-0.165	-3.13	-0.13
health fair/poor	-0.362	-4.73	-0.397	-5.59	1.26	-0.361	-5.04	-0.02
divorced	-0.894	-5.87	-0.899	-6.62	0.08	-0.938	-6.78	0.71
widow(er)	-0.631	-4.94	-0.684	-5.82	1.06	-0.679	-5.68	1.05
previous wave								
ever had severe	-0.030	-0.29	-0.113	-1.18	2.05	-0.096	-1.00	1.76
ever had mild health	0.210	2.13	0.141	1.53	2.02	0.170	1.81	1.30
health good	-0.070	-1.29	-0.086	-1.66	0.83	-0.083	-1.60	0.76
health fair/poor	-0.120	-1.55	-0.129	-1.80	0.31	-0.138	-1.89	0.66
divorced	-0.210	-1.43	-0.134	-1.01	-1.18	-0.177	-1.31	-0.56
widow(er)	-0.153	-1.15	-0.053	-0.43	-2.07	-0.055	-0.44	-2.23
Observations	35,320		44,895			43,291		

Nijman Verbeek / Hausman tests comparing models	**Balanced/ Unbalanced**		**Balanced/ Excluding returns**		**Unbalanced/ Excluding returns**	
	stat	**p-value**	**stat**	**p-value**	**stat**	**p-value**
All coefficients (df=14)	22.4	0.070	27.0	0.019	33.7	0.002
Age (df=2)	1.9	0.379	4.1	0.130	2.4	0.301
Current health (df=4)	4.2	0.380	2.4	0.670	13.9	0.008
Curr. family status (df=2)	1.3	0.535	1.2	0.548	2.9	0.238
Lagged health (df=4)	9.3	0.054	5.9	0.208	5.9	0.210
Lagged fam. St. (df=2)	4.5	0.108	5.0	0.082	3.0	0.223

Notes. *Fixed effects OLS estimates. Sample 1992-2004. Dependent variable: ln(wealth). "Balanced" uses only the observations in the balanced panel; "Unbalanced" uses all observations; "Excluding returns" uses all observations except those of the temporarily out group after they have missed one wave and come back into the panel. Z-diff statistics are the t-values on the differences between the given estimates and the estimates based upon the balanced panel only (in the first column). The Nijman Verbeek / Hausman tests are explained in the text.*

The Hausman test comparing the balanced panel estimates (column "balanced") of Table 8 and the estimates based upon the complete unbalanced panel (column "unbalanced"), does not reject the null hypothesis that non-response is not selective at the 5% level ("Nijman and Verbeek test - all" in Table 8; p-value = 0.071).[18] The same result is obtained for subsets of coefficients; only for the four lagged health variables, the differences between column 1 and column 2 estimates are close to jointly significant (p-value 0.054). The results are in line with expectations: wealth falls with age and with health problems, and long run effects are generally larger than short run effects (since the coefficients on the lagged and current values of the same variable are usually of the same sign). Divorce or widowhood also leads to substantial reductions of household wealth, but here the lagged variables are insignificant, implying that the long run and short run effects are not significantly different.

The final columns ("excluding returns") use the unbalanced panel, excluding the observations of the *temporarily out* group, after they have missed a wave and have come back into the panel. This mimics the situation in which non-respondents in one specific wave would never be interviewed in any follow-up waves – wave non-response automatically becomes attrition. The Nijman Verbeek test shows that in this case, estimates would be significantly biased due to attrition (p-value 0.019) , suggesting that temporary non-respondents are rather special where wealth formation is concerned, and having them in the sample after they have missed an interview is important, to avoid selectivity bias. This is also confirmed by the Hausman test comparing the estimates using the full unbalanced panel, and the unbalanced panel excluding the returnees: these two sets are significantly different also (p-value 0.0023). In particular, the effects of current health variables are significantly different when observations for respondents who return to the panel are retained (the joint test result gives p-value 0.008).

Tables 9 and 10 present the results for home ownership and labour force participation. The model used here is a static logit model with fixed effects:

$$P(y_{it}=1\,|\,x_{it},\alpha_i)=(\exp(x_{it}\beta+\alpha_i))^{-1}$$

where y_{it} is the dependent variable of interest: 1 for home owners (or labour force participants); 0 for non home owners (or non-participants), x_{it} is the vector of explanatory variables (age; current and lagged values of health and marital status) and α_i is an unobserved household (or individual) specific effect. No assumptions are made about α_i.[19] The model is estimated using the conditional logit estimator of Chamberlain (1980), which is the conditional maximum likelihood estimator, conditioning on the sum over *t* of the y_{it} for each individual *i.* This estimator only uses the respondents whose housing situation or labour force status changes (from owner to non-owner,or working to non-working, or vice versa), explaining the much lower numbers of observations used for estimation than in Table 8. We used the standard command for this in Stata (clogit), which can handle an unbalanced panel, assuming that non-response is random, conditional on the explanatory variables. The covariance matrix of the estimator is computed in the same way as for maximum likelihood.

For the models explaining home ownership in Table 9, no significant differences are found between the three sets of estimates, using the balanced panel only, using the complete balanced panel, and using the unbalanced panel excluding the observations in the *temporarily out* group after they have returned into the sample.[20] According to all three sets of estimates, the probability that a household owns its home falls with age and, in particular, with a transition of the head of household's family status from being married into being divorced or widowed. The effect of a divorce is larger than the effect of widowhood, and usually materializes immediately and not with a lag; for widowhood, the long run effect is about 1.5 times larger than the short run effect (and the difference between long run and short run effect – the coefficient on lagged widowhood – is always significant). Health variables play a limited role: all the individual current and lagged health indicators are insignificant at the 5% level.

For labour force participation (Table 10), however, the tests show significant differences between balanced panel estimates and unbalanced panel estimates, irrespective of whether or not we include the *temporarily out* after coming back into the panel. (And the differences between unbalanced panel estimates, with and without the observations on those who were temporarily out, are insignificant; the p-value of the test is 0.0861 (see Table 10)). This implies selective non-response that is not removed by bringing back in temporary non-respondents.

Table 9. Conditional logits for home ownership

	Balanced		Unbalanced			Excluding returns		
	Estimate	t-value	Estimate	t-value	z-diff	Estimate	t-value	z-diff
age	10.337	6.84	9.640	7.01	1.12	10.546	7.33	-0.45
age squared	-0.806	-6.63	-0.743	-6.69	-1.28	-0.817	-7.06	0.32
current wave								
ever had severe	-0.008	-0.05	-0.008	-0.05	-0.00	0.023	0.15	-0.50
ever had mild health	-0.030	-0.18	-0.085	-0.57	0.78	-0.128	-0.82	1.71
health good	-0.180	-1.83	-0.135	-1.51	-1.10	-0.163	-1.76	-0.51
health fair/poor	-0.217	-1.72	-0.253	-2.26	0.63	-0.234	-2.01	0.36
divorced	-1.922	-10.01	-1.867	-11.24	-0.57	-1.916	-11.07	-0.07
widow(er)	-1.216	-6.56	-1.168	-7.05	-0.57	-1.166	-6.85	-0.67
previous wave								
ever had severe	-0.300	-1.71	-0.368	-2.39	0.82	-0.333	-2.07	0.47
ever had mild health	0.288	1.71	0.183	1.21	1.38	0.219	1.40	1.10
health good	0.094	0.97	0.099	1.12	-0.11	0.098	1.08	-0.11
health fair/poor	0.012	0.10	0.040	0.36	-0.48	0.028	0.24	-
divorced	-0.274	-1.50	-0.420	-2.69	1.54	-0.355	-2.16	1.00
widow(er)	-0.459	-2.46	-0.634	-3.81	2.09	-0.587	-3.42	1.75
Observations	5,780		7,008			6,638		

Nijman Verbeek / Hausman tests comparing models	**Balanced/ Unbalanced**		**Balanced/ Excluding returns**		**Unbalanced/ Excluding returns**	
	stat	**p-value**	**stat**	**p-value**	**stat**	**p-value**
All coefficients (df=14)	17.7	0.218	15.8	0.326	14.9	0.383

Notes. *Fixed Effect logit estimates. Sample 1992-2004. Dependent variable: 1 if home owner; 0 otherwise. "Balanced" uses only the observations in the balanced panel; "Unbalanced" uses all observations; "Excluding returns" uses all observations except those of the temporarily out group after they have missed one wave and come back into the panel. Z-diff statistics are the t-values on the differences between the given estimates and the estimates based upon the balanced panel only (in the first column). The Nijman Verbeek / Hausman tests are explained in the text.*

Table 10. Conditional logits for labour force participation

	Balanced		Unbalanced			Excluding returns		
	Coeff	t-value	Coeff	t-value	z-diff	Coeff	t-value	z-diff
Age is (ref 50-53)								
54/55	-0.574	-3.34	-0.599	-4.00	0.29	-0.562	-3.72	-0.15
56/57	-0.853	-5.10	-0.792	-5.39	-0.76	-0.785	-5.27	-0.90
58/59	-1.296	-7.77	-1.222	-8.31	-0.94	-1.212	-8.13	-1.13
60/61	-1.975	-11.80	-1.902	-12.83	-0.95	-1.894	-12.57	-1.12
62/63	-3.084	-18.13	-3.011	-19.96	-0.93	-3.013	-19.62	-0.97
64/65	-3.773	-21.45	-3.746	-23.92	-0.33	-3.730	-23.40	-0.57
66/67	-4.434	-24.27	-4.357	-26.65	-0.95	-4.346	-26.10	-1.18
68/69	-4.744	-24.79	-4.705	-27.30	-0.47	-4.682	-26.64	-0.81
70+	-5.324	-26.40	-5.252	-28.62	-0.86	-5.233	-28.01	-1.20
current wave								
ever had severe	-0.448	-4.04	-0.552	-5.44	2.33	-0.506	-4.93	1.38
ever had mild health	0.034	0.33	0.010	0.10	0.59	0.009	0.09	0.70
health good	-0.092	-1.54	-0.144	-2.61	2.27	-0.142	-2.52	2.44
health fair/poor	-0.708	-8.50	-0.812	-10.64	3.14	-0.805	-10.37	3.23
divorced	-0.153	-0.91	-0.154	-1.04	0.01	-0.182	-1.20	0.40
widow(er)	0.056	0.41	0.066	0.53	-0.16	0.021	0.17	0.65
previous wave								
ever had severe	0.056	0.49	0.051	0.49	0.11	0.037	0.34	0.49
ever had mild health	-0.215	-1.89	-0.181	-1.71	-0.81	-0.209	-1.94	-0.16
health good	-0.005	-0.08	-0.022	-0.41	0.76	-0.032	-0.58	1.34
health fair/poor	-0.487	-5.73	-0.508	-6.53	0.65	-0.507	-6.38	0.68
divorced	0.098	0.57	-0.002	-0.02	1.20	0.048	0.31	0.68
widow(er)	0.285	1.95	0.185	1.39	1.65	0.192	1.41	1.71
Observations	18,358		21,623			19,916		

Nijman Verbeek / Hausman tests comparing models	Balanced/ Unbalanced		Balanced/ Excluding returns		Unbalanced/ Excluding returns	
	stat	p-value	stat	p-value	stat	p-value
All coefficients (df=21)	40.7	0.006	44.07	0.002	30.3	0.086
Age (df=9)	12.2	0.204	12.4	0.193	10.9	0.282
Current health (df=4)	18.7	0.000	15.7	0.004	9.8	0.045
Curr. family status (df=2)	0.0	0.985	0.46	0.796	3.1	0.209
Lagged health (df=4)	1.3	0.860	2.2	0.705	3.6	0.469
Lagged fam. St. (df=2)	3.1	0.217	2.9	0.232	1.8	0.411

Notes. *Fixed effects logit estimates. Sample 1992-2004. Dependent variable: 1 if in paid work; 0 otherwise. "Balanced" uses only the observations in the balanced panel; "Unbalanced" uses all observations; "Excluding returns" uses all observations except those of the temporarily out group after they have missed one wave and come back into the panel. Z-diff statistics are the t-values on the differences between the given estimates and the estimates based upon the balanced panel only (in the first column). The Nijman Verbeek/Hausman tests are explained in the text.*

In this case, it seems worthwhile to investigate to which extent the bias is due to attrition because of mortality. For this purpose, we redid the tests without the observations (in the years when they are still alive) on those who died before they could take the 2004 interview. With this sample of "survivors", the Nijman Verbeek tests do not show evidence of selective non-response or attrition, irrespective of whether the observations of *temporarily out* respondents, after they have come back, are used or not. The null of no selection bias is not rejected at the 5% level: the test statistic is 30.8 with p-value 0.07 without the *temporarily out o*bservations, and 28.4 with p-value 0.13 if the observations on respondents who missed at least one wave are included (these results are not presented in the table). The higher p-value for the latter case suggests that also in this case, bringing back temporarily non-respondents helps to mitigate selection problems. But, as expected, it does not solve the problem of selection bias caused by mortality.

The Nijman Verbeek Hausman tests on individual coefficients, or on groups of coefficients on related variables, show that the main reason for rejecting the null hypothesis is differences in the effects of current health. The onset of a severe health condition or deterioration in self-assessed health always has a negative effect on the probability to participate in the labour market, as expected, but the effects are larger according to estimates using the (complete) unbalanced panel than when using the balanced panel. In spite of these statistically significant differences, the qualitative conclusions from the three sets of estimates in Table 10 are the same. Labour force participation falls monotonically with age and with health problems. A transition into fair or poor health has a long run effect that is about 1.6 times as large as the short run effect, and this is the only health variable for which long run and short run effects are significantly different. Transitions into widowhood or divorce have no significant effects.

6. Conclusions

In this study, we have investigated the effects of unit non-response in follow-up waves on inference based on the Health and Retirement Study (HRS). Our analysis focused on the HRS cohort born 1931–1941 that was interviewed every two years since 1992. We have focused on how bringing respondents, who do not participate in one interview, back into the sample at later waves, can mitigate the attrition bias. In cross-sectional analysis of the distributions of household income or wealth in 2004, we found that bringing back this group helped substantially to reduce selection bias. With this group included, there is basically no evidence of selection bias that would warrant the use of more complicated weighting schemes than the weights provided by HRS. On the other hand, much larger selection effects are found when the *temporarily out* respondents are discarded, mimicking the situation that they would not be available. This shows the value of having (and then, obviously, using) the *temporarily out* group in later waves.

Panel data analysis confirms that not including the *temporarily out* group can bias estimates of models explaining household wealth; with this group included, tests for selective attrition and non-response show no evidence of selection bias. Similar analyses of panel data models explaining other variables confirm that the HRS efforts to keep respondents in the sample, or bring them back into the sample after they have missed one wave, are successful in the sense that selectivity problems are avoided. For home ownership, we never find any evidence of selection bias; for labour force participation, we find evidence of attrition bias due to mortality, but not due to other sources of unit non-response; here the situation also improves by bringing back temporary non-respondents.

These findings have implications for users as well as designers of surveys such as the HRS, including, for example, the English Longitudinal Study of Ageing (ELSA) and the Survey of Health, Ageing and Retirement in Europe (SHARE) which target similar populations in different countries and have similar sample designs. Attempting re-interviews, for those who missed a wave, appears to have high potential for reducing attrition bias. From a user's perspective, we would argue in favour of using the unbalanced sample in longitudinal analysis. We have found that the balanced sample—the sample that excludes those who come back to the study—suffers from significant selection on observables when looking at financial outcomes in 2004.

Acknowledgements

This research was funded by the U.S. Social Security Administration. We are grateful to four reviewers and an associate editor for useful comments.

References

Attanasio O and Hoynes H. (2000) Differential mortality and wealth accumulation. *Journal of Human Resources,* 35, 1-29.

Banks J, Marmot M, Oldfield Z and Smith JP. (2009) The SES health gradient on both sides of the Atlantic. In Wise D. ed. *Developments Economics of Aging*. Pp. 359-406. National Bureau of Economic Research, Cambridge, MA.

Behr A. (2006) Comparing estimation strategies for income equations in the presence of panel attrition. *Jahrbücher für Nationalökonomie und Statistik,* 226, 361-383.

Behr A, Bellgart E and Rendtel U. (2005) Extent and determinants of panel attrition in the European Community Household Panel. *European Sociological Review*, 21, 489-512.

Cameron C and Trivedi P. (2005) *Microeconometrics: methods and applications.* Cambridge University Press, New York.

Chamberlain G. (1980) Analysis of covariance with qualitative data. *Review of Economic Studies,* 47, 225–238.

Falaris EM. (2003) The effect of survey attrition in longitudinal surveys: evidence from Peru, Côte d'Ivoire and Vietnam. *Journal of Development Economics*, 70, 133-157.

Fitzgerald J, Gottschalk P and Moffitt R. (1998) An analysis of sample attrition in panel data: the Michigan Panel Study of Income Dynamics. *Journal of Human Resources*, 33, 251-299.

Hausman JA. (1978) Specification tests in econometrics. *Econometrica,* 46, 1251-1271.

Hawkes D and Plewis I. (2006) Modelling non-response in the National Child Development Study. *Journal of the Royal Statistical Society*, 169, 479-491.

Heeringa SG and Connor JH. (1995) Technical description of the Health and Retirement Survey sample design. Institute for Social Research, University of Michigan, Ann Arbor.

Hill DH and Willis RJ. (2001) Reducing panel attrition: a search for effective policy instruments. *Journal of Human Resources* 36, 416-438.

Horowitz JL and Manski CF. (1998) Censoring of outcomes and regressors due to survey nonresponse: identification and estimation using weights and imputation. *Journal of Econometrics,* 84, 37-58.

Horvitz D and Thompson D. (1952) A generalization of sampling without replacement from a finite population. *Journal of the American Statistical Association*, 47, 663-685.

Hoynes H, Hurd M and Chand H. (1998) Household wealth of the elderly under alternative imputation procedures. In D Wise. ed. *Inquiries in the economics of aging*. Pp. 229-257. University of Chicago Press, Chicago.

Jones AM, Koolman X and Rice N. (2006) Health-related non-response in the British Household Panel Survey and European Community Household Panel: using inverse-probability weighted estimators in non-linear models. *Journal of the Royal Statistical Society*, 169, 543-569.

Kalton G and Brick M. (2000) Weighting in household panel surveys. In R David ed. *Researching social and economic change: the use of household panel studies*. Pp 96-112. Routledge, New York.

Kapteyn A, Michaud PC, Smith JP and van Soest A. (2006) Effects of attrition and nonresponse in the Health and Retirement Study. RAND Labor and Population working paper WR-407, RAND, Santa Monica, CA.

Lillard L and Panis CW. (1998) Panel attrition from the Panel Study of Income Dynamics: household income, marital status and mortality. *Journal of Human Resources,* 33, 437-457.

Little RJ and Rubin DB. (1987) *Statistical analysis with missing data.* Wiley, New York.

Lynn P. (2006) Editorial: attrition and non-response. *Journal of the Royal Statistical Society* 169, 393-394.

Manski CF and Altonji JG. eds. (1998) Attrition in longitudinal surveys. *Journal of Human Resources,* 33(2), Special Issue.

Nicoletti C and Peracchi F. (2005) Survey response and survey characteristics: microlevel evidence from the European Community Household Panel. *Journal of the Royal Statistical Society*, 168, 763-781.

Nijman TE and Verbeek M. (1996) Incomplete panels and selection bias. In Matyas L and Sevestre P. eds. *The Econometrics of Panel Data: Handbook of Theory and Applications (Second Revised Edition)*. Pp 449-490. Academic Publishers, Dordrecht.

Olsen R. (2005) The problem of respondent attrition: survey methodology is key. *Monthly Labor Review*, 128, 63-70.

Rodgers W. (2006) Incentive size effects in a longitudinal study. Unpublished mimeo. Survey Research Center, Ann Arbor, MI.

Vandecasteele L and Debels A. (2007) Attrition in panel data: the effectiveness of weighting. *European Journal of Sociology*, 23, 81-97.

Van Soest A and Hurd M. (2008) A test for anchoring and yea-saying in experimental consumption data. *Journal of the American Statistical Association*, 103, 126-136.

Wooldridge J. (2002) Inverse probability weighted M-estimators for sample selection, attrition, and stratification. *Portuguese Economic Journal,* 1, 117-139.

Appendix. Definition of Variables

We used two panel status variables in this analysis. The first one tracks the status of a respondent's record in 2004, conditional on being interviewed in 1992. We define four states: (1) continuously interviewed between 1992 and 2004; (2) missed some interviews but interviewed in 1992 and 2004; (3) died prior to 2004; and (4) not interviewed in 2004 for reasons other than death. The variables used for this construction are x*iwwave* and x*alive* from the tracker file (where x denotes the wave, e.g. A, B …). Someone who was reported dead in 2004 is defined as (3) (died), even though an exit interview was collected in 2004. Someone presumed alive by the interviewer is defined as alive. The other status variable is a wave-specific variable that uses the same information as the cumulative status variable but tracks the status at each wave (1 = core interview provided, 2 = dead, 3 = no interview provided, known alive). Table A.1 documents the variables we use in the analysis.

Table A.1 Variable Definitions

Demographics	**Type**	**Definition**	**RAND HRS vars**
age	**years**	**age of respondent**	**ragey_b**
age 50-55 spline	years	(age-50)*1(age<56)+5*1(age>55)	
age 56-60 spline	years	(age-55)*1(age>55)	
female	0/1	gender of respondent	Ragender
born outside U.S.	0/1	respondent born outside U.S.	rabplace(11)
African American	0/1	race is African American	raracem(2)
Hispanic	0/1	race is Hispanic	Hispan
married	0/1	respondent married/partnered	rmstat(1,2,3)
widow(er)	0/1	widow or widower	rmstat(7)
divorced	0/1	currently divorced	rmstat(4,5,6)
once divorced	0/1	once divorced but now married	rmdiv>0
single	0/1	never married	rmstat(8)
household size	number	number of household members	Hhhres
Census Division	1/9	Census division of primary residence in 1992[21]	rcendiv
number of siblings	number	number of siblings alive	rlivsib
number of children	number	number of children alive	hchild
dad alive	0/1	father alive	rdadliv
mom alive	0/1	mother alive	rmomliv
Health Status			
health good	0/1	health reported good	rshlt(=3)
health fair/poor	0/1	health reported fair/poor	rshlt(=4,5)
ever had severe cond.	0/1	ever had cancer/lung/heart/stroke	rcancre rhearte rstroke rlunge
ever had mild cond.	0/1	ever had psychic/diabetes/blood pressure	rdiabe rhibpe rpsyche
at least one ADL	0/1	at least one limitation in activities of daily living	radla>0
SES and Employment Status			
high school	0/1	high school education (completed or not)	raeduc(2,3)
some college	0/1	some college education (not completed)	raeduc(4)
college and above	0/1	completed college education or higher degree	raeduc(5)
own house	0/1	own primary residence	hafhous!=6
working	0/1	working for pay	rlbrf(1,2)
retired or partly retired	0/1	self-reported retired/partly retired	rlbrf(4,5)
disabled	0/1	self-reported disabled	rlbrf(6)
not labour force	0/1	not in labour force or unemployed	rlbrf(3,7)
have pension current job	0/1	conditional on working	rjcpen
receive pension income	0/1	receive any income from a pension	rpeninc

(table A.1 cont'd)

Income and Wealth $USD 2004 (BLS CPI used)			
total wealth	$USD2004	IRAs+Stocks+Bonds+Savings+Certificate&Deposits +Primary residence value + other assets - Debt – Mortgage	haira hastck habond hachck hacd hadebt hamln hahous hamort harles hatran haothr
hld income	$USD2004	Household annual gross income	hitot
individual earnings	$USD2004	Individual annual gross earnings	riearn
poverty threshold	0/1	based on CPS poverty definition for household income, does not include institutionalized family members	hinpov

Endnotes

[1] See http://hrsonline.isr.umich.edu

[2] In previous analysis, we have found that the weighted HRS and the March CPS were very similar in 1992 for a subset of outcomes such as education, labour force participation and civil status (Kapteyn et al 2006).

[3] HRS experimented with randomized "end games" for a subset of respondents classified as "hard refusal." The reward for participation in such games could go up to $100. Hill and Willis (2001) find this has an effect on participation in the 1996 wave. For a similar experiment in 2000, Rodgers (2006) reports strong participation effects for re-contacts of respondents who did not provide an interview the previous wave.

[4] To investigate this further, it would be possible to distinguish among several reasons for unit non-response (not located, not contacted, or refused to participate).

[5] Item non-response in open-ended questions on wealth and income components in the HRS is substantial, but follow-up questions provide information on income and wealth brackets. To deal with item non-response in wealth and income components, we follow the large majority of studies using the HRS and use the RAND-HRS imputations (see Hoynes, Hurd and Chand 1998). These multiple imputations use bracket responses as well as information on characteristics of respondents and are based on covariates similar to those used in our analysis.

[6] The estimates do not take account of the complex nature of the two-stage survey design, which might mean that standard errors are underestimated; earlier studies, however, suggest that the design effects in the HRS are quite small (Van Soest and Hurd 2008).

[7] If we exclude the home ownership dummy, an even stronger effect of being in the lower wealth quintiles is found.

[8] Fitzgerald et al (1998) report differences in terms of labour income which are usually only statistically different from zero at the 10% level.

[9] Throughout the paper we use the respondent level weights and not the household weights, since in our analysis the respondent is the unit of observation. (For a variable at the household level such as wealth, the same value is used for members of the same household.)

[10] Fitzgerald et al (1998) refer to this assumption as *selection on observables*.

[11] In addition, we make auxiliary assumptions, e.g. non-response in the CPS is completely random and the cells used to construct the weights are chosen adequately.

[12] To see this, note that the probability that a population member is in the 2004 sample is the product of the inclusion probability in 1992 and the retention probability. Under our assumptions, the HRS 1992 weight is proportional to the inverse of the former and $p(x_{i0})$ is proportional to the inverse of the latter.

[13] About 1.3% of the 2004 sample are institutionalized; they have HRS weight zero and are not included in our computations.

[14] The estimates do not take account of the complex nature of the two stage survey design, which might mean that the size of the tests that we use is larger than the intended 5% level; earlier studies, however, suggest that the difference is small (see also endnote 5).

[15] Standard errors and t-tests are calculated using 500 bootstrap replications. A standard bootstrap procedure in Stata was used, without replacement and treating the weights as fixed.

[16] In order to deal with zeros as well as negative amounts, we use the common inverse hyperbolic sine transformation $y = \log(u + \sqrt{1+u^2})$.

[17] The estimates do not take account of the complex nature of the two stage survey design; see endnote 5.

[18] This may seem surprising given the stylized fact that life expectancy is positively associated with wealth, implying that attrition due to mortality is likely to be selective. If we exclude the observations on those who die during the sample period, we do find selective attrition, suggesting that attrition due to mortality and other temporary or permanent non-response, lead to biases in opposite directions.

[19] We also estimated random effects models, but these were always rejected against the corresponding fixed effects models by a Hausman test (details available upon request).

[20] We also do not find significant differences if we exclude the (early) observations on those who die before they would be interviewed in 2004.

[21] The US Census Bureau defines nine census divisions, used as regional indicators.