VALIDATING MORTALITY ASCERTAINMENT IN THE HEALTH AND RETIREMENT STUDY

November 3, 2016

David R. Weir

Survey Research Center

University of Michigan

This research is supported by the National Institute on Aging (U01 AG009740).

ABSTRACT. Accurate tracking of mortality is of central importance in longitudinal studies of older populations, both as an important health outcome and as a necessary part of accounting for the representativeness of the panel over time. For nationally representative studies where good population life tables exist the quality of mortality ascertainment can be validated by comparison with life tables. This is demonstrated here for the Health and Retirement Study (HRS), using both period and cohort comparisons from 1992-2011. Mortality ascertainment in the HRS is effectively complete.

Introduction

Mortality ascertainment is critically important to longitudinal studies of aging. For epidemiological research, it is a crucial outcome to assess for evaluating determinants of health and disparities in health. Failure to accurately capture deaths makes it difficult to properly define exposure to risk of any outcome, including mortality, leading to potentially biased findings. Finally, for studies that aim to represent defined populations the levels of mortality provide a key gauge of the quality of that representation. It is somewhat surprising, then, that so little research exists on methods for evaluating mortality ascertainment. This paper aims to provide statistical methods for evaluating mortality ascertainment in population-based longitudinal studies and to apply them to the Health and Retirement Study (HRS).

The HRS is an important public resource for the study of aging, with thousands of data users and over 2,000 peer-reviewed journal publications (Sonnega, et al., 2014). Of those published articles, more than 250 include mortality as a subject and effectively all of them depend on the representativeness of the study and the accuracy of exposure to risk over time for their conclusions.

Mortality Ascertainment in the HRS

Currently, the HRS documents mortality through two primary sources. First, at each biennial wave all previously surviving panel members are sought to obtain interviews. If a death is reported, that information is recorded and an attempt is made to obtain an exit interview with next-of-kin. Second, following each wave of data collection, a search is made in the National Death Index (NDI+) for death information, including cause of death. Because NDI charges per search, we exclude persons we know were alive because they provided interviews in the most recent wave, or with whom direct contact was made but no interview taken, and we exclude deaths previously linked to NDI. We submit to NDI all reported deaths to panel members, anyone whose vital status was not definitively established during tracking, and all attritors (previous panel members who have been removed from the sample and so no longer tracked through interviewing).

The NDI linkage was most recently conducted following the 2012 wave and has death information through the end of 2011. The most recent completed HRS wave was in 2014 and so death information from tracking is available through 2013, with partial coverage of 2014. Table 1 summarizes mortality ascertainment by source. For continuing panel members, most deaths are reported in both sources (11666/12220 = 95.5%). An additional 3.5% are reported in tracking but not matched in NDI, and 1.0% are found in NDI but not in tracking. Deaths among attritors by definition cannot be identified through tracking. I examine attrition in more detail later in the paper. Approximately 3.1% (395/12615) of total deaths from 1992-2011 were deaths to attritors identified only through NDI. Overall, tracking identified 95.9% and NDI 96.6% of deaths.

TABLE 1. Mortality ascertainment in the HRS, by source.

	Deaths before 2012			Deaths after 2011		
Ascertainment source	In sample at time of death	Attritors	All	In sample	Attritors	Total
Both	11666		11666			11666
Tracking only	428		428	1827		2255
NDI only	126	395	521		n/a	521
	12220	395	12615	1827		14442
% in NDI	96.5%	100.0%	96.6%	0.0%	0.0%	84.4%
% in tracking	99.0%	0.0%	95.9%	100.0%	0.0%	96.4%

.

The HRS 2012 and 2014 waves identified an additional 1827 deaths, excluding deaths reported to have occurred before 2012. NDI linkage will be sought for these cases but at the present time only the survey information on vital status is available.

Of the 11,666 deaths identified by tracking and matched to NDI, 10,728 (92%) had a date of death reported to HRS in a next-of-kin (exit) interview. Of those exit interviews, 93.3% of reported dates matched exactly on month and year to NDI, and 6.3% matched within a year to NDI. Only 0.4% of exit reported death dates were more than a year different from the matched NDI record. Thus, HRS tracking and exit information can be considered reliable to establish fact and timing of death.

Comparison of mortality rates to US life tables.

The apparent high concordance between HRS tracking information and NDI linkage is encouraging but does not address either the completeness of mortality ascertainment or the representativeness of the HRS for mortality in the US population. To address those questions we turn to comparison with life tables. This is done in two ways: a comparison of period mortality and a comparison of cohort survival rates. The HRS includes an oversample of African-Americans, whose mortality is higher than that of whites. Rather than attempt to use HRS sampling weights to adjust for this we use unweighted HRS data matched by race and sex to life tables by race and sex. In this way, every HRS participant's vital status counts equally in the validation of mortality.

Period mortality: Actual vs expected deaths and standardized mortality ratios.

The most comprehensive approach to comparing mortality rates is through the use of period standardized mortality ratios as this approach can include the broadest coverage of HRS mortality experience. Cohort survival rates studied below must be done on a somewhat more restrictive basis. A standardized mortality ratio is the ratio of actual deaths to the number of deaths that would be expected given the number of person-years of exposure at life table standard rates. In all that follows, we partition the HRS sample into four groups by sex and race, and match those groups separately to life tables by sex and race, then recombine.

We use single-year life tables by single year of age from age 50 to 99, which are currently available through 2011 (NCHS). Prior to 1996 unabridged tables end at age 85; conventional methods of graduation were used to produce single-year life tables consistent with later years. The life tables give central death rates (m_x) by calendar year for single year of age by sex and race. We label these rates as

$$m_{xit}^{us} ,$$

where the superscript ^{us} indicates US life table, and the subscripts indicate single year of age ($_x$), one of four sex/race groups ($_i$), and calendar year ($_t$).

We construct person-years of observation from HRS from individual records in the public tracker file, which has month and year of birth and month and year of death. The HRS sample design includes all spouses regardless of age. Persons entering the sample as younger spouses, before their own birth cohort is fully represented in the study, are systematically different from the full population of their age because they are married and mortality varies by marital status. We thus would not expect their mortality to match life tables. We therefore begin observation on each individual with her first interview taken after her birth cohort became age-eligible. For example, a woman born in 1943 who joined the HRS as a younger spouse in 1992 would begin observation with her interview in 1998 when her birth cohort (War Babies, 1942-47) entered, while her husband born in 1941 begins observation in 1992. Once entered, observation continues until death, date of interview in 2014, or the end of 2013. Observation periods after 2011 cannot be matched to NDI at this time. Note that attrition does not end observation.

For each calendar year in observation, a number of person-months of observation by age is calculated based on month and year of birth. For example, someone born in May of 1941 surviving in observation for all of 1993 would contribute 4.5 months of observation at age 51 in 1993 and 7.5 months of observation at age 52 in 1993. That same person's death in 2003 would be assigned to age 61 if before May, age 62 if after May, and split between 61 and 62 if it occurred in May.

We denote person-years of exposure as

$$y_{xit}^{hrs},$$

where y indicates person-years of observation, the superscript ^{hrs} indicates observed data from the HRS, and the subscripts again indicate single year of age ($_x$), one of four sex/race groups ($_i$), and calendar year ($_t$).

The population studied is everyone who gave at least one age-eligible HRS interview from 1992 through 2008 inclusive. They enter observation on the date of their first interview after becoming age-eligible, and exit observation on the date of their death, or the date of their interview in 2014, or the midpoint of 2014 for those who remained in the sample but did not give an interview in 2008, or the end of 2011 (the most recent year for which NDI data are available) for those who were removed from the sample in 2012 or earlier. For most purposes we limit all analysis to the end of 2011.

The number of persons, person-years of observation, and deaths for each of the sex/race groups is shown in Table 2 for calendar years 1993 through 2011. There are over 350,000 person-years of observation. The total of deaths is less than the total shown in Table 1 because of the sample restrictions (120 deaths occurred at ages 100 or older and are not included because US life tables end at age 99; 145 deaths occurred to persons who never gave an interview; 446 deaths occurred to persons who did give an interview but died before reaching age-eligibility; and 42 deaths occurred to persons whose first interview was in 2010 or later).

TABLE 2. HRS observational frame for mortality, 1993-2011

	N	Person-Years	Deaths
White male	10,997	130,530	4,798
White female	13,544	169,570	5,148
Black male	1,711	19,286	858
Black female	2,569	31,313	1,058
Total	28,821	350,698	11,862

Expected deaths then are the product of the life table rates times the person-years of observation in HRS corresponding to each age, sex/race group, and calendar year. We denote expected deaths in each such cell as

$$d^*_{xit} = m^{us}_{xit} * y^{hrs}_{xit} ,$$

We can then calculate the total number of expected deaths in a calendar year as

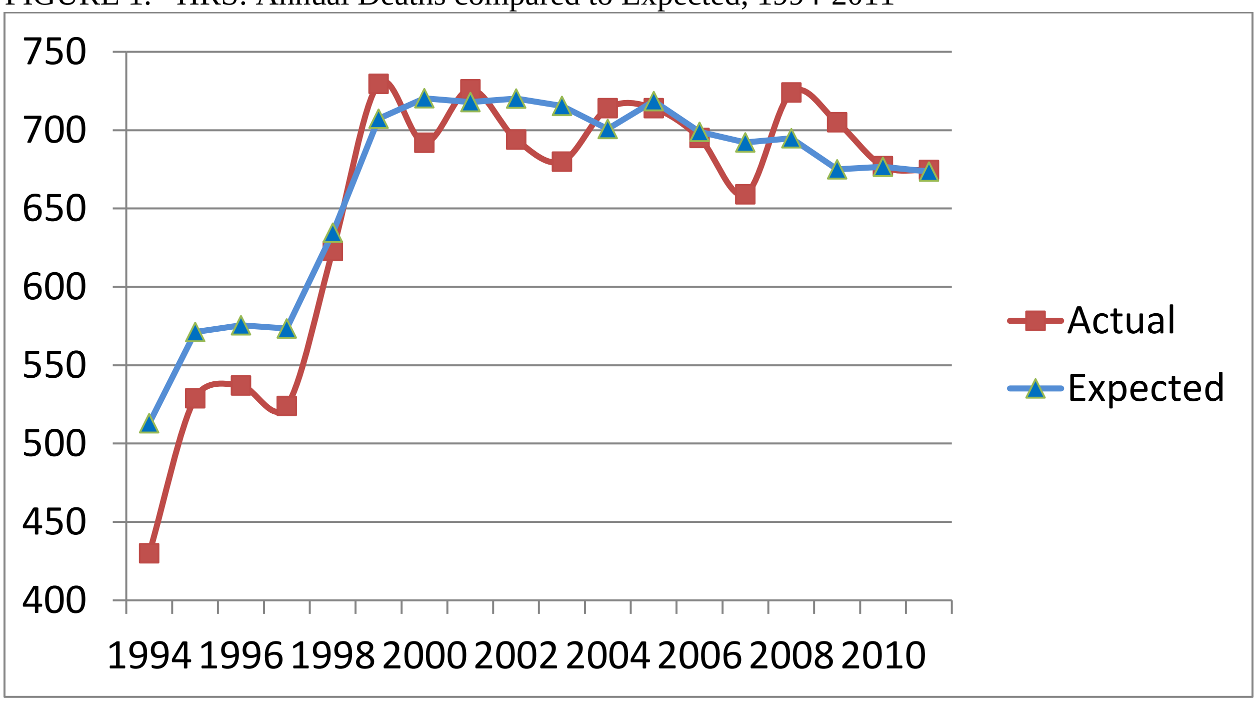
$$D_t^* = \sum_{xi} d_{xit}^*$$

Similarly, each death to an HRS participant is classified by single year of age at death, calendar year of death, and race-sex group. These can also be summed by calendar year:

$$D_t^{hrs} = \sum_{xit} d_{xit}^{hrs}$$

The number of actual and expected deaths by year in the HRS observational frame are shown in Figure 1. The number of actual and expected deaths increased with the addition of the AHEAD cohort in 1993, and again with the addition of the CODA and War Baby cohorts in 1998. Since then, the numbers have been fairly constant with a small boost from the addition of the Early Baby boom cohort in 2004. The number of expected deaths exceeded the number of actual deaths in the early years but the two have been close since 1998.

FIGURE 1. HRS: Annual Deaths compared to Expected, 1994-2011

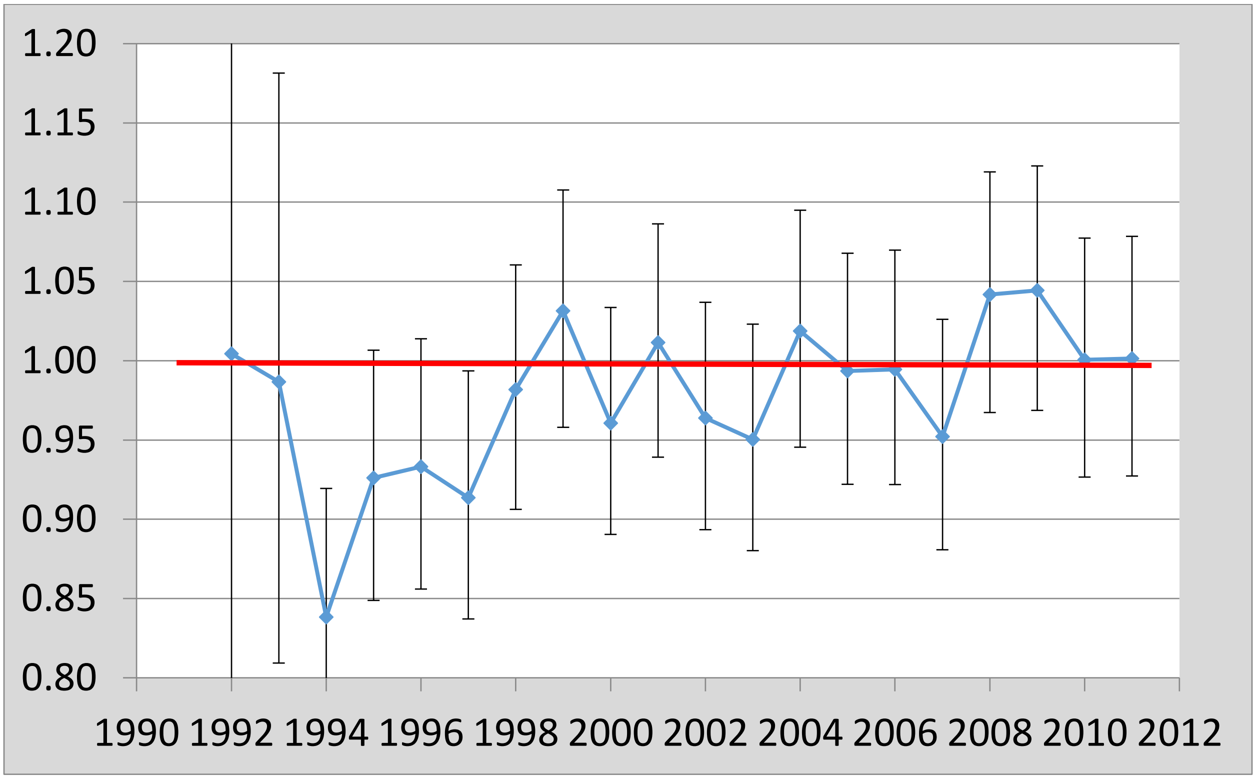


These death counts can be combined to construct the standardized mortality ratio in year t as:

$$SMR_t = \frac{D_t^{hrs}}{D_t^*}$$

Standardized mortality ratios are shown in Figure 2. Confidence intervals based on the Poisson distribution are also shown. The original 1992 HRS cohort had mortality close to expectation in 1992-93 (very few total deaths and very large error bounds). The addition of the much older AHEAD cohort greatly increased the number of expected deaths and produced lower-than-expected mortality in the early years, from 1994-1997, and in some of those years the difference was statistically significant. Since 1998, the SMR has been near one and not statistically significantly different from one.

FIGURE 2. Standardized mortality ratios, 1992-2011



To probe this break around 1998, we compare single-year-of age mortality schedules for HRS with HRS-weighted mortality schedules based on life tables in the two time periods. Figure 3a shows the life tables for 1992-97. Clearly, the pre-1998 HRS death rates are very close to the life table up to around age 75 and lower at ages above that. The most plausible reason for this is that the original AHEAD cohort, persons 70 and above in 1993, recruited community-dwellers only and excluded nursing home residents who were then about 5% of the population above age 70. Death rates in nursing homes are nearly five times that of community dwellers, and deaths to nursing home residents thus are nearly twenty percent of all deaths over age 70. Because relatively few people survive in nursing homes beyond about five years, this bias eroded over time as HRS/AHEAD followed respondents into nursing homes and kept them in observation. Figure 3b shows that since 1998 mortality rates are close to life table rates at all ages and if anything slightly higher at older ages.

FIGURE 3a. HRS mortality by age compared to life tables, 1992-97

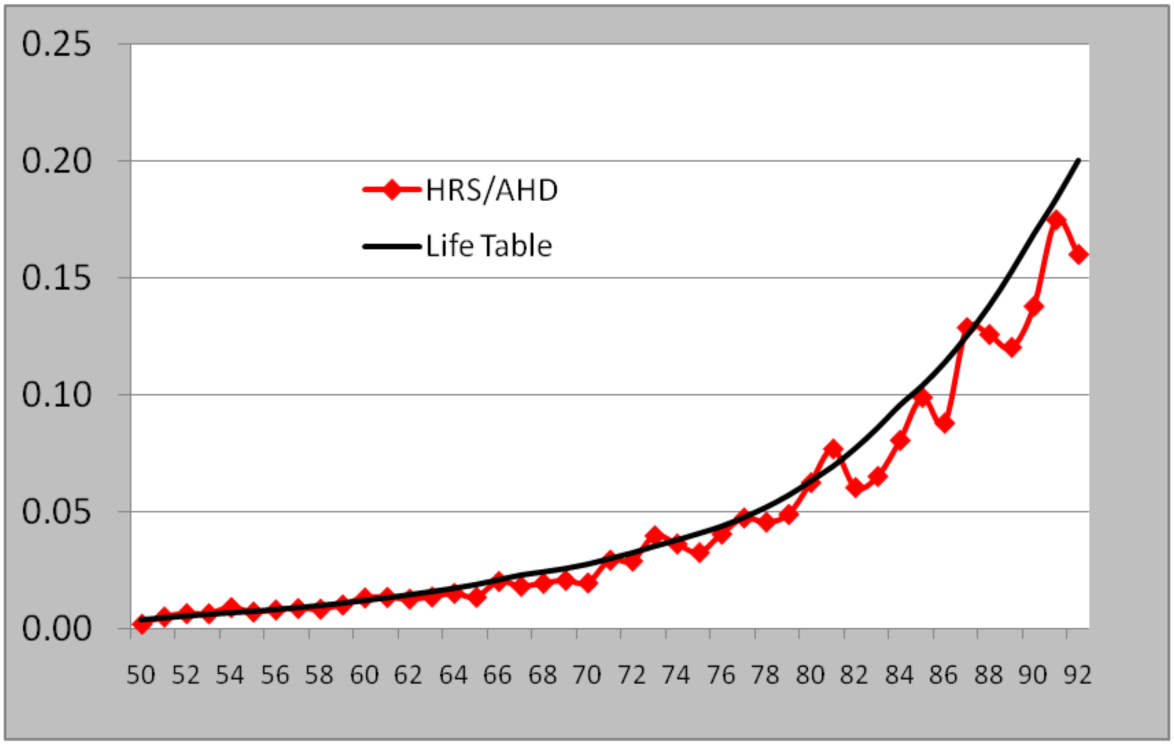
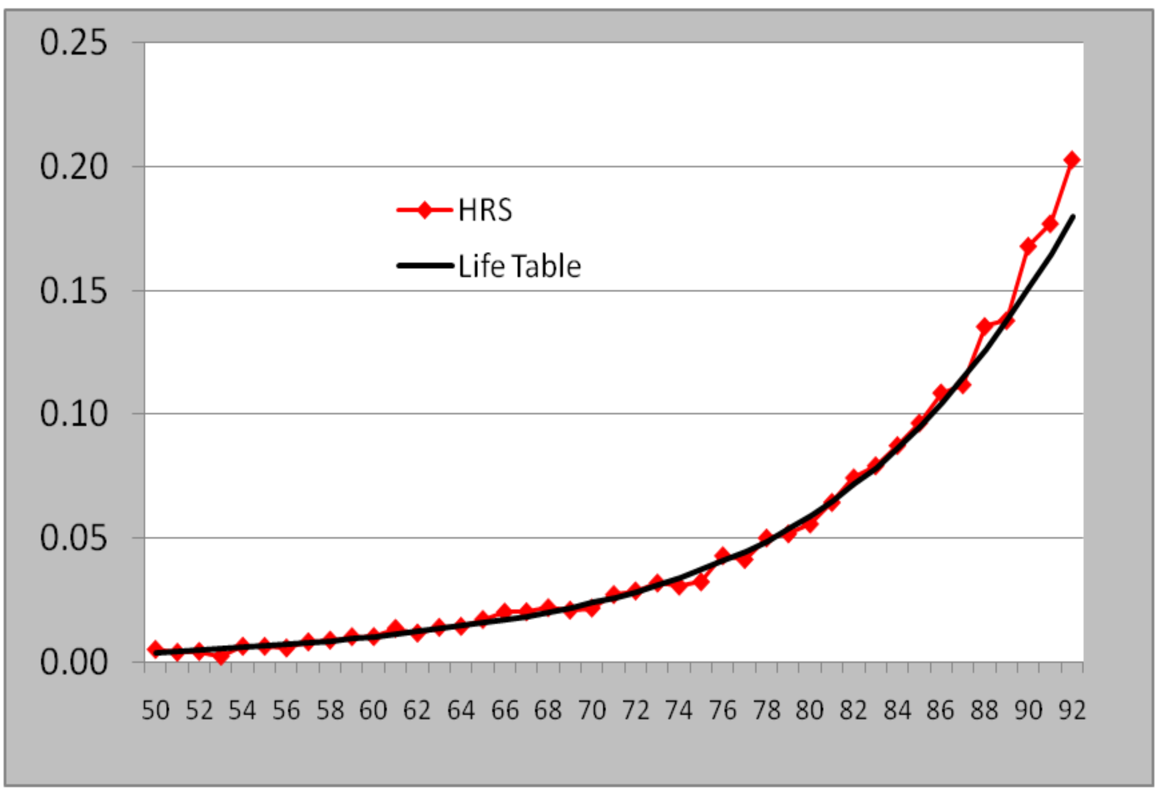


FIGURE 3b. HRS mortality compared to life tables, 1998-2004

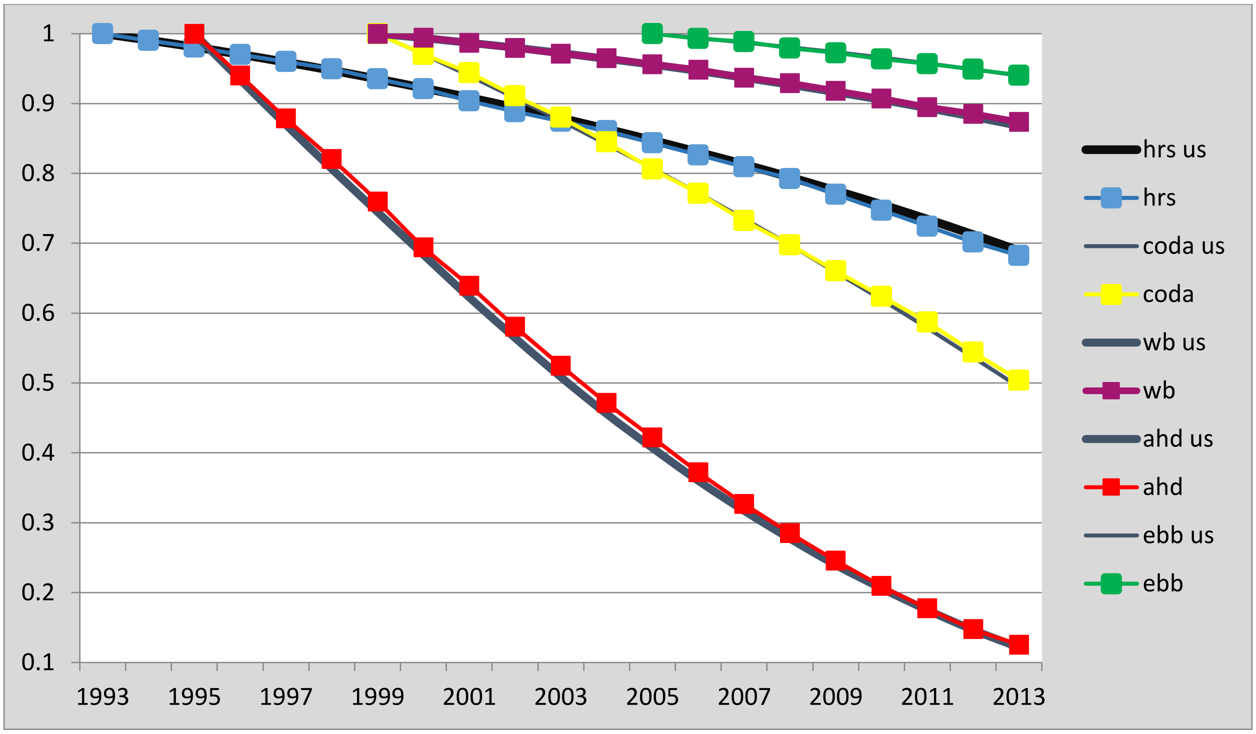


Survival by cohort.

The HRS sample has been built up over time by the successive addition of different birth cohorts. At each new recruitment, spouses of age-eligible individuals also enter, some of whom are themselves age-eligible and some of whom are not. For this analysis of cohorts, we restrict observation to persons who were in the target age-eligible birth cohorts in the year they entered the study so as to limit the age variation in the cohort and because age-ineligibles differ systematically from their age peers. For spouses who were outside the age-eligible range when they entered but subsequently became age-eligible when their birth cohort entered the study, we consider them for this analysis as entering into observation in the year their birth cohort entered, provided they survived and gave an interview in that year. This means that we can follow birth cohorts from a fixed entry point: the HRS cohort born 1931-41 entering in 1992; the AHEAD cohort born 1923 and earlier entering in 1993/94; the CODA cohort born 1924-30 entering in 1998 (with some having entered previously as spouses) and War Babies born 1942-47 entering in 1998 (again with some having entered previously as spouses). In this way, each cohort is a closed population followed from a fixed point in calendar time and the proportion surviving at each subsequent date can be calculated directly.

To create life table expected survival we create synthetic cohort life tables from the annual single-year-of-age period life tables 1992-2011. For each single year of birth four cohort life tables are constructed from the single-year-of-age period life tables by race and sex. Each individual is matched to the corresponding cohort life table by race, sex, year of birth, and year of entry. The actual HRS cohort survival curves and the corresponding life table curves are shown in Figure 4.

FIGURE 4. Cohort survival rates in HRS compared with life tables



In all cohorts actual survival is visually virtually undistinguishable from life table rates. Were there systematic underreporting of mortality in the HRS we would expect to see the actual cohort survival rates falling more slowly than the life table rates. We do not observe such departures, suggesting the study is capturing most deaths accurately. The one exception, again, is that the AHEAD cohort showed somewhat higher survival in the early years. Table 3 summarizes survival to the end of 2011, the last year for which national single-year life tables have been published. The AHEAD cohort is now nearly extinct and the actual survival rate almost identical to expected at about 15%. The original 1992 HRS cohort has the largest deviation at about 1% lower survival than expected, whereas the two adjacent birth cohorts both have slightly lower higher survival than expected.

TABLE 3. Cohort survival from entry to the end of 2011, compared with life tables

	AHD	CODA	HRS	WB	EBB
Birthyears	<1923	1924-30	1931-41	1942-47	1948-53
Entry year	1993	1998	1992	1998	2004
Expected survival	14.7%	53.7%	71.2%	88.1%	94.9%
Actual survival	14.8%	54.4%	70.2%	88.5%	94.9%

Attrition and mortality

For persons remaining in the HRS sample the combination of nearly complete tracking and nearly complete NDI linkage assures complete mortality coverage. For attritors, persons removed from the HRS and no longer tracked, we must rely on NDI linkage alone. HRS makes every effort to retain persons in the sample, including those who refuse to do an interview in one or more waves. In HRS, removal from the sample is primarily the result of insistence on the part of the participant, though the study has periodically removed some persons who have never insisted on removal but have repeatedly demonstrated unwillingness to participate. Table 4 below indicates the number of persons leaving the HRS sample by wave due to death and due to attrition. Particularly high rates of attrition can be seen before the 2006 and 2012 waves of HRS when case reviews were done.

TABLE 4. Death and attrition from HRS, by wave

	Death	Attrition
1994	236	22
1996	1,140	11
1998	1,367	246
2000	1,459	218
2002	1,598	259
2004	1,319	218
2006	1,433	591

2008	1,289	142
2010	1,645	103
2012	1,213	440
2014	1,349	364
Total	14048	2614

Of the 38,000 persons ever in the HRS, 6.8% have been permanently removed via attrition, compared with 36.8% who have died while in the sample. As we saw in Table 1, 395 of the attritors were matched to NDI through 2011. Based on the high rate of NDI matching for in-sample deaths we might conclude very few deaths to attritors are missing. However, attritors are slightly less likely to have provided Social Security numbers, which raise by about 10 percentage points the match rate to NDI. As a check on coverage, we might ask whether their mortality rates are substantially lower than expected from life tables. Because attritors may differ in many ways from the general population we cannot make strong inferences from the comparison to life tables as we did for the full HRS sample, but substantially lower mortality would be a concern.

The challenge for studying the mortality of attritors is correctly defining their period of post-attrition observation. In order to attrit, someone has to survive to the point of refusal so for the study of attrition-related mortality we cannot include observation time while in the sample prior to the date of attrition. There is, however, no clearly defined date of attrition corresponding to, e.g., the date of first interview. We defined attrition as occurring at the beginning of the calendar year of the wave in which the person was first removed from tracking and excluded any deaths occurring prior to that date.

Table 5 summarizes the comparison for all years combined. In the full age-eligible HRS, which includes attritors both before and after attrition, actual deaths were 9,084; 28 short of the total of 9112 expected deaths for an SMR of 99.7%, which is not statistically different from 100%. For post-attrition observations actual deaths were 351 compared to expected of 370, so 19 short for an SMR of 94.8%. That is about the same ratio of NDI matching in the tracked cases. It is possible, then, that NDI is not perfectly complete for attritors and this very small shortfall accounts for much of the overall difference in HRS.

TABLE 5. Actual and expected deaths, all of HRS and post-attrition observations only

	Actual	Expected	SMR	Lower CI	Upper CI
All	9084	9111.6	0.996	0.976	1.0175
Post-attrition	351	369.7	0.948	0.851	1.0500

From Table 5 we can also infer the impact of alternative approaches to dealing with attrition. Currently, HRS counts attritors as in observation for mortality and obtains death information from NDI. If HRS did not have access to NDI to search for attritors, but kept them in observation, then actual deaths would be 8733 (9084-351) and SMR would be 95.8%, significantly below one. However, if HRS ended observation on attritors on the date of attrition the impact on ascertainment would be negligible. Expected deaths would fall to 8741.9 and SMR would be 99.9%, even slightly closer to one. Thus, either consistent approach to the treatment of attritors leaves HRS mortality ascertainment intact. Maintaining attritors in mortality observation allows analysts to include them in analyses of mortality effects of baseline or other early measurements.

Conclusion

When comprehensive population life tables are available it is a straightforward task to evaluate mortality ascertainment in longitudinal studies of those populations. We applied such methods to the US Health and Retirement Study. We conclude from this review of mortality surveillance that it is essentially complete in HRS. The slightly low mortality in the first few years of the study is due to well-known sample selection decisions that excluded nursing home residents at baseline and not to inadequate mortality registration for persons in the study. Therefore, individual observations on survival for persons in the study can be considered accurate, exposure time in the panel can be considered valid, and the study can be considered representative. We encourage other longitudinal studies to conduct similar evaluations.

SOURCES

National Center for Health Statistics (NCHS) Vital Statistics of the United States 1992 (II:6); 1993 (IIIA:6); 1994 (IIA:6); 1995 (IIA:6)

National Center for Health Statistics (NCHS) Vital and Health Statistics Series 2, no. 129 (1996)

National Center for Health Statistics (NCHS) National Vital Statistics Reports 47:28 (1997); 48:18 (1998); 50:6 (1999); 58:10 (2000-2005); 58:21 (2006); 59:09 (2007); 61:03 (2008); 62:07 (2009); 63:07 (2010); 64:11 (2011).

REFERENCES

Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JWR, Weir DR. "Cohort Profile: the Health and Retirement Study (HRS)." International Journal of Epidemiology. 2014; 43 (2): 576-585