Proxy interviews and bias in the distribution of cognitive abilities due to non-response in longitudinal studies: a comparison of HRS and ELSA

David Weir, Jessica Faul, Kenneth Langa
University of Michigan
dweir@isr.umich.edu

Abstract

Cognitive impairment is an important topic for longitudinal studies of ageing, and one that directly affects ability to participate. We study bias in measured cognition due to non-response in the Health and Retirement Study (HRS) and the English Longitudinal Study of Ageing (ELSA). The much greater use of proxy interviews for impaired respondents in the HRS virtually eliminates attrition bias in measured cognition, whereas there is a noticeable bias in ELSA where proxies are infrequently used. Using Medicare claims data for the HRS we are able to compare cognitive impairment among dropouts post-attrition with that for continuing participants. There again we see that the use of proxy interviews virtually eliminates a bias that would otherwise appear.

Introduction

Longitudinal studies are concerned with bias due to attrition, because over time it may compromise the ability to fully represent the diversity of the populations they aim to study.

Moreover, if attrition is systematically related to outcomes of interest or to respondent characteristics correlated with these outcomes, then estimates of the relationships between characteristics and outcomes may also be biased, especially when examined longitudinally. For studies of older populations, attrition bias on cognitive function is of particular concern. Unlike most other dimensions of study for which there is no strong *a priori* theoretical basis to predict a direction of bias, level of cognitive function is fundamentally related to participation in surveys. Surveys are complex conversations that become progressively more difficult as cognitive abilities decline, with the result that the cognitively impaired are less likely to participate or may even be excluded from participation. At the same time, cognitive impairment is one of the most important topics to be studied in longitudinal studies of ageing

populations, because of its debilitating effects and the tremendous burden it can place on families and societies (Langa et al 2001; Ferri et al 2005). Cognitive abilities decline with age generally, and serious impairments and dementia rise sharply in prevalence with age (Brayne et al 1999; McArdle, Fisher and Kadlec 2007; Plassman et al 2007). Accurately capturing the prevalence and burden of dementia is an important aim of longitudinal studies of ageing; however, the prevalence of cognitive impairment will be underestimated in analyses that do not take into account attrition bias on cognitive function where it exists. (Boersma et al 1997; Brayne et al 1999; Chatfield and Brayne 2005).

Prior research on other longitudinal studies of older populations has tended to find the expected correlation of attrition with cognitive deficits (Anstey and Luszcz 2002; Matthews and Chatfield 2004; Chatfield and Brayne 2005; Van Beijsterveldt and Van Boxtel 2002). While such findings, together with the strong theoretical expectation of bias, motivate a study of the issue in other studies, one cannot

assume that such bias is similar across all studies. Different studies place different demands on respondents and different studies have different approaches to retaining participants. The goal of this paper is to assess to what extent attrition biases the representativeness of the Health and Retirement Study (HRS) and the English Longitudinal Study of Ageing (ELSA) with respect to their population distributions of cognitive ability and whether the use of proxy interviews reduces this bias.

Methods

Samples and response status

Data for these analyses come from the HRS and ELSA - both biennial, longitudinal, nationally representative surveys of, US adults aged 51 and older, and English adults aged 50 and older, respectively (Juster and Suzman 1995; Marmot et al 2003). The samples for these analyses come from all available waves of data from each study – 1992 to 2008 from HRS, and waves 1 to 3 from ELSA (collected in 2002 to 2007)(for more details on sample size and response rates to the HRS and ELSA see Cheshire et al 2011). For our initial set of analyses we categorized the response status of sample members in any given wave into three categories: interviewed, non-response (consisting of both refusals and permanent attritors), and death. In subsequent analyses, response status was expanded to four categories to distinguish self-interviews from interviews by proxy.

In longitudinal studies it is important to be clear about the definitions of non-response and attrition. The term attrition properly denotes a permanent departure of a surviving participant from the study sample, never to return. It excludes exits due to death. It gained common usage in the past in longitudinal studies like the Panel Study of Income Dynamics (PSID) in which a single wave of refusal resulted in being dropped from the sample. That is no longer best practice in longitudinal surveys. Indeed, the return of participants after missing a wave is crucial to maintaining their representativeness (Kapteyn et al 2006). As a result, the concept of attrition now has several disadvantages for understanding possible bias or unrepresentativeness over time in a longitudinal study. Firstly, it is subject to arbitrary definition by

the field organization and the variable consent standards imposed by different ethics boards. Should someone who has refused four waves in succession be dropped permanently (attrited) or kept on in hopes of a return with potentially costly future contacts? What type of refusal should be taken as a permanent refusal? More importantly for our interests here, a narrow focus on permanent attrition would miss the potentially larger problem of non-response.

Non-response is sometimes defined by survey organizations to focus only on response patterns of persons remaining in the sample. In that definition, permanent attrition is excluded from non-response. In this paper we define non-response more broadly to include both permanent attrition and non-response of persons in the sample. We therefore begin with the concept of non-response without separation between those who are permanently removed from the study and those who remain eligible. It is important, however, not to confuse mortality with non-response. Mortality affects the population we are trying to represent over time, and it should affect the sample similarly. Both HRS and ELSA use linkage to national death registries to ascertain mortality for attritors and to validate reported deaths of participants. For each survey year, we consider as eligible to respond not only those participants who were in the active sample, but also those who had previously been removed from the sample. We remove from the eligible pool those who were in the active sample, but reported deceased during the survey year, and also those who had been removed from the active sample and whose deaths as reported in the national death registries occurred prior to the midpoint of the survey year. Those who were eligible and in the active sample could either provide an interview in that year, or not. By our definition, non-respondents also include those permanent attritors who had been removed from sample and so were not contacted for interview but remained alive.

Cognitive measures

In both HRS and ELSA, cognitive function is assessed through several questions. Detailed information and documentation for the full set of cognitive measures in each study, including their derivation, reliability, and validity, are available at the

HRS (Ofstedal, Fisher and Herzog 2005) and ELSA (Taylor et al 2007) websites. For these analyses, only performance on the episodic memory tasks was considered. Performance on these tasks was chosen because it is a sensitive measure of cognitive change and has been used in previous comparisons between HRS and ELSA (Langa et al 2009). The episodic memory tasks integrated in these surveys consists of testing for verbal learning and recall, where the participant is asked to memorize a list of ten common words. Scores were calculated by totaling the number of target words respondents were able to remember immediately after hearing the list (immediate recall) and after a series of distraction items (delayed recall). Scores could range from 0, with no word correctly remembered, to 10, with every word correctly remembered for each task.

Proxy interviews - Unlike most longitudinal surveys, the HRS integrates the use of proxy informants into the sample design. This design was adopted in order to minimize the effects of attrition and non-response due to ageing and ageing-related health-conditions. Whenever possible, interviewers are instructed to interview sample members; however, some individuals are unable to complete an interview because of physical or cognitive limitations, or because they are unwilling to participate themselves. In the HRS, most proxy interviews are designated as such at the beginning of the interview. Either the respondent or their spouse or other family member indicates, prior to the interview, that the respondent is unable to participate. In other cases, the respondent is willing to be interviewed and a self-interview is started, but the interviewer has concerns about the respondent's ability to provide accurate information. Specifically, an unusually long interview time during the initial part of the survey, more than a threshold number of "don't know" responses, and poor performance on the cognition items triggers the interviewer to seek a proxy and begin the interview again. Proxy interviews are conducted with someone who is familiar with the financial, health, and family situation of the sampled individual. In practice, this is generally the spouse or partner of the sampled respondent. In the absence of a spouse, the proxy is often a daughter or a son, or less frequently, another relative or a caregiver.

In the HRS, the proxy interview is a separately programmed and worded interview. For most questions, this involves only wording changes (e.g. from "you" to "him" or "her"), but some questions that are thought to be inappropriate to ask of proxies are omitted entirely. These include questions intended to assess depressed mood; the test of cognitive status; subjective expectation questions; and questions asking for subjective evaluations of the person's job or retirement.

Because the cognitive performance tests cannot be conducted with a proxy respondent, a different set of measures is used in the proxy interview to assess the respondent's present cognitive status and change in status between waves. Proxy respondents are asked to rate the respondent's overall memory and change in memory compared to the prior wave, as well as their behavior in terms of overall judgment and organization of daily life. With regard to memory, proxy respondents are asked a series of questions about the respondent's change in memory for various types of information in the last two years. These questions are adapted from the short form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE; Jorm and Jacomb 1989; Jorm 1994). While the cognitive measures for proxy and self-respondents are not the same and so cannot be used interchangeably, they can each be used to arrive at comparable broad categories of cognitive impairment, and these categories can be used to study the impact of cognitive impairment on, e.g. the use of formal medical care and informal caregiving (Langa et al 2001; Langa et al 2004) or the impact of other health events on cognitive impairment (Iwashyna et al 2010).

Proxy respondents are also used in ELSA, although the rules for who is eligible for a proxy interview have changed over time. In the first three waves of ELSA used in this paper, proxies were used only in cases where the respondent was away in a hospital or nursing home throughout the whole fieldwork period, or because they had refused a self-interview. Beginning with Wave 4 (not available for these analyses), respondents were also eligible for a proxy interview if they could not be interviewed in person because of a physical or cognitive impairment.

Additional data

Additional data come from Medicare claims records for 1991-2007 linked to 16,977 HRS respondents. These files contain information collected by Medicare (the US national health insurance program for the elderly and disabled) to pay for health care services provided to a Medicare beneficiary. Only respondents who were continuously enrolled in the Medicare fee-for-service program in the preceding five years were considered. Thus, the approximately 12 percent of respondents who had ever opted to receive their Medicare benefits through a managed care organization were excluded, because managed care providers do not report information about utilization and diagnosis in the Medicare claims files. Respondents were classified as having received a dementia diagnosis if they had a claim reporting at least one diagnosis code (ICD-9-CM) included in the list of codes that comprise the Chronic Condition Warehouse definition of "Alzheimer's Disease and Related Disorders or Senile Dementia" in any of the Medicare claims files, including: inpatient, outpatient, part B physician supplier, Skilled Nursing Facility (SNF), hospice, and durable medical equipment files. This claims-based diagnostic measure does have reasonable sensitivity and specificity for dementia (0.85 and 0.89 for dementia, and 0.64 and 0.95 for Alzheimer's disease, respectively; see Taylor et al 2009).

Analysis

We evaluate the role of attrition on bias in cognition in four parts. We begin with a description of the studies and patterns of non-response in the HRS and ELSA. Second, we offer a measure of the bias in cognition arising from non-response in both surveys, by comparing baseline values of cognition between respondents and all survivors at each follow-up wave, both with and without using sampling weights. Next, we ask how important proxy interviews are to response rates and the containment of attrition bias, by repeating the above comparisons using only self-respondents and treating proxy interviews as non-response. Finally, we use the HRS linkage to Medicare records, to answer the question of whether the development of diagnosed dementia post-baseline increases the likelihood of subsequent non-response.

Results

Descriptive statistics

Table 1 presents descriptive statistics from baseline interviews in the two studies. The HRS sample was built up over time, with different birth cohorts entering at different times. We show baseline statistics for each cohort, and for all cohorts combined. ELSA began in 2002 with a sample of persons 50 years of age and older. Cognitive scores vary with age, being lowest in the two oldest cohorts of the HRS sample. Overall, cognition is slightly higher in HRS than in ELSA and the HRS sample of baseline interviews is slightly younger than the ELSA baseline (because of the addition of younger cohorts in 1998 and 2004 after the study sample had begun to age). Both studies have slightly more women than men, reflecting the gender composition of older populations in which women live longer. We note that for the purposes of the analyses that follow, modest differences in baseline composition by age or cognition between the two studies is of no consequence.

Table 1. Descriptive statistics of samples studied from HRS and ELSA (unweighted).

		Age		Cognition		Percent
Cohort	N	Mean	(s.d.)	Mean	(s.d.)	Female
HRS92	9,794	55.5	(3.2)	10.8	(3.6)	52.9%
AHEAD	7,399	77.6	(5.9)	7.4	(3.9)	61.3%
CODA	2,301	70.6	(2.0)	9.4	(3.4)	59.1%
WB	2,061	53.2	(2.8)	11.4	(3.2)	45.9%
EBB	2,690	52.9	(1.7)	10.6	(3.2)	48.6%
All HRS	24,245	63.2	(11.4)	9.7	(3.9)	55.0%
ELSA	11,392	65.3	(10.4)	9.1	(3.9)	54.5%

Notes. *Abbreviations are HRS (Health and Retirement Study), ELSA (English Longitudinal Study of Ageing, baseline interviews in 2003), HRS92 (first Health and Retirement Study cohort introduced in 1992, age-eligibles are born 1931-41), AHEAD (second Health and Retirement Study cohort introduced in 1993 as Asset and Health Dynamics of the Oldest-Old, age-eligibles born before 1924), CODA (Children of Depression Age; third Health and Retirement Study cohort introduced in 1998, age-eligibles born 1924-30), WB (War Babies; fourth Health and Retirement Study cohort introduced in 1998, age-eligibles born 1942-47), EBB (Early Baby Boomers; fifth Health and Retirement Study cohort introduced in 2004, age-eligibles born 1948-53). Cognition is measured by the sum of words recalled immediately and after delay from a list of ten words.*

Attrition and non-response

Table 2 shows the response rates of survivors in HRS and ELSA. Because the HRS sample was built up over time, with different birth cohorts entering at different times, we show these rates separately for each entry cohort.   There are some modest differences across the HRS cohorts, with the older entrants having better response rates. Most of these contrasts are statistically significant but they are not important for our analysis, and are small relative to the differences between HRS and ELSA. We also provide data for all HRS cohorts combined to facilitate comparison with

ELSA. At wave 2 the HRS combined response rate was 92.4%, indicating a loss of less than 8% from baseline. Response rates are much lower in ELSA where nearly 20% of the surviving baseline sample members did not give an interview at wave 2. The losses are smaller at subsequent waves, due in part to the practice of bringing back people who missed a wave. By wave 3, non-response in HRS was 10%, compared with 27.5% in ELSA. The large differences in response rates between HRS and ELSA are highly statistically significant.

Table 2. Response rate of surviving members of the age-eligible baseline cohort, by wave of follow-up, HRS cohorts compared with ELSA.

Entry Cohort	Base-line Year	Follow-up Wave							
		1	2	3	4	5	6	7	8
HRS 92	1992	91.8	88.8	86.2	83.7	83.4	82.1	81.0	80.4
AHEAD	1993	93.7	91.8	90.2	88.7	87.3	86.7	86.9	
CODA	1998	93.8	91.6	90.0	89.4	88.1			
WB	1998	92.3	91.6	88.7	87.4	86.1			
EBB	2004	89.8	87.3						
All HRS		92.4	89.9	88.0	85.9	85.0	83.2	82.1	
ELSA	2002	80.7	73.1						
HRS-ELSA		**11.7****	**16.8****						

*Notes. For abbreviations, see Table 1. Response rates are unweighted percentages. For differences between studies, estimates in **bold** are greater than zero at P<.05 significance level in a one-tailed test; estimates with * are significant at p<.01, and ** at p<.001.*

Bias in distribution of cognition

Our concern here is with a bias in sample composition due to higher rates of non-response among persons with lower levels of cognition, not with any possible bias in the measures of cognition. The magnitude of sample composition bias on cognition will depend on both the amount of non-response and the extent to which non-respondents differ from respondents in their level of cognitive function. In fact, bias, defined as the difference between cognition among the interviewed and the cognition of all survivors, will be equal to the product of the non-response rate and the difference in mean cognition between responders and non-responders. There are numerous ways to look for bias in longitudinal data. Here, we measure it by comparing the baseline values of cognition between respondents and all potential respondents. This measure asks the question: were those who stay in the study and respond in future waves different at baseline from the average?

Table 3 shows our measure of bias in cognition at each follow-up wave of the two studies. The bias measure is the difference in baseline cognition between responders at wave t and all survivors to wave t. A zero means there is no bias. A positive number indicates that responders were somewhat higher on cognition at baseline, and a negative number would indicate that they were lower. Our hypothesis is that respondents will be selected from among those with better cognition, and our tests of statistical significance therefore are based on a one-tailed test of the null hypothesis of no difference against that alternative.

Table 3. Bias in mean level of cognition, by wave of follow-up, HRS cohorts compared with ELSA (using baseline weights).

Entry Cohort	Follow-up Wave							
	1	2	3	4	5	6	7	8
HRS 92	0.05	0.07	0.06	0.06	0.05	0.05	0.04	0.04
AHEAD	0.03	0.02	0.01	0.00	0.04	0.06	0.08	
CODA	-0.01	0.04	0.00	0.03	0.03			
WB	0.07	0.09	0.06	0.06	0.06			
EBB	0.08	0.08						
All HRS	**0.05**	**0.06**	0.04	0.04	0.05	0.06	0.05	0.04
ELSA	**0.20****	**0.21****						
ELSA-HRS	**0.15****	**0.15****						

Notes: Bias is measured as the difference in mean baseline cognition score between responders in a follow-up wave and all survivors to the date of the follow-up wave, using baseline sampling weights. For abbreviations, see notes to Table 1. Estimates in **bold** are greater than zero at P<.05 significance level in a one-tailed test; estimates with * are significant at p<.01, and ** at p<.001.

For the individual HRS cohorts, the magnitude of bias is not statistically significantly different from zero. However, in general a small bias emerges immediately and then does not grow very much over time, even though response rates continue to fall. When all the HRS cohorts are combined, the magnitude of bias is marginally significant statistically at the first and second waves only. In ELSA, by contrast, the magnitude of bias is nearly four times greater than in HRS, and it is statistically significantly different from both zero and the HRS bias level at p<.001.

ELSA has approximately twice as much non-response as HRS (Table 2) but three to four times as much bias in cognition measurement due to non-response (Table 3). Since bias is equal to the product of non-response and the difference between responders and non-responders, that suggests that non-response in ELSA is more selective of low-cognition respondents than is non-response in HRS.

One possible solution to biased non-response is to adjust sampling weights for characteristics related to non-response. HRS sampling weights adjust for only a small number of demographic characteristics (age, marital status, race and ethnicity and cohort of entry). ELSA weights are based on a wider range of variables, but neither study explicitly models non-response propensity on cognitive ability. Details on the calculation of sample weights in the HRS (Heeringa and Connor 1995) and ELSA (Taylor et al 2007) can be found online. The conventional sampling weights for HRS are for the community-dwelling population only. To properly evaluate the effects of attrition and non-response it is important to include nursing home residents for which the HRS has now begun to issue sampling weights. Table 4 shows the impact of using sampling weights. In both countries the use of current-wave sampling weights for respondents helps to reduce the bias in cognition from non-response, but it does not fully eliminate the bias. At wave 2 the bias in ELSA continues to be greater than both zero and the HRS bias at significance levels of p<.001.

Table 4. Bias in mean level of cognition, by wave of follow-up, HRS cohorts compared with ELSA, using current wave sampling weights.

Entry Cohort	Follow-up Wave							
	1	2	3	4	5	6	7	8
HRS 92	0.04	0.04	0.03	0.02	0.01	0.05	0.03	0.03
AHEAD	-0.04	0.03	-0.01	0.04	0.05	0.11	0.12	
CODA	-0.05	0.02	-0.03	0.04	0.04			
WB	0.04	0.11	0.03	0.04	0.07			
EBB	0.08	0.07						
All HRS	0.01	**0.04**	0.02	0.03	0.03	0.06	0.04	0.03
ELSA	**0.11***	**0.17****						
ELSA-HRS	**0.10***	**0.13****						

*Notes. Bias is measured as the difference in baseline cognition score between responders in a follow-up wave and all survivors to the date of the follow-up wave. For abbreviations see notes to Table 1. Estimates in **bold** are greater than zero at P<.05 significance level in a one-tailed test; estimates with * are significant at p<.01, and ** at p<.001.*

The role of proxy interviews

The HRS makes extensive use of proxy interviews—nearly 10% of interviews in most waves are completed by proxies because respondents are either unable or unwilling to do the interview. By contrast, fewer than 3% of ELSA interviews in waves 2 and 3 were completed by proxies. It is very likely that the use of proxies in HRS contributes to the overall higher response rates shown in Table 1 above. In order to determine whether the use of proxies might also contribute to the lower levels of bias in HRS, we repeat the comparisons of Tables 1 and 2 above, with proxy interviews treated as non-response, instead of as completed interviews. Tables 5 and 6 do the same comparisons as in Tables 2 and 3, except that we treat HRS proxy interviews as if they were non-responders, and only self-interviews count as response.

Without proxies, HRS response rates as shown in Table 5 are considerably lower than rates in Table 2, especially for the oldest cohort (AHEAD). For all cohorts, the response rate is about eight percentage points lower. ELSA uses very few proxies, so its self-interview response rates are similar to the overall rates in Table 2. Nevertheless, HRS response rates are still 4.8 percentage points higher than ELSA at wave 2, and 9.2 percentage points higher at wave 3, and those differences remain strongly significant statistically. The use of proxies, then, is not the only factor in higher HRS response rates.

Table 5. Self-interview response rate of surviving members of the age-eligible baseline cohort, by wave of follow-up, HRS cohorts compared with ELSA

Entry Cohort	Baseline Year	Follow-up Wave							
		1	2	3	4	5	6	7	8
HRS 92	1992	86.4	83.7	80.1	76.9	76.5	76.2	76.9	76.3
AHEAD	1993	81.2	77.5	74.7	71.9	71.1	71.9	69.9	
CODA	1998	87.1	83.6	83.1	83.4	81.9			
WB	1998	85.6	83.9	82.8	83.1	82.5			
EBB	2004	86.2	84.0						
All HRS		84.9	82.2	79.3	77.2	76.7	75.3	75.6	
ELSA	2002	79.8	71.9						
ELSA-HRS		5.1**	10.3**						

*Notes. For abbreviations, see Table 1. Self-interview response rates treat interviews with proxies as non-response. For differences between studies, estimates in **bold** are greater than zero at P<.05 significance level in a one-tailed test; estimates with * are significant at p<.01, and ** at p<.001.*

In Table 6, we see that the amount of bias on cognition in HRS would be much greater without proxy interviews, and in most cases statistically significant. Indeed, at the second follow-up, bias would be nearly identical to that in ELSA if it were not for the proxy interviews taken in HRS, and the differences between studies are not significant at either the first or second follow-ups. Thus, while proxies are only a partial explanation for the overall higher response rates in HRS, they appear to be nearly a complete explanation for the much smaller bias in cognition.

Table 6. Bias in mean level of cognition, by wave of follow-up, HRS cohorts excluding proxy interviews compared with ELSA.

Entry Cohort	Follow-up Wave							
	1	2	3	4	5	6	7	8
HRS 92	**0.09***	**0.12****	**0.14****	**0.14****	**0.14****	**0.11****	**0.09***	0.09
AHEAD	**0.25****	**0.35****	**0.41****	**0.40****	**0.41****	**0.33****	**0.39****	
CODA	0.11	**0.21***	**0.19***	**0.18**	**0.15**			
WB	0.12	**0.16***	0.10	0.11	0.11			
EBB	0.09	0.10						
All HRS	**0.14****	**0.19****	**0.21****	**0.20****	**0.19****	**0.16****	**0.15****	0.09
ELSA	**0.20****	**0.21****						
ELSA-HRS	0.06	0.02						

*Notes. Bias is measured as the difference in baseline cognition score between responders (excluding proxies) in a follow-up wave and all survivors to the date of the follow-up wave. For abbreviations see notes to Table 1. Estimates in **bold** are greater than zero at P<.05 significance level in a one-tailed test; estimates with * are significant at p<.01, and ** at p<.001.*

We can directly assess the importance of proxies at mitigating bias by looking at non-response rates by level of cognition in the prior wave. This is done in Figure 1. In ELSA, non-response is very high among those who scored poorly on the cognitive measure, and declines rapidly to about the midpoint of the range of cognition. High cognitive function is associated with a somewhat higher response compared with average function, but average and high cognitive groups have much better response than those with low cognitive function. In HRS there is virtually no association of response with prior-wave cognitive function, particularly at the low end of cognition. However, if we again treat proxy interviews as if they were non-respondents, we see an association of cognition with non-response that very much mirrors the ELSA pattern, albeit at somewhat higher overall response rates.

Figure 1. Non-response as a function of prior wave cognition, ELSA, HRS, and HRS excluding proxies
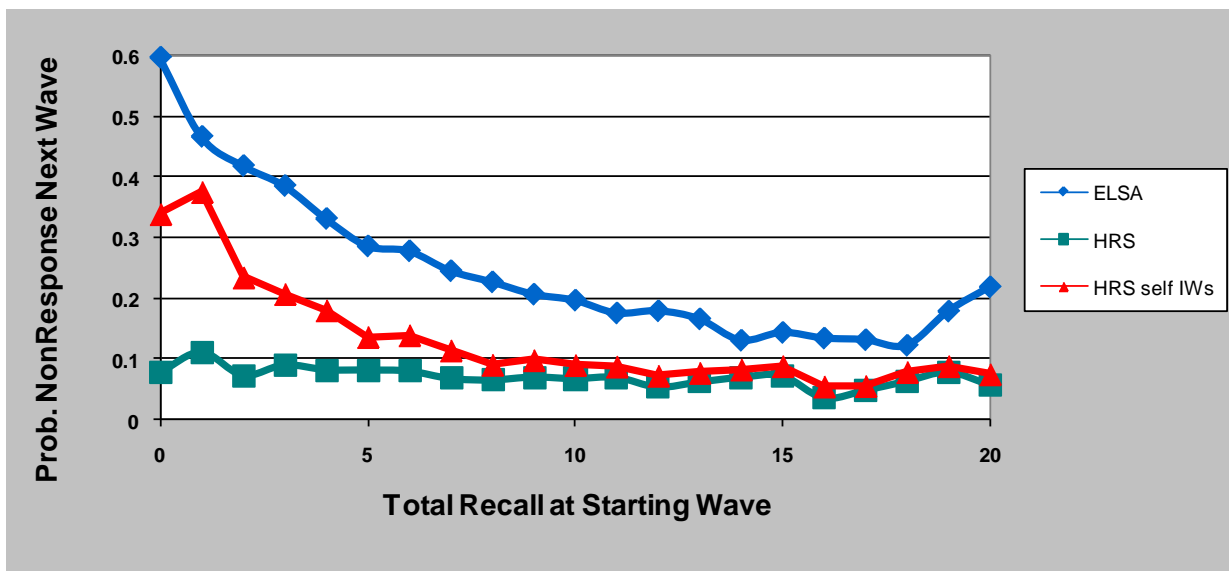
Using Medicare records to study post-attrition outcomes

Although the ability to study the association of non-response with characteristics measured in previous waves is a notable strength of longitudinal studies compared with new cross-sections, there is still the question of whether events occurring after the last interview taken with a respondent (and therefore unobserved in the panel survey) might have influenced subsequent non-response. The HRS now has the ability to study such events through its administrative linkage to Medicare records. The HRS asks everyone who reports coverage by Medicare, which is available to nearly everyone from the age of 65 on, to consent to a linkage to Medicare. Once given, the consent applies to all past and future years of Medicare data unless revoked. When someone declines to do an interview, uses a proxy informant, or asks never to be interviewed again, the consent to link to Medicare records they gave during a previous interview remains valid. This allows us to continue to observe health events for some participants after the date of their last interview.

From the Medicare records we determine the earliest date of a claim bearing a dementia-related diagnosis. For every survey year, we construct a variable indicating whether the person had ever had a dementia diagnosis in his or her Medicare records in that year or any prior year. To allow enough years of Medicare observation to establish a diagnosis, we limit the analysis to persons 70 years of age and older in years 1998 and after.

Table 7 shows the number of surviving participants in each year, the percentage of those who gave an interview, the percentages of interviewed and non-interviewed with Medicare

linkage (excluding those with managed care participation in the preceding five years), the percentage with dementia diagnosis among those with linkage by respondent type, and the estimate of bias in the percent with dementia from relying only on the interviewed sample. The crucial comparison is the rate of dementia diagnosis among respondents as compared with non-respondents. From 1998 through 2002 they are very close, with virtually no bias in dementia rates due to non-response. At the bottom of the table we show the combined figures for 2000 and 2002. Those years provided the sample frame for the ADAMS study of dementia, which drew a stratified sample from respondents in those years. There was no differential in those years between Medicare diagnosis rates among HRS respondents and those among non-respondents. In 2004 and 2006 a gap appears on the order of a three to five percentage point difference, giving rise to what appears to be a small bias toward better cognition among respondents.

Table 7. Medicare diagnosis of dementia among HRS respondents and surviving non-respondents aged 70 and older by survey year.

Year	Number of Eligibles 70+	Interview rate	CMS-FFS linkage rate		Dementia diagnosis rate		Bias
			Among the interviewed	Among Non-respondents	Among the interviewed	Among Non-respondents	
1998	8435	91.8%	70.4%	46.1%	12.4%	13.9%	-0.1%
2000	8393	89.5%	67.6%	45.7%	13.6%	12.4%	0.1%
2002	8759	86.9%	66.9%	43.6%	14.6%	15.2%	0.1%
2004	9122	84.4%	66.4%	39.0%	15.7%	18.7%	0.5%
2006	9558	82.8%	64.2%	31.7%	16.0%	21.5%	0.9%
ADAMS (2000-02)	17152	88.1%	67.3%	44.5%	14.1%	13.9%	0.0%

Notes. The interview rate is the number of interviews with persons aged 70 and older divided by the total number of survivors aged 70 and older in that year. The CMS-FFS linkage rate is the proportion of persons successfully matched to CMS claims records with no managed care participation in the preceding five years. Dementia diagnosis is based on ICD-9 diagnoses recorded in the Medicare claims data. Bias is equal to the non-response rate (one minus column 3) times the difference between interviewed and non-respondents (column 6 minus column 7). The last row of the table combines years 2000 and 2002, from which the ADAMS dementia sub-study sample was drawn.

To show again the importance of proxy interviewing, Table 8 compares claims-based dementia diagnosis rates for self-interviews, proxy interviews, and the non-responders, and computes a hypothetical bias estimate by treating proxy interviews as non-responders. Without the proxy interviews, the bias from using self-interviews only is about three times as large as what was shown in Table 8 when proxies were included with the interviews.

Table 8. Dementia diagnosis rates by type of interview, and bias if proxy interviews treated as non-responders.

<div align="center">

Dementia diagnosis rate

</div>

Year	Self-interviews	Proxy interviews	Non-respondents	Bias if proxies treated as non-responders
1998	11.5%	27.0%	13.9%	-1.0%
2000	12.8%	25.9%	12.4%	-0.7%
2002	13.7%	29.8%	15.2%	-1.1%
2004	14.7%	37.2%	18.7%	-1.6%
2006	15.2%	45.6%	21.5%	-2.2%

Notes. Dementia diagnosis is based on ICD-9 diagnoses recorded in the Medicare claims data. Bias is calculated treating proxy interviews as non-responders.

The small bias found in Table 8 for 2004 and 2006 may be more apparent than real. The inference of bias assumes that respondents and non-respondents are similar in all other respects. In fact, they differ in age, which is an important determinant of dementia rates, and has a modest correlation with response rates, and so is a confounder of the relationship of dementia to subsequent non-response. Comparing respondents to non-respondents, the non-respondents with linkage were about a year and a half older than respondents with linkage. Thus, the unadjusted comparison in Table 8 is likely to overstate the difference in age-adjusted rates of dementia between respondents and non-respondents.

We show the effect of controlling for age by means of logistic regressions in Table 9. In 2000 and 2002, the odds ratio for giving an interview was .88 for those with a dementia diagnosis compared to those without, and was not statistically significantly different from 1.0 (no effect).   When age was included in the model, the OR for dementia became exactly 1.0.   In 2004 and 2006, the OR for an interview with someone with a dementia diagnosis was .75 and this was statistically significant, meaning the unadjusted effect of a dementia diagnosis was to lower the odds of giving another interview. When age was included, the OR rose to .95 and was no longer statistically significant.   Thus, the relatively small apparent bias in the most recent waves of HRS is largely an artifact of the age composition of the groups used to make that inference. Other confounders may also be at work, but controlling for age alone is sufficient to explain the small difference in interview rates between those with and without a diagnosis of dementia.

Table 9. Logistic regression of probability of giving an interview on Medicare diagnosis of dementia, with and without controls for age, 2000/02 and 2004/06 (persons 70 and older)

	2000-02		2004-06	
	Model 1	Model 2	Model 3	Model 4
dementia	0.881	1.002	**0.750****	0.952
(z-stat)	(1.40)	(.02)	(3.88)	(0.62)
age		**0.976****		**0.960****
(z-stat)		(4.92)		(9.63)

*Notes. The sample is limited to persons 70 and older who were linked to Medicare claims (excluding managed care participants). The dependent variable is whether or not a surviving individual gave an interview. The predictor variable of dementia indicates whether there was a diagnosis of dementia indicated in the claims prior to the year of interview. Reported coefficients are odds ratios, and z-statistics are in parentheses. Estimates in **bold** are greater than zero at P<.05 significance level in a one-tailed test; estimates with * are significant at p<.01, and ** at p<.001.*

Discussion

We use several different analytical approaches to assess the extent of sample composition bias on cognitive function due to attrition and non-response in longitudinal surveys. We find that the use of proxy interviews in the HRS essentially eliminates such bias, and demonstrate the magnitude and importance by comparison with ELSA, in which proxies are used much less frequently and in which attrition bias in cognition is much greater.

It is possible that the choice of cognition measure we used to assess bias could affect the findings. Other studies that have demonstrated a selective non-response bias related to cognition have used the Mini-Mental State Examination (MMSE) as their measure of cognitive performance (Brayne et al 1999; Anstey and Luszcz 2002; Matthews and Chatfield 2004). The MMSE tests some other domains of function besides the episodic memory domain used here. However, independent analyses examining the relationship between cognitive functioning and attrition in the HRS using other cognitive domains, also found that poor cognitive status increased the likelihood of a proxy interview but did not have a significant effect on the overall response rate (Ziniel 2008).

Using linked Medicare administrative records for the HRS only, we were able to go beyond conventional analytic approaches, that rely on observations at earlier waves to assess attrition bias, by a direct comparison of post-attrition outcomes of attritors to those who stay in observation. Here, too, we find that proxy interviewing, as implemented in the HRS, is both essential and sufficient to eliminate bias.

Proxy interviews are not a perfect substitute for interviewing a respondent because not all the measures obtained from a respondent can be obtained through a proxy. Nevertheless, compared with the failure to fully represent the frail and cognitively impaired in longitudinal studies of ageing, the loss of a few measures is of less importance. The use of proxies should be standard practice in studies that aim to fully represent the range of functioning in older populations.

While longitudinal studies differ in their objectives, sampling strategies, and methodologies, as well as in their levels of non-response over time, the issue of assessing the effect of panel non-response is salient to all of them. Attrition and attrition-related characteristics are often cited as a potential source of bias in panel studies; however, what is clear from these results is that this potential does not affect all studies to the same degree. Analysis of non-response in other longitudinal studies has also shown that while non-response in samples of older persons is not random, non-random non-response does not always produce bias of any consequence (Mihelic and Crimmins 1997). In the case of the HRS, we find that employing the use of proxy respondents ameliorates the potential biasing effect of cognition-related non-response. Indeed, others have noted that it is possible to retain the very old and very frail in a panel study with appropriately designed field methods (Tennstedt et al 1992; Deeg 2002). Insight into respondent-related determinants of response, both at baseline and subsequent follow-ups, should be used to inform sample retention and refusal conversion protocols, to minimize attrition bias in longitudinal studies.

Acknowledgements

References

Anstey KJ and Luszcz MA. (2002) Selective non-response to clinical assessment in the longitudinal study of aging: implications for estimating population levels of cognitive function and dementia. *International Journal of Geriatric Psychiatry,* 17, 704-709.

Brayne CE, Spiegelhalter DJ, Dufouil C, Chi LY, Dening TR, Paykel ES, O'Connor DW, Ahmed A, McGee MA and Huppert FA. (1999) Estimating the true extent of cognitive decline in the old old. *Journal of the American Geriatrics Society,* 47, 1283-88.

Boersma F, Eefsting JA, van den Brink W and van Tilburg W. (1997) Characteristics of non-responders and the impact of non-response on prevalence estimates of dementia. *International Journal of Epidemiology,* 26, 1055-62.

Chatfield MD, Brayne CE and Matthews FE. (2005) A systematic literature review of attrition between waves in longitudinal studies in the elderly shows a consistent pattern of dropout between differing studies. *Journal of Clinical Epidemiology* 58, 13-19.

Cheshire H, Ofstedal MB, Scholes S, and Schroeder M. (2011) A Comparison of Response Rates in the English Longitudinal Study of Ageing and the Health and Retirement Study. *Longitudinal and Life Course Studies* 2, 127-144.

Deeg DJH. (2002) Attrition in longitudinal population studies: does it affect the generalizability of the findings? *Journal of Clinical Epidemiology* 55, 213-15.

Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, Hall K, Hasegawa K, Hendrie H, Huang Y, Jorm A, Mathers C, Menezes PR, Rimmer E and Scazufca M. (2005) Alzheimer's Disease International. Global prevalence of dementia: a Delphi consensus study. *Lancet* 366, 2112-7.

Heeringa SG and Connor J. (1995). Technical Description of the Health and Retirement Study Sample Design. Online version; originally published as HRS/AHEAD Documentation *Report* DR-002. Survey Research Center at the Institute for Social Research, University of Michigan. http://hrsonline.isr.umich.edu/sitedocs/userg/HRSSAMP.pdf

Iwashyna TJ, Ely EW, Smith D and Langa KM. (2010) Long-term cognitive impairment and functional disability among survivors of severe sepsis. JAMA 304, 1787-1794.

Jorm AF. (1994) A short form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): development and cross-validation. *Psychological Medicine* 24, 145-153.

Jorm AF and Jacomb PA. (1989) The Information Questionnaire on Cognitive Decline in the Elderly (IQCODE): sociodemographic correlates, reliability, validity and some norms. *Psychological Medicine* 19, 1015-1022.

Juster FT and Suzman R. (1995) An overview of the Health and Retirement Study. *Journal of Human Resources* 30, S7–S56.

Kapteyn A, Michaud P, Smith J and van Soest A. (2006) *Effects of attrition and non-response in the Health and Retirement Study*. RAND Working Paper No.WR-407.

Langa KM, Chernew ME, Kabeto MU, Herzog AR, Ofstedal MB, Willis RJ, Wallace RB, Mucha LM, Straus WL and Fendrick AM. (2001) National estimates of the quantity and cost of informal caregiving for the elderly with dementia. *Journal of General Internal Medicine* 16, 770-78.

Langa KM, Larson EB, Wallace RB, Fendrick AM, Foster NL, Kabeto MU, Weir DR, Willis RJ and Herzog AR. (2004) Out-of-pocket health care expenditures among older Americans with dementia. Alzheimer Disease and Associated Disorders 18, 90-98.

Langa KM, Llewellyn DJ, Lang IA, Weir DR, Wallace RB, Kabeto MU and Huppert, FA. (2009) Cognitive health among older adults in the United States and in England. *BMC Geriatrics* 9:23. http://www.biomedcentral.com/1471-2318/9/23

Marmot M, Banks J, Blundell R, Lessof C and Nazroo J. (2003) *Health, wealth, and lifestyles of the older population in England: the 2002 English Longitudinal Study of Ageing*. Institute for Fiscal Studies, London.

Matthews FE, Chatfield M, Freeman C, McCracken C, Brayne C and MRC Cognitive Function and Ageing Study (CFAS). (2004) Attrition and bias in the MRC cognitive function and ageing study: an epidemiological investigation. *BMC Public Health* 4:12. http://www.biomedcentral.com/1471-2458/4/12/

McArdle JJ, Fisher GG and Kadlec KM. (2007) A latent growth curve analysis of age trends in tests of cognitive ability in the elderly U.S. population, 1992 - 2004. *Psychology and Aging* 22, 525-45.

Mihelic AH and Crimmins EM. (1997) Loss to follow-up in a sample of Americans 70 years of age and older: the LSOA 1984–1990. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* 52B, 37-48.

Ofstedal MB, Fisher GG and Herzog AR. (2005) Documentation of cognitive functioning measures in the health and retirement study. HRS/AHEAD Documentation Report DR-006. Survey Research Center at the Institute for Social Research, University of Michigan. http://hrsonline.isr.umich.edu/sitedocs/userg/dr-006.pdf

Plassman BP, Langa KM, Fisher GG, Heeringa SG, Weir DR, Ofstedal MB, Burke JR, Hurd MD, Potter GG, Rodgers WL, Steffens DC, Willis RJ and Wallace RB. (2007) Prevalence of dementia in the United States: the Aging, Demographics, and Memory Study. *Neuroepidemiology* 29, 125-32.

Taylor Jr. DH, Ostbye T, Langa KM, Weir D and Plassman BL. (2009) The accuracy of Medicare claims data as an epidemiological tool. *Journal of Alzheimer's Disease* 17, 807-15.

Taylor R, Conway L, Calderwood L, Lessof C, Cheshire H, Cox K and Scholes S. (2007) Technical report (wave 1): health, wealth and lifestyles of the older population in England: the 2002 English Longitudinal Study of Ageing. National Institute for Social Research. http://www.ifs.org.uk/elsa/publications.php?publication_id=4360

Tennstedt SL, Dettling U and McKinlay JB. (1992) Refusal rates in a longitudinal-study of older-people - implications for field methods. *Journal of Gerontology* 47, S313–18.

Van Beijsterveldt CE, van Boxtel MP, Bosma H, Houx PJ, Buntinx F and Jolles J. (2002). Predictors of attrition in a longitudinal cognitive ageing study: the Maastricht Aging Study (MAAS). *Journal of Clinical Epidemiology* 55, 216-223.

Ziniel SI. (2008) *Cognitive Aging and Survey Measurement*. Dissertation: University of Michigan.