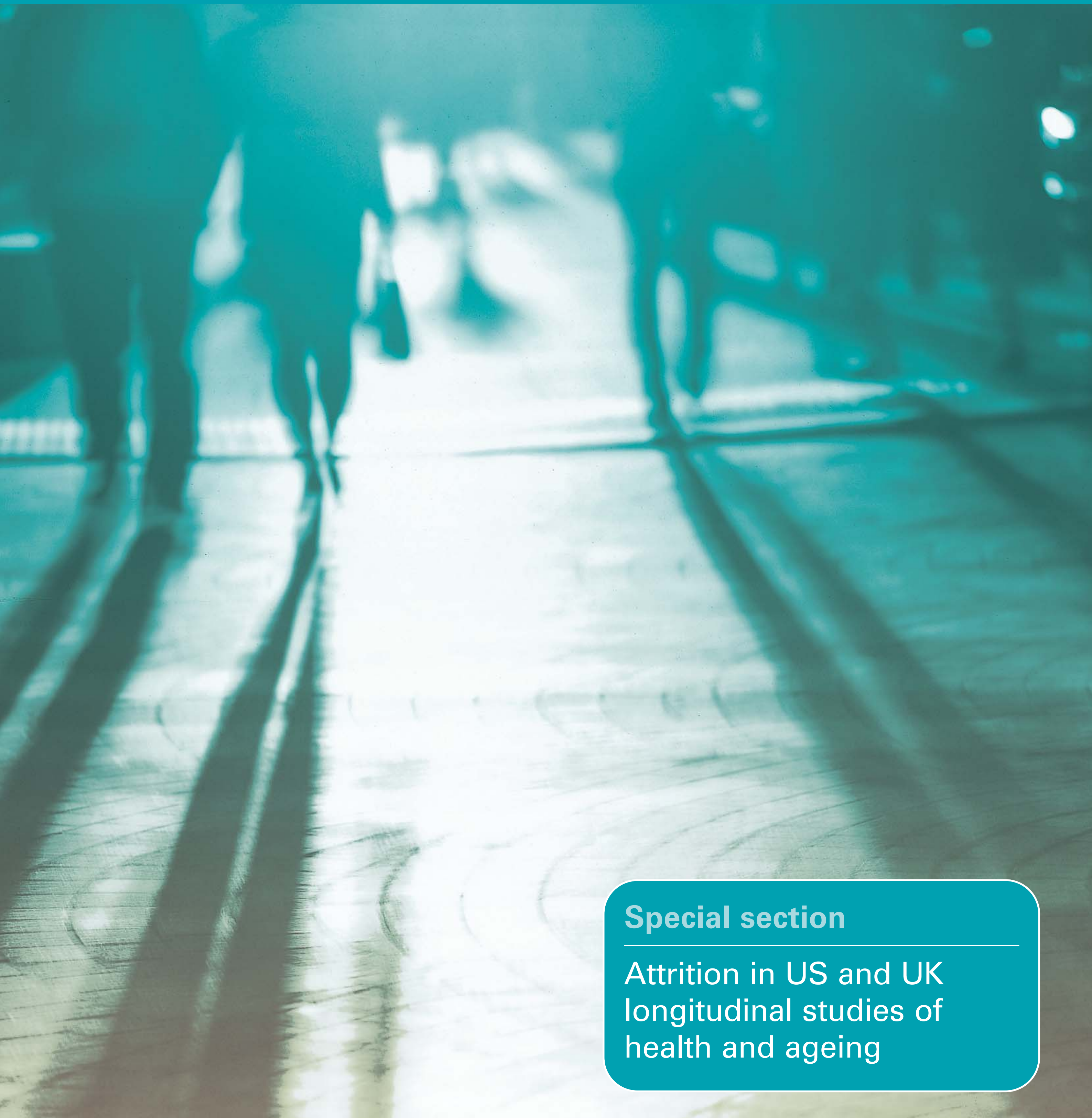
VOLUME 2 | ISSUE 2 | **MAY 2011**

ISSN 1757-9597

Longitudinal and Life Course Studies: International Journal



Special section

Attrition in US and UK longitudinal studies of health and ageing

Published by

SLLS Society for Longitudinal and Life Course Studies

Supported by

Longview Promoting Longitudinal and Life Course Studies

Longitudinal and Life Course Studies

EDITORIAL BOARD

Open Journal System

The ***LLCS*** journal is produced using the Open Journal System, which is part of the Public Knowledge Project. OJS is open source software made freely available to journals worldwide, for the purpose of making open access publishing a viable option for more journals. Already, more than 8,000 journals around the world use this software.

CONTENTS

The Editors are most grateful to the Nuffield Foundation for supporting the journal through its development phase.

Published by the Society for Longitudinal and Life Course Studies
Address: Institute of Education, 20 Bedford Way, London, WC1H 0AL
Email: info@slls.org.uk
http://www.slls.org.uk/

AUTHOR GUIDELINES SUMMARY

Submission of Papers

All papers, written in the English language, should be submitted via the *LLCS* website as a Microsoft Word 2003 file, or in a compatible format. If there is a good reason why this is not possible, authors are requested to contact the Journal Manager **before** submitting the paper. All subsequent processes involving the author are carried out electronically via the website.

Preparation of Texts

Length. The paper should normally be **approximately 5000 words**, with longer papers also accepted up to a maximum of 7000 words. This word count excludes tables, figures and bibliography.

Font and Line Spacing. Please use Calibri (or similar sans serif) font, 12pt, with 1.5 line spacing in all main text, single space in figure or table captions.

Page Layout. All text should be justified to the left hand margin (all margins of at least 2.5cm) with no indents at the start of paragraphs and a line space separating paragraphs. All headings and sub-headings used within the body of the paper should also be left justified. Please do **NOT** use automatic text formatting.

Weblinks. To help our readers to look up cited references or other information available on the web, authors should ensure that all such references are activated.

Ensuring an Anonymous (Blind) Review. Please submit papers with a full detailed title page. Once a paper has been submitted to *LLCS* via the website, it will be 'anonymised' by the removal of all author name(s) and institution names on the title page, and any identifying electronic document properties will also be removed. Authors do not need to remove their name(s) from the main text or references but any reference to their work or themselves should be expressed in the third person.

Abstract. The abstract (no subheads or paragraphs) should be no more than 250 words (not part of the main word count).

Keywords. These should be included just below the author list (minimum 3 maximum 10).

Abbreviations. Words to be abbreviated should be spelt out in full the first time they appear in the text with the abbreviations in brackets. Thereafter the abbreviation should be used.

References. Please use the Harvard Style and refer to examples in the full Guidelines

Authors not complying with these reference guidelines will be asked to make all the necessary alterations themselves, if their paper is accepted for publication.

Notes. As a general rule, supplementary notes should be avoided, but if thought to be essential, they should *not* appear on the page as 'Footnotes', but instead, be included as Endnotes.

Supplementary material. Supplementary material may be uploaded with the submission, and if the paper is published, will be visible to the reader via a link on the RHS of the screen.

Submissions containing graphs, tables, illustrations or mathematics. All graphs, tables and illustrations should be embedded in the submitted text, and have clear, self-explanatory titles and captions. Mathematical expressions should be created in Word 2003 (or a compatible package) with equation editor 3.0, unless the author has good reason to use other software, in which case please contact the Journal Manager. All biological measures should be reported in SI units, as appropriate, followed, in the text, by traditional units in parentheses.

Author Citation. If the paper is accepted for publication, a version of the paper with all authors cited in full on the title page will be used. Only individuals who have contributed substantially to the production of the paper should be included.

Copy Editing. All accepted manuscripts are subject to copy editing, with reference back to author with suggested edits and queries for response. Particular care should be taken by the author to check references.

Proofs. The corresponding author will be asked to view a layout proof of the article on the website and respond with final amendments within 3 days.

(full Author Guidelines at: http://www.journal.longviewuk.com/index.php/llcs/about/submissions#authorGuidelines)

SLLS International Conference 2011

Bielefeld, Germany   Venue: Ravensberger Spinnerei

26th- 28th September 2011

This is the second conference of the *Society for Longitudinal and Life Course Studies (SLLS).* The conference will take place in Bielefeld, the location of the World Congress of Sociology in 1994. The venue is the Ravensberger Spinnerei near the centre of the city. Situated in a small park, the venue consists of several buildings of industrial architecture (a former spinning mill). In addition there is a small movie theatre and a beer garden. Hotels and all other places of interest can be reached by walks within a few minutes. The university building can be reached by tram within 15 minutes.

The conference comprises keynote lectures and parallel group and poster sessions. The topics the conference will address extend across the whole range of longitudinal and life course studies, for example:

> Life course stages from infancy to old age (including physical, psychological, social developmental and ageing processes and functioning); household, and income dynamics; intergenerational transfers and returns to learning; gene-environment interaction; 'mixed', and comparative methods; innovative methodology in design, measurement, analysis and research practice (quantitative and qualitative); longitudinal research and policy.

Keynote addresses by:

Hans-Peter Blossfeld *(National Education Panel Study, University of Bamberg, Germany)*
Jutta Heckhausen *(University of California, USA)*
Marjo-Riitta Jarvelin *(*Dept. of Epidemiology and Public Health, Imperial College, London)

Click HERE for up-to-date information about the conference.

There will be a **discounted conference fee for SLLS members** so if you have not joined yet, please consider doing so soon. Please click here for full details and a membership form to complete on-line.

Longitudinal and Life Course Studies 2011 Volume 2 Issue 2 pp 98 - 100 *SPECIAL SECTION* *ISSN 1757-9597*

Overview: a fishy story - the roles of rods and nets in maintaining representative longitudinal survey samples

Peter Lynn
University of Essex, Institute for Social and Economic Research
plynn@essex.ac.uk

If you talk to a keen angler, he or she will be replete with stories about favourite locations for angling, special bait, or a particularly effective way of casting; in short, those aspects of their craft that increase the probability of a catch. With a modest stretch of the imagination, angling can be thought of as an analogy for gaining survey co-operation. Survey researchers and interviewers are replete with their own stock of practices that they believe increase the probability of a sample member taking part in their survey. Both the angling enthusiast and the survey researcher may be able to draw on some evidence regarding the extent to which particular methods tend to be successful and the mechanisms through which this success is achieved. But both will also be influenced by subjective personal preferences and beliefs, often with little or no basis in evidence.

The angling enthusiast is unlikely, however, to regale too many stories about how to avoid his or her fish escaping once it has been caught. There will be the occasional tale of a particularly wily fish leaping out of the keepnet to freedom, but by and large the problem of how to keep hold of a fish once caught has been nailed. The same is not true of the survey researcher. The problem of how to keep sample members in a longitudinal study is a very real and worrisome one, which is often debated in survey organisations, amongst survey funders, and at scientific workshops and conferences.

Survey researchers have for many years been setting about the task of obtaining evidence regarding how and why a variety of factors might affect their chances of making a catch, i.e. getting a selected sample member to participate in their survey. But only in the last couple of decades has considerable attention been paid to the factors that might affect the probability of keeping hold of that initial catch, i.e. avoiding sample attrition (Campanelli and O'Muircheartaigh 1999, Fitzgerald et al 1998, Laurie et al 1999, Lepkowski and Couper 2002). In the UK, the importance of sample attrition was recognised by a special conference held in May 2004 (Lynn 2006) and by the Survey Design and Measurement Initiative (SDMI) of the Economic and Social Research Council (Lynn and Erens 2010). Two of the six research projects commissioned in 2007 under SDMI were specifically focussed on attrition, while a third addressed non-response more generally but with a sub-project focussed on attrition. The importance of research into ways of dealing with non-response on longitudinal surveys has never been greater.

There are perhaps two main reasons why longitudinal survey researchers are concerned about sample attrition. The first is simply that a reduction in sample size will reduce the precision of estimates. Ultimately, there may be a fear that the sample size reduces to the extent that funders decide that the study is not worth continued support. The second is that non-response and attrition may introduce bias to estimates. This will happen if non-respondents differ systematically from respondents in terms of the key measures (Watson and Wooden 2009).

This special issue of *Longitudinal and Life Course Studies* is devoted to non-response and attrition in longitudinal surveys of ageing. All four contributions use data from the (American) Health and Retirement Study (HRS). Three of the four involve comparisons with the English Longitudinal Study of Aging (ELSA). The authors tackle a number of important issues. Banks et al and Cheshire et al both attempt to identify and explain differences in response rates between the two surveys. They are both successful at identification but less successful at explanation. The latter result is rather inevitable when you only have two observations (the surveys) and several potential explanatory factors. Nevertheless, the exercise is not fruitless, as both papers highlight important differences between the two surveys and suggest promising avenues for further research.

One of the important differences between HRS and ELSA is examined in the paper by Weir et al, namely the ready acceptance by HRS of a proxy interview in cases where a personal interview with the sample member is not possible. Survey strategy regarding the use of proxy interviews is often discussed but under-researched (Moore 1988). There are trade-offs to be made between overall response and personal (as opposed to proxy) response, particularly on a longitudinal survey where accepting a proxy response at one wave may make it harder to obtain a personal response in the future. This might not matter were there not also a trade-off between non-response error and measurement error. Weir et al show that the acceptance of proxies explains part of the response rate difference between the two surveys but, perhaps more importantly, explains all of the differential bias in terms of cognitive ability. This demonstrates the important role that proxy interviews can play and is a good example of a study that moves beyond looking at effects on response rates and examines effects on the bias of substantive measures of interest.

The paper by Michaud et al also is also concerned with bias in substantive measures, specifically estimates from realistic panel data models. Their interest is in the effect of converting previous wave non-respondents to become current wave respondents. Burton et al (2006) showed that refusal conversion attempts are worthwhile for longitudinal surveys, in the sense that the successfully converted sample members often then remain respondents for many subsequent waves. But that work did not assess the impact on substantive estimates. Michaud et al compare estimates using the full HRS data with those that would be obtained if observations subsequent to a wave non-response were excluded. They conclude that panel model estimates of wealth would be substantially biased if these sample members had not been subsequently converted to become respondents.

Collectively, the papers in this special issue should serve to remind longitudinal researchers that error in their substantive estimates can be influenced by many aspects of survey design and implementation. Decisions about survey procedures - such as whether and in what circumstances to accept or seek proxy responses, or whether and how to seek responses from previous wave non-respondents - make a difference. These decisions do not merely make a difference to response rate; they can also affect bias in estimates. Different decisions may have different implications for bias. This point should certainly be of concern to researchers interested in comparing estimates from surveys that have used different procedures, or indeed researchers drawing upon data from multiple waves of a survey that has changed its procedures over time.

Procedures that matter are not only those relating to the use of proxy respondents and to attempting to interview previous wave non-respondents. The papers by Banks et al and Cheshire et al also highlight differences between HRS and ELSA in procedures such as the use of respondent incentives, sample design, field issue policy, data collection mode and between-wave contacts. All these and more could potentially introduce differential non-response error and/or differential measurement error. Researchers have recently begun to look beyond the use of procedures that are standardised across the whole sample, and to examine whether procedures tailored to the circumstances of particular subgroups might be more effective in combating attrition. A couple of recent studies have investigated a number of ways in which this might be done (Fumagalli et al 2010, McGonagle et al 2009). This seems like a promising avenue to pursue.

We still have a long way to go to understand the nature and causes of all errors in longitudinal survey data. We should constantly re-assess and re-evaluate all survey features and procedures. And with respect to non-response error, it is both the features of our fishing rod and the features of our keepnet that are important.

References

Burton J, Lynn P and Laurie H. (2006) The long-term effectiveness of refusal conversion procedures on longitudinal surveys. *Journal of the Royal Statistical Society Series A (Statistics in Society),* 169, 459-478.

Campanelli P and O'Muircheartaigh C. (1999) Interviewers, interviewer continuity, and panel survey nonresponse. *Quality and Quantity,* 33, 59-76.

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

Fumagalli L, Laurie H and Lynn P. (2010) Experiments with methods to reduce attrition in longitudinal surveys. ISER Working Paper 2010-04, Colchester: ISER, University of Essex. Downloaded on 3-10-2010 from www.iser.essex.ac.uk/publications/working-papers/iser/2010-04.pdf.

Laurie H, Smith R and Scott L. (1999) Strategies for reducing nonresponse in a longitudinal panel survey. *Journal of Official Statistics,* 15, 269-282.

Lepkowski J and Couper MP. (2002) Nonresponse in the second wave of longitudinal household surveys. In RM Groves et al. eds. *Survey Nonresponse*, Wiley, New York.

Lynn P. (2006) Attrition and non-response. *Journal of the Royal Statistical Society Series A (Statistics in Society),* 169, 393-394.

Lynn P and Erens B. (2010) The ESRC Survey Resources Network: Opportunities for the advancement of survey methods. *International Journal of Market Research,* 52, 295-302.

McGonagle K, Couper M and Schoeni R. (2009) An experimental text of a strategy to maintain contact with families between waves of a panel study. *Survey Practice*, June 2009. Downloaded on 3-10-2010 from http://surveypractice.org/2009/06/29/panel-contacts.

Moore JC. (1988) Self/proxy response status and survey response quality, a review of the literature. *Journal of Official Statistics,* 4, 155-172.

Watson N and Wooden M. (2009) Identifying factors affecting longitudinal survey response. In P Lynn, ed. *Methodology of Longitudinal Surveys*, Wiley, Chichester.

Attrition and health in ageing studies: evidence from ELSA and HRS

James Banks
University of Manchester, Institute for Fiscal Studies
Alastair Muriel
Institute for Fiscal Studies ali_m@ifs.org.uk
James P. Smith
The RAND Corporation

(Received September 2010 Revised February 2011)

Abstract

This paper investigates the characteristics associated with attrition in the English Longitudinal Study of Ageing (ELSA) and the Health and Retirement Study (HRS), with a particular focus on whether attrition is systematically related to health outcomes and socio-economic status. Our focus is on the attrition of living respondents, not attrition through death, and respondents who died are therefore excluded from our analysis. We have three main results. Firstly, raw attrition is substantially higher in ELSA than in HRS, but whether this is primarily due to differences in the administration of the two surveys, or to other unobserved differences between England and the U.S. is not clear from the available evidence. Second, these differential attrition rates do not change the core conclusions regarding comparisons between the two countries of health and socio-economic status. Finally, very few observable characteristics predict attrition in either study among respondents in their seventies. Among respondents aged 55-64, wealth appears to predict attrition in the U.S. (but not in England), and low education predicts attrition in England (but not the U.S.). Since the more serious attrition problem exists in ELSA, we conduct additional analysis of attrition in that survey. We find that respondents' level of numeracy strongly predicts attrition, but this does not account for the education gradient in attrition in ELSA.

Introduction

In recent years, we have witnessed the development of large longitudinal studies of ageing in many countries around the world. The U.S. Health and Retirement Study (HRS), launched in 1992, provided a template for studies such as the English Longitudinal Study of Ageing (ELSA), the Survey of Health, Ageing and Retirement in Europe (SHARE), the Korean Longitudinal Study of Ageing (KLOSA), the Chinese Health and Retirement Survey (CHARLS), the Longitudinal Ageing Survey in India (LASI) and several more surveys in the field or in development in other countries.

These new ageing studies, which share a comparable template, provide rich sources of information for researchers interested in the dynamics of health, socio-economic status, retirement and wellbeing among ageing populations. Their panel nature allows us to investigate the nature and determinants of within person and within household experiences in retirement and health onsets, and the manner in which these central life domains co-relate. There are now more than twenty-five countries in the world which have initiated such comparable longitudinal ageing studies and more countries are certainly on the way.

An important concern with all panel studies, and particularly those focused on an older population, is the potential for bias caused by individuals non-randomly dropping out of the survey over time. If attrition from a survey is systematically related to outcomes of interest or to variables correlated with these outcomes, then not only will the survey cease to be representative of the population of interest, but estimates of the relationships between different key outcomes, especially in a longitudinal context, may also be biased.

The issue of non-response in longitudinal surveys – both initial non-response and subsequent attrition – has a distinguished history in survey research and statistics (Sudman and Bradburn 1974; Groves and Couper 1998; Little and Rubin 1987). Most of the existing literature has focused on non-ageing panels in the United States, especially during earlier time periods when attrition rates typically were considerably lower. (Becketti et al 1988; Fitzgerald et al 1998; Lillard and Panis 1998; Zapel 1998).

In this paper we present results of an investigation into observable characteristics associated with attrition in ELSA and the HRS, with a particular focus on whether attrition is systematically related to health outcomes and socio-economic status (SES). Investigating the links between health and SES is one of the primary goals of the ELSA and HRS, so attrition correlated with these outcomes is a critical concern.

We begin by looking at raw rates of attrition in the two surveys, and show that panel attrition is a far greater problem in ELSA than in HRS. We consider several possible explanations for ELSA's poorer retention rates, including the greater 'maturity' of HRS (which has been running for ten years longer than ELSA), differences in sampling rules and procedures used in the two surveys, the 'quality' of the two respective survey organizations, and differences in financial incentives offered to respondents. However, the available evidence does not allow us to state definitively whether these explanations, taken together, can account for the disparity in attrition rates between the two surveys.

Having documented raw attrition rates in ELSA and HRS, we then consider the possible bias such attrition could introduce into estimates of disease prevalence derived from the two surveys. In recent papers, we have used data from these surveys to demonstrate that middle aged and older Americans are substantially less healthy than their English counterparts, across a range of important illnesses (Banks et al 2006; Banks, Muriel and Smith 2010). In the same research, we highlighted a substantial socio-economic gradient in health in both countries, a gradient which is present whether education, income or financial wealth is used as a measure of SES. This gradient persists (in both countries) even after controlling for behavioral risk factors. However, if attrition is systematically related to health and/or SES in ELSA or HRS, this attrition may have implications for our estimates of disease prevalence or for the SES gradient in health. Our earlier research focused on two age groups in England and the United States - those aged 55 to 64, and an older group aged 70 to 80, since focusing on reasonably tight age groups helps to ensure that observed health and socio-economic gradients are not driven purely by variation in health or socio-economic status by age. Since one of the aims of this paper is to ascertain the robustness of our earlier results to patterns of attrition, it is those same age groups on which we focus in this paper.

Having established that attrition does not change the core conclusion of this previous work – that Americans have higher rates of disease prevalence at older ages than the English – we go on to a broader investigation of observable characteristics which systematically predict attrition in the two surveys. We find few observable characteristics that predict attrition in either study among those in their seventies. In the group aged 55-64, wealth appears to predict attrition in the U.S. (but not in England), and low education predicts attrition in England (but not the U.S.). Since the more serious attrition problem exists in ELSA, we conduct additional analysis of attrition in that survey. We find that respondents' level of numeracy strongly predicts attrition, but does not account for the education gradient in attrition in ELSA.

Many modern longitudinal surveys have adopted the practice of attempting to convince attritors from prior waves to return as participants in the panel. This retrieval of prior wave attritors may be important in maintaining the long run viability of the panel. Given the rising importance of returnees in panel studies, we present a 'returnee' analysis for both the HRS and ELSA surveys.

This paper is divided into six sections. Section 1 summarizes the data used in our analysis, while the following section describes the most salient patterns of attrition in HRS and ELSA. The third section evaluates some possible reasons for the much higher rate of attrition in ELSA compared to HRS. Section 4 sets out comparative patterns of disease prevalence in the two countries, and explores how these patterns might be altered when we take into account attrition. Section 5 presents models that attempt to identify personal attributes that appear to predict subsequent attrition in both countries. The final section contains our main conclusions.

1. Data

This research presents evidence from two comparably designed ageing studies in the U.S. and England respectively. The studies were purposely designed to be very comparable in terms of population sampling, periodicity, broad content, and in many cases even the specific wording of questions. The Health and Retirement Study (HRS) is a nationally representative sample of the population aged 50 and over in the United States (Juster and Suzman 1995). The initial HRS cohorts were sampled in the early 1990s and subsequent cohorts have been added to establish and maintain full age representation of the post age fifty population. Follow-ups have taken place at two-year intervals since 1992. In this research we use a sample of non-Hispanic Whites in both countries, to ensure greater comparability between the U.S. and English samples. For example, it is well known that African-Americans suffer much worse health outcomes in the U.S. (Hayward et al 2000) and we want our cross-country comparisons to be independent of the quite distinct racial and ethnic composition of the countries.

Questions were asked in each HRS wave about self-reports of general health status, the prevalence and incidence of many chronic conditions, functional status and disability, and medical expenditures. Other related health variables include depression scales, health insurance, smoking and physical exercise.

Data from England come from the English Longitudinal Study of Ageing (ELSA). In ELSA, around 12,000 respondents from three separate years of the Health Survey for England (HSE) survey (those who responded in the years 1998, 1999 and 2001) were recruited to provide a representative sample of the English population aged 50 and over. Detailed employment, income, wage, and asset modules have been fielded and the quality of the baseline data appears to be quite high. The first wave of ELSA was conducted in 2002 and three waves are available for this research.

ELSA content is especially rich in the health domain. Its health module collects information on self-reported general health, specific diagnoses of disease (hypertension, heart disease, diabetes, stroke, chronic lung diseases, asthma, arthritis and osteoporosis, cancer, and emotional and mental illness including depression, memory and cognitive assessment, disability and functioning status (e.g. ADLs and IADLs), health behaviors (smoking, alcohol consumption, and physical activity), and symptoms of heart disease (dizziness and chest pain). While not identical, many of these modules closely parallel those available in HRS. For prior lifetime prevalence, both surveys collect data on individual self-reports of specific diseases of the general form 'Did a doctor ever tell you that you had …'.

Both HRS and ELSA are known to have directly comparable high quality measurement of several dimensions of socio-economic status - importantly, for our purposes, education and income - as well as demographic variables. A unique aspect of both these surveys is that they also contain high-quality wealth modules using a comprehensive and detailed set of questions on the important components of wealth (Juster and Smith 1997).

Finally, both surveys also track the mortality of survey participants, even among those who left the survey in the years before their death. Each survey is matched to the relevant country's National Death Index (which includes information about date and cause of death of all respondents regardless of their participation in subsequent waves of the survey).[1] These matches with the national death indexes are highly successful – over 95% of individuals give permission for their records to be linked and are successfully matched.

This mortality information is important for our analysis, since it allows us to distinguish between those who dropped out of each survey voluntarily (despite still being alive) from those who simply died. In this paper we define individuals as having 'attrited' if they do not respond to the survey, but are not dead according to the mortality data.

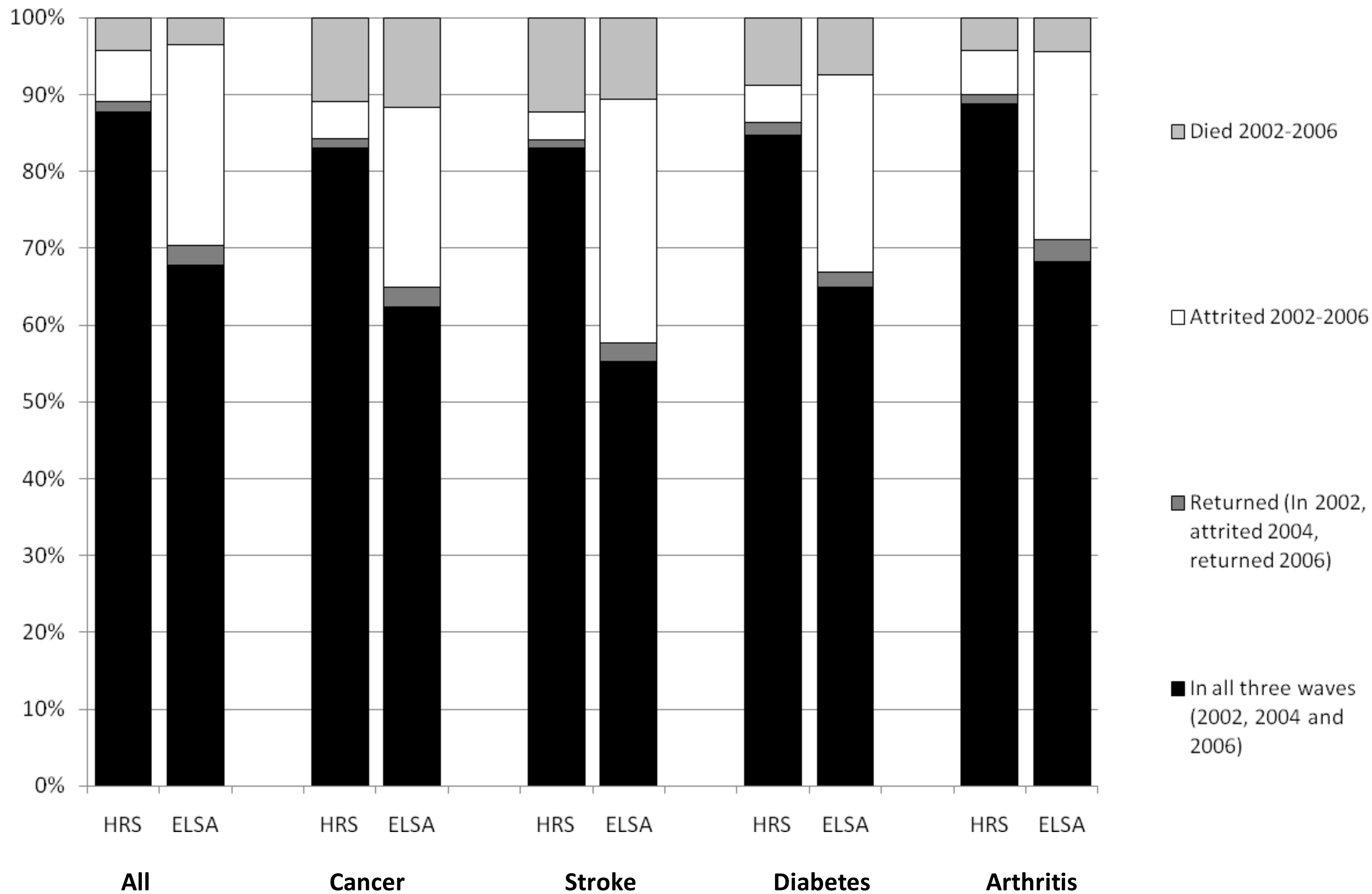
For the purpose of maintaining comparability, in this paper we use the 2002, 2004 and 2006 waves from ELSA as well as from the HRS, since these are the only years for which ELSA data is currently available. We will discuss below the implications of this choice for the conclusions that we derive from the HRS.

2. Patterns of Attrition in ELSA and the HRS

Losses from the sample in panel surveys of the elderly can be traced to two main sources - mortality and sample attrition. Sample attrition is more complicated given that both HRS and ELSA have, as part of their design, an attempt to bring back into the panel respondents who previously attrited. Another complication in comparing these surveys is that by the time of the beginning of the ELSA panel, HRS was a more mature panel in that some respondents had been interviewed for as many as six waves. In this section, we highlight the most salient patterns of sample lost in HRS and ELSA over the same period of time - calendar years 2002 to 2006 - the maximum window allowed in ELSA.

Figure 1 compares rates of attrition and mortality among 55-64 year olds in ELSA and the HRS, between 2002 and 2006 (these years comprise the first and third waves of ELSA). For HRS respondents, the years 2002 and 2006 correspond to different wave numbers depending on which cohort they belonged to, extending from the sixth and eighth wave for the original HRS cohort (51-61 years old in 1992) who would have largely aged out of the 55-64 year old age group in 2002, to the third to fifth wave for those cohorts added to the HRS in 1998. (Further details regarding the HRS's cohort design are given in Section 3, below.)

Figure 1. Retention, attrition and death in ELSA and the HRS, 2002-2006
55-64 year olds, by health condition at baseline (2002)



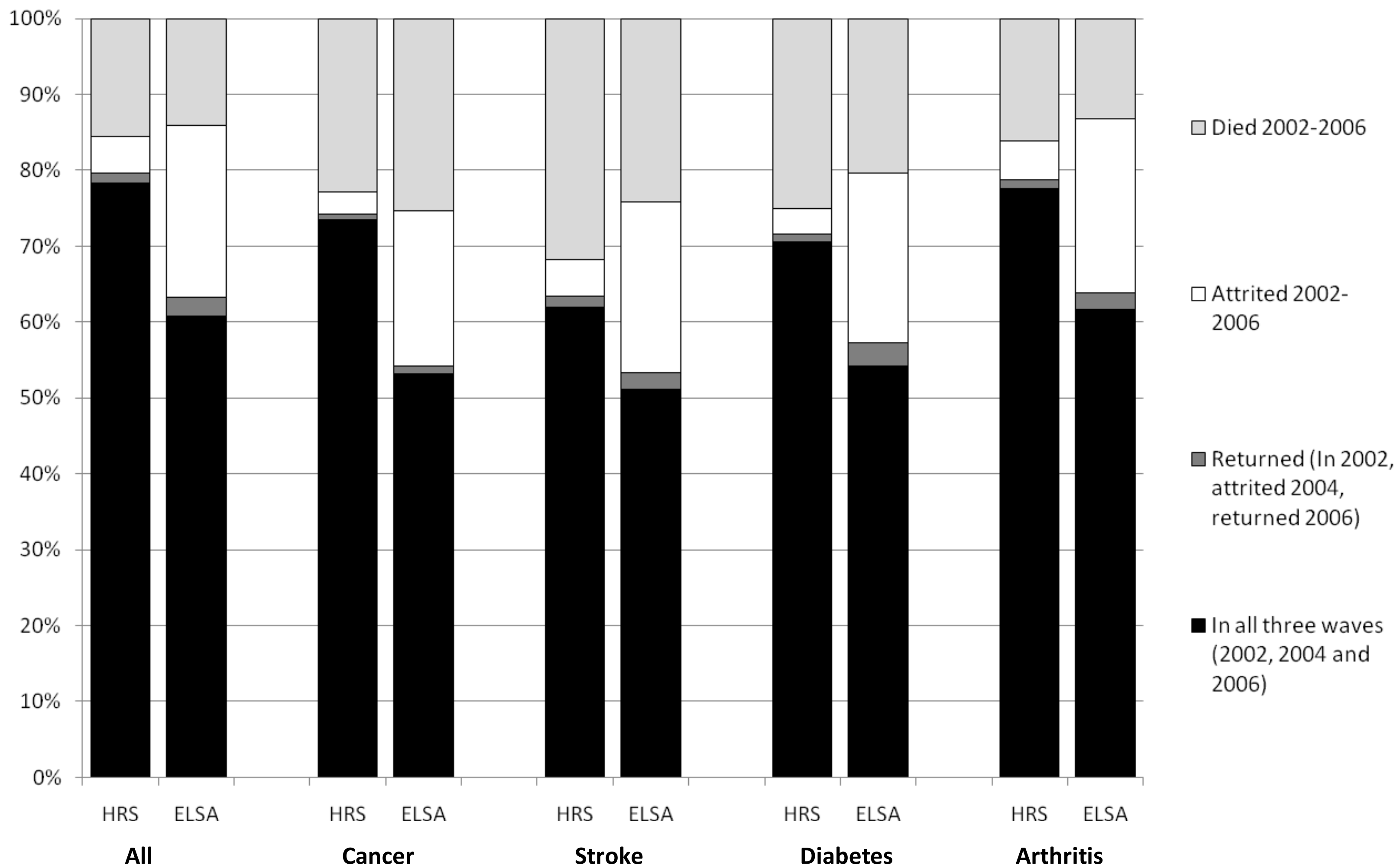
We divide individuals who responded in 2002 into four mutually exclusive categories: (1) those who responded to all three survey waves (2002, 2004 and 2006); (2) those who responded in 2002 but did not respond in 2006 (having dropped out of the survey either in 2004 or in 2006). We refer to these individuals as 'attriters'; (3) those who responded in 2002, did not respond in 2004, but returned to the survey in 2006. We refer to these individuals as 'returners'; (4) those who responded in 2002, but subsequently died. It should be noted that the category whom we label 'attriters' may become 'returners' in future waves if they come back into the survey. Our categories only apply for events that occurred within the survey window 2002-2006.

Looking at the two leftmost bars in Figure 1, which show attrition rates for the 55-64 year old sample in both HRS and ELSA, it is immediately apparent that attrition is substantially higher in ELSA. 88% of HRS respondents in this age group responded to all three waves, while in ELSA this fraction is just 68%. Moreover, this large difference in retention is not accounted for by differential mortality in the two countries, which shows broadly similar overall mortality rates in both countries (around 4% in this age group). It is differential attrition which accounts for the disparity – with ELSA having an attrition rate across two waves nearly four times higher than the rate in HRS (less than 7% of HRS respondents drop out of the survey between 2002 and 2006, compared with more than 26% in ELSA). Our final category, the 'returners' who drop out in 2004 but return in 2006, comprise 2.6% of the ELSA sample, and 1.3% of the HRS sample. As a fraction of those who did attrit between 2002 and 2004, HRS was also more successful in 'recovering' individuals who drop out of the survey. 15% of HRS attriters subsequently return, compared with less than 10% of attriters in ELSA.

The remaining bars in Figure 1 illustrate how rates of retention, attrition and mortality vary according to disease prevalence in 2002, among four diseases for which we have comparable information in the HRS and ELSA – cancer, stroke, diabetes and arthritis. Not surprisingly, among individuals who had been diagnosed with cancer at baseline, retention rates are lower than for the full sample (83% in the HRS, 62% in ELSA), and the same holds true for individuals diagnosed with stroke (83% retention in the HRS, 55% in ELSA). However, much of this lower retention stems from mortality, rather than attrition, with mortality rates of 10-12% in both countries among respondents diagnosed with stroke or cancer. Rates of attrition among cancer sufferers are actually lower than those for the full sample in both countries (5% in the HRS, 23% in ELSA). Among individuals diagnosed with stroke, attrition rates are lower than the rate for the full sample in the HRS (4% of stroke victims attrited), but higher in ELSA (32% attrition among stroke victims). For the two less severe conditions, diabetes and arthritis, retention rates are broadly unchanged from those observed in the full sample (around two thirds retention in ELSA, and four fifths in the HRS).

Figure 2 provides the same breakdown of retention, attrition and mortality for our older age group – individuals aged 70 to 80. Retention rates are once again higher in the HRS, which retained over 78% of 70-80 year olds between 2002 and 2006, compared with less than 63% in ELSA. Again mortality rates are broadly similar in both countries (around 15%), with the difference in retention driven entirely by ELSA's higher attrition – nearly 23%, compared with just 6% in the HRS.

Figure 2. Retention, attrition and death in ELSA and the HRS, 2002-2006
70-80 year olds, by health condition at baseline (2002)



Turning to rates arrayed by baseline disease prevalence, we again find lower retention rates among individuals who had been diagnosed with cancer at baseline – 74% of cancer sufferers are retained in the HRS, and 53% in ELSA. For stroke victims, retention rates are lower still in both surveys – 62% in the HRS, 51% in ELSA. As with the younger age group, these differences in retention are driven by mortality rather than attrition. The mortality rate among cancer sufferers is 23% in the HRS, 25% in ELSA. Among those diagnosed with stroke, the mortality rate is 32% in the HRS, 24% in ELSA. Similarly, diabetes is not associated with significantly higher attrition in this age group, but is associated with higher mortality – 25% of diabetes sufferers in the HRS had died by 2006, as had 20% of those in ELSA. Arthritis, however, is not associated with either higher attrition or mortality.

In summary, we have demonstrated that attrition is a far greater problem in ELSA than in the HRS, with attrition rates nearly four times higher in the English study. This is true in both the 55-64 and 70-80 year old age groups. When we examine how rates of attrition vary by baseline disease prevalence, we find similar patterns in both countries, with significantly lower retention among cancer and stroke victims. However, this difference appears to be driven largely by mortality, rather than different rates of attrition.

It is important to point out that the attrition rates in ELSA are not high by the standard of other ageing panels in Europe. For example, in the SHARE survey of twelve continental European countries, the combined lost to sample from attrition and mortality between the first and second waves alone was forty percent.[2] Since mortality rates are if anything lower in continental Europe, this higher sample lost is due to even greater rates of attrition in SHARE.

Our primary interest in this paper concerns the effects of differential attrition and not lost to sample due to mortality, which has been investigated elsewhere (Attanasio and Hoynes 2000; Banks, Muriel, and Smith 2010). With that objective in mind, in Table 1 we repeat the division of individuals in these age groups into those who remain in all three waves, those who attrit, and those who return – but we now remove individuals who died between 2002 and 2006 from the sample. Table 1 also adds an extra category: ‘healthy’ individuals, who are not suffering from any of the conditions listed in the table, and are also free from heart conditions[3]. Removing deaths from the sample in this way makes clear that, among the 55-64 year old age group, attrition appears to be slightly *lower* among individuals who are suffering from health problems at baseline (with the exception of stroke victims in ELSA, whose attrition rate is 36% compared with 27% for the healthy sample). In the 70-80 year old age group, there is no clear association between baseline health and attrition. In general, attrition among individuals with baseline health problems is of similar magnitude to attrition among the healthy sample.

Table 1. Attrition in HRS and ELSA (2002-2006), by pre-existing health condition (excluding deaths)

		In all three (2002, 2004 and 2006) waves	Attrited 2002-2006	In 2002, attrited 2004, returned 2006	N
55 to 64 year olds	**(%) :**				
All	HRS	91.7	7.0	1.4	4,368
	ELSA	70.3	27.0	2.7	3,482
Healthy*	HRS	90.1	8.6	1,3	1,719
	ELSA	70.8	26.6	2.7	1,502
Cancer	HRS	93.2	5.4	1.5	411
	ELSA	70.7	26.4	2.9	173
Stroke	HRS	94.6	4.2	1.2	165
	ELSA	61.8	35.5	2.6	76
Diabetes	HRS	92.9	5.2	1.9	547
	ELSA	70.2	27.8	2.0	196
Arthritis	HRS	92.8	6.0	1.3	2,221
	ELSA	71.3	25.5	3.1	988
70 to 80 year olds	**(%) :**				
All	HRS	92.9	5.6	1.5	3,482
	ELSA	70.7	26.5	2.8	2,210
Healthy*	HRS	92.9	5.6	1.6	827
	ELSA	71.7	24.8	3.5	537
Cancer	HRS	95.3	3.9	0.9	571
	ELSA	71.3	27.3	1.3	150
Stroke	HRS	90.8	7.0	2.2	272
	ELSA	67.4	29.7	2.9	137
Diabetes	HRS	94.2	4.6	1.3	550
	ELSA	68.1	28.2	3.7	216
Arthritis	HRS	92.6	6.0	1.5	2,278
	ELSA	71.0	26.4	2.6	889

** **Note**. 'Healthy' individuals are those free from all conditions listed in this table and free from heart problems. As explained in endnote 2, heart problems are not included directly in Table 1, as we do not have a perfectly comparable measure of heart complaints between the two surveys.*

3. Explaining the higher attrition rate in ELSA

Why is the attrition rate in ELSA so much higher than in HRS? Numerous factors contribute to a panel survey's retention rate, so it is worth considering potential explanations for ELSA's high rate of attrition compared with HRS. One obvious place to start is the differing 'maturity' of the two panels in 2002. More mature panels may be characterized by lower rates of attrition since the least committed respondents may have long since gone.

Gauging the relative 'maturity' of the ELSA and HRS panels is not entirely straightforward. On the HRS side, in 2002 some respondents (the original birth cohort ages 51-61 in 1992 who would be 61-71 years old in 2002) were in their sixth wave of participation. The AHEAD cohort was initially sampled in 1995 when they were seventy years old or over, so that the 2002 wave was their fifth wave. In 1998, two new cohorts were added – the Children of the Depression Age (CODA-62-69 years old in 1992) and the War Babies cohort (born between 1942 and 1947 and between ages 55-62 in 2002). These two new cohorts were in their third wave of participation in 2002. In summary, HRS respondents in 2002 had previously participated in the survey from anywhere between three to six waves. It is possible, therefore, that individuals with a high propensity to drop out had already left the HRS by 2002.[4]

Measuring the 'maturity' of the ELSA sample is not entirely straightforward either. As explained above, all ELSA respondents were recruited from three prior waves of the Health Survey for England (HSE) so that 2002 actually represented their second wave of participation in a survey with varying years of periodicity depending on the year of their HSE interview (1998, 1999 or 2001). Since the first ELSA wave ultimately achieved a household response rate of 70% of age-eligible households (Marmot et al 2003), the residual 30% non-response might also be considered a form of attrition. However, treating individuals' initial HSE interview as their 'first' wave would also be a misleading basis for comparison with the HRS, since HSE respondents were agreeing to a single interview (the HSE is not a longitudinal survey), while HRS respondents were explicitly signing up to take part in a longitudinal survey. For this reason, we use the first (2002) ELSA wave as our basis comparison, since 2002 is the first year that both English and American respondents were agreeing to take part in longitudinal surveys.

The problem remains, however, that HRS respondents had (on average) been members of the panel for more waves than ELSA respondents by 2002, having initially joined the survey in an earlier calendar year. We can look for evidence of the effect this has on attrition by examining retention rates among new HRS cohorts, who were being interviewed for the first time. While there was no new HRS cohort in 2002, there was a new cohort of 51 to 56 year olds added to the HRS in 2004. One problem with using this new cohort is their relatively young age, since younger working respondents tend to exhibit higher attrition. The attrition rate for this cohort between 2004 and 2006 was just 10.6%. A new cohort was also added to the HRS in 1998 – and attrition rates for this cohort were lower still: just 7.3% to the next wave in 2000, perhaps indicating that attrition rates in surveys in western countries have risen over time. Among ELSA respondents aged 51 to 56, between their first and second waves (2002 to 2004), attrition was 19%. The 'mature survey' explanation, therefore, cannot by itself explain much of the disparity in attrition rates.

Another possible explanation would centre on different levels of mobility in the two countries. A key challenge for any household panel study is simply keeping track of families as they move over time. But mobility at older ages is actually much higher in the United States than it is in England (Banks, Oldfield and Smith 2009), which would argue for higher attrition in the U.S. than in England. This, clearly, cannot explain the higher rates of attrition in ELSA.

Differential 'respondent burden' is another oft-cited reason for non-response (Groves and Couper 1998; Zabel 1998). Given how closely ELSA's questionnaire was modeled on the HRS, this explanation is unlikely – average interview length is almost identical (around one and a half hours) in both surveys, and many of ELSA's questions and modules are directly based on HRS counterparts. Nonetheless, it remains possible that respondents in the two countries simply react differently to (objectively reasonably similar) questionnaires, so we cannot rule out the issue of respondent burden entirely.

A more significant difference between the two surveys, however, is found in the financial incentives (or 'rewards') that are offered to respondents to thank them for taking part. Both ELSA and the HRS offer such rewards, but the HRS offers a considerably larger sum: $100 per person, compared to £10 per person (around $15 at current exchange rates) in ELSA, for the waves that we are considering here. Both theoretical considerations (e.g. Hill and Willis 2001) and experimental evidence (Rodgers 2002) suggest that larger financial incentives may drive improved retention, though as Laurie and Lynn (2009) note, much of our experimental evidence regarding the effectiveness of financial incentives derives from cross-sectional, rather than longitudinal, surveys. Nonetheless, it is certainly possible that ELSA respondents may be under-incentivized compared to HRS respondents, contributing to the different levels of attrition seen in the two surveys.

ELSA and HRS also differ somewhat in their sampling methodology in the treatment of individuals and households. ELSA is a sample of *households*, so that if a household is randomly chosen for interview, *all* age-eligible individuals in that household (everyone aged 50 and over) will be added to the ELSA sample. HRS, in contrast, is a sample of *families*, so that when an individual aged over 50 is selected for interview, their *partner* (if they are part of a couple) will also be sampled for the HRS. But other members of the household will not be added to the HRS sample, regardless of whether or not their age would make them eligible.

Table 2 addresses this issue directly, by investigating attrition rates at the household level (for all respondents in the survey, regardless of their age) between 2002 and 2004, according to the number of respondents in the household. In this Table, we exclude all households in which a death occurred. We divide households into three categories: (1) households which do not attrit at all (no household members leave the survey): (2) households which *partially* attrit (some but not all household members leave the survey) and (3) households which *completely* attrit (all members of the household leave at the same time).

Table 2. Attrition at the Household Level

	Number of respondents in household							
	1		**2**		**3**		**All**	
	ELSA	HRS	ELSA	HRS	ELSA	HRS	ELSA	HRS
No household members attrit	75.2%	92.7%	73.0%	95.1%	43.3%	NA	74.0%	94.9%
Some household members attrit	0.0%	0.0%	8.0%	1.4%	46.7%	NA	4.2%	0.7%
All household members attrit	24.8%	5.3%	19.0%	3.5%	10.0%	NA	21.8%	4.4%
Total	100%	100%	100%	100%	100%	NA	100%	100%
Number of households	3,802	5,850	3,921	5,352	30	0	7,761	11,202

***Note**. Deaths have been excluded from the sample, so do not count as attrition.*

Overall the attrition of *at least one* household member is quite high in ELSA when there are three or more respondents in the household, though such households are also less likely than one- or two-respondent households to see *all* household members attrit. Since this situation involves only thirty households in ELSA, however, it cannot account for differential attrition between the surveys.

The more relevant case is when there are two respondents in the household - typically the wife and husband. Sampling partners was an innovation of both HRS and ELSA and stands in sharp contrast to typical panels such as the Panel Study of Income Dynamics, which rely on a single respondent who answers questions for both partners (Fitzgerald et al 1998). Existing research has shown that the quality of health information reported for the partner is much lower (Weir and Smith 2007; Smith 2007). Is this gain in data quality about the partner offset by greater difficulty in keeping people in the sample when there are two of them?

The data in Table 2 suggest there is no additional attrition loss by making both partners panel members. In both HRS and ELSA, overall attrition in two person households is almost identical to that in one person households. Attrition decisions are certainly correlated between spouses, since if one person attrits the probability that the other partner also attrits is about 70% in both ELSA and HRS. This often occurs when one spouse may deny access to the other in the interviewing process. However, different sampling procedures of households and families in ELSA and HRS design, fail to explain any of the differential attrition between the two surveys.

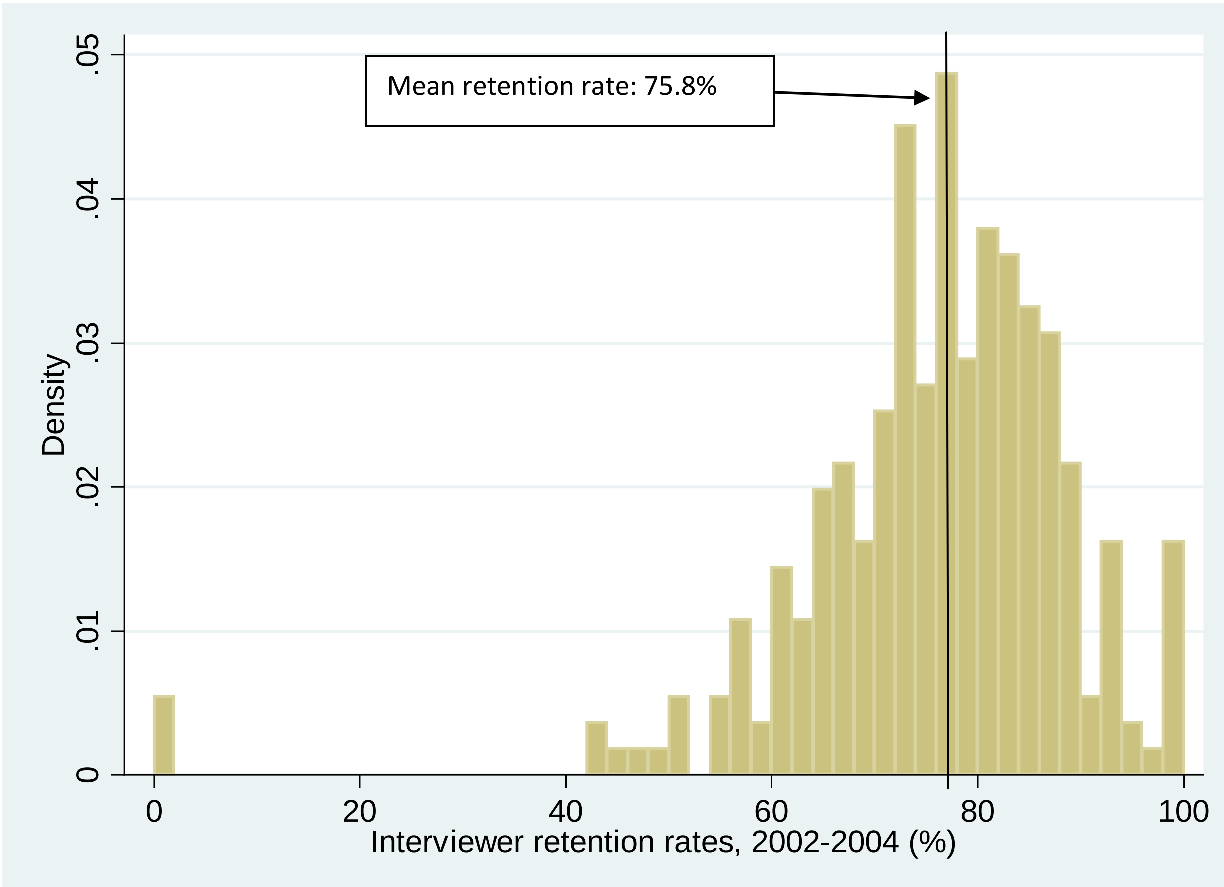
In addition to this difference in sampling, the two surveys also differ in their default mode of interview. All ELSA interviews are conducted face to face, but HRS interviews can take place either by phone or face to face, with the majority taking place by phone for the non-baseline waves. We might speculate that this could reduce attrition in several ways – whether because some respondents may find it more convenient to answer questions by telephone, or because it may be easier to trace households who have moved when all one needs is a telephone number, rather than an address. While attrition rates in the HRS do not vary by mode of interview, it remains possible that the convenience of a telephone interview may help to reduce attrition overall, since individuals who are willing to accept a phone interview may have refused a face to face interview (and so attrited from the survey).

With the exception of the financial incentives for participation and differing interview modes, few of the other structural differences in survey design that we were able to examine appear likely to account for the substantial difference in attrition between these two surveys. We therefore move on to investigate the extent to which other differences in survey implementation could play a role. After all, one hypothesis that has to be considered is that the HRS survey interview team may be better trained, more experienced or otherwise better equipped to retain sample members compared with the ELSA team. We attempt to cast some light on this question by using information on the retention rates of individual interviewers in ELSA.

For this analysis we use ELSA data linking respondent identifiers to the interviewer who administered their 2002 survey questionnaire. For each interviewer, we can therefore observe the fraction of their 2002 respondents who remained in the survey in 2004. For each ELSA respondent we calculate their interviewer's 'leave one out' retention rate (that is, the interviewer's retention rate for all respondents *apart from* the individual we are currently considering). This is a number between zero and one, with zero implying that no other respondents questioned by this interviewer remained in the survey (100% attrition), and one implying that all were retained (0% attrition). We take this as an imperfect but useful indicator of 'interviewer quality'.[5]

Figure 3 shows a histogram of interviewer retention rates (at the interviewer level) between 2002 and 2004. We see that although the distribution is reasonably dense around the mean retention rate of 75.8%, the distribution is quite wide.[6] The bottom 10% of interviewers see less than two thirds of their respondents retained in the survey, while the top 10% see retention of close to 90% and above.

Figure 3. Interviewer effects in ELSA (retention rates)



One way of gauging whether the ‘quality’ of interviewers could account for differential attrition is to eliminate the bottom tail of the ELSA interviewer distribution, and calculate the impact this would have on ELSA’s retention rate. As a first calculation, we compute the ELSA retention rate if the bottom 25% of interviewers had the same mean retention rate as the top 75%. This would improve ELSA’s overall (weighted by number of interviews) retention rate by just three percentage points, from 77.6% to 80.6%. As a more extreme truncation of the interviewer distribution, we compute the ELSA retention rate if the bottom 50% of interviewers had the same mean retention as the top 50%. This would increase ELSA’s overall retention by less than seven percentage points, from 77.6% to 84.3%. In fact, for ELSA to match HRS’s retention rate of 95.5%, we would need to remove the bottom ninety percent of the interviewer distribution, and allocate them the same mean retention rate as the top ten percent. This is an extremely large change in the distribution of interviewer retention, suggesting that differential interviewer quality is not the primary reason for the ELSA’s higher between wave attrition. This should not be surprising since both survey organizations (NatCen in England and ISR in America) are highly respected in their fields.

In summary, while ELSA’s attrition rates are significantly higher than those of the HRS, we do not have sufficient evidence to precisely identify the causes. Mobility, maturity of the survey and respondent burden do not seem promising explanations for the gap. Differing sampling methods, interview modes and especially financial incentives seem likely to play a role but the precise magnitude of their effects are difficult establish in the absence of experimental variation. Existing research on respondent financial incentives would seem to suggest, however, that this is too large an attrition gap for financial incentives alone to explain (Groves and Couper 1998; Rodgers 2002).

An alternative possibility is that conducting panel surveys with high initial response rates and low rates of attrition is simply more difficult in England (and by extension Western European countries, in light of the SHARE experience) than in the US. Even in the United States, initial response rates of new HRS cohorts have been declining and attrition rates have been rising somewhat, indicating that the scientific challenges in conducting high quality panel surveys are becoming more daunting. In contrast, these challenges appear to be much less severe in developing countries where attrition rates appear to be considerably lower (Thomas et al 2001).

4. The impact of attrition on estimates of disease prevalence

We turn next to the impact attrition has on a key outcome of interest – estimates of disease prevalence in the two countries. One of the primary uses of HRS and ELSA involves conducting longitudinal analysis of health status. A concern for both surveys, but particularly ELSA, is the impact that attrition has on key outcomes of interest, such as health and the SES-health gradient. In this section, we examine the effect of attrition on estimates of disease prevalence.

In previous work (Banks et al 2006), we compared the prevalence of a number of diseases (stroke, lung disease, cancer, hypertension, diabetes and heart problems) among middle age adults (aged 55-64 years old) in England and in the United States. We found that Americans were much less healthy than their English counterparts. These differences were large along all points of the socio-economic status distribution, and were present in biological measures of health as well as self-reported disease prevalence.

In a recent extension of this work (Banks, Muriel and Smith 2010), we examined disease prevalence for an older age group (70-80 year olds), and explored patterns in new onsets of disease (‘incidence’) among both 55-64 and 70-80 year olds. Using data from ELSA and HRS, Table 3 summarizes the main results. We find that disease incidence and prevalence are both higher among the Americans in age groups 55-64 and 70-80, indicating that Americans suffer not only from higher past cumulative disease risk (as indicated by their higher disease prevalence), but also experience higher immediate risk of new disease onset or incidence compared to the English.

Table 3. Disease prevalence and incidence in older adults in the US and England

Prevalence in 2002 (%)	Stroke	Lung	Cancer	HBP	Diabetes	Heart	Heart attack	Sample size
Age 55-64								
England	2.28	5.62	5.48	33.40	5.88	8.21	4.05	3,775
US	3.52	8.22	9.57	42.65	12.07	15.50	NA	4,437
Age 70-80								
England	7.17	8.28	7.80	47.67	10.38	20.99	10.01	2,706
US	8.42	12.26	17.92	59.00	17.23	32.06	NA	4,013

Incidence 2002-2006 (%)	Stroke	Lung	Cancer	HBP	Diabetes	Heart	Heart attack	Sample size
Age 55-64								
England	1.70	2.00	2.99	10.17	3.33	2.61	1.85	2,645
US	2.07	3.08	4.26	10.03	6.00	6.25	3.31	3,965
Age 70-80								
England	4.68	2.78	4.80	9.83	4.44	4.80	3.38	1,688
US	5.51	3.89	5.88	8.31	4.66	9.28	5.42	3,214

Notes. *NA-not available*.
Source: England-English Longitudinal Survey of Ageing (ELSA; United States-Health and Retirement Survey (HRS). See Banks, Muriel, and Smith (2010).

Table 4 examines the extent to which the health status of the baseline sample (respondents in 2002) is altered when we restrict the sample in various ways to take account of sample retention. The top row of each panel ("All 2002 respondents") shows disease prevalence for the full 2002 sample – the same numbers shown in Table 3. In the rows below, we show baseline disease prevalence among four subsamples: (1) individuals who responded in 2002 and were still alive in 2007; (2) individuals who responded in 2002 and were dead by 2007; (3) individuals who responded in 2002 and attrited from the survey by 2006 (but did not die); and (4) the balanced sample - individuals who responded to all three waves, in 2002, 2004 and 2006.

Among our first subgroup, those who remained alive to 2007, we see not surprisingly somewhat lower baseline disease prevalence in both surveys (this is true of both the 55-64 and 70-80 age groups). Among those who died between 2002 and 2007, baseline disease prevalence is substantially higher – especially among the most serious conditions (stroke, heart attack, lung disease, cancer) where 2002 prevalence is often more than twice as high as in the whole sample.

Our key concern is the effects of attrition on disease prevalence and incidence. Starting with the 55-64 year old age group in ELSA (the top panel of Table 4), baseline disease prevalence among attriters is almost identical to prevalence in the full sample. For the same age group in the HRS, attriters appear if anything to be slightly healthier than the full sample. Turning to the older age group (the lower panels of Table 4), we again observe very small differences in estimated disease prevalence between the attritor and full sample in either ELSA or HRS.

The final row of each panel of Table 4 displays disease prevalence among the three wave balanced panel (individuals who responded to the 2002, 2004 and 2006 surveys). Among 55-64 year olds, in both ELSA and HRS, disease prevalence in the balanced panel remains similar to that in the full baseline sample, but is slightly lower for all conditions, largely due to mortality bias and not attrition. In this age group, even the 'survivor' bias is small, since mortality is not high. Among 70-80 year olds, this bias is slightly larger – with lower disease prevalence among the balanced panel for almost all conditions, with 'survivor' bias accounting for almost all the difference.

Table 4. Disease prevalence in the US and England (2002) – the impact of attrition

Prevalence in 2002 (%)	Stroke	Lung	Cancer	HBP	Diabetes	Heart	Heart attack	Sample size
Age 55-64 – England								
All 2002 respondents	2.28	5.62	5.48	33.40	5.88	8.21	4.05	3,775
Responded 2002 and -								
Alive in 2007	2.11	5.18	4.91	33.14	5.65	7.69	3.63	3,643
Dead in 2007	6.67	17.78	18.52	39.26	12.59	20.74	14.07	132
Attrited by 2006	2.81	6.73	4.72	34.17	5.93	8.04	3.52	998
Balanced Panel, all waves	1.84	4.55	5.06	32.97	5.57	7.61	3.69	2,548
Age 55-64 – US								
All 2002 respondents	3.52	8.22	9.57	42.65	12.07	15.50	NA	4,437
Responded 2002 and -								
Alive in 2007	3.24	7.52	8.88	41.66	11.37	14.08	NA	4,255
Dead in 2007	7.74	18.06	24.47	52.49	26.04	33.04	NA	189
Attrited by 2006	0.77	6.01	6.86	38.97	8.76	11.26	NA	290
Balanced Panel, all waves	3.43	7.57	8.97	41.78	11.44	14.28	NA	3,965
Prevalence in 2002 (%)	**Stroke**	**Lung**	**Cancer**	**HBP**	**Diabetes**	**Heart**	**Heart attack**	**Sample size**
Age 70-80 – England								
All 2002 respondents	7.17	8.28	7.80	47.67	10.38	20.99	10.01	2,706
Responded 2002 and -								
Alive in 2007	6.23	6.74	6.83	47.39	9.16	19.96	9.54	2,322
Dead in 2007	12.66	16.38	13.15	48.88	15.14	27.05	13.40	384
Attrited by 2006	7.15	6.02	6.83	46.83	9.76	22.44	11.06	634
Balanced Panel, all waves	5.86	7.28	7.03	48.00	9.38	18.94	8.76	1,621
Age 70-80 – US								
All 2002 respondents	8.42	12.26	17.92	59.00	17.23	32.06	NA	4,013
Responded 2002 and -								
Alive in 2007	6.83	9.66	16.33	57.01	15.26	28.94	NA	3.398
Dead in 2007	17.87	23.20	26.85	64.05	27.85	46.88	NA	628
Attrited by 2006	8.76	6.48	10.63	62.24	14.45	37.45	NA	184
Balanced Panel, all waves	6.63	9.87	16.74	56.64	15.30	28.47	NA	3,213

While this analysis reveals that a balanced panel does have somewhat lower disease prevalence than the full sample, this bias is driven by mortality, rather than attrition. In addition, the key result is that, no matter how we restrict the sample, Americans have higher disease prevalence than the English in both age groups. This is true whether we look at the full 2002 sample, only those who remain alive, or at the full three wave balanced panel. The choice of these differential samples does not alter that result.

5. Predictors of attrition

While individuals who drop out of ELSA and the HRS appear to differ little from the full non- mortality sample in terms of their health, a relation could appear in multivariate analysis or there could be systematic attrition, based on socio-economic status (SES) and other baseline characteristics, that one may want to relate to these health outcomes. In this section, we examine these issues by estimating full multivariate models of attrition in HRS and ELSA.

Table 5 contains estimated marginal effects from multivariate probit models with associated z statistics in parenthesis. This model is estimated on a dependent variable ('attrited') which is equal to one if an individual dropped out of the survey between 2002 and 2006, and equal to zero if they remained in the survey (responding to both the 2002 and 2006 waves). To highlight the role of attrition, individuals who died have been removed from the sample. Separate attrition models were estimated for those aged 55-64 and those aged 70-80 in 2002 in each country.

The model includes measures of individuals' socio-economic status (quintiles of baseline wealth[7] and baseline household income and education level[8]) and a dummy variable for labour market status (1 = working). 2002 baseline health status is measured in several ways - a set of dummy variables for the presence of specific diseases at baseline, and separate indicators that a respondent's self-reported health is excellent or very good, good, fair, with the poor response being the left out group. There are also a set of demographic controls for marital status (married, separated, divorced, and widowed, with never married being the excluded class) and housing tenure (1= home owner). Finally, the model includes a full set of single year age dummies within each age interval which are interacted with sex. For ease of exposition, these age/gender effects are not displayed in Table 5.

Table 5. Models of Attrition – ELSA and HRS
(Probits, marginal effects with z statistics in parenthesis below)

Sample	(1) ELSA sample aged 55-64	(2) HRS sample aged 55-64	(3) ELSA sample aged 70-80	(4) HRS sample aged 70-80
VARIABLES	Attrited from sample between 2002 and 2006			
Income				
Income quintile 1	0.015	-0.016	0.046	-0.026
	(0.53)	(1.09)	(1.17)	(1.73)
Income quintile 2	-0.004	-0.028	-0.006	-0.024
	(0.15)	(2.18)*	(0.18)	(1.89)
Income quintile 3	0.001	-0.000	0.010	-0.014
	(0.04)	(0.01)	(0.29)	(1.15)
Income quintile 4	-0.008	0.016	0.019	0.001
	(0.34)	(1.53)	(0.61)	(0.10)

(Table 5 cont'd)

Wealth				
Wealth quintile 1	0.055	-0.016	0.071	0.020
	(1.42)	(1.21)	(1.31)	(0.99)
Wealth quintile 2	0.017	-0.033	0.023	0.001
	(0.61)	(3.03)**	(0.64)	(0.11)
Wealth quintile 3	0.002	-0.027	0.018	-0.006
	(0.06)	(2.68)**	(0.55)	(0.49)
Wealth quintile 4	-0.004	-0.021	0.043	0.006
	(0.17)	(2.25)*	(1.36)	(0.59)
Highest qualification (ELSA only)				
Degree	-0.160	NA	-0.068	NA
	(6.41)**	NA	(1.63)	NA
Below degree	-0.140	NA	-0.053	NA
	(5.85)**	NA	(1.48)	NA
A level	-0.113	NA	-0.091	NA
	(3.72)**	NA	(1.71)	NA
O level	-0.065	NA	-0.030	NA
	(3.12)**	NA	(0.94)	NA
CSE	0.034	NA	-0.078	NA
	(0.93)	NA	(2.03)*	NA
Foreign qual.	-0.080	NA	-0.044	NA
	(2.85)**	NA	(1.36)	NA
Years of education (HRS)				
Ed. 13 to 15 (HRS)	NA	0.006	NA	-0.007
	NA	(0.64)	NA	(0.76)
Ed. 16+ (HRS)	NA	-0.013	NA	-0.002
	NA	(1.34)	NA	(0.25)
In work	0.031	0.004	0.015	-0.000
	(1.69)	(0.50)	(0.36)	(0.01)
Baseline health conditions				
Angina	-0.035	NA	-0.012	NA
	(1.02)	NA	(0.41)	NA
Heart attack	-0.035	NA	0.033	NA
	(0.77)	NA	(0.92)	NA
Heart failure	-0.095	NA	0.139	NA
	(0.88)	NA	(1.23)	NA
Heart prob. (HRS)	NA	-0.003	NA	0.015
	NA	(0.31)	NA	(1.69)

(Table 5 cont'd)

Stroke	0.070	-0.038	-0.001	0.011
	(1.26)	(1.94)	(0.02)	(0.74)
Lung disease	0.066	-0.002	-0.050	-0.014
	(1.85)	(0.18)	(1.36)	(1.16)
Cancer	0.022	-0.008	0.022	-0.018
	(0.61)	(0.64)	(0.58)	(1.91)
High blood press.	0.005	-0.005	-0.006	0.010
	(0.31)	(0.63)	(0.32)	(1.36)
Diabetes	-0.003	-0.017	0.004	-0.013
	(0.10)	(1.53)	(0.13)	(1.35)
Arthritis	-0.035	-0.017	-0.011	0.005
	(1.96)*	(2.29)*	(0.56)	(0.68)
Marital status				
Married	0.022	-0.047	0.020	-0.005
	(0.61)	(1.91)	(0.43)	(0.18)
Separated	-0.052	-0.040	0.119	0.082
	(0.71)	(1.35)	(0.90)	(0.87)
Divorced	-0.018	-0.132	-0.114	0.007
	(0.43)	(0.62)	(2.02)*	(0.21)
Widowed	-0.057	-0.005	-0.065	0.014
	(1.28)	(0.19)	(1.41)	(0.47)
Self-reported health				
Health ex./v. good	-0.036	-0.013	-0.052	-0.008
	(1.00)	(0.66)	(1.27)	(0.48)
Health good	-0.057	0.000	0.027	-0.003
	(1.63)	(0.01)	(0.66)	(0.19)
Health fair	-0.039	0.004	-0.008	0.010
	(1.12)	(0.17)	(0.19)	(0.57)
Health: missing	0.286	-	0.405	-
	(3.04)**	-	(3.47)**	-
Tenure: owner	0.007	-0.032	0.006	0.002
	(0.20)	-(2.19)*	(0.14)	(0.15)
Observations	3431	4255	2189	3395

Notes. *Robust z statistics in parentheses*
** significant at 5 % level, ** significant at 1% level*

Perhaps the most striking result of these probit models is that even when estimated in a multivariate context, health variables - whether through disease prevalence or self-reported health - in either country and in both age groups, do not predict subsequent attrition from the survey. The only exception to that summary is that respondents suffering from arthritis are less likely to attrit in both countries among those 55-64 years old. Our results from the previous section (finding little evidence of a systematic relationship between health and attrition) are apparently robust to the introduction of a standard set of controls for other attributes[9].

Turning next to the variables measuring socio-economic status, we find very different patterns in the two countries. Among 55-64 year olds in ELSA, there is strong evidence that the least educated individuals are more likely to drop out of the survey than their more educated peers. There are no education effects for this age group in HRS. In contrast, the least wealthy respondents in HRS in this age group are the most likely to attrit with no statistically significant income or wealth effects on attrition in ELSA. Among older ELSA and HRS respondents, there appear to be no strong SES correlates of attrition - neither education, income, nor wealth.

In the HRS, there is some evidence that housing tenure predicts attrition among 55-64 year olds, with individuals who own their home slightly less likely to attrit, holding all other attributes constant, possibly reflecting the higher mobility of renters (and consequent increased difficulty in tracing them for future survey waves) compared with owner-occupiers in the U.S. (Banks et al 2009). In ELSA, by contrast, housing tenure does not predict attrition – perhaps reflecting the fact that mobility is low among both renters and owner-occupiers in England (also shown in Banks et al 2009), with a much smaller differential between the two groups than is seen in the U.S.

The strongest predictors of attrition in ELSA actually have nothing to do with personal attributes at all – they are the 'self-reported health missing' dummy variables, indicating that an individual did not answer the self-reported health questions in ELSA's health module[10]. Since refusing to answer questions is likely to indicate that an individual was not wholly committed to the survey, it is perhaps unsurprising that such individuals are less likely to respond to requests for a repeat interview in subsequent waves.

In summary, the only strong indication of SES bias in attrition in ELSA comes from the 55-64 year old age group where it appears that less-educated individuals in this age group are more likely to drop out of the survey. Given ELSA's much higher attrition rate, it is worth investigating the reasons why differential attrition by education might arise.

One possibility is that less-educated respondents simply found the ELSA survey more burdensome to answer than higher-educated respondents did. Having agreed to take part in ELSA's first wave, it is possible that these individuals didn't fully appreciate the demands of the interview and questionnaire. ELSA is a long survey that probes domains of life (and especially the economic domain) that were not addressed in the prior HSE wave. If this was the explanation, we would expect the bulk of attrition of lower-educated respondents to take place between ELSA's first and second waves (2002 to 2004), since such respondents may have had little idea of the survey's contents in 2002 (and so agreed to participate initially), but would have known what to expect from the survey by the second wave in 2004. Between 2004 and 2006, we would therefore anticipate that the education effect would diminish, as all respondents now know what to expect from the survey. This explanation fails our test. We re-estimated models in Table 5 separately for attrition from 2002 to 2004, and attrition from 2004 to 2006. For both waves, the coefficients on the education variables are of similar magnitude and statistical significance[11].

Another possibility is that the education gradient is accounted for by less-educated respondents also having lower levels of numeracy. The ELSA interview involves many questions with numerical answers (notably the income and wealth questions, but also many other sections of the questionnaire), which less numerate respondents may find quantitatively demanding and be less comfortable answering. In our final empirical analysis we examine this possibility directly, as well as investigating the role of other ELSA-specific interview variables on subsequent retention.

The 2002 ELSA questionnaire asked respondents up to five basic questions involving successively more complex numerical calculations. The six possible questions are presented in Appendix 1. Answers to all questions are unprompted (i.e. respondents are not given a menu of possible answers to choose from). Each respondent initially receives questions q2, q3 and q4. If all of these are answered incorrectly the respondent receives question q1 and that is the end of their numeracy module. Otherwise the respondent receives question q5. If the respondent reports a correct answer to any (or all) of questions q3, q4 and q5, they receive the final and most difficult question q6 that requires an understanding of compound interest.

Using these questions we allocate individuals into one of four groups according to their broad numerical ability. This has the advantage of allowing us to choose groups that have some prevalence in the population, since a simple count of correct answers does not take into account the relative difficulty of the questions and may lead to some clusters where there are many observations, with relatively few individuals at the extremes. The precise coding is indicated in Appendix 1 and has been described and analysed in more detail in Banks and Oldfield (2007).

In Table 6, we repeat the probit model of Table 5 (using the same covariates), but with the addition of a number of variables to capture numeracy, interview outcomes other than the subsequent attrition and possible interviewer effects. Only the marginal effects for education and these additional variables are reported, in order to ascertain whether the education effect is diminished when these factors are taken into account.

The results show that numeracy is strongly predictive of attrition, with the two most numerate groups more than 10% less likely to attrit amongst 55-64 year olds, and more than 12% less likely to attrit among the 70-80 year olds. However, the inclusion of numeracy does little to diminish the size or significance of the education effect, suggesting that numeracy is not the principal explanation for the education gradient in attrition in ELSA. Attrition by numerical ability is, however, a serious cause for concern to which we return briefly in our conclusions.

The probit model estimated in Table 6 also includes additional variables relating to the administration of the ELSA interview for each respondent. These are certain procedural factors available in ELSA, which may be ‘early warning signs’ of subsequent attrition. For example, we know whether or not a respondent completed all the elements of the face to face interview or returned a partial interview, and we also know details relating to the return of their ‘self-completion questionnaire’ to the ELSA survey team. This questionnaire is given to all respondents at baseline, but many respondents (particularly those who are single) are left to fill this questionnaire in at their leisure, and return it to ELSA by post[12]. We have included a dummy variable for whether an individual failed to return this questionnaire completely, and another to indicate whether they returned it only after being sent a postal reminder by the ELSA team. As we might expect, failure to return the questionnaire is strongly predictive of subsequent attrition, being associated with a 17% (for 55-64 year olds) or 21% (for 70-80 year olds) increase in attrition. For many ELSA attritors, the decision to leave the survey may have occurred immediately after the baseline interview. Requiring a postal reminder, however, is not predictive of attrition (provided the individual did eventually return their questionnaire).

Another procedural parameter included in Table 6 relates to the success rate of the interviewer who conducted a respondent’s first ELSA interview, the construction of which was described in the previous section. This variable (‘Interviewer retention’ in Table 6) has a large and highly significant association with attrition in both the 55-64 and 70-80 year old age groups – with individuals interviewed by someone who successfully retained many of their other subjects also more likely to remain in the survey themselves. In order to ensure that these interviewer retention rates are not simply capturing unobserved area effects, we experimented with adding regional identifiers, rural and urban dummy variables, and a combination of both[13]. Whilst the area effects were indeed significant predictors of attrition, in all cases the interviewer retention rate remained strongly statistically significant, with a large marginal effect.

Table 6. The effect of numeracy and interview effectiveness - ELSA

Sample	(1) ELSA sample aged 55-64	(2) ELSA sample aged 70-80
VARIABLES	Attrited between 2002 and 2006	
Numeracy indicator (base = group 1)		
Numeracy group 2	-0.043	-0.076
	(1.65)	(2.88)**
Numeracy group 3	-0.094	-0.127
	(3.38)**	(4.18)**
Numeracy group 4	-0.102	-0.146
	(3.15)**	(3.44)**
Interview and self-completion outcomes:		
Partial interview	0.004	-0.035
	(0.05)	(0.37)
Self completion not returned	0.185	0.214
	(5.19)**	(5.17)**
Self completion with reminder	0.039	0.068
	(0.86)	(1.10)
Interviewer retention rate	-0.447	-0.261
	(5.00)**	(2.44)*
Highest qualification level:		
Degree	-0.146	-0.032
	(5.55)**	(0.73)
Below degree	-0.126	-0.018
	(5.09)**	(0.49)
A level	-0.091	-0.068
	(2.87)**	(1.21)
O level	-0.044	-0.006
	(2.04)*	(0.19)
CSE	0.043	-0.062
	(1.16)	(1.57)
Foreign qualifications	-0.061	-0.022
	(2.11)*	(0.65)
F-test on region effects	*0.02*	*0.01*
Observations	*3431*	*2189*

Notes. *Other control variables are as in Table 5, namely: income and wealth quintiles, single year age dummies interacted with sex dummies, dummies for employment and health conditions at baseline, marital status, self-reported health and housing tenure. In addition, a full set of regional dummies was included.*
*Robust z statistics in parentheses; * p < 0.05; ** p<0.01*

Our final line of investigation considers factors that are associated with *return from* attrition. As demonstrated in Figures 1 and 2, a subset of individuals who dropped out of these surveys between 2002 and 2004 subsequently return in 2006, with a return rate slightly higher in HRS than in ELSA. In order to search for attributes significantly correlated with return from attrition, we run a probit model using a sample of respondents who attrited between 2002 and 2004, with a dependent variable equal to one if an individual returned to the sample in 2006. Given the smaller sample size of attritors, we pool the entire sample aged 50 and above in each survey, and run a probit model of return from attrition on the following variables: income and wealth quintiles, education, a quadratic in age, dummy variables for employment and health conditions at baseline, marital status, self-reported health and housing tenure, all defined as in Table 5. None of these variables was statistically significant, with the exception of college-education in the HRS (marginal effect = -0.083, z = 2.19), and being divorced in ELSA (marginal effect = 0.124, z = 2.20). For brevity, the full table of results is not presented (but is available from the authors, on request). These results suggest that it would be difficult for survey agencies to target the potential returnees from the pool of attritors based on their observable attributes in previous waves.

6. Conclusions

In this paper, we investigate the relative importance of sample attrition in two of the most important existing longitudinal studies of ageing - the English Longitudinal Survey of Ageing (ELSA) and the Health and Retirement Study (HRS). While attrition exists in both surveys, it is considerably higher in ELSA than in HRS. We explored several possible reasons for these differences, including some which seem unlikely to account for the gap (different rates of household mobility in the two countries, and different respondent burdens from the questionnaires), and several which may account for some (or all) of the gap, including survey maturity, differing sampling methods and survey administration, and differential financial incentives offered to respondents. Indeed, the large difference in financial incentives offered by the HRS and ELSA (the former offering a reward over 6 times the size of the latter) seems likely to play a significant role in explaining the difference in attrition – though the size of the effect cannot be tested without experimental or quasi-experimental variation which is not present in either survey.

The impact of sample attrition on the parameters of interest is not context free. In our application, we examine the impact of attrition on estimates of disease prevalence in the two countries. We find that sample attrition does not significantly affect conclusions regarding comparisons of disease prevalence, in part because in both univariate and multivariate contexts, attrition does not appear to be related to prior disease prevalence. Indeed, we find few attributes that are predictive of attrition in either survey. Attrition is negatively related to prior wave wealth in the HRS and negatively related to prior wave education and numerical ability in ELSA, suggesting that across these two dimensions, at least in these two older age groups, more care must be exercised in analysing the nature of the SES health-wealth gradient in HRS and the SES health-education gradient in ELSA. Housing tenure (specifically being a renter as opposed to an owner-occupier) predicts attrition in the HRS among individuals aged 55-64, suggesting that the high degree of mobility among renters in the U.S. may pose problems for survey administration. In neither survey do we find any attributes that appear to successfully identify who, among the prior wave attritors, the survey was able to bring back into the fold in subsequent waves.

Acknowledgements

This research was funded by grants from the National Institute of Aging R37-AG025529**.** Additional co-funding from the Economic and Social Research Council through the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS (CPP) is gratefully acknowledged. Excellent programmer assistance was provided by Iva MacLennan and David Rumpel.

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. (2006) Disease and disadvantage in the United States and in England. *Journal of the American Medical Association (JAMA),* 295, 2037-2045.

Banks J and Oldfield Z. (2007) Understanding pensions: cognitive function, numerical ability and retirement saving. *Fiscal Studies,* 28, 143-170.

Banks J, Muriel A and Smith JP. (2010) Disease prevalence, disease incidence, and mortality in the United States and in England. *Demography* (forthcoming).

Banks J, Oldfield Z and Smith JP. (2009) Housing mobility at older ages in England and the United States. Paper presented at the 2009 American Economic Association meetings.

Becketti S, Gould W, Lillard L and Welch F. (1988) The Panel Study of Income Dynamics after fourteen years: an evaluation. *Journal of Labor Economics,* 6, 472-92.

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.

Groves R and Couper M. (1998) *Nonresponse in household interview surveys*. New York, Wiley.

Hausman J and Wise D. (1979) Attrition bias in experimental and panel data: the Gary Income Maintenance Experiment. *Econometrica,* 47,455-74.

Hayward M, Crimmins E, Miles T and Yang Y. (2000) The significance of socioeconomic status in explaining the race gap in chronic health conditions. *American Sociological Review,* 65, 910-930.

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

Juster FT and Smith JP. (1997) Improving the quality of economic data: lessons from the HRS and AHEAD. *Journal of the American Statistical Association,* 92, 1268-78.

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

Laurie H and Lynn P. (2009) The use of respondent incentives on longitudinal surveys. In P Lynn ed. *Methodology of Longitudinal Surveys*. Wiley, West Sussex.

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 R and Rubin D. (1987) *Statistical analysis with missing data*. Wiley, New York.

Smith JP. (2007) The impact of socioeconomic status on health over the life-course. *Journal of Human Resources,* 42, 739-764.

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

Rodgers, WL. (2002) Size of incentive effects in a longitudinal study. Presented at the 57th Annual Meeting of the American Association for Public Opinion Research, St. Pete Beach, Florida.

Sudman S and Bradburn N. (1974) *Response efforts in surveys*. (1974) Aldine, Chicago.

Thomas D, Frankenberg E and Smith JP. (2001) Lost but not forgotten: attrition and follow-up in the Indonesia Family Life Survey. *Journal of Human Resources* 36, 556-592.

Weir D and Smith JP. (2007) Do panel surveys really make people sick? U.S. arthritis trends in the Health and Retirement Study. *Social Science and Medicine* 65, 2263-2267.

Zabel J. (1998) An Analysis of Attrition in the PSID and the Survey of Income and Program Participation. *Journal of Human Resources*, 33, 479-506.

Appendix

Derivation of numeracy classification variables.

Box 1a. Numeracy items in ELSA questionnaire

q1) If you buy a drink for 85 pence and pay with a one pound coin, how much change should you get?

q2) In a sale, a shop is selling all items at half price. Before the sale a sofa costs £300. How much will it cost in the sale?

q3) If the chance of getting a disease is 10 per cent, how many people out of 1,000 would be expect to get the disease?

q4) A second hand car dealer is selling a car for £6,000. This is two-thirds of what it cost new. How much did the car cost new?

q5) If 5 people all have the winning numbers in the lottery and the prize is £2 million, how much will each of them get?

q6) Let's say you have £200 in a savings account. The account earns ten per cent interest per year. How much will you have in the account at the end of two years?

Box 1b. Construction of broad cognitive function categories

Classification	Response to questions	Proportion of sample
Group I	Either: q2, q3, q4 all incorrect Or: q2 correct; q3, q4, q5 all incorrect	16.24%
Group II	At least one of q2, q3, q4, q5 incorrect; q6 incorrect	46.46%
Group III	Either: q2, q3, q4, q5 correct; q6 incorrect Or: At least one of q2, q3, q4 correct; q5 q6 correct	26.08%
Group IV	q2, q3, q4, q5, q6 correct	11.22%

Endnotes

[1] In each country, the linked data (the National Death Index in the US and the National Health Service Central Register Database in England) is the data which, at the population level, is used to compute official life tables.

[2] http://www.share-project.org

[3] Heart problems are not included directly in Table 1, as we do not have a perfectly comparable measure of heart complaints between the two surveys. ELSA asks about a series of specific conditions: angina, congestive heart failure, heart murmur and heart attack. The HRS, in contrast, asks a generic question about heart problems in general.

[4] Even this is not a full summary since new spouses can join the survey at any wave.

[5] We use the word 'imperfect' since occasionally ELSA respondents who prove difficult to contact (or reluctant to respond) are handed to some of the most experienced members of the interview team, so that the 'best' interviewers are often allocated the hardest cases. Since cases are therefore not completely randomly assigned, the attrition rate of an interviewer is therefore not a perfect guide to interviewer quality. This is less of a problem at baseline.

[6] 75.8% is the mean retention rate without weighting to take into account the number of interviews conducted by each interviewer. When we weight by number of interviews, the mean retention rate matches ELSA's overall retention rate of 77.6%.

[7] Wealth quintiles are defined within age groups, and are based on the net total non-pension wealth of the respondent and their spouse (if present).

[8] These variables differ between the models for the U.S. and England, reflecting the different education systems in the two countries.

[9] Nor is this lack of significant baseline health effects due to our choice of ten year age bands – when we re-estimate the model pooling all respondents aged 50 and over (results available upon request), we again find no significant effects of baseline health conditions, with the exception of arthritis.

[10] There were no HRS respondents in this sample with missing self-reported health.

[11] Results available from the authors on request.

[12] For couples who are interviewed simultaneously, however, one member of the couple is asked to fill in the questionnaire while the other undergoes the face to face interview, so that no posting is required.

[13] We thank an anonymous referee for this suggestion.

Longitudinal and Life Course Studies 2011 Volume 2 Issue 2 pp 127- 144 SPECIAL SECTION ISSN 1757-9597

A comparison of response rates in the English Longitudinal Study of Ageing and the Health and Retirement Study

Hayley Cheshire
National Centre for Social Research
ELSAdata@natcen.ac.uk
Mary Beth Ofstedal
University of Michigan
Shaun Scholes
University College, London
Mathis Schroeder
German institute for Economic Research, Berlin

(Received October 2010 Revised February 2011)

Abstract

Survey response rates are an important measure of the quality of a survey; this is true for both longitudinal and cross-sectional surveys. However, the concept of a response rate in the context of a panel survey is more complex than is the case for a cross-sectional survey. There are typically many different response rates that can be calculated for a panel survey, each of which may be relevant for a specific purpose. The main objective of our paper is to document and compare response rates for two long-term panel studies of ageing, the English Longitudinal Study of Ageing (ELSA) and the Health and Retirement Study (HRS) in the United States. To guide our selection and calculation of response rates for the two studies, we use a framework that was developed by Peter Lynn (2005) and present several different types of longitudinal response rates for the two surveys. We discuss similarities and differences in the study designs and protocols and how some of the differences affect comparisons of response rates across the two studies.

Introduction

Response rates are often used to gauge the quality of a survey. They provide a single measure that is taken to reflect the representativeness of the respondents who participated and the overall quality of the survey. To maximize response rates, panel surveys work hard to retain participants over time and to limit the loss of statistical power from attrition out of the study. Some fieldwork strategies, like the use of incentives, are often used to encourage participation amongst specific groups, which helps to address concerns about the possible effect of non-response bias on some measures derived from the survey data.

The use of response rates as a basis for comparison across different studies has some key limitations. Despite recent efforts to develop guidelines (The American Association for Public Opinion Research, 2008) there is a lack of standardisation in the way that survey organizations and researchers use survey outcome data to calculate response rates. Second, particularly in the case of panel surveys, which involve more than one data collection round, there are many different ways that response rates can be calculated. In recognition of this, Lynn (2005) has developed a framework of best practice in the recording of outcomes and the estimation and presentation of response rates for surveys with multiple data collection events. Lynn proposes that no single response rate can summarize the overall level of response in a panel survey and that different response rates are relevant for different analytic and evaluative purposes. The following rates are outlined by Lynn (2005):

Longitudinal response rates

Longitudinal response rates are useful for analysts who make use of multiple data collection rounds for longitudinal analysis at the individual (or micro) level. The "complete" response rate is defined by Lynn as the proportion of sample members who participate in every data collection round, of those who were eligible for all rounds. This response rate gives an indication of the completeness or representativeness of the sample that is used in longitudinal analyses involving all waves. Longitudinal response rates may also be calculated for any subset of (two or more) waves.

Cross-sectional response rates

Cross-sectional response rates are needed for those who restrict their analysis to a single round of data collection and/or to a set of discrete waves. Response rates may either be unconditional (i.e. the proportion of sample members who respond in a given wave of all those who are eligible in that wave) or conditional on prior wave response (i.e. the proportion of sample members who respond in a given wave of those who responded in the immediately prior wave or some other prior wave).

The primary objective of this paper is to use Lynn's framework to document response rates for two large and influential panel studies of ageing, the English Longitudinal Study of Ageing (ELSA) and the Health and Retirement Study (HRS) based in the United States. The ELSA study was designed to be comparable to the HRS from the outset and, as a result, the two studies share many core design features. The samples for both studies are nationally representative of community dwelling, middle-aged and older adults (age 50+ for ELSA and age 51+ for HRS).[i] Both are longitudinal panel surveys and conduct a core interview with the same study participants every two years. The questionnaires for the core interview contain substantial overlap on a wide range of topics including employment history, retirement experiences, plans and expectations, economic status, family and household composition, support transfers, health, disability and use of health services.

This is the first time that Lynn's framework has been applied to these two studies. The paper focuses on variations in sample design and fieldwork protocol employed by each study, while at the same time discussing the utility and limitations of the Lynn framework. The paper does not discuss at any length the implications of response rate differences across studies, and no attempt is made to measure response bias. Both issues warrant further investigation, but go beyond the scope of this paper. The issue of response bias is addressed in several other papers in this Special Issue of the journal.

The paper first provides a description of the sample designs for the ELSA and HRS studies before moving to the eligibility criteria adopted for inclusion in the response calculations. Each type of response rate is then presented separately before leading to a discussion of cross-study differences.

Design of the ELSA and HRS studies

English Longitudinal Study of Ageing

The ELSA sample was designed to represent people aged 50 and over (persons born on or before 29 February 1952), living in private households in England at the time of wave 1. A total of 11,391 interviews were achieved with age-eligible sample members (or core members) at the first wave. The sample was selected from households that had previously responded to the Health Survey for England (HSE). The HSE uses a multi-stage stratified random sampling procedure. Three HSE years, 1998, 1999 and 2001 were selected as the sampling frame for ELSA wave 1. Each of these HSE years had a general population "core" sample that was nationally representative. HSE 1999 also had an ethnic minority boost sample but this was not followed up for ELSA.

The HSE response rates are found to be relatively constant from year to year.[ii]

HSE households were only issued to the field at ELSA wave 1 if they included at least one individual who was age-eligible, and who according to administrative records remained alive and gave permission to be re-contacted in the future. No indication was given to respondents at the time of their HSE interview that they would be specifically approached for the ELSA study. The decision to follow-up participants from HSE resulted in ELSA inheriting a large degree of non-response prior to its first wave. In HSE cooperating households, age-relevant information had been collected in order to establish which households were eligible for ELSA, but this was not available for HSE non-cooperating households as no interview had taken place. As a result, for response rate calculations it was necessary to assume that the same proportion of people in HSE non-cooperating households would have been eligible for ELSA. Further detail on the ELSA sampling design can be found in Appendix A or the ELSA Technical Reports (Taylor et al 2007; Scholes et al 2008; Scholes et al 2009). Details about the HSE are also available from its Technical Reports (Erens and Primatesta 1999; Erens, Primatesta and Prior 2001; Prior et al 2003).

Table 1 illustrates the gap in time between the HSE interview and ELSA wave 1. The HSE interview represented the first time ELSA participants were approached and their HSE data can be linked by analysts to data collected as part of the ELSA study. The HSE interview is therefore treated as ELSA Wave 0 and is included as the first stage in the calculation of response rates. The first wave of ELSA fieldwork started in March 2002. Those sampled from HSE 1998 had the largest gap of four years between their HSE and ELSA interview. Those sampled from HSE 2001 had the smallest gap of one year, and therefore unsurprisingly had the highest household contact rate at ELSA wave 1 compared with HSE 1998 and 1999 (Taylor et al 2007).

Table 1. Timing of HSE and ELSA waves

BIRTH COHORT		**HSE**			**ELSA**		
		1998	1999	2001	2002	2003	2004
1900 – 1952	HSE 1998	*			*		*
1900 – 1952	HSE 1999		*		*		*
1900 – 1952	HSE 2001			*	*		*

Health and Retirement Study

The Health and Retirement Study (HRS) has a nationally representative sample of over 20,000 men and women over the age of 50 in the United States. The study began in 1992 as a longitudinal study of a pre-retirement cohort of individuals born in 1931-1941 and their spouses of any age. This birth cohort is referred to as the original HRS cohort. In 1993 a parallel study, the Study of Asset and Health Dynamics of the Oldest Old (AHEAD), was launched. The AHEAD sample was comprised of a cohort of persons born before 1924 and their spouses of any age. In 1998, the study design was modified to convert the HRS sample from a set of specific cohorts into a steady state sample that represents the community-dwelling U.S. population over age 50. This was achieved by combining the HRS and AHEADcohorts into a single data collection effort, based on a common questionnaire and common field protocols, and adding new cohorts in 1998 to fill in the age range over 50 (the CODA cohort consisting of persons born between 1924 and 1930 and the War Baby (WB) cohort born between 1942 and 1947). The steady state design is maintained by adding a new six-year cohort of persons in their early to mid-50s every six years (2004, 2010, etc).

Sample design

In 1992, a large household screening operation based on an area multi-stage probability sample design was undertaken to identify eligible sample members for the HRS, AHEAD and WB cohorts. A brief screening interview was attempted with approximately 69,500 households across the United States. The screening interview contained a listing of all adult members of the household, their year of birth, and partner status. About 14% of addresses in the sample were found to be unoccupied or non-residential and are considered non-eligible. Of the remaining addresses, screening interviews were completed with over 99% (Heeringa and Conner 1995).

The AHEAD sample was supplemented at the oldest ages (age 80+) with individuals selected from the Medicare Beneficiary list maintained by the then Health Care Financing Administration (HCFA), now the Center for Medicare and Medicaid Studies (CMS) (Heeringa 1995). The Medicare list provided roughly one-half of the sample with the remainder coming from the 1992 household screen. HRS had slightly better success recruiting respondents from the sample obtained through the household screener compared to the Medicare list frame (response rates for the two groups were 82% and 77%, respectively). The CODA sample (first interviewed in 1998) was drawn entirely from the Medicare list sample.

In 2004, a second household screening effort was undertaken with 38,385 households to identify eligible sample members for the Early Baby Boom (EBB) (born 1948-1953) and Middle Baby Boom (MBB) (born 1954-1959) cohorts. The EBB cohort was recruited into the study in the 2004 wave and the MBB cohort was added in 2010. The 2004 household screening interview was very similar to that used in 1992. 13% of households were determined to be vacant or non-residential (non-eligible) and, of the remaining households, a screening interview was completed with 91%.

In response rate calculations for the HRS, the screener response rate is factored into the baseline response rate for each entry cohort. Specifically, the baseline response rate is calculated as the product of the screener response rate and the interview response rate, where the interview response rate is the percentage of known eligible sample members who completed an interview. More detail on how eligibility is defined for each study is provided in a later section.

Minority individuals (Blacks and Hispanics) are oversampled in the HRS at a rate of about 2 to 1. In addition, the HRS, AHEAD and WB cohorts contain an oversample of Florida residents.

Table 2 identifies the years in which core interviews were conducted with each cohort. Each of the cohorts has been followed up at roughly two-year intervals since their introduction into the study. Although there is some variation across waves, since 1998 the fieldwork has generally started in February or March of the designated year and ended in January or February the following year.

Table 2. Data collection years for the HRS, by study cohort

BIRTH COHORT	COHORT	DATA COLLECTION WAVE (1992-2006)														
		92	93	94	95	96	97	98	99	00	01	02	03	04	05	06
1890 – 1923	AHEAD		*		*			*		*		*		*		*
1924 – 1930	CODA							*		*		*		*		*
1931 – 1941	Original HRS	*		*		*		*		*		*		*		*
1942 – 1947	War Baby							*		*		*		*		*
1948 – 1953	Early Baby Boom													*		*

Response rates: conceptualisation, calculation and comparison

Conceptual issues in comparisons of response rates across studies

Table 3 puts the two studies together and shows the years in which each study began and the cycle of data collection through to 2006. Subsequent rounds of data collection have taken place in 2008 and 2010, but those waves are not included in this paper.

Table 3. Data collection years for ELSA and HRS

Year	ELSA	HRS
1992		HRS cohort (Wave 1)
1993		AHEAD cohort (Wave 1)
1994		HRS cohort (Wave 2)
1995		AHEAD cohort (Wave 2)
1996		HRS cohort (Wave 3)
1997		
1998	Original cohort (Wave 0) (HSE Interview Years)	HRS cohort (Wave 4) AHEAD cohort (Wave 3) CODA & WB cohorts (Wave 1)
1999		
2000		HRS cohort (Wave 5) AHEAD cohort (Wave 4) CODA & WB cohorts (Wave 2)
2001	Original cohort (Wave 0) (HSE Interview Year)	
2002	Original cohort (Wave 1)	HRS cohort (Wave 6) AHEAD cohort (Wave 5) CODA & WB cohorts (Wave 3)
2003		
2004	Original cohort (Wave 2)	HRS cohort (Wave 7) AHEAD cohort (Wave 6) CODA & WB cohorts (Wave 4) EBB cohort (Wave 1)
2005		
2006	Original cohort (Wave 3) Refresher cohort (Wave 1)	HRS cohort (Wave 8) AHEAD cohort (Wave 7) CODA & WB cohorts (Wave 5) EBB cohort (Wave 2)

As illustrated by Table 3, the variations in survey design and fieldwork timing adopted by ELSA and HRS present difficulties around how best to calculate and compare response rates, and there are several different approaches that could be taken.

One option for response rate comparison would be to focus on a calendar year (or set of years) in which interviews were conducted in both studies, (i.e. 2002, 2004 or 2006) and calculate response rates based on the group of respondents who were interviewed and eligible for interview in that year. Analysts might opt to use this approach in order to compare outcomes from each study at the same point in time. However, comparisons of response rates in this way are problematic, as the HRS is at a very different stage of its lifecycle than ELSA and participants have varying levels of study experience.

An alternative, and the one adopted in this paper, is to take a wave by wave approach, for example comparing the first wave of ELSA with the first wave of HRS, and repeating this for subsequent waves. It is then possible to track the success of each study in obtaining cooperation from respondents in the baseline wave, as well as in maintaining the panel of original sample members at each successive wave.

For ELSA, the full cohort of individuals age 50+ has been interviewed since the start of the study, making this approach fairly straightforward. (The refresher cohorts added in subsequent waves of ELSA are not included in response rate calculations for this paper.) However this is not the case with HRS. As noted previously (and shown in Tables 2 and 3), the HRS sample is made up of different birth cohorts, most of which entered the study in a different year. Thus, wave 1 for the original HRS sample occurred in 1992, whereas wave 1 for the Early Baby Boom cohort took place in 2004. Or, stated differently, 2004 was wave 1 for the EBB cohort, wave 4 for the CODA and WB cohorts, wave 6 for the AHEAD cohort, and wave 7 for the HRS cohort. In order to obtain comparable longitudinal response rates for the HRS, we first calculated wave-specific response rates (i.e. wave 1, wave 2, wave 3, etc) for each cohort, then took the weighted average of the wave-specific response rates (weighted by the sample size of each cohort) to get an overall response rate for each wave. This means that the response rates for different waves are based on different combinations of cohorts. For example, in Tables 5-8, response rates for waves 1 and 2 are based on all five cohorts, those for waves 3 to 5 are based on four cohorts (all except the EBB cohort, for which only two waves had been conducted through 2006), the response rate for waves 6-7 are based on only two cohorts (HRS and AHEAD), and that for wave 8 is based only on the original HRS cohort.

To further add to the complexity, it is also necessary to treat ELSA Wave 0 (HSE interview) as the first wave for comparison with the equivalent Wave 1 for HRS for some response rates. This is to take account of the fact that respondents had prior experience of the HSE interview, and therefore ELSA wave 1 could not be classified as the *first* contact attempt. It is important to note however that ELSA wave 1 is still referred to as the baseline wave of ELSA, as only households with at least one productive interview with an age-eligible sample member at wave 1 were followed up for interview at wave 2.

Eligibility criteria

Lynn's (2005) framework highlights the importance of establishing clear and consistent definitional criteria before calculating response rates. A key issue relates to the survey outcome of interest (i.e. what is deemed a completed data collection event or "response"). This is made complicated when data collection events involve multiple components. For example, ELSA and HRS both have a biennial core interview followed by a self-administered questionnaire, but every four years ELSA also has a separate visit by a qualified nurse. In the calculation of response rates, should the data collection event therefore be considered "complete" only if all components have been successful? Also, if respondents start the interview but stop halfway, or if they decline to answer some questions, should these situations be treated as complete outcomes?

A second issue relates to eligibility status (i.e. who should be considered as part of the population of interest). In general, not all sample members of a survey may be eligible for all data collection rounds or events. HRS has carried out supplemental data collection exercises using postal or internet surveys to a sub-set of the main sample, so they may not all be eligible at the same point in time. Furthermore, deaths or geographical moves within the study area may affect eligibility status from one study wave to the next.

For each study, eligibility criteria were set in order to classify sample members according to their status at each wave. They were categorized as either respondents, non-respondents or ineligible. For analysis, respondents and non-respondents are always included in the response rate denominator, while those deemed ineligible are removed completely from the calculations. The definitional criteria used for each group are given below.

Respondents

Each study defines a "response" in a given wave as full or partial completion of a core interview or proxy interview, but not necessarily the supplemental components (e.g. self-completion or nurse visit).

Ineligible

As expected, both studies treat people who have died as ineligible. ELSA also classifies those who have moved outside of Britain as ineligible (due to moving out of the population of interest), where as HRS does not apply any geographical ruling to its eligibility criteria. Unlike HRS, ELSA also excludes those who have moved into an institution or care home from the response rate denominator, in order to allow consistent comparisons with earlier waves of ELSA. To explain this further, moves to an institution or care home were recorded by ELSA interviewers after wave 1, but the actual interview for those in care homes was first introduced at wave 3. ELSA Technical Reports also therefore treat those in care homes as ineligible in the published response rates.

The difference in eligibility criteria across studies described above is likely to have a negligible effect on response rates due to small numbers. By ELSA wave 3, only 89 age-eligible sample members interviewed at wave 1 were identified by interviewers to have moved out of Britain, and 76 had moved into an institution (most likely a residential/nursing home). In total this equates to only 1.4% of those who had successfully completed a wave 1 interview.

Non-respondents

This group consists mainly of those who have refused at a given wave or who could not be contacted. In addition, both studies include those who have asked to be removed from the sample, or who have moved and cannot be traced, as eligible for the study. For ELSA, only moves not known to be outside of Britain are classified as non-response.

Unknown eligibility

Each study has a sub-group of sample members whose eligibility is 'unknown' at a given wave due to non-contact or unsuccessful tracing. To compensate for this, ELSA estimated a proportion of cases with unknown eligibility to be ineligible at each wave using age-sex mortality rates and annual rates of moves into an institution. Those "unknown" cases not re-classified as ineligible remain as non-respondents. In contrast the HRS, for which the percentage with unknown status is typically very low, assumed all those with unknown status to be eligible in response rate calculations (so treat them as non-respondents). The justification for applying the age-sex mortality rates to "unknown eligibles" in ELSA is due to incomplete mortality information. Mortality checking is carried out for ELSA sample members who provided consent to linkage to the National Health Service Central Register at wave 1, but there are groups from HSE that are not covered by this process. For example, ELSA never established who resided in non-cooperating HSE 1998, 1999 or 2001 households, and as a result, no information is available to link to official mortality records. In addition, responding households at HSE that did not take part in ELSA wave 1 are not included in the mortality checking process (Taylor et al. 2007).

Response rate definitions

For ELSA, the response calculations are based on the original age-eligible sample members identified for wave 1 in 2002. The ELSA wave 3 refreshment sample of those aged 50-53 has not been included for ease of analysis and comparability with ELSA Technical Reports. For HRS, the response rate calculations are based on all age-eligible sample members at a given wave from all cohorts.

We focus on the following set of response rates outlined in Lynn's framework: 1) unconditional cross-sectional response rates, 2) cross-sectional response rates conditional on completion in the prior wave, 3) unconditional longitudinal response rates, and 4) longitudinal response rates conditional on completion in the baseline wave. Each of these response rates is defined in Table 4.

Table 4. Description of response rates from Lynn's framework

Response rate	Description	Numerator	Denominator
RR_i	Unconditional cross-sectional response rate for wave *i*	Response at wave *i* (*i=1,…,k*)	Eligible at wave i (*i=1,…,k*)
$RR_{i\|i-1}$	Conditional cross-sectional response rate for wave *i*, given completion in immediately prior wave (*i-1*)	Response at wave *i* (*i=1,…,k*)	Eligible at wave i (*i=1,…,k*) and response at wave *i-1*
$RR_{\{1,2,\ldots,k\}}$	Unconditional longitudinal response rate for any combination of waves *1,2,…,k*	Response at every wave of interest (*1,2,…,k*)	Eligible at every wave of interest (*1,2,…,k*)
$RR_{\{1,2,\ldots,k\}\|1}$	Conditional longitudinal response rate for any combination of waves *1,2,…,k*, given completion in baseline wave (wave *1*)	Response at every wave of interest (*1,2,…,k*)	Eligible at every wave of interest (*1,2,…,k*) and response at wave *1*

The first two response rates are cross-sectional. As such, they represent the percentage of sample members who participated in a given wave of the survey. The unconditional cross-sectional response rate represents the percentage who participated in wave *i*, of all sample members who were eligible to participate in that wave (regardless of whether an interview was attempted). As such, it gives an indication of the percentage of the target population that is represented in the survey at each wave. The conditional cross-sectional response rate represents the percentage who participated in wave *i*, of those who were eligible to participate in wave *i* and who had participated in the previous wave (*i*-1). This is sometimes referred to as a "re-interview" response rate, and reflects the success of a survey at retaining respondents from one wave to the next. Cross-sectional response rates are most useful for analyses using a single wave of survey data.

The lower pair of response rates from Table 4, that we focus on in the paper, are longitudinal response rates. Longitudinal response rates represent the percentage of sample members who participated in multiple waves of the survey (minimum of two, up to the total number of waves that have been conducted), of those who were eligible in all of those waves. The unconditional longitudinal response rate represents the percentage of sample members who participated in each wave of a series of waves, of all sample members who were eligible to participate in each of those waves. As with the unconditional cross-sectional response rate, it gives an indication of the percentage of the target population that is represented in a sequence of waves in the survey. The conditional longitudinal response rate represents the percentage of sample members who participated in each of a series of waves, of those who were eligible in all of those waves and who participated in the baseline wave. As such, it reflects the success of the survey in retaining the original panel in subsequent waves. Generally longitudinal response rates correspond with a series of consecutive waves starting with the baseline wave (e.g., waves 1-4), but they could also be calculated for other sequences of waves (e.g. waves 1, 3, 5, or waves 4-8). Longitudinal response rates are most useful for analyses that make use of multiple waves of survey data; the rate that is most relevant for a given analysis will correspond with the waves used in the analysis.

Response rates for each study

Tables 5-8 present response rates for each of the four types described in Table 4. Appendix B includes examples based on ELSA, which illustrate how the different types of response rates were calculated.

Cross-sectional response rates

Table 5 presents unconditional cross-sectional response rates for ELSA and HRS for each wave. These rates represent the percentages of all eligible sample members in the designated wave who completed an interview in that wave. For HRS, the unconditional cross-sectional response rate for wave 1 (RR_1) is typically referred to as the baseline response rate, whereas for ELSA this rate corresponds with response at HSE (wave 0). As a result the RR_2 rate reported for ELSA actually represents that achieved at wave 1 (baseline).

For HRS the baseline response rate was 78%. At ELSA wave 0 the achieved rate was 70.2% and the baseline rate at wave 1 was 46.5%. The ELSA wave 1 rate includes non-cooperating HSE households in the denominator (see Appendix B for example calculation). Because non-responding households at wave 1 were not followed in either of the studies, response rates for each consecutive wave are necessarily lower than the baseline response rates. The unconditional cross-sectional response rate at wave 2 was 72.5% for HRS. By wave 8 of HRS, about two-thirds of the original target sample who were still eligible at wave 8 completed an interview in that wave.

Table 5. Unconditional cross-sectional response rates for each wave

Study	RR_1	RR_2	RR_3	RR_4	RR_5	RR_6	RR_7	RR_8
ELSA*	70.2%	46.5%	39.2%	36.1%				
HRS	78.0%	72.5%	72.0%	70.3%	68.4%	69.1%	67.6%	66.4%

**For ELSA, $_1$=wave 0; $_2$=wave1, etc.*

Table 6 presents conditional cross-sectional response rates, which are based on those who responded in the previous wave. The numerator for each rate includes individuals who completed an interview in both waves *i* and waves *i*-1 and the denominator includes those who were interviewed in wave *i*-1 and still eligible in wave *i*.

The conditional cross-sectional response rates provide an indication of the success of each survey at retaining respondents from one wave to the next. For example, $R_{2|1}$ represents the percentage of baseline respondents who completed an interview at wave 2 (of those still eligible at wave 2). These rates are all considerably higher than the unconditional cross-sectional response rates in Table 5.

Focusing first on the wave 2 rates ($R_{2|1}$), HRS has the higher rate of 92.6%. ELSA's conditional rate of 64.7% represents response at ELSA wave 1 conditional upon participation at wave 0 (HSE years 1998, 1999 or 2001). ELSA's comparatively low rate is largely due to the inclusion of those who did not consent to re-contact after their HSE interview. These cases were in fact not issued to field at the start of ELSA wave 1 but still need to be included in the denominator (see discussion section). It is worth noting that the decision to sample from three different HSE years is unlikely to have affected the overall conditional $R_{2|1}$ rate for ELSA, as similar household response rates were reported for each HSE year (see ELSA wave 1 technical report).

For both ELSA and HRS, the conditional cross-sectional rates tend to increase over the length of the study, as the sample that was interviewed in the prior wave becomes increasingly selective of more cooperative individuals.

Table 6. Conditional cross-sectional response rates for each wave

Study	**$RR_{2\|1}$**	**$RR_{3\|2}$**	**$RR_{4\|3}$**	**$RR_{5\|4}$**	**$RR_{6\|5}$**	**$RR_{7\|6}$**	**$RR_{8\|7}$**
ELSA*	W1|W0= 64.7%	W2|W1= 81.5%	W3|W2= 85.6%				
HRS	92.6%	94.1%	94.5%	94.6%	95.1%	95.5%	95.6%

**For ELSA, $_1$=wave 0; $_2$=wave1, etc.*

Longitudinal response rates

Table 7 presents unconditional longitudinal response rates. These rates are cumulative; they represent the percentage of respondents who completed an interview in a set of consecutive waves starting with the first wave (1 through *k*), among all sample members who were eligible in all of those waves (including those who were in the original sample but did not complete an interview at the first wave). As such they represent the proportion of the target population that has participated in all of the designated waves of the study.

The proportion of the target sample that was represented in both waves 1 and 2 in the surveys was 45.8% in England and 70.4% in the U.S. The rates decline across waves, though the decline is fairly gradual for both studies. The rates in **bold** represent the "complete" response rates based on all waves through 2006. For ELSA, the complete response rate indicates that about 35% of eligible respondents participated in the HSE and Waves 1-3 of ELSA. For HRS, 56% of eligible respondents participated in Waves 1-8 of the HRS.

Table 7. Unconditional longitudinal response rates

Study	**$RR_{1,2}$**	**$RR_{1,2,3}$**	**$RR_{1,2,3,4}$**	**$RR_{1,2,3,4,5}$**	**$RR_{1,2,3,4,5,6}$**	**$RR_{1,2,3,4,5,6,7}$**	**$RR_{1,2,3,4,5,6,7,8}$**
ELSA*	45.8%	38.6%	**34.5%**				
HRS	70.4%	67.7%	64.3%	61.3%	60.7%	58.4%	**56.1%**

**For ELSA, $_1$=wave 0; $_2$=wave1, etc.*

Table 8 presents conditional longitudinal response rates. This rate represents the percentage of respondents, that participated at wave 1 and were eligible in all of the other waves, who completed an interview in a set of consecutive waves, starting with the baseline wave (1 to *k*). In other words, it is the proportion of the eligible baseline sample that completed an interview in all of the designated set of waves.

For the ELSA conditional longitudinal rate it made sense to adopt ELSA wave 1 as the first wave of the study (baseline) rather than wave 0, as this rate reflects how successful the study has been in maintaining the original panel who had actually completed an ELSA baseline interview.

The conditional longitudinal rates for waves 1 and 2 are, by definition, the same as the conditional cross-sectional response rates for wave 2 shown in Table 5 (i.e. conditional on wave 1 participation). For subsequent waves, however, the two sets of rates differ. As with the unconditional longitudinal rates, the conditional longitudinal rates gradually decline across waves. Still, the rates remain quite high, reflecting the success of the studies in retaining those who originally participated in the study in subsequent waves. Over 70% of the baseline respondents participated in each of the first three waves of ELSA (of those who were eligible for all three waves), and over two-thirds of the original HRS respondents participated in eight consecutive waves.

Table 8. Conditional longitudinal response rates (conditional on participation at baseline)

Study	$RR_{1,2\|1}$	$RR_{1,2,3\|1}$	$RR_{1,2,3,4\|1}$	$RR_{1,2,3,4,5\|1}$	$RR_{1,2,3,4,5,6\|1}$	$RR_{1,2,3,4,5,6,7\|1}$	$RR_{1,2,3,4,5,6,7,8\|1}$
ELSA*	81.5%	70.8%					
HRS	92.6%	87.6%	83.3%	79.6%	74.7%	71.9%	68.6%

**For ELSA, $_1$=wave 1; $_2$=wave 2, etc.*

Discussion

The results in Tables 5-8 reflect substantial differences in response rates across the two studies. While Lynn provides a clear model for calculating different types of response rate, there are issues relating to sample design and fieldwork practice across HRS and ELSA which need to be carefully considered. Some of the key issues are noted below.

Study design features and protocols that may influence response rates

Sample design

In calculating response rates for ELSA and HRS, it became apparent that differences in study design complicate both the calculation and interpretation. Unconditional rates are dependent on the initial response to the survey, and as such are influenced by the sample frame and sampling procedures. This paper has highlighted the difficulties around trying to compare unconditional rates based on follow-up from another survey (ELSA) with other multi-stage sampling techniques employed in HRS.

HRS inherited very little non-response during recruitment for the original sample, due to a highly successful screening field effort. In contrast, ELSA carried forward a high proportion of non-response from HSE which then affected the overall unconditional response rates. The following sample breakdown of age-eligible sample members for HSE (Wave 0) shows the magnitude of this effect:

Table 9. Breakdown of age-eligible sample members from HSE households

Total number of productive HSE individual interviews	18,651
Total number of non-responding individuals in HSE cooperating households	1,270
Estimate of non-responding individuals in HSE non-cooperating households	6,630
Total individuals (response rate denominator)	26,551

The estimate of the number of age-eligible individuals in HSE non-cooperating households accounts for nearly 25% of the denominator. Hence, there is a large group of people who did not themselves take part at HSE and so were not followed up for ELSA, but who are still considered eligible for response rate calculations. A small proportion of these are estimated to have died or moved into an institution prior to calculating rates for each wave, so the denominator is reduced accordingly (see eligibility section). It is important to bear in mind therefore that the denominators used in the ELSA response calculations represent estimated rather than actual eligibility. In contrast, the success of HRS household screening to obtain the initial sample has limited the amount of estimation required.

Furthermore, those age-eligible individuals who were interviewed for HSE but refused to be re-contacted after their HSE interview, are also still considered eligible in response calculations for ELSA, despite never being approached for the study. This needs to be kept in mind when evaluating ELSA's cross-sectional rate for wave 1 conditional upon participation at wave 0. If the 1,681 age-eligible individuals who refused re-contact after wave 0 are excluded from the denominator, the rate would increase from 64.7% to 71.6% (see stage 5 in Appendix A).

Lynn's framework is used to illustrate the proportion of sample members interviewed from the target population, but no account is taken of the impact of fieldwork management on response rate. All of the response rates presented in this paper include 'all those eligible' in the denominator, but in reality, a sizeable proportion of eligible sample members are ***not*** issued to field at the start of each wave. Fieldwork agencies tend to rely more on field response rates in order to track success of fieldwork efforts, as this rate is based on all those actually issued to field. Both ELSA and HRS have had to remove a number of cases from the sample due to refusal to be re-contacted, but they are still considered eligible for the rates presented in this paper. Generally there are cross-study differences in how non-respondents are managed, as HRS was less restrictive in their handling of prior wave refusals after the second wave. At wave 2, both studies chose to issue households from wave 1 with at least one productive interview with an age-eligible individual. However, at the start of wave 3 some element of subjectivity was introduced to the decision to issue prior wave refusals. For ELSA, 91% of those who completed a wave 1 interview and were still eligible at wave 3 were issued to field. For HRS, the comparable figure was over 99%.

Overall, conditional response rates seem to provide the most standardized basis for analysis of cross-study study performance. By limiting the denominator to those interviewed at the previous wave, it is possible to get a sense of how successful each study has been in maintaining its original panel of members.

The possible influence of some fieldwork practices across studies on response rate is covered below. Our understanding of study differences can be enhanced further by looking at how different types of non-respondents are handled across studies and how this may impact on interpretation of the overall rates.

Interview Mode

The core HRS questionnaire has been designed for administration either in person or by telephone. Up to 2002, follow-up interviews with all participants under age 80 were conducted by telephone and, since 2006, half of those under 80 (those not assigned to the enhanced face-to-face sample in that wave) complete the interview by telephone in each wave. The remainder are interviewed face-to-face. Although the ELSA questionnaire has similar content, it has only been administered face-to-face. In longitudinal studies, using the same mode, each wave helps to avoid potential mode effects across waves. The face-to-face mode has helped ELSA interviewers to establish a good rapport with sample members over time and has allowed the inclusion of some cognitive and physical measures which require the presence of an interviewer. For HRS, the decision to implement telephone interviewing for part of the sample was a cost-saving decision. However, the practice may have served to encourage participation amongst would-be refusers, who find the telephone mode more convenient or less invasive.

Incentives

There is strong support for the use of incentives in surveys, as incentives increase response rates in a linear fashion and may act as a motive in itself for participation (Singer 2002). Prepaid incentives are also found to be more effective than promised or contingent incentives (Jackle and Lynn 2007). With this in mind, differences in the incentive amounts offered to HRS and ELSA respondents may have had some influence on willingness to participate. ELSA offers £10 to sample members for completion of a face-to-face interview. In contrast, the amount offered to HRS participants for the core interview has increased over time, from $20 in 1992 to $40 in 2006. In HRS, an additional incentive of $40 is given to participants in the enhanced face-to-face sample (which includes physical measures, biomarkers and psychosocial self-administered questionnaire). An extra incentive of this kind was not offered to ELSA sample members who completed a follow-up nurse visit (wave 2) or self-completion (waves 1 to 3). Although both studies have employed the use of differential incentives amongst highly resistant respondents, the amounts differ substantially, with HRS offering up to $100 to this group in comparison to £20 offered by ELSA. The method of administering incentives also differs across the studies. In the baseline wave of HRS, incentives are paid at the time of interview, and in follow up waves, incentives are included with the initial contact letter, prior to scheduling the interview. In contrast, ELSA has only administered incentives on a conditional basis (after the interview), which may affect the willingness of participants to respond.

Proxy interviews

A further key difference that impacts on the response rates in the two studies, is the way proxy respondents are used. ELSA has a more restrictive policy regarding proxies in comparison to HRS. ELSA allows proxies if cognitive impairment, physical or mental ill health prevented a respondent from doing a face-to-face interview. Likewise if the respondent was away in hospital or temporary care throughout the whole fieldwork period, a proxy interview was permitted. HRS has a somewhat more lenient policy towards proxy interviews. In addition to health-related restrictions, which make up the bulk of reasons for proxy interviews in HRS, proxy interviews are accepted for respondents who are unavailable or unwilling to be interviewed but who grant permission for someone to complete the interview on their behalf. ELSA rates for complete proxy interviews were 1% and 2% at waves 2 and 3 respectively, whereas the rates in the HRS varied between 5% in 1992 and almost 14% in 1995 (AHEAD). Since combining the cohorts in 1998, the HRS proxy rate has been between 7 and 11%.

There are pros and cons to using proxy respondents in surveys. Accepting proxy interviews not only helps to improve response rates, but it may also reduce selection bias, as individuals who are interviewed by proxy tend to be different in important ways from those who complete a self-interview. This is particularly true for surveys of older adults, for whom poor health tends to be a key factor in non-participation. The paper by Weir, Faul and Langa in this Special Issue examines this in relation to measures of cognition. On the other hand, if proxy respondents answer questions differently to how respondents would themselves answer, the use of proxy interviews may increase measurement bias.

Between wave contacts

HRS and ELSA differ with respect to the number and types of contacts that are made with respondents between core interviews. HRS conducts supplemental postal and internet studies between core interview waves and most respondents receive at least one request to participate in a minimum of one supplemental study. In addition, HRS typically sends a newsletter to respondents shortly before the start of each round of core data collection. In contrast, ELSA does not send a newsletter between waves or conduct between-wave supplemental studies, although ELSA typically sends holiday greeting cards and a newsletter prior to the start of a new wave. Between wave contacts are often thought to be beneficial for keeping participants engaged and interested in the study. At a minimum, sending something to respondents, whether it is a card, newsletter or questionnaire, can be useful for identifying potential movers, based on mail that is returned as undeliverable. However, whether benefits extend beyond that is unclear. In an experimental study based on an Internet panel survey in the Netherlands (the LISS panel), investigators found that sending participants different types of materials containing the information about and/or highlights of findings from the study (e.g. newsletters, post-cards, e-cards, ring binders) had no effect on participation in subsequent waves of the study (Scherpenzeel and Vis 2010). With regard to additional survey components (supplemental studies), analysis of interview outcomes in HRS suggests that the mode and content of the supplemental requests may have a role in influencing continued participation in the core interview (Ofstedal and Couper 2008). Although the HRS supplemental studies were not assigned experimentally and results should be interpreted with some caution, respondents who were invited to participate in the internet survey and the diabetes mail survey had higher response rates in the next core interview wave than those who were not asked to participate. In contract, HRS participants who were invited to participate in the Consumption and Activities Mail Survey (CAMS), which focuses primarily on household expenditures, were less likely to participate in the next core interview wave.

Tracking movers

Both ELSA and HRS use similar fieldwork protocols for locating or tracking participants who move between waves. Both studies make use of information collected in a previous wave on contact persons and information obtained from public, commercial and/or administrative databases, such as telephone and address listings. A number of these methods are carried out proactively (i.e. before the start of fieldwork for a given wave) by centralized staff, whereas others are carried out during data collection by field interviewers and/or staff who are specifically trained in tracking methods (Couper and Ofstedal 2009). The percentage of participants who are not located (lost to tracking) in each wave depends on the level of mobility among sample members, as well as the quality of the resources available for tracking movers. Both of these factors may differ across ELSA and HRS. For the general population, mobility rates tend to be somewhat higher in the United States compared to most countries in Western Europe (Couper and Ofstedal 2009). Nevertheless, in both studies, the fraction of respondents who are not located each wave is very small. In the 2006 wave of HRS, 0.6% of total core sample members and 5.6% of core non-respondents were not successfully traced. These figures are similar to ELSA which at wave 3 (2006) had 1.4% of total eligible sample members that had moved and could not be traced (equivalent to 7% of wave 3 non-respondents issued to field). Rather, most of the non-response at each wave (between 75% and 80%) is due to refusal.

Conclusion

The primary purpose of this paper was to document response rates for ELSA and HRS using the framework for longitudinal studies proposed by Lynn (2005). In doing so, this paper has shown that even with a specified framework, a strict comparison of response rates across studies can be problematic without considering differences in sample design, eligibility criteria and fieldwork protocol. As noted in the discussion, the higher response rates observed in HRS can be explained, at least in part, by differences in sample design, respondent incentives, protocols relating to the use of proxy respondents, and interview mode.

For purposes of comparison across studies, it may be of practical interest to supplement the rates proposed by Lynn (2005) with other types of response rates. For example, the impact of fieldwork management can be represented by using the field response rate, based only on those cases actually issued to field in a given wave. The conditional cross-sectional response rate in Lynn's framework (presented in Table 6) is most similar to the field response rate, except that it excludes sample members who did not complete an interview in the prior wave and, thus, will always be higher than the field response rate. However, a downside of the conditional cross-sectional rates in this paper (conditional on participation in the immediately prior wave) is that they do not capture respondents' movement in and out of the study. A critical element of longitudinal studies is bringing people back in after they missed a wave (e.g. see the Kapteyn et al article in this Special Issue) and this could be tracked by using yet another response rate: the cross-sectional response rate conditional on *baseline* response.

Response rates provide only part of the picture with regard to selection bias; the other part depends on the extent to which non-respondents differ from respondents on characteristics of interest (Groves and Couper 1998). Our exclusive focus on response rates is, thus, a limitation of this paper. Where they are possible, analyses comparing non-respondents and respondents would help inform the degree to which respondents are representative of the target population and the extent to which non-response bias is likely to be an issue. Such comparisons are typically not feasible for the baseline wave, as information on non-respondents tends to be extremely limited or absent altogether. However, a key strength of panel surveys is that they allow for comparisons of those who drop out versus continue to participate in subsequent waves, and several of the papers in this Special Issue address such comparisons. Additional research along these lines is needed in order to make informed judgments about quality within and across studies.

Acknowledgements

The English Longitudinal Study of Ageing (ELSA) has been developed through collaboration between three primary institutions: University College London (UCL), the Institute of Fiscal Studies (IFS) and the National Centre for Social Research (NatCen), with academics at the Universities of Manchester, Cambridge, Nottingham, Exeter and East Anglia.

Funding for the first four waves of ELSA was provided by the US Institute on Aging (NIA) and a consortium of British Government departments, specifically: Department for Education and Skills, Department of Environment, Food and Rural Affairs, Department for Work and Pensions, HM Treasury, HMRC (formerly Inland Revenue), Department for Communities and Local Government and Office for National Statistics. Ethical approval was granted by the Multi-centre Research and Ethics Committee (MREC).

The HRS (Health and Retirement Study) is sponsored by the National Institute of Aging (grant number NIA U01AG009740). The study is conducted by the Survey Research Center at the University of Michigan.

References

The American Association for Public Opinion Research. (2008) *Standard definitions: final dispositions of case codes and outcome rates for surveys, 5th edition*. The American Association for Public Opinion Research, Lenexa, Kansas.

Couper MP and Ofstedal MB. (2009) Keeping in contact with mobile sample members. In P Lynn. ed. *Methodology of longitudinal surveys*. Wiley, New York.

Erens B and Primatesta P. eds. (1999) *Health Survey for England 1998, Vol. 2: methodology and documentation.* The Stationery Office, London. (http://www.archive.official-documents.co.uk/document/doh/survey98/hse-08.htm)

Erens B, Primatesta P and Prior G. eds. (2001) *Health Survey for England. The Health of Minority Ethnic Groups 1999, Vol. 2: methodology and documentation*. The Stationery Office, London. (http://www.archive.official-documents.co.uk/document/doh/survey99/hse99-14.htm)

Groves RM and Couper MP. (1998) *Non-response in household interview surveys*. John Wiley and Sons , New York.

Heeringa SG. (1995) *Technical description of the Asset and Health Dynamics of the Oldest-Old (AHEAD) study sample design*. Survey Research Center, University of Michigan, Ann Arbor, Michigan. (http://hrsonline.isr.umich.edu/sitedocs/userg/AHDSAMP.pdf)

Heeringa SG and Conner J. (1995) *Technical description of the Health and Retirement Study sample design*. Survey Research Center, University of Michigan, Ann Arbor, Michigan. (http://hrsonline.isr.umich.edu/sitedocs/userg/HRSSAMP.pdf)

Jackle A and Lynn P. (2007) *Respondent incentives in a multi-mode panel survey: cumulative effects on non-response and bias*. ISER Working Paper 2007-01. University of Essex, Colchester.

Lynn P. (2005) *Outcome categories and definitions of esponse ates for anel urveys and ther urveys involving multiple data collection events from the same units*. Unpublished manuscript. University of Essex, Colchester. (http://www.nonresponse.org/uploadi/editor/1129102463Outocome%20categories%20same%20units.pdf)

Ofstedal MB and Couper MP. (2008) *Piling it on: the effect of increasing respondent burden on participation in a panel study*. Paper presented at the Panel Survey Methods Workshop, Colchester, July 2010.

Prior G, Deverill C, Malbut K and Primatesta P. eds. (2003) *Health Survey for England 2001: methodology and documentation*. The Stationery Office, London. (http://www.archive2.official-documents.co.uk/document/deps/doh/survey01/md/md-00.htm)

Scholes S, Taylor R, Cheshire H, Cox K and Lessof C. (2008) *Retirement, health and relationships of the older population in England: the 2004 English Longitudinal Study of Ageing, Technical Report.* National Centre for Social Research, London. (http://www.ifs.org.uk/elsa/publications.php)

Scholes S, Medina J, Cheshire H, Cox K, Hacker E and Lessof C. (2009) *Living in the 21st century: the 2006 English Longitudinal Study of Ageing, Technical Report*. National Centre for Social Research, London. (http://www.ifs.org.uk/elsa/publications.php)

Scherpenzeel A and Vis C. (2010) *Encouraging and maintaining participation in an internet panel: effects of letters, incentives and feedback*. Paper presented at the Panel Survey Methods Workshop, Mannheim, July 2010.

Singer E. (2002) The use of incentives to reduce non-response in household surveys. In Groves R, Dillman D, Eltinge J and Little R. eds. *Survey Non Response*. Pp 163-177. Wiley, New York.
Taylor R, Conway L, Calderwood L, Lessof C, Cheshire H, Cox K and Scholes S. (2007) *Health, wealth and lifestyles of the older population in England: the 2002 English Longitudinal Study of Ageing, Technical Report*. National Centre for Social Research, London. (http://www.ifs.org.uk/elsa/report03/w1_tech.pdf).

APPENDIX A

Detailed sampling design for ELSA study

Creation of the ELSA wave 1 sample from the Health Survey for England (HSE) is best described in five stages:

Stage 1: 31,051 households were issued at the start of fieldwork across HSE 1998, 1999 and 2001.

Stage 2: In the early stages of the HSE interview, all responding households were asked to provide the date of birth for every resident regardless of whether each went on to complete a full individual HSE interview. This meant that all age-eligible individuals could be identified in responding households. In contrast, non-responding households were not included in the ELSA sampling frame because there was no available information about residents that would have made it possible to identify those who were aged 50+ at the time of ELSA wave 1.

A sampling frame was constructed from the HSE responding households using information about the residents at the time of HSE interviewing. Overall, 23,132 households responded to HSE 1998, 1999 and 2001 and so formed the foundation of the ELSA sample while a further 7,919 households did not respond to HSE and so were not included in the sampling frame.

Stage 3: From the available HSE information two sample member types were identified for the ELSA wave 1 interview in 13,203 households.

- First, potential age-eligible **sample members** (SM) were identified. These were defined as individuals who were living within an HSE responding household and were born before 1 March 1952. In total 19,924 sample members were identified.

- Second, potential **younger partners** (YP) were defined as the cohabiting younger spouses/partners of sample members, who were living within the household at the time of the HSE interview and were born after 29 February 1952. 1,269 younger partners from HSE were identified.

9,929 households that responded to HSE were not eligible for inclusion in the final ELSA sample because they did not contain an age-eligible individual.

Stage 4: A mortality check was conducted for those potential sample members and younger partners who gave their permission (95%) to be 'flagged' with the National Health Service Central Register (NHSCR) run by the Office for National Statistics (ONS). This register keeps track of registrations with general practitioners but also with official death registrations and with people who leave the UK health system. No check was conducted on the HSE 2001 sample as little time had passed since that interview. 401 households were dropped as a result of deaths between HSE and ELSA wave 1.

Stage 5: Potential sample members and younger partners were not included in the final ELSA sample if *all* HSE respondents aged 50 years or older within the household had refused, when asked, to being re-contacted in the future. Even though these people had not directly refused to take part in ELSA (they would not have been aware of the study at the time of HSE) it would have been unethical to have re-contacted them. Overall, 1,224 of the 12,802 eligible HSE households were removed on this basis (9.6%). This equated to a loss of 1,681 age-eligible individuals.

To summarise, the ELSA wave 1 sample was only selected from households that responded to HSE (Stage 2). Furthermore, households were only issued to field if they included at least one age-eligible individual (Stage 3) who, according to administrative records, remained alive (Stage 4) and gave permission to be re-contacted in the future (Stage 5).

APPENDIX B

Example response rate calculations for the ELSA study

Unconditional cross-sectional response rate (Table 5)

ELSA wave 1 (W1):

In order to derive unconditional response rates it was necessary to classify age-eligible ELSA sample members according to their status at HSE (W0). A distinction is made between those from HSE cooperating and HSE non-cooperating households.

HSE cooperating households, respondents in W0
Respond in W1 = 11,205
Non-respond in W1 = 6,125
Ineligible in W1 = 1,321
Total = 18,651

HSE cooperating households, individual non-respondents in W0
Respond in W1 = 186
Non-respond in W1 = 1,027
Ineligible in W1 = 57
Total = 1,270

HSE non-cooperating households
Non-respond in W1 = 5,947
Ineligible in W1 = 683
Total = 6,630

The number of productive outcomes in wave 1 was 11,391. The number estimated to be eligible was 11,205 + 186 + 6,125 + 1,027 + 5,947 = 24,490. Hence, as shown in Table 5, the estimated unconditional response rate in wave 1 was 11,391 / (11,391 + 6,125 + 1,027 + 5,947) = 0.465 × 100 = 46.5%.

For the calculation of ELSA unconditional response rates we have included non-cooperating HSE households. However, if we base the calculation solely on cooperating HSE households the unconditional rate increases to 61.4% (11,391) / (11,391+7,152) = 0.614 × 100.

Also it is worth bearing in mind that included in the non-response figures for wave 1 are individuals who were not issued at wave 1 because they refused to be re-contacted after their HSE interview (see ELSA sample design section). This group therefore had no opportunity to be interviewed at wave 1, but still need to be included in the denominator for unconditional rates because they were part of the original target population.

Conditional cross-sectional response rate (Table 6)
ELSA wave 1:
HSE denominator
Total productive interviews at W0 = 18,651.
Total ineligible by time of W1 = 1,321.
Total denominator for W1 response calculation = 17,330.

ELSA wave 1
Productive interviews completed at W0 and W1 = 11,205.
Productive interview at W0 only = 6,125.

As shown in Table 6, the estimated cross-sectional response rate in wave 1 conditional on successfully responding in wave 0 was 11,205/17,330 = 0.647 × 100 = 64.7%.

Unconditional longitudinal response rate (Table 7)
ELSA wave 1:

The denominator for the wave 1 (longitudinal) unconditional response rate focused on those original age-eligible sample members in waves 0 and 1 (irrespective of their outcome status at either wave or whether issued to field in wave 1). The numerator focused on those eligible sample units that responded in both waves 0 and 1. The response rate, therefore, indicates the proportion of eligible sample units that responded in *every* wave up to and including wave 1.

Productive interviews completed at W0 and W1 = 11,205.
Total number estimated to be eligible for interview atW0 and W1 = 24,490.

As shown in Table 7, the estimated (longitudinal) unconditional response rate in wave 1 is 11,205/24,490 = 0.458 × 100 = 45.8%.

Conditional longitudinal response rate (Table 8)
ELSA wave 3:

The wave 3 longitudinal response rate (defined for respondents in waves 1, 2 and 3) conditional upon having successfully responded in wave 1 was calculated as follows:

Number who successfully responded in ELSA waves 1-3 = 7,168.
Number who took part at wave 1 and were estimated to be eligible for interview in waves 2 and 3 = 10,126.

As shown in Table 8, the estimated longitudinal response rate in wave 3 conditional upon response in wave 1 was 7,168/10,126 = 0.708 × 100 = 70.8%.

Endnotes

[i] Nursing home residents were excluded from the original samples for HRS and ELSA.

[ii] For the three HSE surveys chosen, the household response rate ranged from 74% to 76% and the adult individual response rate ranged from 67% to 70%.

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

				wave			
Interview status	**1992**	**1994**	**1996**	**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

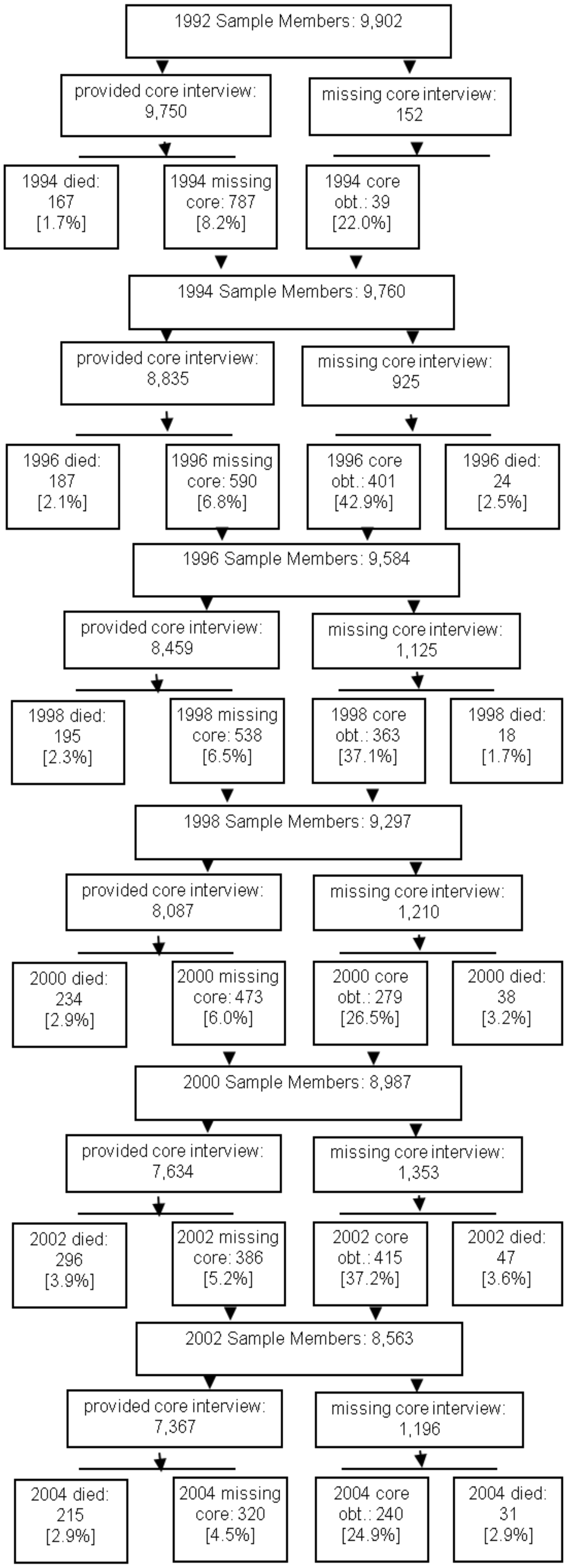


Table 3 summarizes baseline characteristics of respondents who gave an interview in 1992 by type of participation over the time period 1992–2004. Characteristics associated with mortality reflect the well-known positive association between health and socio-economic status (SES). Older individuals, African-Americans, 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

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

The parameter vectors to be estimated are $\beta_j,\ j = t, d, o,$ 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:

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

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.

Longitudinal and Life Course Studies 2011 Volume 2 Issue 2 pp 170 - 184 SPECIAL SECTION ISSN 1757-9597

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

(Received September 2010 Revised February 2011)

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).

	Follow-up Wave							
Entry Cohort	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	Follow-up Wave							
Cohort	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.

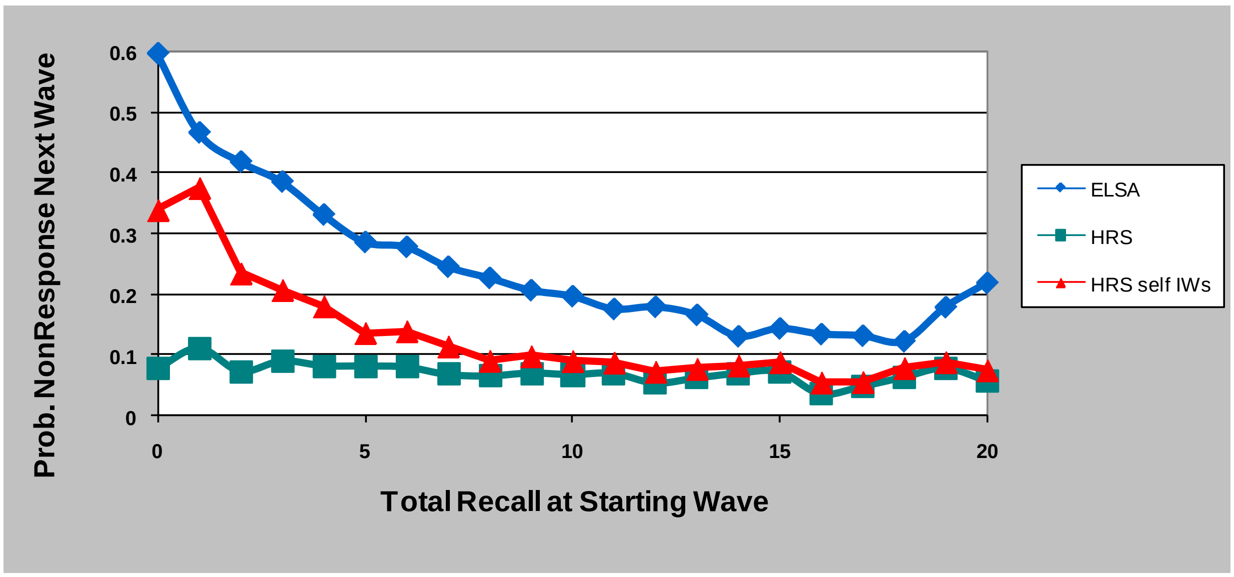
Table 6. Bias in mean level of cognition, by wave of follow-up, HRS cohorts excluding proxy interviews compared with ELSA.

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.

			CMS-FFS linkage rate		**Dementia diagnosis rate**		**Bias**
Year	Number of Eligibles 70+	Interview rate	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.

	Dementia diagnosis rate			
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

The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

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.

Longitudinal and Life Course Studies 2011 Volume 2 Issue 2 pp 185 - 201 ISSN 1757-9597

Social class returns to higher education: chances of access to the professional and managerial salariat for men in three British birth cohorts

Erzsébet Bukodi
Centre for Longitudinal Studies, Institute of Education, London
e.bukodi@ioe.ac.uk

John H Goldthorpe
Nuffield College, Oxford, and Centre for Longitudinal Studies, Institute of Education, London

(Received November 2010 Revised March 2011)

Abstract

In economics there is a well-established tradition of research into the earnings returns to education. We aim to make a sociological contribution by focusing on the social class returns: specifically, by examining the returns to higher education as indicated by chances of access to the professional and managerial salariat, while taking into account the effects of cognitive ability and class origins and also differences in access to professional and managerial positions. We draw on data for men from three British birth cohort studies covering children born in 1946, 1958 and 1970. We find that, while over the period covered the growth of the salariat ensured that absolute returns to both higher and lower tertiary qualifications were largely maintained, despite the growing numbers with such qualifications, returns relative to those to higher secondary qualifications diminished. Also, the advantages offered by lower tertiary qualifications as compared with higher secondary qualifications differ according to men's class origins. Overall, there is no evidence of any increase in education-based, meritocratic selection to the salariat. Rather, the growth of the salariat appears to be associated with some decline in its selectivity in terms of both qualifications and cognitive ability, with this decline being more marked in its managerial than in its professional components.

Introduction

In economics there is a well-established tradition of research into the earnings returns to education. Of late, attention has tended to focus on the earnings returns to *higher* education. In most advanced societies the issue has arisen of the balance between the supply of and the demand for higher-educated personnel. Does continuing 'skill-biased' technological change ensure a steadily rising demand for the higher educated so that, even with a growing supply, earnings returns are maintained or even increased? Or does the expansion of higher education lead, at some point, to supply outstripping demand so that a problem of 'over-qualification' is created and earnings returns fall? Apart from their academic interest, such questions are ones of obvious policy relevance.

So far, sociologists have been little involved in the debates that have ensued. However, we believe - and this is the motivation for the present paper - that sociologists do have a worthwhile contribution to make, and especially by way of broadening the standard approach from economics (Müller and Jacob 2008). In this respect, the following points may be made.

First, while it is understandable, given their disciplinary paradigm, that economists should concentrate their attention on the earnings returns to education, there is no reason, from either an academic or a policy point of view, why *only* earnings returns should be considered. In this paper we focus instead on social class returns: that is, on the returns to higher education *in the form of chances of access to the professional and managerial salariat*. We conceptualise class, in a way that is now becoming widely accepted, including in British official statistics (Office of National Statistics – (ONS) 2005a,b), in terms of positions defined by relations within labour markets and production units and more specifically by employment relations (Erikson and Goldthorpe 1992; Goldthorpe 2007, vol. 2: chs. 5-7; McGovern et al 2008; Rose and Harrison eds. 2010). Understood in this way, class can be shown to be related not only to individuals' earnings or current incomes, but further to their *income security*, their *short-term income stability* and their *longer-term income prospects* (see e.g. Elias and McKnight 2003; Goldthorpe and McKnight 2006; McGovern et al 2008; Lucchini and Schizzerotto 2010; Bihagen and Nermo 2010). In the case of access to the professional and managerial salariat, what is thus entailed is access to class positions whose holders are generally advantaged, relative to those in other class positions, as regards their levels of income and, *in addition*, as regards low risks of long-term or recurrent unemployment, low dependency on variable earnings (as resulting, say, from piece-rate or time-rate payment), and the expectation of rising earnings until late into working life due to incremental salary scales and relatively well-defined career opportunities.

Second, as regards the actual way in which the effect of education on earnings is exerted, economists mainly rely on human capital theory: through education, individuals invest in their human capital and then gain returns on this investment from the earnings they obtain in the labour market. From a sociological standpoint, this approach appears unduly abstract (Granovetter 1981) in that it leaves out of account the social relations that the labour market involves: that is the fact that individuals' earnings come from the *jobs* they are offered by employers and which they take up via an employment contract (or, in the case of 'independents', that they create for themselves in relations with clients or customers). A focus on the class returns to education gives primacy to access to jobs, since it is through their jobs that individuals become situated in the social relations of economic life that define their class positions. Further, though, individuals in this way also enter into the different groupings of jobs that constitute *occupations*; and the possibility can then be raised that the importance that employers attach to educational qualifications in relation to different occupations may vary, and even across occupations that imply similar class positions (Jackson, Goldthorpe and Mills 2005). In considering access to the salariat, we will therefore ask whether differences are apparent in the part played by education, as between the two broad occupational groupings that the salariat comprises: that is, those of professionals and managers.

Third, a main concern of economists is to go beyond the empirical association existing between education and earnings to estimate the *causal effect* of education on earnings, where a causal effect is understood as the impact of some intervention or 'treatment'. This understanding of causation does, however, give rise to problems in that education is to a significant degree a matter of choice rather than simply a 'treatment' that is received, and also, in that this choice is likely to be influenced by factors that may have their own direct effects on earnings: in particular, ability or various resources - economic, cultural and social - associated with individuals' families of origin. In seeking to deal with these problems, economists are led to treat ability and social origins as factors that have in some way to be statistically controlled so that the 'parameter of interest' - that taken to give the causal effect of education on earnings - can be estimated without bias (for an informative review, see Blundell, Dearden and Sianesi 2005). Our approach differs in two ways. First, we would regard the statistical analyses that we present as being no more than descriptive.[1] But second, we include measures of ability and of social class origins in these analyses not simply as controls but because they too are of substantive interest to us. We wish to know how these factors are associated with chances of access to the salariat, considered both independently of and in interaction with education.

Research questions, data and variables

The foregoing considerations lead us to focus our attention on three main issues as they arise in the British case.

(i) In a historical context in which the salariat has expanded and at the same time the numbers obtaining higher, or 'tertiary', education have increased, how, if at all, has the relationship between holding tertiary level qualifications and access to the salariat changed?

(ii) How far and in what ways does our understanding of this relationship - and of any changes in it - have to be modified when individuals' cognitive abilities and their social class backgrounds are brought into the analysis?

(iii) How far do differences arise in the importance of tertiary level qualifications and other factors as regards access to the professional and to the managerial divisions of the salariat?

A limitation of the paper is that we address these questions in the case of men only. Treating the same questions in the case of women would be clearly more complicated, as a result of problems arising from their selection into employment and, if only on grounds of space, calls for a separate paper.

Our data come from three British birth cohort studies: the Medical Research Council National Survey of Health and Development (NSHD), the National Child Development Study (NCDS) and the British Cohort Study (BCS). These studies aim to follow through their life-courses, children born in Britain in one week in 1946, 1958 and 1970 respectively (for further details, see Ferri, Bynner and Wadsworth eds. 2003: Appendix 1). It is by reference to the experience of men in these three birth cohorts, that we aim to assess the extent of changes over time in the social processes with which we are concerned.[2]

For each cohort, we have information, recorded in months, on respondents' employment histories, including details of occupation in each job coded to the 3-digit level of the OPCS SOC90 classification. In this paper, we focus on the employment histories of men up to age 34, the latest age for which, as of now, we have information for respondents in all three cohorts. Further, in the regression analyses through which we chiefly address the research questions we have set out, we restrict our attention to those men who, under a model developed in previous work (Bukodi and Goldthorpe 2009), may be regarded as having achieved a stage of 'occupational maturity': i.e. a stage in their working lives at which subsequent job changes have a declining probability of entailing occupational change. In this way, we believe, we can best compare 'like with like' across the three cohorts.[3]

We then determine men's class positions at age 34 on the basis of their current occupation and employment status (employer, self-employed, employee etc) according to the National Statistics Socio-Economic Classification (NS-SeC) which can be regarded as a new and improved instantiation of the Goldthorpe class schema (ONS 2005a,b; Goldthorpe 1997, 2007 vol 2, ch 5). We identify the salariat, access to which is the primary dependent variable of our analyses, with Classes 1 and 2 of the 7-class 'analytical' version of NS-SeC, as shown in Table 1.

Table 1. National Statistics Socio-economic Classification, seven-class version *

Class 1	Higher managerial and professional occupations
Class 2	Lower managerial and professional occupations
Class 3	Intermediate occupations
Class 4	Small employers and own account workers
Class 5	Lower supervisory and technical occupations
Class 6	Semi-routine occupations
Class 7	Routine occupations

** For the detailed composition of the classes by occupational group and employment status, based on SOC90, see ONS (2005b: Table 15).*

When we move on to consider chances of access to the professional and managerial divisions of the salariat, we make this distinction on the basis of the 13-category 'operational' version of NS-SeC. The professional division is identified with categories L3 and L4 which cover all professional and also higher technical occupations (e.g. scientific, electronic and IT technicians), and the managerial division with categories L1, L2, L5 and L6, which cover all managerial and also higher administrative occupations (ONS 2005b: Table 2).[4]

The independent variables of our analyses are educational qualifications plus ability, social class origins and number of occupations held up to the stage of occupational maturity.

As regards qualifications, we work with the four following categories:

1. Lower secondary or below.
2. Higher secondary: 5+ O-Levels or A-levels, National Vocational Qualifications, Level 3.
3. Lower tertiary: university diplomas, National Vocational Qualifications, Level 4.
4. Higher tertiary: university degrees, National Vocational Qualifications, Levels 5 and 6.

We code men to these categories according to the highest level of qualification that they had achieved by occupational maturity.

Our measures of ability - in effect, of cognitive ability - derive from tests taken by members of the three cohorts while at school. For the 1946 cohort, the tests were administered at age 8, for the 1958 cohort at age 11, and for the 1970 cohort at age 10. It appears generally accepted that standardised scores on these tests (z-scores) give a good approximation to IQ scores (Schoon 2010). Their chief attraction for us is that they provide the best measures that we have available of ability as distinct from educational attainment, although we would recognise that performance on the tests is likely to have been in some degree influenced by education up to the point at which they were taken.

As our indicator of men's class origins, we take their father's class position, also as coded to NS-SeC, during their childhood. For men in the 1946 and 1958 cohorts, father's class is determined at age 11, and for men in the 1970 cohort, at age 10.[5]

Finally, number of occupations to maturity refers to all occupations held for at least three months, falling into different categories of the 3-digit SOC90 classification.

The historical context

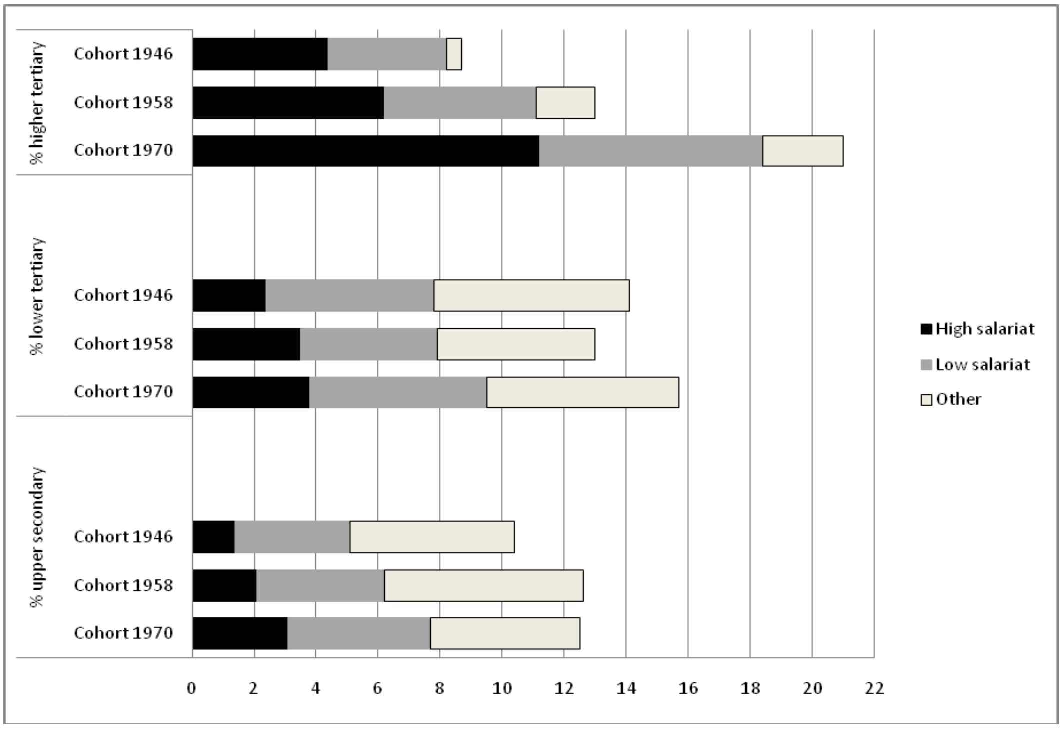
In Britain, as earlier indicated, the historical period to which our research relates saw both an expansion of the salariat and a growth in the numbers of men with tertiary-level qualifications. At the beginning of the 1970s, the proportion of all men in employment who could be counted as members of the salariat was around 25% but by 2005 had risen to around 40%. This change is shown up in the proportion of men in the 1946, 1958 and 1970 cohorts who, at age 34, were found in NS-SeC Classes 1 and 2: that is, 32%, 38% and 49%, respectively - these figures reflecting the well-known tendency for changes in the occupational structure to be effected in large part through cohort replacement.

In 1963 the Robbins Report initiated a rapid growth in the provision of higher education in Britain, both through 'autonomous' universities and through polytechnics, colleges of further education and other public sector institutions - the so-called 'binary system'. The proportion of 18-19-year-olds in higher education was around 7% at the start of the 1960s but by 1990 had risen to 20%, while participation in higher education among older age-groups, especially on a part-time basis, also increased (Halsey 2000). In the 1946, 1958 and 1970 cohorts, the proportions of men with higher tertiary qualifications were 9%, 13% and 21%, respectively, and with lower tertiary qualifications, 14%, 13% and 16%.

To give an initial indication of how these developments relate to each other, Figure 1 shows the proportions of men across the cohorts holding tertiary qualifications - together with, for purposes of comparison, the proportions with higher secondary qualifications - and at the same time the percentage of men with these differing levels of qualification who were found in the higher and lower levels of the salariat as represented by NS-SeC Classes 1 and 2.

It can be seen that for men in the 1946 cohort, higher tertiary (HT) qualifications provide a virtual guarantee of access to the salariat - only 6% of those with degrees or equivalent failing to achieve professional or managerial occupations. For men with HT qualifications in the 1958 and 1970 cohorts, access to the salariat is slightly less assured, but what remains notable is that even as the proportion of all men with such qualifications rises sharply, the very large majority - upwards of 80 per cent - still find positions within the salariat. Moreover, there is no decline across the cohorts in the proportion of men with HT qualifications obtaining positions in the higher salariat.

Figure 1: Percentage of men with differing levels of education and (as shown by boxes) proportions found in the higher and lower salariat, by cohort



As regards men with lower tertiary (LT) qualifications, the story is somewhat different, chiefly in that there is not, as with those holding HT qualifications, a sustained increase in their number across the cohorts. However, the proportion of men with LT qualifications being found in the salariat rises from the 1946 to the 1958 cohort and then holds constant - at around 60 per cent for the 1970 cohort, in which the proportion of all men with LT qualifications is highest. And at the same time, the chances of men with LT qualifications gaining access to the higher rather than the lower salariat are better for men in the 1958 and 1970 cohorts, than for those in the 1946 cohort.

What is then suggested is that, over the period covered, the growth of the salariat meant that demand for higher-educated personnel tended to run ahead of supply. Thus, while having a HT qualification continued to give very good chances of gaining a position in the salariat, even as the numbers of men with this level of qualification increased, the chances of access to the salariat for men with LT qualifications *also* improved. And further, as can be seen from the bottom panel of Figure 1, men with only higher secondary (HS) qualifications likewise had rising chances of entering the salariat as between the 1958 and 1970 cohorts - chances which in fact come to equal those of men with LT qualifications.[6]

Results: access to the salariat

In this section our main concern is with the chances of men being found in the salariat as opposed to other class positions in relation to their level of qualifications, cognitive ability, class origins, and number of different occupations held, up to occupational maturity - all measured as previously indicated. The results of binomial logistic regression analyses are reported in Table 2, in the form of average marginal effects. What the coefficients show is the net effect on the probability of being found in the salariat of a unit change in an independent variable, when averaged over the populations represented by our birth cohorts. We see such population-averaged coefficients as appropriate since, as we earlier emphasised, we aim here essentially at description rather than at inferring the causal effects of educational qualifications or of other variables. At the same time, we do wish to make group comparisons within and across our cohorts, which would be problematic with the more usual subject-specific logit coefficients, on account of potential confounding with residual heterogeneity (Allison 1999; Mood 2010).[7]

In the case of the 1946 cohort, access to the salariat appears in some degree 'meritocratic'. A HT qualification gives a substantial improvement, of around 30%, in the probability of being found in the salariat as compared to a HS qualification, while ability is also a significant independent factor (i.e. over and above its effect via qualifications) and class background is not (i.e. over and above its effects on qualifications and also perhaps on performance on cognitive tests). However, it has at the same time to be noted that men in the 1946 cohort with LT qualifications have no better chances of accessing the salariat than those with only HS qualifications.

With the 1958 cohort changes are evident in various respects. An LT qualification does now give an advantage, of around 15%, over a HS qualification; but class origin effects also become significant, over and above those of ability and qualifications, and chiefly in that men of salariat - i.e. Class 1 and 2 - origins appear to have better chances of accessing the salariat themselves than do men of non-salariat origins. Further, number of occupations held prior to occupational maturity also has a significant, if small, effect.

Table 2. Probabilities of men being found in salariat versus non-salariat class positions by cohort, average marginal effects with standard errors estimated under binomial logistic regression models

	Cohort								
	1946			1958			1970		
Educational qualifications									
lower secondary or less	-0.311	0.025	**	-0.175	0.020	**	-0.280	0.026	**
upper secondary (ref.)									
lower tertiary	0.013	0.032		0.149	0.027	**	0.019	0.031	
higher tertiary	0.318	0.021	**	0.299	0.022	**	0.221	0.019	**
Cognitive ability									
score	0.084	0.010	**	0.086	0.007	**	0.073	0.009	**
missing (dummy)	0.025	0.027		0.015	0.019		0.026	0.015	
Class origins									
class 1 (ref.)									
class 2	0.037	0.054		-0.033	0.029		-0.001	0.029	
class 3	0.033	0.049		-0.058	0.029	*	-0.015	0.039	
class 4	-0.040	0.048		-0.188	0.027	**	-0.065	0.030	*
class 5	0.017	0.046		-0.112	0.027	**	-0.070	0.028	*
class 6	-0.008	0.047		-0.156	0.027	**	-0.074	0.032	*
class 7	-0.071	0.046		-0.149	0.026	**	-0.142	0.029	**
Missing	0.043	0.064		-0.115	0.027	**	-0.075	0.028	**
Number of occupations to maturity	0.004	0.003		0.007	0.002	**	-0.005	0.004	
N	2457			4742			4005		

* Significant at $p < 0.05$; ** significant at $p < .01$

In the light of previous research, focussing on occupational attainment as measured in terms of both earnings and status (Bukodi 2009; Bukodi and Goldthorpe 2009), we would interpret these results as reflecting in large part the distinctive experience of members of the 1958 cohort. These men entered the labour market at a time of severe economic recession, high rates of unemployment (double-digit from 1981 to 1988), and labour market turbulence. Adverse effects are in fact evident in the level of their first occupations and in the greater instability of their subsequent occupational careers, as well as in the occupational level they had attained at maturity. The greater advantage attaching to LT qualifications in this cohort and also the significant effect of occupational changes, we would therefore see as reflecting the fact that LT qualifications more often than either HT or HS qualifications, are acquired in the course of men's working lives and can then provide a basis for upward mobility into the salariat from perhaps quite low-level positions on entry into the labour market.[8] At the same time, it could be expected that in difficult labour market conditions, when the comparative advantages of higher educational qualifications are likely to be reduced (Moscarini and Vella 2008), individuals at all levels of qualification will be likely to look more to their families of origin for support in their working lives, and thus, that the extent of family resources will become more important for individuals' chances of obtaining more desirable class positions.

With the 1970 cohort, it can then be seen that some reversion to the pattern of results with the 1946 cohort occurs. In accessing the salariat, LT qualifications no longer give any advantage over HS qualifications, nor is the effect of number of occupations to maturity significant. However, class origin effects still matter, if to a lesser extent than with the 1958 cohort, and in particular the effects of salariat origins. In this connection it may be noted that men in the 1970 cohort also experienced recession conditions in their early working lives - i.e. during the early 1990s - although with a less severe impact on the labour market than those of the 1980s. Finally, from the standpoint of the 1970 cohort, one important secular trend is suggested: as regards access to the salariat, the relative advantage of HT as against HS qualifications declines across the cohorts - i.e. from around 30% to 20% - even though HT qualifications remain more important than any other factor included in our model.

The question might at this point be raised of how far our findings would differ if we restricted our attention to access only to the higher-level salariat, as represented by NS-SeC Class 1. We have in fact carried out the appropriate analyses (available on request). While, not surprisingly, HT qualifications appear in absolute terms as yet more important than in regard to access to the salariat as a whole, their declining importance across the cohorts relative to HS qualifications is still clearly in evidence. Indeed, the most notable feature of the results we obtain is the extent to which, in their overall pattern, they follow those reported in the text above, including the specific '1958 effects' that we have noted.

We next go on to some elaboration of the results of Table 3, in terms of the probabilities of access to the salariat of different groups as defined by the main independent variables of our logistic regression model (Cox and Wermuth 1996, 115-9; Long 2009). The probabilities are calculated under a version of the model modified in the following ways: ability is treated in terms of quintiles rather than scores, class origins are dichotomised into salariat/non-salariat, and qualifications*ability and qualifications*class origins interaction terms are introduced (and prove to be significant). Number of occupations is evaluated at its mean for each cohort. In Figure 2 we show results for men of salariat origins and in Figure 3 for men of non-salariat origins. Three main points emerge.

First of all, Figures 2 and 3 taken together reflect results already indicated in Table 2. For men with HT qualifications the chances of access to the salariat do not increase from the high levels already existing with the 1946 cohort, while for men with LT qualifications these chances show an improvement between the 1946 and 1958 cohorts that largely holds up with the 1970 cohort. At the same time, for men with only HS qualifications, a decline in their chances of access to the salariat between the 1946 and 1958 cohorts is quite strongly reversed with the 1970 cohort. Overall, then, there is an increasing tendency for men who have only LT or HS qualifications to attain salariat positions. However, what can now further be seen from Figures 2 and 3 is that this tendency is, if anything, more marked among men in the lower ability quintiles. In other words, so far as both qualifications *and* ability are concerned, *the salariat would appear to become less selective over time*.

Figure 2. Predicted probabilities, with 95 % confidence intervals, of being found in salariat positions, by educational qualifications and cognitive ability quintiles, men of salariat background

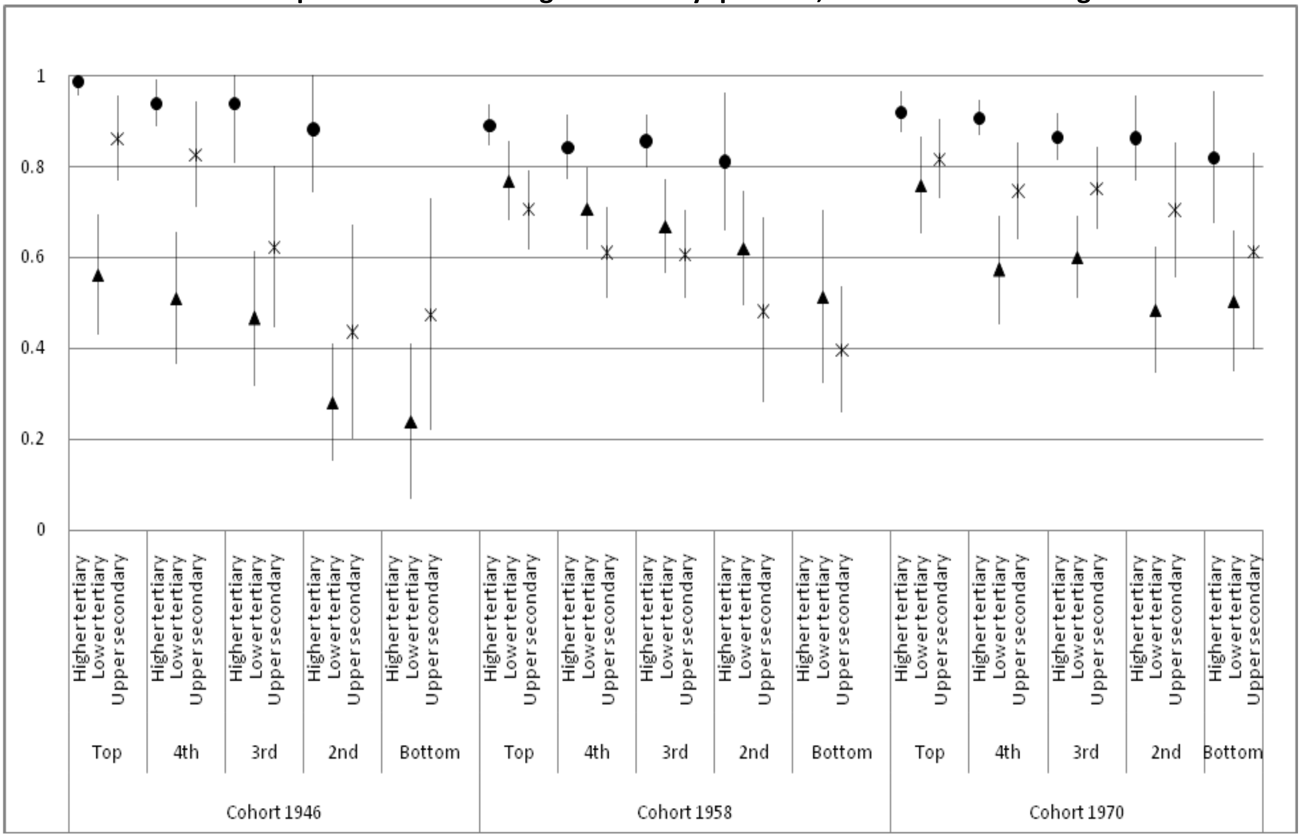
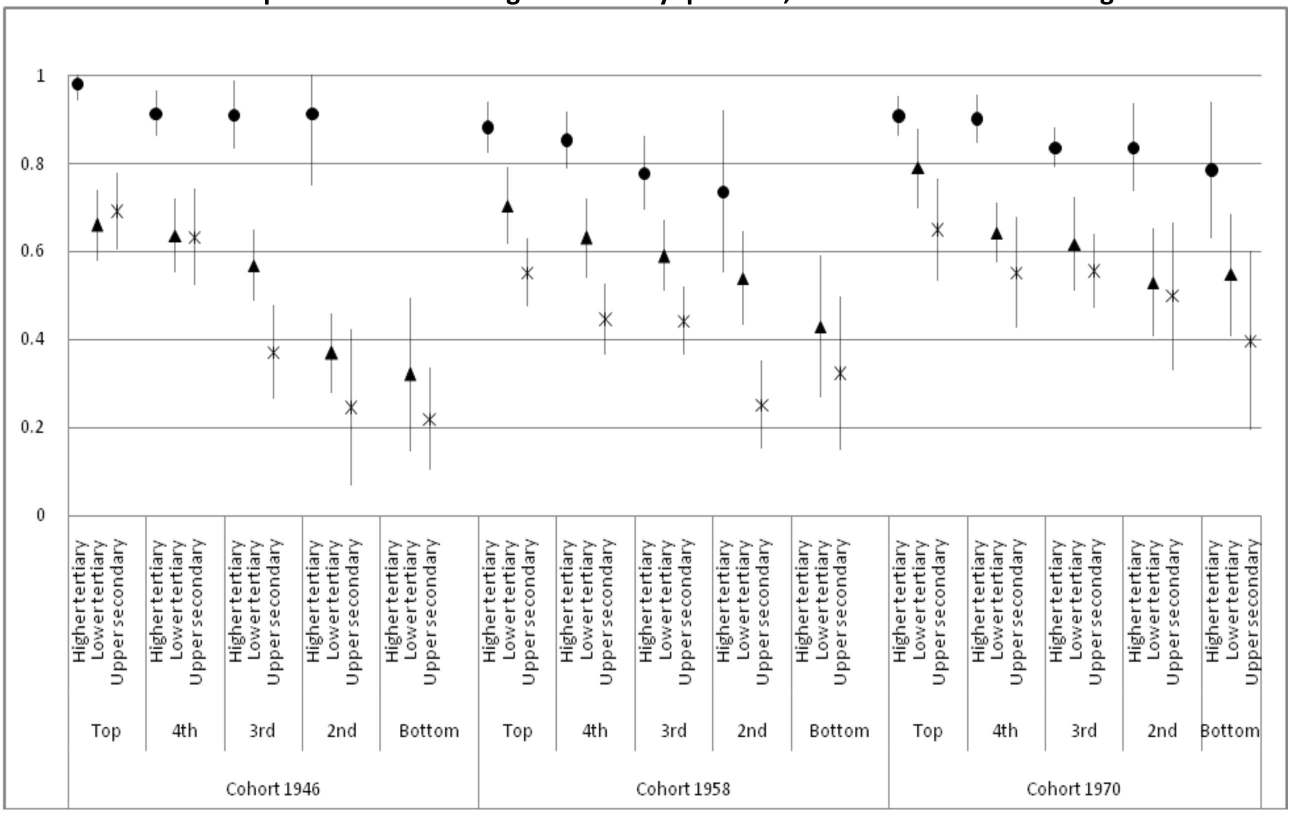


Figure 3. Predicted probabilities, with 95% confidence intervals, of being found in salariat positions, by educational qualifications and cognitive ability quintiles, men of non-salariat background



Second, focusing now on men with HT qualifications, it is evident that in all three cohorts, these men have very good absolute chances of access to the salariat, and that HT qualifications, if acquired, tend to offset the negative effects of both relatively low ability and disadvantaged class origins.[9] HT qualifications are in fact generally associated with a probability of being found in the salariat in the range of 80% to virtually 100% - with the possible exception of men in the 1958 cohort who are in the lower ability quintiles and of non-salariat origins. As regards the relative advantage of having HT as compared to HS qualifications, this is substantial, although tending to be greater at lower ability levels and for men of non-salariat background.

Third, in the case of men with LT qualifications, it can be seen that while they have lower absolute chances of access to the salariat than men with HT qualifications, the difference notably narrows for the 1958 cohort and especially for men of salariat origins. LT qualifications do not appear to compensate for low ability to the same extent as HT qualifications but, as compared to HS qualifications, LT qualifications do appear to give more advantage to men of non-salariat background. Figure 2 shows that men of salariat origins with LT qualifications have in general a lower probability of being found in the salariat than their counterparts with HS qualifications, except in the 1958 cohort. But Figure 3 shows that men of non-salariat origins with LT qualifications have in general a higher probability of moving up into the salariat than their counterparts with HS qualifications in both the 1958 and the 1970 cohorts, and also in the 1946 cohort if they are in the lower ability quintiles.

In sum, one might say the following. In absolute terms, HT qualifications represent the dominant route into the salariat for men in all three cohorts alike, and are less modified in their effects than are LT or HS qualifications, by differences in either individuals' ability or class origins. However, in relative terms, the advantage of HT qualifications declines across the cohorts. For men without HT qualifications, the relation between qualifications and chances of access to the salariat is more complex. For men of non-salariat origins without HT qualifications, LT qualifications - often obtained, we know, in the course of their working lives - give generally better chances of upward mobility into the salariat than do HS qualifications. But for men of salariat origins without HT qualifications, HS qualifications always provide better chances of access to the salariat - i.e. in their case of inter-generational *immobility* within the salariat - than they do for corresponding men of non-salariat origins, and these chances are in fact better than those associated with LT qualifications, except in the case of the 1958 cohort.

Results: access to the professional and managerial divisions of the salariat

In Table 3 we show the results of fitting models analogous to those used in the previous section but with our attention now being limited to those men in the three cohorts who, at occupational maturity, were found in the salariat and with the dependent variable being access to its professional as opposed to its managerial division.

Table 3: Probabilities of men being found in professional rather than managerial positions, by cohort, average marginal effects with standard errors estimated under binomial logistic regression models

	Cohort								
	1946			1958			1970		
Educational qualifications									
lower secondary or less	-0.137	0.058	*	-0.103	0.034	**	-0.111	0.037	**
upper secondary (ref.)									
lower tertiary	0.142	0.043	**	0.125	0.039	**	0.014	0.042	
higher tertiary	0.139	0.047	**	0.236	0.035	**	0.184	0.035	**
Cognitive ability									
score	0.002	0.021		0.002	0.016		0.041	0.015	**
missing (dummy)	-0.046	0.054		0.037	0.039		0.012	0.027	
Class origins									
class 1 (ref.)									
class 2	0.137	0.061	*	-0.094	0.043	*	-0.018	0.038	
class 3	0.218	0.053	**	0.010	0.048		-0.012	0.055	
class 4	0.086	0.064		-0.067	0.073		-0.043	0.045	
class 5	0.202	0.049	**	-0.059	0.047		0.000	0.042	
class 6	0.111	0.061		0.017	0.060		0.017	0.052	
class 7	0.190	0.048	**	0.010	0.047		0.012	0.048	
Missing	0.182	0.077	*	-0.107	0.048	*	-0.014	0.040	
Number of occupations to maturity	-0.019	0.009	*	0.000	0.002		-0.008	0.007	
N	946			1733			1919		

** Significant at p < 0.05; ** significant at p < 0.01*

The positive coefficients across the cohorts for both HT and LT qualifications indicate that, for the men in question, higher educational qualifications in general increase the probability of their being found in professional rather than managerial positions - usually by upwards of 10% to upwards of 20%. However, some differences across the cohorts are also revealed. For the 1946 cohort, the coefficients for HT and LT qualifications are very similar, but for the 1958 cohort that for HT qualifications is clearly stronger, and for the 1970 cohort the advantage given by LT qualifications over HS qualifications is not significant.

Three other points may be noted as also suggesting some change between the professional and managerial divisions of the salariat as regards their patterns of recruitment. First, and perhaps most interestingly, it is only with the 1946 cohort that class origin effects are of importance. In this case, all the coefficients except that for Class 4, that of small 'independents', are significantly positive with reference to Class 1, that of the higher salariat. That is to say, for men in this cohort who gained access to the salariat, coming from a *less* advantaged class background increased the probability - by, it appears, some 10-20% - of their being professionals rather than mangers; or, alternatively put, coming from a Class 1 background distinctively favoured entry into management rather than the professions. That this effect is restricted to the 1946 cohort is the result, we would suspect, of falling numbers of family-run business enterprises or at all events of a declining tendency of sons to enter into the management of such enterprises. Second, in the 1946 cohort, number of occupations held has a significant negative association with becoming a becoming a professional, or, in other words, is positively associated with becoming a manager, but this is not the case in the two later cohorts. This reflects, perhaps, an increasing 'professionalisation' of management and some decline in the chances of men 'working their way up' into managerial positions. And, third, with the 1970 cohort, ability, over and above qualifications, has a significant, even if not very large, positive effect on the chances of becoming a professional rather than a manager.

In Figures 4 and 5 we present probabilities of being found in professional positions for the different groups of men defined by our main independent variables under a model analogous to that used for Figures 2 and 3. The probabilities of being found in managerial positions will of course be the complements of those plotted here.

As might be expected from Table 3, the two figures, for men of salariat and non-salariat backgrounds, show much similarity: i.e. class origins do not appear to have any very large or systematic effects on the chances of men who access the salariat being found in one or other of its divisions. Insofar as a class origin effect is suggested, then, again as would be expected from Table 3, it is in the case of the 1946 cohort.

What of main interest emerges, from both figures alike, is the increase in importance, from the 1946 cohort to the two later cohorts, of HT relative to LT qualifications as regards entry into the professional division of the salariat, and the fact that this increase seems most marked among men in the higher ability quintiles. Thus, one finds that in the 1970 cohort men of salariat and non-salariat origins alike who have a HT qualification *and* who are in the top ability quintile have a very high probability - upwards of 70% - of being professionals rather than managers.[10]

Overall, then, we find various indications that the decline in selectivity into the salariat as a whole that we previously noted has tended to be more marked in its managerial than in its professional component. Over the period covered by our birth cohorts, management may have become more professionalised in the sense of recruitment becoming more dependent on higher level qualifications obtained prior to entry and less on worklife mobility. But similar tendencies are in fact evident among the professions themselves; and insofar as in this period the expansion of the salariat outran the supply of higher educated personnel, it would appear to be the managerial division that chiefly 'took the strain'.

Figure 4. Predicted probabilities, with 95% confidence intervals, of men being found in professional positions by educational qualifications and cognitive ability quintiles, men of salariat background

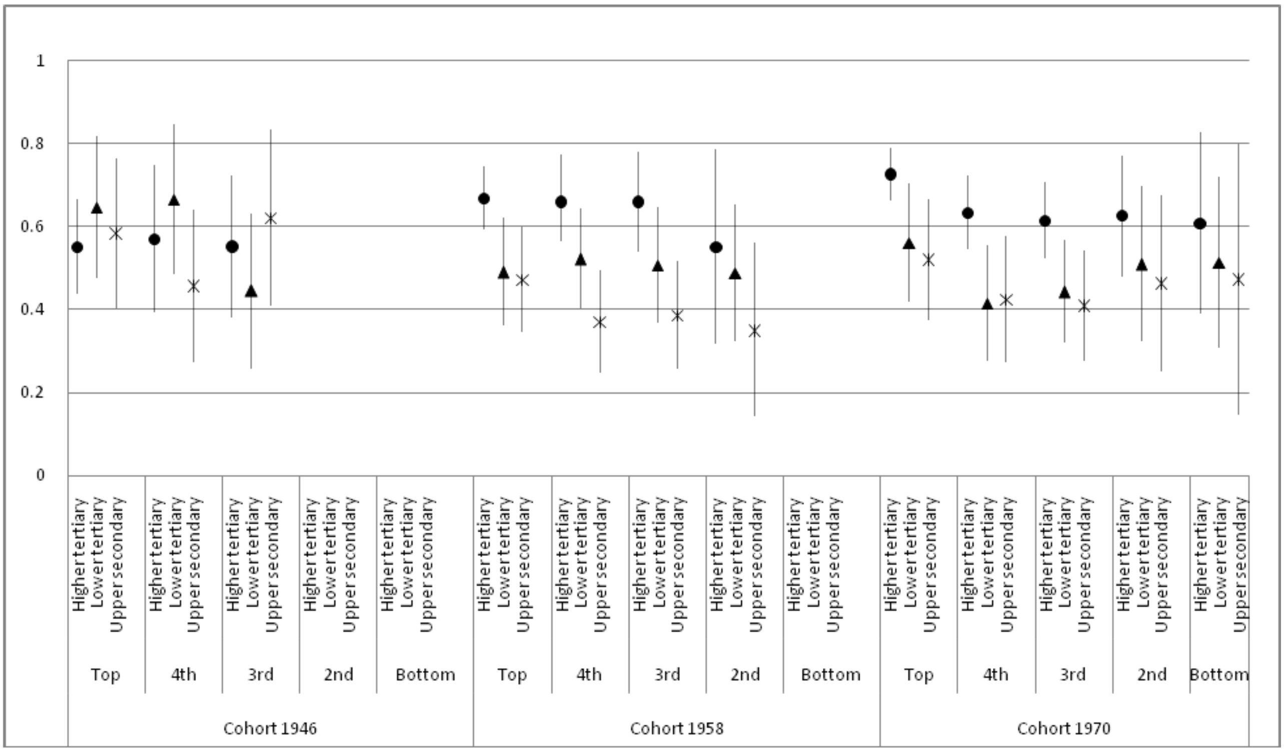
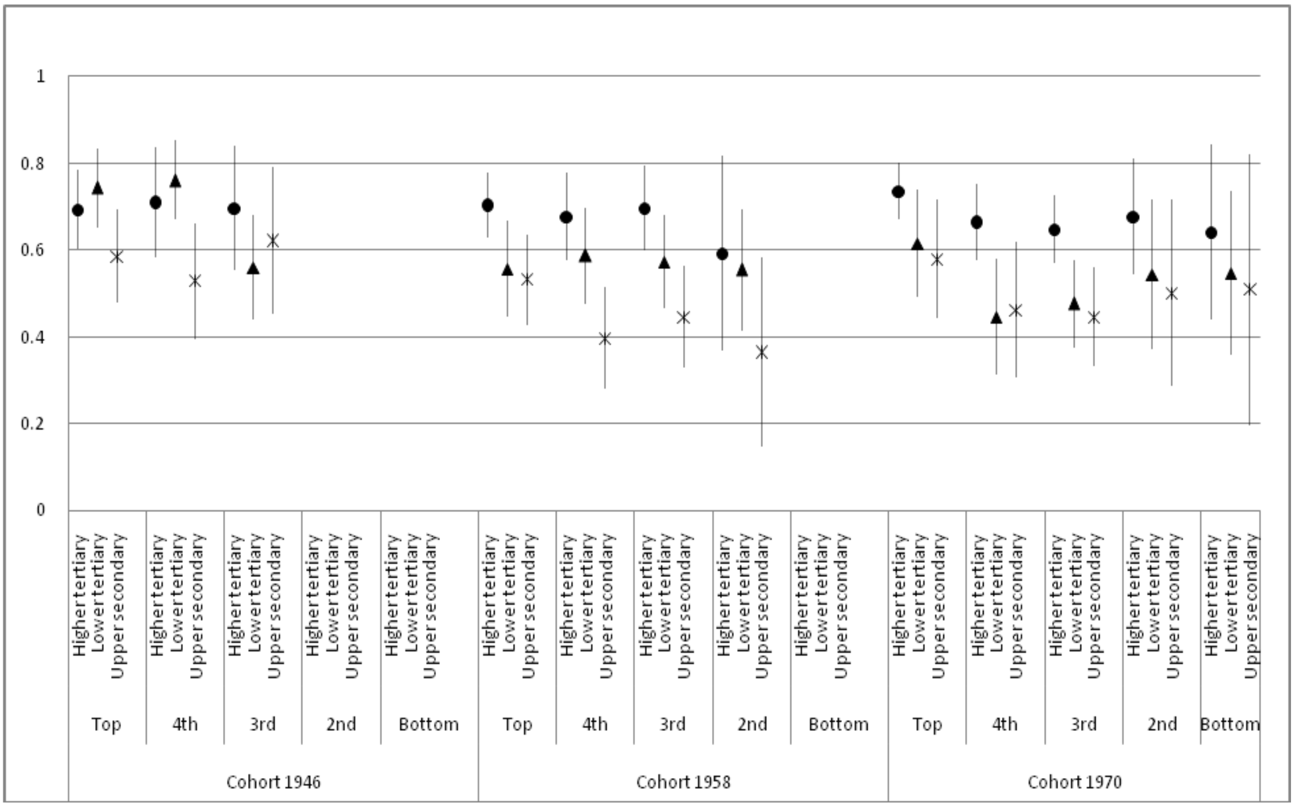


Figure 5. Predicted probabilities, with 95% confidence intervals, of men being found in professional positions by educational qualifications and cognitive ability quintiles, men of non-salariat background



Conclusions, implications and further research

We initially posed three specific research questions. In the light of the results we have reported, we would now sum up our answers as follows.

First, how far, for men in our three cohorts, has the relationship changed between holding tertiary level qualifications and gaining access to the salariat? In this regard, it is important to distinguish between absolute and relative returns. Across the cohorts, absolute returns to tertiary education in terms of access to the salariat generally increased. Men in the 1946 cohort with HT qualifications had a very high probability - virtually a guarantee - of entering the salariat, and for men in the two succeeding cohorts this probability declined only slightly, despite the growing numbers with such qualifications. Moreover, the probability of men with LT qualifications entering the salariat increased sharply between the 1946 and 1958 cohorts and did not then significantly decrease between the 1958 and 1970 cohorts.

The crucial underlying factor here is the growth of the salariat, as effected, one may suppose, in large part by cohort replacement. This growth outran the supply of highly qualified personnel despite the expansion of higher education. Thus, we find that across the cohorts, men with only HS - or even lower - qualifications *also* show increased chances of moving into the salariat. However, since demand is here the driving force, a different story has to be told about the relative returns to tertiary level qualifications. While HT qualifications always give a clear advantage over HS qualifications in the chances of access to the salariat, this advantage declines - by about a third - from the 1946 to the 1970 cohort; and LT qualifications give a significant advantage over HS qualifications only for men in the 1958 cohort, whose distinctive experience under adverse labour market conditions we have emphasised.

Second, how far does our understanding of the - changing - relationship between holding tertiary level qualifications and accessing the salariat have to be modified when individuals' cognitive abilities and social class origins are brought into the analysis? In general, we find that the chances of access to the salariat of men who have achieved HT qualifications are only slightly modified by the independent effects of either ability or class origins. However, a different situation obtains where lower level qualifications are involved. For men of salariat background, LT qualifications give an advantage over HS qualifications in accessing the salariat only in the case of the 1958 cohort; in the other two cohorts HS qualifications are more advantageous at all ability levels. In contrast, for men of non-salariat background, LT qualifications tend to be more advantageous in gaining access to the salariat than HS qualifications. This is generally the case for men in the 1958 and 1970 cohorts and also for those in the 1946 cohort with lower ability levels. In other words, the surest route into the salariat for all men is that via HT qualifications. But LT qualifications - often achieved in the course of working life - tend to give better chances than HS qualifications as regards upward mobility into the salariat, while the reverse applies as regards maintaining inter-generational stability within the salariat.

Further, we can say that there is no evidence of any secular trend across the cohorts in the direction of greater education-based meritocracy (Goldthorpe and Jackson 2008). Class origins do not have a significant independent effect on the chances of men in the 1946 cohort entering the salariat, but with the 1958 cohort, men of salariat background are advantaged over men of non-salariat background and this advantage persists, albeit at a lower level, with the 1970 cohort.

Third, how far do differences arise in the importance of tertiary qualifications and other factors, as regards access to the professional and to the managerial divisions of the salariat? If we focus our attention on those men who have in fact entered the salariat, we see that it is, in all three cohorts, the possession of HT and LT, rather than of lower level qualifications, that chiefly increases the probability of their being found in professional rather than in managerial positions. Class origins and worklife occupational mobility are of additional significance only in the case of men in the 1946 cohort, for whom higher salariat origins and also greater occupational mobility are associated with entry into management rather than the professions.

At the same time, we find that across the cohorts the probability of being a professional rather than a manager is increasingly associated with having an HT rather than an LT qualification, and especially in the case of men with higher ability levels. In other words, as over the period covered, excess demand for higher educated personnel led to some overall decline in the selectivity of the salariat, in terms of both qualifications and ability, it would appear that it is among managers that this decline has been most marked.

The findings reviewed in the foregoing serve, we believe, to show that thinking in terms of the social class returns to higher education, rather than simply earnings returns, and treating ability and class origins as more than factors to be controlled, can provide a larger and more differentiated account of what is involved in the changing relationship between the demand for and the supply of higher-educated personnel. At the same time, though, we are aware that there are a number of ways in which the analyses we have presented will need to be extended in future work.

First, our analyses are obviously incomplete in being limited to men. In order to obtain a full picture of the class returns to higher education, in the context of changing demand and supply, women must be included, which will involve taking up the difficult problems that arise concerning their selection into employment.

Second, we need to widen the range of the independent variables of our analyses. For example, we would like to know more about the part played in access to the salariat, and its professional and managerial components, by different *trajectories* of worklife occupational mobility (Bukodi 2009), especially in conjunction with different kinds of qualification. And further we would think it important to include variables relating to other individual characteristics apart from cognitive ability, such as personality and life-style characteristics (Osborne-Groves 2004; Jackson 2006).

Third, and perhaps most importantly, it has to be recognised that the results we have reported, necessarily reflect conditions that obtained in Britain in a specific historical period - *but conditions that, we know, have subsequently changed in significant ways*. On the one hand, from the last decades of the twentieth century, the rate of growth of the salariat - and especially of its higher level as represented by NS-SeC Class 1 - would appear to have slackened off (Goldthorpe and Mills 2008). On the other hand, in the early 1990s the 'binary' system of higher education came to an end, and a further major expansion began, aimed at the creation of a 'mass' system (Halsey 2000). At the same time, women were increasingly realising their full academic potentialities. By the millennium, the proportion of

18-19-year-olds in higher education had risen to over 30 per cent. We would therefore wish to take our analyses forward in time, to cover men and women born from the 1980s onwards and their careers in a period in which, rather than the demand for higher-educated personnel being in excess of supply, the reverse could be thought more likely the case, with consequent problems of 'over-qualification' (Green and Zhu 2008) and 'bumping down', and in which in turn, recession conditions may have yet more negative effects than previously. The series of birth cohort studies on which we draw in this paper was, unfortunately, interrupted between 1970 and 2000. However, possibilities exist for constructing at least a partially comparable 'quasi-cohort' from alternative data sources, such as the British Household Panel Study (Blanden and Machin 2004), that we are currently exploring.

Acknowledgements

Our research was supported by a small research grant from the British Academy (reference: SG090060). We are indebted to Diana Kuh for access to the MRC NSHD data-set. For helpful comments on earlier versions of the paper, we thank Colin Mills and Ian Plewis.

References

Allison PD. (1999) Comparing logit and probit coefficients across groups. *Sociological Methods and Research*, 28, 186-208.

Bihagen E and Nermo M. (2010) The effectiveness of ESeC and EGP in clustering occupations: a study of occupational wage growth in Sweden. In D. Rose and E. Harrison. eds. *Social Class in Europe*. Routledge/ESA, London.

Blanden J and Machin S. (2004) Educational inequality and the expansion of UK higher education. *Scottish Journal of Political Economy*, 51, 230-49.

Blundell R, Dearden L and Sianesi B. (2005) Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey. *Journal of the Royal Statistical Society*, Series A, 168, 473-512.

Bukodi E. (2009) Education, first occupation and later occupational attainment: cross-cohort changes among men and women in Britain. *CLS Working Paper 2009/4*. Centre for Longitudinal Studies, Institute of Education, London.

Bukodi E and Goldthorpe JH. (2009) Class origins, education and occupational attainment: cross-cohort changes among men in Britain. *CLS Working Paper 2009/3*. Centre for Longitudinal Studies, Institute of Education, London.

Cox DR and Wermuth N. (1996) *Multivariate dependencies: models, analysis and interpretation*. Chapman and Hall, London.

Despotidu S and Shepherd P. (1998) *1970 British Cohort Study: twenty-six year follow-up, 1996*. Social Statistics Research Unit, City University, London.

Elias P and McKnight A. (2003) Earnings, unemployment and the NS-SeC. In. D Rose and DJ Pevalin. eds. *A researcher's guide to the National Statistics Socio-economic Classification*. Sage, London.

Erikson R and Goldthorpe JH. (1992) *The constant flux: a study of class mobility in industrial societies*. Clarendon Press, Oxford.

Ferri E, Bynner J and Wadsworth M. (2003) *Changing Britain, changing lives: three generations at the turn of the century*. Institute of Education, London.

Freedman D. (1997) From association to causation via regression. In VR Kim and SP Turner. eds. *Causality in Crisis*. Notre Dame University Press, Notre Dame.

Goldthorpe JH. (1997) The "Goldthorpe" class schema: some observations on conceptual and operational issues in relation to the ESRC review of government social classifications. In D Rose and K O'Reilly. eds. *Constructing classes: towards a new social classification for the UK.* ESRC and ONS, Swindon and London.

Goldthorpe JH. (2007) *On sociology* (2nd ed). Stanford University Press, Stanford.

Goldthorpe JH and McKnight A. (2006) The economic basis of social class. In SL Morgan, DB Grusky and GS Fields. eds. *Mobility and inequality: frontiers of research in sociology and economics*. Stanford University Press, Stanford.

Goldthorpe JH and Jackson M. (2008) Education-based meritocracy: the barriers to its realisation. In A Lareau and D Conley. eds. *Social class: how does it work?* Russell Sage Foundation, New York.
Goldthorpe JH and Mills C. (2008) Trends in inter-generational class mobility in modern Britain: evidence from national surveys, 1972-2005. *National Institute Economic Review*, 205, 1-18.
Granovetter M. (1981) Toward a sociological theory of income differences. In I Berg. ed. *Sociological Perspectives on Labor Markets*. Academic Press, New York.
Green F and Zhu Y. (2008) Overqualification, job dissatisfaction, and increasing dispersion in the returns to graduate education. *Department of Economics Discussion Papers*, University of Kent, Kent.
Halsey AH. (2000) Further and higher education. In AH Halsey and J Webb. eds. *Twentieth century British social trends*. Macmillan, London.
Hawkes D and Plewis I. (2006) Modelling non-response in the National Child Development Study. *Journal of the Royal Statistical Society*, Series A, 169, 479-91.
Jackson M. (2006) Personality traits and occupational attainment. *European Sociological Review*, 22, 187-199.
Jackson M, Goldthorpe JH and Mills C. (2005) Education, employers and class mobility. *Research in Social Stratification and Mobility*, 23, 3-34.
Long JS. (2009) *Group comparisons in logit and probit using predicted probabilities*. Department of Statistics, Indiana University.
Lucchini M and Schizzerotto A. (2010) Unemployment risks in four EU countries: a validation study of the ESeC. In D Rose and E Harrison. eds. *Social class in Europe*. Routledge/ESA, London.
McGovern P, Hill S, Mills C and White M. (2008) *Market, class, and employment*. Oxford University Press, Oxford.
Mood C. (2010) Logistic regression: why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, 26, 67-82.
Moscarini G and Vella F. (2008) Occupational mobility and the business cycle. *IZA Discussion Paper 3369*.
Müller W and Jacob M. (2008) Qualifications and the returns to training across the life course. In K-U Mayer and H Solga. eds. *Skill formation: interdisciplinary and cross-national perspectives*. Cambridge University Press, Cambridge.
Nathan G. (1999) *A review of sample attrition and representativeness in three longitudinal surveys*. Governmental Statistical Service Methodology Series, 13, London.
Office of National Statistics. (2005a) *The National Statistics Socio-economic Classification: origins, development and use*. Palgrave Macmillan, London.
Office of National Statistics. (2005b) *The National Statistics Socio-economic Classification: user manual*. Palgrave Macmillan, London.
Osborne-Groves M. (2004) Personality and the inter-generational transmission of earnings. In S Bowles and H Gintis. eds. *Unequal chances: family background and economic success*. Princeton University Press, Princeton.
Rose D and Pevalin DJ. eds. (2003) *A researcher's guide to the National Statistics Socio-economic Classification*. Sage, London.
Rose D and Harrison E. eds. (2010) *Social class in Europe*. Routledge/ESA, London.
Schoon I. (2010) Childhood cognitive ability and adult academic attainment: evidence from three British cohort studies. *Longitudinal and Life Course Studies*, 1, 241-58.
Wadsworth M, Kuh D, Richards M and Hardy R. (2005) Cohort profile: the 1946 National Birth Cohort (MRC National Survey of Health and Development). *International Journal of Epidemiology*, 35, 49-54.

Endnotes

[1] We do in any case have doubts about the approach to causation that the economists follow: specifically, about whether the 'parameter of interest' can be regarded as having 'a life of its own' outside of the data from which it is estimated (cf. Freedman, 1997).

[2] As with all longitudinal studies, the problem of missing data arises. However, a number of analyses of attrition and non-response have been undertaken and the results are generally

encouraging in suggesting that no major biases are being created (Despotidu and Shepherd 1998; Nathan 1999; Hawkes and Plewis 2006; Wadsworth et al 2006).

[3] Kaplan-Meier survival estimates made under our model show that up to around age 30 very few men appear as having achieved occupational maturity but that, after this age, the proportion increases rather sharply. However, while at age 34 over 80 per cent of men in the 1958 and 1970 cohorts are treated as having reached occupational maturity, this is the case with only 60 per cent in the 1946 cohort (see further Bukodi and Goldthorpe 2009).

[4] It should be noted that, in some contradiction with the term 'salariat', large employers - i.e. employers with over 25 employees but who are not employees of their own incorporated businesses - are included in the managerial division of Class 1; and that self-employed professionals are included in the professional divisions of Classes 1 and 2. However, such large employers, mainly proprietors of construction firms, garages, stores etc, are a very small minority - around 5 per cent - of all in Class 1, while in the case of professionals, self-employment is often, as, say, with GPs, Church of England clergy or some financial professionals, an essentially technical status reflecting legal or fiscal considerations.

[5] The quality of information on fathers' occupations and employment status in the 1958 cohort is less good than in the other two cohorts. In this case, we therefore proceed by first taking their Socio-Economic Group codings which are available and from which a reasonable approximation to NS-SeC can be derived, and by then improving on this approximation as regards NS-SeC Classes 5 and 6 by cross-classifying SEG codings with codings to the Registrar General's Social Classes which are also available. Full details can be obtained from the authors on request.

[6] In fact, if we look at men in our fourth, residual educational category (i.e. men with only lower secondary qualifications at best) the proportion entering the salariat can again be seen to increase across the three cohorts - from 18% to 21% to 27%.

[7] If we were to use these latter coefficients as the basis for such comparisons, we would not, as it happens, be led to conclude anything very different from what is said in the text below - suggesting that, under the model we use, problems of residual heterogeneity are not severe.

[8] For example, among men in the 1958 cohort the proportion holding LT qualifications increases from 8 per cent at age 24 to 12 percent at age 34, while the proportion holding HT qualifications increases from 10 per cent to 12 percent and the proportion holding HS qualifications stays constant at 17 per cent.

[9] It may, however, be noted that in both Figure 2 and Figure 3 no points are recorded for men with HT qualifications in the lowest ability quintile since in both cases the numbers involved are negligible.

[10] We have undertaken analyses restricted to higher-level - i.e. Class 1 - positions within both the professional and managerial divisions of the salariat, and again these show no distinctive features, apart from the yet greater importance for access of HT qualifications.

Longitudinal and Life Course Studies 2011 Volume 2 Issue 2 pp 202 - 227 ISSN 1757-9597

Educational attainment, labour market conditions and the timing of first and higher-order births in Britain

Andrew Jenkins
Institute of Education, University of London
a.jenkins@ioe.ac.uk

(Received December 2010 Revised March 2011)

Abstract

This paper presents analyses of the effects of women's education and the rate of aggregate unemployment on birth hazards using data from the 1958 and 1970 British cohort studies. The hazard of first birth was negatively associated with higher levels of education. Once controls for unobservables were included, there was no relationship between education and the hazard of second births. Lagged unemployment was found to be negatively related to first birth hazards but this was only statistically significant among the later cohort, while for higher order births there was evidence of a positive association with unemployment.

1. Introduction

This paper analyses fertility in Britain, with special reference to the effects of women's education and the state of the labour market. The research uses longitudinal micro-data on two cohorts who had different experience of education, combined with macro-data on labour market conditions to examine how these factors impacted on the timing of births. Hazard models were estimated separately for each of the two cohorts. The models analyse the timing of the first three births, focusing on the associations of birth hazards with education level and a time-varying unemployment covariate.

The research develops the literature on this topic in several important ways. Both first births and higher-order births will be considered. Much of the literature has looked just at the timing of the first birth, which gives a very incomplete picture of fertility as a whole. Secondly, many published papers do not allow for unobserved heterogeneity. This is vital as there may, for example, be unobserved characteristics of women which drive their decisions about both education and fertility. Controlling for such heterogeneity in the sample is then crucial in order to attempt to make inferences about causation rather than just associations among the variables, and our estimates use techniques which allow us to control robustly for unobserved heterogeneity. Thirdly, I use rich data on two cohorts of women which allow a full range of relevant controls, including factors from the childhood of the cohort members, to be incorporated into the model.

2. Literature Review

In Britain, as in other developed economies, successive cohorts have tended to wait longer before starting a family. The stylized facts are that, among women born in England and Wales in the 1950s, fewer than a quarter were still childless by age 30, but for women born in the early 1970s, about 40 per cent were still childless by the time they reached 30. In earlier cohorts, too, a higher proportion of women became mothers by the end of their reproductive lives. Around 13 per cent of women born in 1950 remained childless. This rose to 18 per cent of women born in 1960 and it is estimated that about the same proportion of those born in 1970 will not have children (ONS 2007; Bray 2008). The statistical association between education and the timing of first births is quite well-established. What is less clear is the more difficult question of whether education can be said to have any causal link with birth hazards; the evidence on the direction of the relationship between education and higher-order births has also proved more mixed than that for first births.

Education of women is widely regarded as a central factor in the trend to fewer and later births in developed economies. A higher level of education is associated with beginning child-bearing at a later age and with fewer children, on average, by the end of a woman's reproductive years. This can be explained very broadly in terms of the greater opportunity cost of forgone earnings, which will be higher for well-educated women who have greater earning power. This would be a rationale for avoiding childbearing while still studying, as well as for deferring, if not avoiding, it once on the labour market (e.g. Hotz et al 1997). However, education increases income, through own earnings and possibly through assortative mating, which could have a positive effect on fertility, reducing or outweighing the substitution effect, especially at later ages and stages in the reproductive span (Gustaffson 2001). In particular, higher earning power may make it easier to afford the costs of reducing forgone earnings through the purchase of childcare (Ermisch 1989) and of owner-occupied housing. Another route which connects low education to early and extended childbearing, is that women who are (or will be) better educated, are better equipped to avoid unintended births.

It is well-established in the demographic literature, that a higher level of education is associated with later timing of the first birth (e.g. Gustafsson 2001). The relationship between educational attainment and transition rates to higher order births remains less clear. The empirical evidence is mixed but several studies have reported that women with higher levels of education have made more rapid transitions to second and/or third births. Kravdal (1992) found a positive association between education and third births for Norway, while Kreyenfeld (2002) reported a similar association for second births in Germany. For Britain, Wright et al (1988), using data from the 1980 Women and Employment Survey, found no evidence that education exerted any influence on progression to third births.

More recently, Rendall and Smallwood (2003) examined parity progression by education level, using data from the Office for National Statistics (ONS) Longitudinal Study for women born between 1954 and 1958. They presented descriptive findings rather than model-based analyses but the results are, nonetheless, interesting. Average age of entry to motherhood was some five years later for highly qualified women but, for any given age at birth of first child, highly qualified women were relatively more likely to have another child and tended to do so more quickly than less well educated mothers. Aassve et al (2006) use data from the British Household Panel Study (BHPS) and estimate a sophisticated model with simultaneous hazards for births, union formation and dissolution, employment and non-employment events, in the presence of unobserved heterogeneity. On fertility, they find a negative relationship between education and the birth hazard (constrained to apply equally to each order). As the authors acknowledge, the complexity of their approach is bought at some cost in terms of the specification of each individual process; notably their specification does not allow for different processes for each birth i.e. the effect of education is assumed the same for each birth. In summary, then, there is overwhelming evidence that better educated women delay entry into motherhood. Some research has found that, once they begin childbearing, well-educated women proceed relatively quickly to second and higher-order births.

Cohorts reaching adulthood since the early 1970s have also experienced a labour market in which unemployment rates have been at times exceptionally high and in general, volatile. This applies both in Britain and in many other European economies. In the case of Britain, unemployment in the 1980s and early 1990s reached levels which had only previously been observed in the depressed decade of the 1930s (Crafts 2007). After the mid-1990s, unemployment fell back to much lower levels. High unemployment might either deter or promote births, depending on whether the income effect of not being able to afford a(nother) child, dominates the substitution effect of the reduction in the mother's earning opportunity costs.

A number of papers have considered the relationship between unemployment and fertility. Ahn and Mira (2001) analyse the relationship between fertility and aggregate unemployment in Spain in the 1970s and 1980s. The Spanish (male) unemployment rate was below 5 per cent in the mid-1970s, but climbed to around 20 per cent in the 1980s. The authors find that unemployment increased the average age at marriage. This higher age at marriage also reduced fertility (they consider the timing of the first three births) although the estimated effects of joblessness on birth hazards conditional on marriage were not statistically significant. Gutierrez-Domenech (2002) applied Cox hazard regression models to the timing of marriage and of births amongst two cohorts of Spanish women. The lagged unemployment rate was negatively and significantly related to the transition into marriage in both cohorts. The birth hazard models were fitted separately for each of the first three births and it was found that, after controlling for other factors, lagged unemployment was negatively associated with the hazard of each birth. This relationship was statistically significant for the first two births to the later cohort (born 1961-77) but was never significant for the earlier cohort (born 1945-60). Kreyenfeld (2000) used data from the German Socio-Economic Panel to study the relationship between unemployment and fertility in the former East Germany following unification. She focused on individual-level unemployment, i.e. whether the woman was herself unemployed at the time of conception, rather than the effect of aggregate unemployment. In a piecewise constant hazard model of first births, it was found that unemployed women were more likely to conceive. The model included age and education level as controls, and the positive association with unemployment applied only to women educated below degree level and not to those with degrees. Del Bono (2001) explored whether unemployment affects fertility through its influence on expectations of the future condition of the labour market. This study used data on a single cohort of young British women, observed up to age 33. It was shown that more favourable expected job opportunities raised the hazard of the first birth.

Studies of this topic for Scandinavian countries include Hoem (2000), who used data on Swedish women born in 1950 or later to analyse times to first birth in the 1980s and 1990s. It was found that the employment rate in the women's local municipality was positively associated with time to first birth in hazard regression models. Hoem used register data and so had few controls for the women's family background. Santow and Bracher (2001) drew on data from the 1992 Swedish Family Survey with a better range of controls, but obtained broadly similar results. They also applied hazard regression models to the time of conceptions leading to first birth. The age-specific unemployment rate was negatively related to the time of first birth conception. The estimated unemployment effects were quite substantial: relative to the base case of unemployment below four per cent, when unemployment was between four and nine per cent, first birth conception rates were reduced by a fifth, and when unemployment exceeded 10 per cent first birth conception rates were lowered by two-fifths. Santow and Bracher also report that their results were largely unaffected by lagging the unemployment variable by one or two years. Kravdal (2002) estimated hazard models separately for first and higher-order births based on Norwegian register data for the period 1992 to 1998. All unemployment variables were lagged by 12 months and included individual unemployment as well as both male and female unemployment rates at municipal level. The woman's own unemployment raised the hazard of the first birth 12 months later, but reduced the hazard of higher-order births. The municipal unemployment rates, both male and female, were associated with lower birth hazards for first and higher-order births. During the period studied, unemployment in Norway varied only between a minimum of two per cent and a maximum of six per cent and so was quite low compared to many other European economies.

Dex et al (2005) fitted Cox hazard regression models to cohort data on time to first birth in Britain, Sweden and the United States (US). Unemployment was measured as aggregate male unemployment rate. Higher unemployment was associated with a significantly lower hazard of motherhood in Sweden and the US but a significantly higher hazard in Britain.

One of the notable features emerging from the literature review, is that studies of Scandinavian countries have tended to find a pro-cyclical relationship there with birth hazards. In good economic times, when unemployment is low, the birth hazard tends to increase. That is, the income effect appears to be out-weighing the substitution effect, and this, it may be conjectured, is related to the welfare state in Scandinavian countries, with generous support in cash and kind for those combining parenthood and employment, and low private opportunity costs of parenthood.

This review of the literature also suggests that most analyses of aggregate unemployment and fertility have tended to focus just on the first birth, and few papers have reported on higher order births. As for education and fertility, while many studies report negative associations between education and first birth, it is far from clear whether education has any causal impact or whether unobservable factors are determining both fertility and education levels of women.

3. Research Questions and Method

In a stylized economic model of fertility, abstracting from issues about partnership, it is usual to assume that having a child yields utility to the mother, but that there is also disutility from earnings lost (including the present value of lost future earnings) during time spent out of the labour force while bearing and looking after the young child. So the observed fertility and labour market behaviour of women will depend, in broad terms, on the strength of their preferences for children versus market work, and on the wages and employment opportunities available in the paid labour market (Dex et al 2005). In this framework, education could have a number of effects. Perhaps most important, acquiring more education raises the market wage, encouraging participation in paid work. On the other hand, education may increase home productivity too, thereby encouraging women to look after children in the home. There is also an income effect – higher earnings of educated women, especially lifetime earnings, make more children potentially affordable. So it is not possible to determine on the basis of theory alone, the direction in which education will influence fertility. From an empirical perspective a further key issue is whether an observed association between education level and fertility can be taken to be a causal effect, or whether it is in fact just reflecting heterogeneity in the population. For example, some women with a strong preference for high earnings and relatively little preference for children, may choose both to invest a lot of time in education and to have few, or no children; here education is not having any causal effect on their behaviour. It is vital, then, to make allowance for such heterogeneity when building good models.

For aggregate unemployment, as with education, while theory provides a useful framework for thought, it does not provide firm predictions about the direction of effects. Increases in aggregate unemployment may have a substitution effect, encouraging women to have children while prospects for paid work are poor. This is closely related to the 'discouraged worker' effect, where people drop out of the labour market during adverse times. Conversely, there is an income effect: in a recession it will become harder to earn sufficiently to afford children, so that women are encouraged to take jobs, work longer hours or anticipate that they may need to do so. This corresponds to the 'added worker' effect (for discussion of literature on added and discouraged workers, see e.g. Killingsworth and Heckman 1986). What is actually observed will depend on which of these effects – the discouraged or the added worker effect - is the stronger.

As noted in the literature review in countries such as Sweden, which have generous support for working mothers, birth hazards have a pro-cyclical pattern. In Britain, maternity leave arrangements and other support for working mothers has tended to become more generous over time (see the information in Appendix 1 for details). So it may be the case that more recent cohorts in Britain have exhibited a greater tendency to have pro-cyclical birth hazards than earlier cohorts. Since different cohorts do not experience the same labour market at the same age, it is not possible to test this formally, but it is of interest to consider whether the results which will emerge here are broadly consistent with such a pattern.

So the research questions are, firstly, is there an association between education and the hazard of births? Secondly, to consider any differences between first and higher order births. Thirdly, the objective is to control for unobserved heterogeneity among the sample of women, enabling firmer inferences to be made about the causality of education[i]. Fourthly, what impact does aggregate unemployment have on fertility behaviour – does the added worker effect dominate the discouraged worker effect, or *vice versa*? And finally, as there are strong trends in the fertility behaviour of women in Britain, the research will consider two cohorts of women, 12 years apart, and determine whether or not the results differ between these two cohorts, after allowing for a range of observed predictors.

In terms of appropriate methods for the analysis, it is necessary to allow for the fact that not all women had a birth by the time they were most recently observed in the data. In other words, some women's birth histories were censored. Duration modelling is now well-established as the appropriate technique to deal with the analysis of times to an event in the presence of censoring (Allison 1984; Kiefer 1988). The basic insight is rather than focusing on factors directly affecting occurrence of the event, instead to look at factors which influence the risk of the event occurring (Newman and McCulloch 1984). Duration models were applied to the data on births. Here, interest centres on the probability that a person who has occupied a state for a certain length of time t leaves it in the next short interval of time. Formally, the hazard is defined as:

$$h(t) = \lim_{dt \to 0} \frac{P(t \le T < t + dt \mid T \ge t)}{dt} \qquad (1)$$

The hazard of a birth at time t given that it has not occurred prior to t will be estimated. Letting *f(t)* be the probability density function and *S(t) = 1 - F(t)* the survivor function, then the hazard function is also often written as:

$$h(t) = \frac{f(t)}{S(t)} \qquad (2)$$

Duration models were estimated separately for each of the two cohorts on which I have data. An exponentiated quadratic was used as the functional form for the baseline hazard. This is preferable to other commonly used functional forms for the hazard, such as the Weibull model, as the quadratic allows for the possibility of a non-monotonic hazard, which is plausible when modelling the hazard for births. Explanatory variables in the model include education, unemployment as a time-varying covariate, and other factors, which economic theory and the empirical research literature suggest may be important, and which are described in more detail in the data section of the paper.

One of the advantages of longitudinal data is that it should be possible to control for omitted variables and unobservables much more effectively than when using cross-sectional data (Davies 1987). There has been considerable debate in the literature on duration models, on the best way to take account of unobservables. It is well known that neglecting to control for the presence of unobserved heterogeneity can lead to mis-specification of the baseline hazard, and that this could in turn bias the parameter estimates on explanatory variables (e.g. Blossfeld et al 2007). A widely-used method for taking account of unobservable factors is to assume a parametric distribution for the heterogeneity, and this distribution is usually chosen as some convenient functional form which will make the resulting mixing distribution analytically tractable (Lancaster 1990). Unfortunately, as Heckman and Singer (1984) note, empirical results can be sensitive to the functional form chosen for the parametric heterogeneity term. They proposed a non-parametric maximum likelihood procedure, in which the distribution of unobservables is approximated by a discrete distribution, and both the probability masses and their locations are estimated from the data. I also adopt this non-parametric approach and write the jth conditional hazard, h_j, for the jth birth as:-

$$h_j = \exp\{\gamma_{0j} + \gamma_{1j}t_j + 0.5\gamma_{2j}t^2_j + Z\beta_j + f_j\theta\} \qquad (3)$$

where t_j is the length of the jth spell; Z is a vector of covariates, which may include time-varying covariates; θ is the person-specific unobserved heterogeneity component; and the γ, β, and f terms are transition-specific parameters to be estimated. The first spell begins at age 16; subsequent spells begin at the time of previous birth plus nine months. All the estimations were carried out in the specialist statistical program for duration modelling, CTM (Continuous Time Models). The distribution of the unobservable term was estimated using the Heckman- Singer non-parametric maximum likelihood procedure. Here a one-factor structure is assumed, such that the unobservable for spell j is $f_j\theta$ and the covariance between $f_i\theta$ and $f_k\theta$ is $f_i f_k Var(\theta)$. By modelling the unobservables in this fashion, I allow for the unobserved heterogeneity to be correlated across spells. To obtain the estimates of the non-parametric distribution, I began by estimating the location and weights to be placed on just two mass points and continued to add mass points until two converge on the same location.

4. Data

The analyses of fertility use data from two British birth cohorts: the National Child Development Study (NCDS), a cohort of individuals all born in the same week in March 1958 and the British Cohort Study (BCS70), who were all born in a single week in April 1970. Members of each cohort have been surveyed at various points in their lives. For NCDS, detailed birth histories were collected in 1991 when cohort members were aged 33, in 2000, when they had reached the age of 42, and again in 2004 at the age of 46. For this project the data from the 2004 survey were combined with data from the 1991 and 2000 NCDS sweeps, making the fertility histories virtually complete. Data on birth histories for BCS70 were also collected in 2000 and 2004. Information from these two sweeps was joined together, taking the record up to age 34. Some women with incomplete birth history data were omitted from the quantitative analysis. The main omitted group was those NCDS women for whom information was only available on births which occurred from age 33, but not before that age. In other words, some left-censored cases were omitted. For both cohorts, cases where mothers had given birth to twins or triplets were also omitted. For NCDS, the sample used for analysis consists of 5,631 women and there are 5,105 BCS70 women in the analyses. For over three-quarters of the NCDS sample, there is a full birth history up to age 46, while for a further 15 per cent there is a history up to age 42, and for the remaining nine per cent, a birth history which is truncated at age 33. For over four-fifths of the BCS sample, there is a full birth history up to age 34, while for the remaining 17.5 per cent, a birth history which is truncated at age 29 or 30.

Explanatory variables

Education is widely regarded as a key factor in understanding fertility behaviour and it was important to include it in the analysis. There are various ways of conceptualising education, each of which has some advantages and some disadvantages. Here, education attainment was treated as a fixed variable, based on years of completed education by age 30. Using a fixed covariate simplifies the specification and effectively treats the destination education level as if it were anticipated. An alternative would be to treat education as a time-varying covariate. In practice, relatively few women in the datasets substantially increased their years of completed education after their teens or early twenties. Our specification effectively rules out the possibility that low educational attainment is the result of early motherhood. There is evidence to support this. Studies of women who have children at a young age, suggest that early motherhood is a marker rather than a driver of subsequent disadvantage in the labour market. For example, Ermisch and Pevalin (2003, 2005) analysed the 1970 British birth cohort data and found that having a child as a teenager had little effect on a woman's qualifications, earnings and employment at age 30. Hawkes' (2003) study on British twins, showed that the apparent effect of early motherhood on educational attainment was much smaller once antecedent factors had been controlled for. The conclusion from this is, that it is not unreasonable to treat women's education as a fixed covariate. For our analyses, education was categorised as low (11 years of completed education by age 30), medium (12 or 13 years of education) and high (more than 13 years of education). Leaving school at age 16, the minimum school leaving age for both cohorts, would imply 11 years of education so the women in the low education category have no time spent in education beyond the minimum. Having 12 or 13 years of education would mean some secondary education beyond the minimum, but no tertiary education;

those in the high education category have more than 13 years of education, so would usually have some tertiary education. Descriptive statistics on the education levels of women in the two cohorts are shown in Table 1. Among the NCDS women, over two-thirds were at low education level, approximately 17 per cent had medium education and 15 per cent had a high level of education. Larger proportions of BCS70 women were reported having education at the medium or high levels, reflecting the secular increase in enrolment and attainment (Makepeace et al 2003). Among the BCS70 cohort of women, about half were classified as low education, of the remainder, slightly more were in the medium education category than the high education category.

Table 1: Education Levels of the NCDS and BCS70 women

Education Level	**NCDS**		**BCS70**	
	1958 cohort		**1970 cohort**	
	N	%	N	%
Low	3,831	68.0	2,572	50.4
Medium	940	16.7	1,340	26.3
High	860	15.3	1,193	23.4
TOTAL	5,631	100.0	5,105	100.0

In Figures 1 to 4, Kaplan-Meier survival curves are plotted for the first two births among each cohort by education level, to illustrate how the timing of births differs for women with differing amounts of education. For the first birth, measured from age 16, Figure 1 and Figure 2 both show very clear differences in survival profiles by education level, with the highly educated taking longer to have a first child than those with a medium level of education, who in turn tend to take longer to begin child-bearing than those whose education level was categorised as low. Comparing Figure 1 and Figure 2, the earlier cohort, NCDS, tend to make the transition to motherhood at younger ages than those in the more recent cohort, BCS70. For the second birth, measured in months since first birth, there is some indication that women with high education make a more rapid transition to second birth although the survival curves for each level of education are very close together (Figures 3 and 4).

Figure 1. Kaplan-Meier Survival Curve for First Birth by Education Level - NCDS

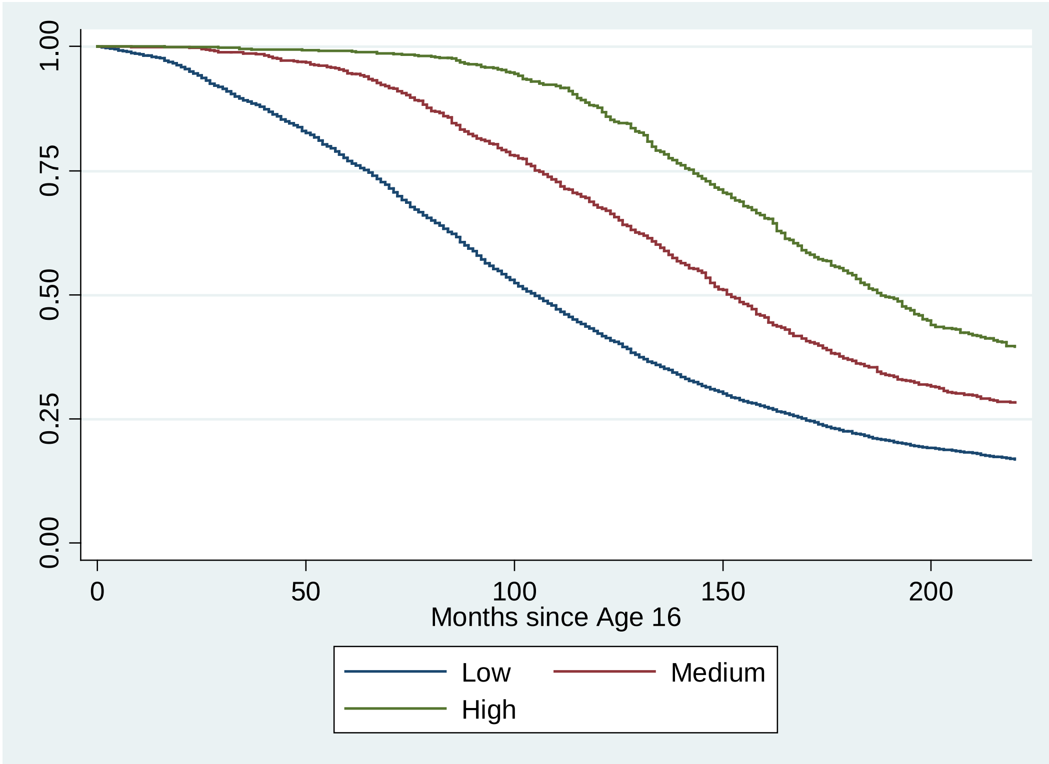


Figure 2. Kaplan-Meier Survival Curve for First Birth by Education Level BCS70

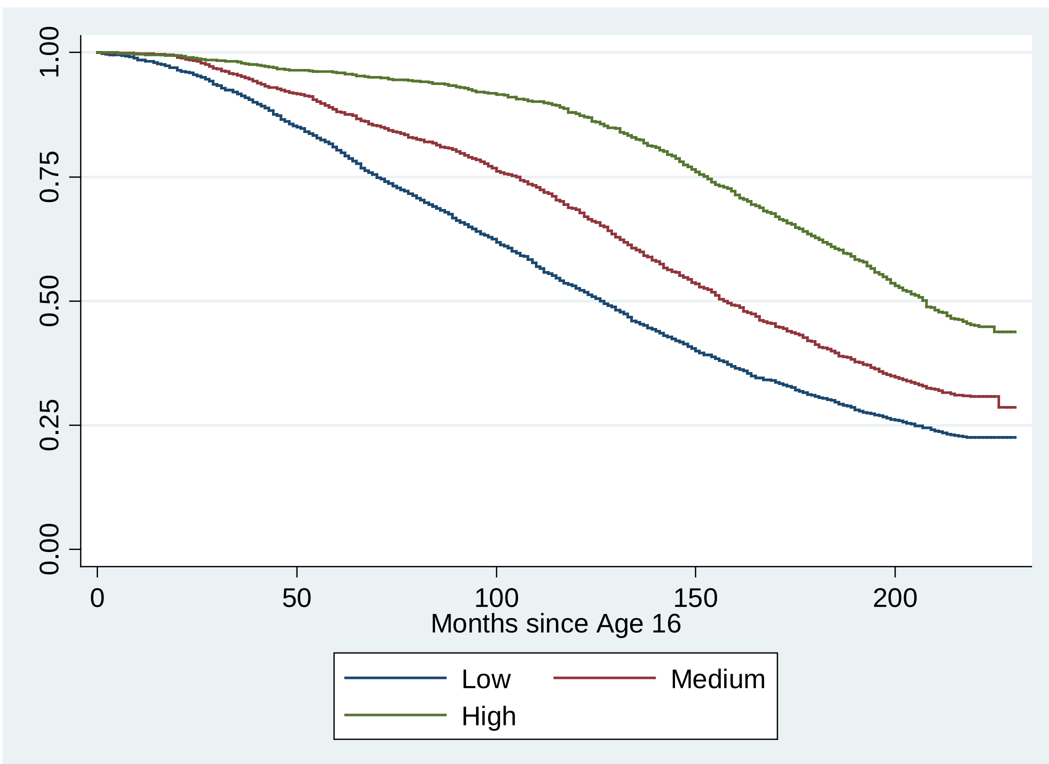


Figure 3. Kaplan-Meier Survival Curve for Second Birth by Education Level – NCDS

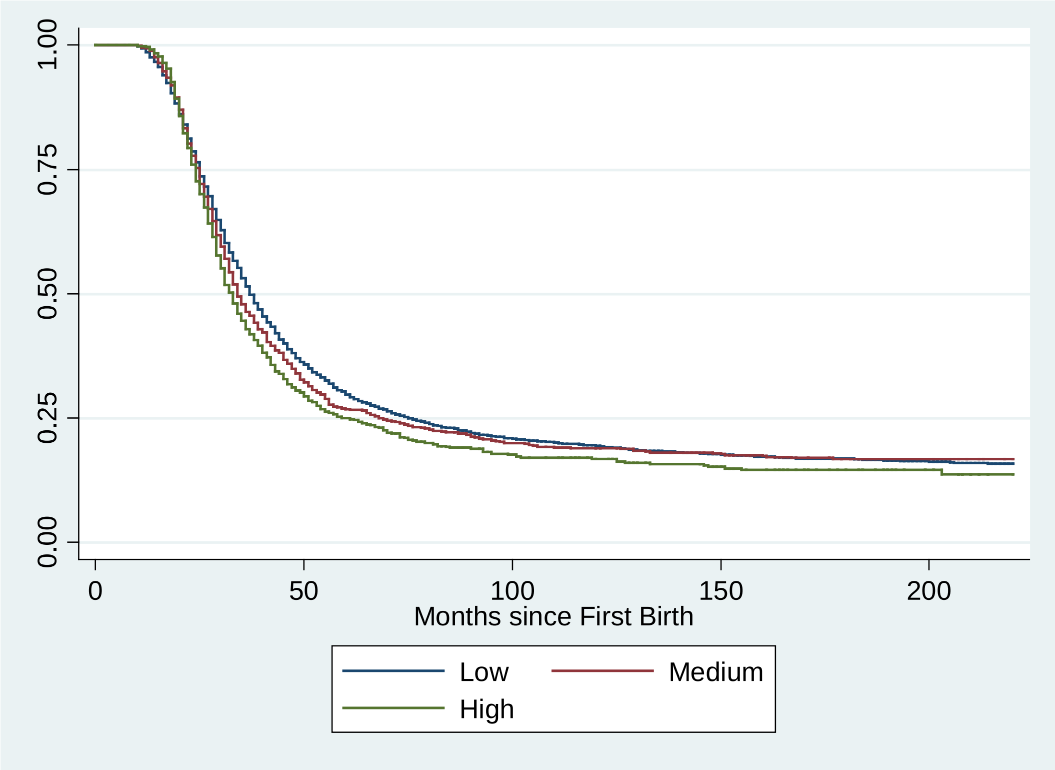
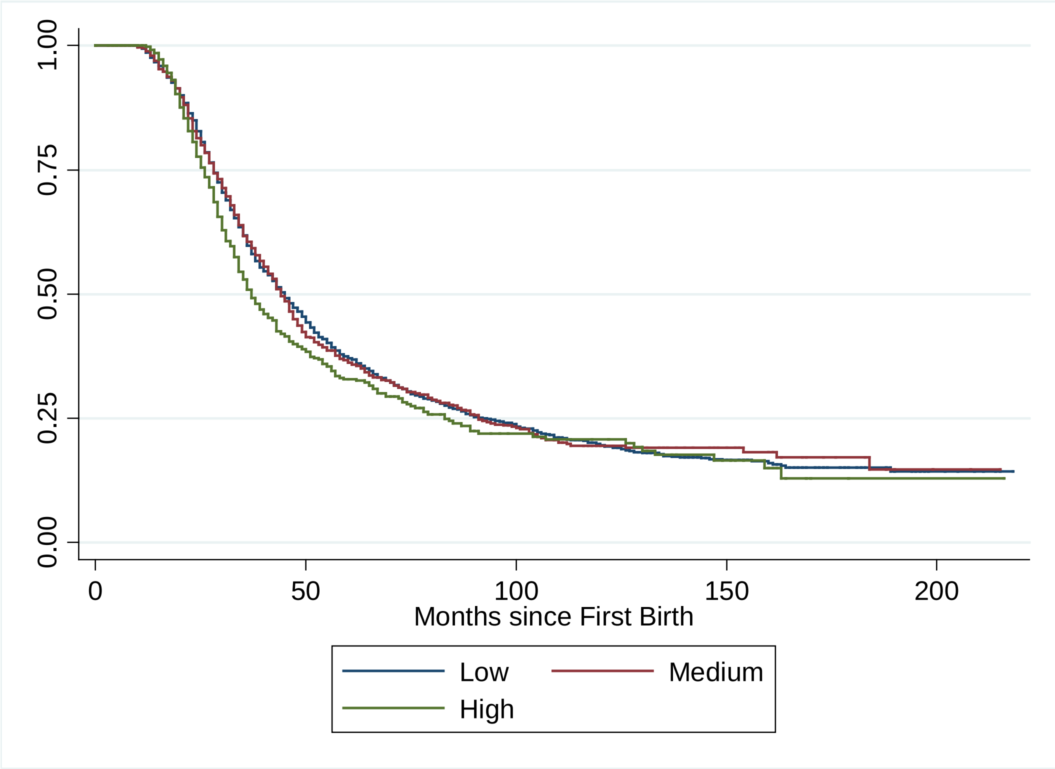


Figure 4. Kaplan-Meier Survival Curve for Second Birth by Education Level – BCS70



As a monthly series for births is being used, it is best to utilise a monthly series on unemployment, and here there were some tricky data issues, as for unemployment a monthly series is required which goes back to the early 1970s, when the 1958 cohort began to enter the labour market. The claimant count is the only series which meets this criterion. However, one serious problem with the claimant count is that it is affected by changes to the rules for eligibility to unemployment benefits. When unemployment was very high in the 1980s, several changes were made to the eligibility rules. I have therefore adjusted the claimant count with the aim of constructing a series which is consistent through time. Information from Lawlor (1990) was used on how many people were removed from the claimant count during the 1980s by various rule changes, and these numbers were added back in to create an adjusted claimant count series. As in Boyer and Hatton (2002), minor changes – those which altered the claimant count by 20,000 or less – were not incorporated in the adjusted series.

I use the male claimant count rather than the female or all persons claimant count because additional long-term changes in women's eligibility to contribute towards unemployment benefits, mean that the adjusted series for male claimants is a better indicator of the state of the labour market than a series which includes female claimants. The impact of the eligibility changes for male unemployment rates is apparent in Figure 5. Unemployment rates were exceptionally high for much of the 1980s and again in the early 1990s. By the year 2000, the adjusted series was around 6 per cent – approximately in line with estimates from the Labour Force Survey. The adjusted male claimant count, then, should give a more realistic picture of conditions in the labour market than the raw claimant count, and so it is the adjusted series which will be used in the analyses.

Figure 5. Claimant Count Unemployment Rates - Males

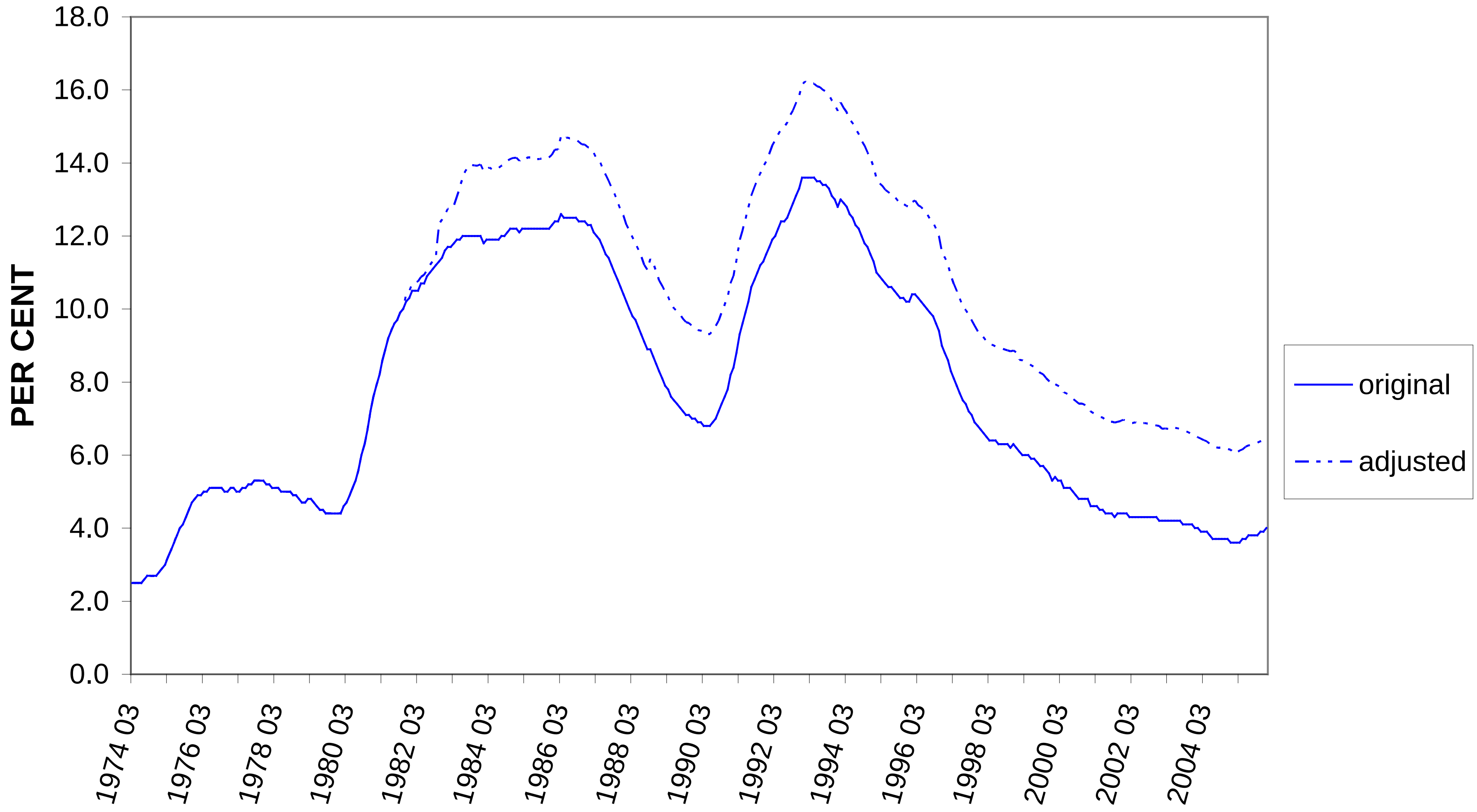


Table 2. Percentages with first birth by age 25 in NCDS and BCS70 samples by socio-economic characteristics

Education	**Low**	**Medium**	**High**			
NCDS	57.7	32.5	12.3			
BCS70	47.4	31.6	12.3			
Age 10 or 11 ability test scores: quintiles	Lowest	Fourth	Third	Second	Highest	Missing
NCDS	64.7	56.7	47.5	40.1	29.9	48.6
BCS70	49.5	43.5	35.2	29.2	19.9	34.5
CM received free school meals	Yes	No				
NCDS	65.6	44.9				
BCS70	52.6	32.7				
CM's father's social class (I = highest)	V	IV	III	II	I	Missing
NCDS	67.5	56.1	48.6	30.5	18.9	55.9
BCS70	62.1	40.5	38.5	23.2	12.8	38.5
CM's mother's age left f/t education	Before 15	15 to 16	16 to 17	17 or more		
NCDS	51.7	53.7	33.3	23.1		
CM's mother's years of f/t education	Less than ten	Ten	Eleven	Twelve plus		
BCS70	40.8	40.5	24.7	22.1		
CM's religion	None	Anglican	Catholic	Other Christian	Non-Christian	
NCDS	47.3	47.5	43.8	44.2	52.5	
BCS70	35.4	37.3	33.5	31.6	35.9	
CM number of siblings	None	One	Two	Three	Four plus	Missing
NCDS	37.5	37.8	48.7	50.6	59.8	44.7
BCS70	33.6	28.9	36.4	44.0	55.0	37.1

***Note**. CM denotes cohort member*

A set of further variables to act as controls in the model was also chosen. Potential explanatory variables which are clearly endogenous, such as marital status or partnership status, were not included in the models. Variables were selected so that, as far as possible, they were similar for each cohort. Scores on ability tests taken in childhood are available at various ages for each cohort. Age 11 scores for NCDS and age 10 scores for BCS70 were used. A range of variables was selected which reflect aspects of the socio-economic background of the cohort members, such as their father's social class, mother's education level, their religion, number of siblings, and whether they experienced poverty as a child, measured by receipt of free school meals. Table 2 reports some descriptive statistics for these explanatory variables. These show the proportions having their first birth by age 25 in each cohort, broken down by each potential explanatory variable. Overall 46.6 per cent of the NCDS women and 35.1 per cent of the BCS70 women had had a first birth by age 25.

Table 2 also shows that the percentage with a first birth by age 25 was inversely related to scores on ability tests in childhood. For the NCDS cohort, nearly two-thirds of those in the lowest quintile on the ability test scores had had at least one child by age 25, compared with just 30 per cent for the highest quintile. For the BCS70 cohort, this percentage fell from about half of those in the lowest quintile on the test scores, to 20 per cent in the highest quintile. Women who had experienced poverty in childhood, measured by receipt of free school meals at age 10 (BCS70) or 11 (NCDS) appeared to begin childbearing at younger ages. For example, among the NCDS women who were likely to have experienced poverty in childhood, almost two-thirds had a first birth by age 25, compared to only 45 per cent of those who had not received free school meals at age 11. Women whose fathers were in higher socio-economic status (SES) groups and women with more educated mothers, were less likely to have had a first birth by age 25, and this applied across both cohorts.

As for religion, those women reporting that they were Anglican and those who said they belonged to a non-Christian religion, had the highest likelihood of the birth of a first child by age 25. Generally, those cohort members who came from larger families also tended to start having children themselves at a younger age. However, for NCDS women, there was little difference between those who had no siblings compared to those who had one, while for BCS70 women, those with one sibling were somewhat less likely than those with no siblings, to have their first child by 25.

5. Results

The estimated hazard models are reported in Table 3 for the NCDS cohort and in Table 4 for the BCS70 cohort. In each of these tables, model A does not control for unobserved heterogeneity, while model B is more complex and specifies non-parametric heterogeneity terms. The factor loading term in the tables refers to the unobserved heterogeneity. All models were estimated in CTM (Yi et al 1987). I estimated the models for as many transitions as were feasible. In practice, this was the first three births for each cohort. However, for the later (BCS70) cohort, only a small proportion of women had already had a third birth by their early to mid-thirties, and such women may not be very typical, so the discussion here concentrates on comparisons of the first two births.

There was a negative association between education level and the hazard of the first birth for women in the NCDS cohort and in the BCS70 cohort. The coefficients on the education variables became much larger in absolute value, once unobserved characteristics of the women were taken into account. The absolute size of the estimated education coefficients was larger for the earlier cohort. As for the hazard of second births, there was some evidence of a positive association with higher levels of education for the NCDS women, but this effect disappeared once controls for unobservable factors were incorporated into the models. There was no evidence of any statistically significant associations between education and hazards of second births among BCS70 women. Overall, since our results show later timing of the first birth for more educated women, and no evidence of faster entry to higher order births, the implication is that more educated women would have fewer children, on average, over the life course.

The unemployment rate (adjusted as described earlier to allow for changes in eligibility rules) was lagged by 12 months and was entered into the models as a time-varying covariate. For NCDS women, there was a negative association ('pro-cyclical') between the lagged unemployment rate and the hazard of first births, but this was not statistically significant. There was a positive association between lagged unemployment and the hazard of second births ('anti-cyclical') and this became much larger and strongly significant after controlling for unobserved heterogeneity. Results for the third birth to NCDS women were similar. For the BCS70 women, there was also a negative association of the unemployment variable and the hazard of first births('pro-cyclical'); this was statistically significant and little affected by whether or not controls for unobservables were included in the model. The hazard of second birth was also negatively related to the lagged unemployment rate and was significant, at least in models which included controls for unobserved heterogeneity. There were, then, quite considerable differences between the two cohorts, in the relationships between the lagged unemployment rate and birth hazards, with NCDS looking like 'discouraged workers' at least for second and third births, while BCS70 look more like 'added workers' with income effects dominating substitution effects.

Table 3. Hazard models of the timing of first three births: NCDS results

	MODEL A *Without controls for unobserved heterogeneity*			**MODEL B** *Including controls for unobserved heterogeneity*		
First Birth	**Coeff**	**Std err**	**t-stat**	**Coeff**	**Std err**	**t-stat**
Factor Loading				8.904	0.788	11.29
Intercept	0.235	0.093	2.52	-5.310	0.706	7.52
Gamma_1	3.054	0.133	22.92	4.955	0.259	19.15
Gamma_2	-2.012	0.077	26.09	-1.936	0.117	16.49
Education (base, low)						
Medium	-0.381	0.046	8.36	-1.175	0.105	11.16
High	-0.618	0.059	10.52	-1.999	0.150	13.34
Free School Meals (FSM) at age 11	0.236	0.050	4.67	0.448	0.111	4.03
Father's SES						
SES I	-0.222	0.094	2.37	-0.512	0.180	2.84
SES II	-0.177	0.056	3.18	-0.482	0.115	4.19
SES III	-0.100	0.040	2.48	-0.228	0.086	2.64
SES IV	-0.052	0.049	1.05	-0.056	0.107	0.52
SES data missing	-0.130	0.077	1.68	-0.138	0.163	0.84
Mother's education (base, left school before age 15)						
Mother left school aged 15 to 16	0.057	0.037	1.55	0.161	0.079	2.04
Mother left school aged 16 to 17	-0.072	0.059	1.21	-0.114	0.124	0.92
Mother left school aged 17 or more	-0.047	0.067	0.71	-0.126	0.126	1.00
Ability Test Score Age 11 (base, lowest quintile)						
Highest quintile	-0.256	0.053	4.85	-0.846	0.118	7.16
Second quintile	-0.208	0.048	4.30	-0.805	0.108	7.44
Third quintile	-0.172	0.048	3.58	-0.709	0.110	6.42
Fourth quintile	-0.091	0.046	1.96	-0.391	0.102	3.82
Ability test: missing data	-0.202	0.054	3.76	-0.569	0.119	4.80

(Table 3 (NCDS) cont'd)

Religion (base, no religion)						
Anglican	0.057	0.033	1.73	0.064	0.069	0.93
Roman Catholic	-0.065	0.049	1.32	-0.294	0.098	2.99
Other Christian	0.035	0.047	0.74	0.005	0.094	0.06
Non-Christian religion	0.328	0.161	2.04	0.377	0.300	1.25
Number of siblings (age 16, base one sibling)						
No siblings	-0.007	0.079	0.09	-0.080	0.150	0.53
Two siblings	0.153	0.047	3.26	0.294	0.095	3.08
Three siblings	0.112	0.051	2.19	0.428	0.108	3.97
Four or more siblings	0.256	0.049	5.22	0.672	0.109	6.14
Siblings: missing data	0.040	0.048	0.83	0.203	0.095	2.14
Unemployment (lagged)	-0.012	0.007	1.67	-0.003	0.008	0.37

	Without controls for unobserved heterogeneity			*Including controls for unobserved heterogeneity*		
Second Birth	**Coeff**	**Std err**	**t-stat**	**Coeff**	**Std err**	**t-stat**
Factor Loading				3.031	0.335	9.06
Intercept	-0.485	0.112	4.32	-3.138	0.321	9.79
Gamma_1	0.501	0.134	3.75	0.749	0.136	5.50
Gamma_2	-3.207	0.162	19.80	-3.429	0.163	21.09
Education (base, low)						
Medium	0.027	0.047	0.57	-0.108	0.055	1.96
High	0.117	0.056	2.07	-0.059	0.066	0.89
Free School Meals (FSM) at age 11	-0.130	0.066	1.96	-0.098	0.074	1.34
Father's SES						
SES I	0.104	0.094	1.11	0.085	0.111	0.77
SES II	0.048	0.059	0.81	0.015	0.068	0.22
SES III	-0.002	0.048	0.05	-0.019	0.054	0.35
SES IV	-0.049	0.060	0.81	-0.049	0.068	0.72
SES data missing	0.052	0.093	0.56	0.054	0.107	0.51
Mother's education (base, left school before age 15)						
Mother left school aged 15 to 16	-0.015	0.043	0.35	-0.016	0.049	0.33
Mother left school aged 16 to 17	-0.015	0.065	0.23	-0.026	0.075	0.35
Mother left school aged 17 or more	-0.042	0.068	0.63	-0.035	0.079	0.44
Ability test score age 11 (base, lowest quintile)						
Highest quintile	-0.043	0.062	0.70	-0.112	0.072	1.57
Second quintile	-0.069	0.058	1.17	-0.136	0.066	2.05
Third quintile	-0.088	0.060	1.46	-0.166	0.069	2.42
Fourth quintile	-0.059	0.059	0.99	-0.097	0.066	1.46
Ability test: missing data	-0.010	0.067	0.16	-0.060	0.075	0.80
Religion (base, no religion)						
Anglican	0.111	0.039	2.87	0.144	0.044	3.30
Roman Catholic	0.065	0.056	1.15	0.037	0.063	0.58
Other Christian	0.025	0.053	0.48	0.046	0.060	0.77
Non-Christian religion	0.164	0.171	0.96	0.250	0.194	1.29
Number of siblings (age 16, base one sibling)						
No siblings	-0.033	0.082	0.41	0.002	0.095	0.02

(Table 3 (NCDS) cont'd)

Two siblings	0.076	0.051	1.49	0.121	0.059	2.05
Three siblings	0.124	0.059	2.10	0.184	0.067	2.74
Four or more siblings	0.140	0.057	2.44	0.229	0.066	3.47
Siblings: missing data	-0.021	0.054	0.39	0.009	0.061	0.15
Unemployment (lagged)	0.008	0.005	1.57	0.054	0.007	8.33

	Without controls for unobserved heterogeneity			*Including controls for unobserved heterogeneity*		
Third Birth	**Coeff**	**Std err**	**t-stat**	**Coeff**	**Std err**	**t-stat**
Factor Loading				5.664	0.671	8.45
Intercept	-1.721	0.183	9.43	-6.539	0.642	10.19
Gamma_1	-0.350	0.230	1.52	-0.032	0.234	0.14
Gamma_2	-2.188	0.304	7.21	-2.450	0.306	8.02
Education (base, low)						
Medium	-0.309	0.086	3.59	-0.454	0.097	4.66
High	-0.217	0.098	2.22	-0.380	0.110	3.44
Free School Meals (FSM) at age 11	0.383	0.094	4.06	0.443	0.109	4.07
Father's SES						
SES I	0.111	0.155	0.72	0.125	0.178	0.70
SES II	-0.218	0.103	2.12	-0.260	0.118	2.20
SES III	-0.104	0.076	1.37	-0.126	0.088	1.43
SES IV	-0.149	0.094	1.58	-0.166	0.109	1.52
SES data missing	-0.320	0.154	2.08	-0.329	0.172	1.91
Mother's education (base, left school before age 15)						
Mother left school aged 15 to 16	0.009	0.071	0.13	0.011	0.081	0.13
Mother left school aged 16 to 17	0.217	0.106	2.05	0.241	0.121	1.99
Mother left school aged 17 or more	0.289	0.114	2.53	0.321	0.131	2.46
Ability test score age 11 (base, lowest quintile)						
Highest quintile	-0.313	0.104	3.01	-0.377	0.119	3.17
Second quintile	-0.218	0.093	2.34	-0.256	0.107	2.39
Third quintile	-0.296	0.097	3.06	-0.346	0.112	3.09
Fourth quintile	-0.136	0.091	1.50	-0.139	0.105	1.33
Ability test: missing data	-0.073	0.102	0.71	-0.109	0.119	0.92
Religion (base, no religion)						
Anglican	0.036	0.064	0.56	0.089	0.073	1.21
Roman Catholic	0.187	0.088	2.14	0.227	0.100	2.26
Other Christian	0.010	0.090	0.11	0.040	0.102	0.39
Non-Christian religion	0.409	0.222	1.84	0.667	0.259	2.58
Number of siblings (age 16, base one sibling)						
No siblings	-0.079	0.152	0.52	-0.093	0.167	0.56
Two siblings	0.165	0.088	1.89	0.174	0.099	1.76
Three siblings	0.194	0.097	2.00	0.204	0.111	1.85
Four or more siblings	0.324	0.092	3.53	0.411	0.106	3.87
Siblings: missing data	0.122	0.090	1.36	0.115	0.101	1.14
Unemployment (lagged)	0.001	0.010	0.07	0.028	0.011	2.62
Log likelihood	-10,339.32			-9,740.96		

Note. *Gamma_1 is the coefficient on the linear term and Gamma_2 on the quadratic term in the hazard specification.*

Table 4. Hazard model of the timing of first three births: BCS results

	MODEL A *Without controls for unobserved heterogeneity*			MODEL B *Including controls for unobserved heterogeneity*		
First Birth	**Coeff**	**Std err**	**t-stat**	**Coeff**	**Std err**	**t-stat**
Factor Loading				3.690	0.178	20.78
Intercept	0.481	0.172	2.79	-1.512	0.254	5.96
Gamma_1	2.456	0.150	16.37	2.367	0.168	14.11
Gamma_2	-1.535	0.141	10.86	-0.620	0.169	3.66
Education (base, low)						
Medium	-0.273	0.040	6.82	-0.442	0.060	7.39
High	-0.695	0.054	12.87	-1.281	0.077	16.54
Free School Meals (FSM) at age 10	0.293	0.050	5.89	0.550	0.073	7.51
FSM data missing	0.153	0.074	2.07	0.069	0.107	0.65
Father's SES						
SES I	-0.420	0.129	3.27	-0.509	0.189	2.69
SES II	-0.387	0.097	3.99	-0.495	0.150	3.31
SES III	-0.288	0.088	3.28	-0.316	0.137	2.32
SES IV	-0.331	0.097	3.40	-0.263	0.150	1.75
SES data missing	-0.314	0.097	3.25	-0.231	0.151	1.53
Mother's education (base, less than 10 yrs f/t education)						
10 yrs of f/t education	0.081	0.076	1.07	0.103	0.108	0.95
11 yrs of f/t education	-0.118	0.088	1.35	-0.063	0.124	0.51
12 or more yrs of f/t education	-0.017	0.089	0.19	-0.032	0.128	0.25
Mother's education data missing	0.066	0.080	0.83	0.108	0.115	0.94
Ability test score age 10 (base, lowest quintile)						
Highest quintile	-0.291	0.065	4.45	-0.555	0.095	5.85
Second quintile	-0.218	0.059	3.68	-0.520	0.085	6.11
Third quintile	-0.218	0.057	3.84	-0.333	0.084	3.94
Fourth quintile	-0.062	0.054	1.14	-0.154	0.080	1.91
Ability test: missing data	-0.242	0.057	4.27	-0.406	0.082	4.94
Religion (base, no religion)						
Anglican	0.117	0.045	2.59	0.004	0.065	0.05
Roman Catholic	-0.040	0.062	0.64	-0.159	0.088	1.81
Other Christian	0.000	0.051	0.00	-0.086	0.073	1.17
Non-Christian religion	0.104	0.104	1.00	0.129	0.148	0.87
Number of siblings (age 16, base one sibling)						
No siblings	0.083	0.056	1.49	0.171	0.082	2.07
Two siblings	0.150	0.053	2.81	0.210	0.078	2.70
Three siblings	0.260	0.073	3.54	0.320	0.107	2.98
Four or more siblings	0.505	0.085	5.97	0.612	0.124	4.93
Siblings: missing data	0.077	0.045	1.71	0.150	0.065	2.33
Unemployment (lagged)	-0.043	0.009	4.59	-0.048	0.009	5.12

(Table 4 (BCS) cont'd)

	Without controls for unobserved heterogeneity			*Including controls for unobserved eterogeneity*		
Second Birth	**Coeff**	**Std err**	**t-stat**	**Coeff**	**Std err**	**t-stat**
Factor Loading				3.670	0.456	8.05
Intercept	-0.473	0.194	2.44	-2.779	0.438	6.34
Gamma_1	2.628	0.228	11.53	3.213	0.245	13.12
Gamma_2	-5.994	0.396	15.13	-6.737	0.409	16.48
Education (base, low)						
Medium	-0.030	0.051	0.59	-0.088	0.059	1.49
High	0.045	0.065	0.68	-0.115	0.077	1.49
Free School Meals (FSM) at age 10	-0.024	0.068	0.35	0.054	0.078	0.69
FSM data missing	0.141	0.098	1.43	0.177	0.108	1.64
Father's SES						
SES I	0.424	0.173	2.46	0.398	0.201	1.99
SES II	0.290	0.134	2.16	0.287	0.152	1.89
SES III	0.087	0.125	0.69	0.069	0.141	0.49
SES IV	0.133	0.137	0.97	0.109	0.155	0.70
SES data missing	-0.026	0.136	0.19	-0.048	0.154	0.31
Mother's education (base, less than 10 yrs f/t education)						
10 yrs of f/t education	0.099	0.099	1.00	0.149	0.113	1.32
11 yrs of f/t education	0.296	0.112	2.66	0.375	0.128	2.92
12 or more yrs of f/t education	0.107	0.115	0.92	0.155	0.132	1.17
Mother's education data missing	0.145	0.106	1.37	0.184	0.120	1.53
Ability Test Score Age 10 (base, lowest quintile)						
Highest quintile	0.041	0.081	0.50	-0.022	0.096	0.23
Second quintile	-0.066	0.075	0.88	-0.162	0.087	1.87
Third quintile	-0.053	0.074	0.72	-0.114	0.085	1.33
Fourth quintile	-0.089	0.072	1.24	-0.144	0.084	1.72
Ability test: missing data	-0.063	0.073	0.86	-0.149	0.084	1.76
Religion (base, no religion)						
Anglican	0.142	0.057	2.46	0.181	0.066	2.75
Roman Catholic	-0.111	0.080	1.39	-0.141	0.091	1.55
Other Christian	0.083	0.066	1.26	0.105	0.076	1.39
Non-Christian religion	0.120	0.128	0.94	0.129	0.147	0.88
Number of Siblings (age 16, base one sibling)						
No Siblings	-0.211	0.072	2.93	-0.206	0.081	2.53
Two siblings	0.081	0.065	1.24	0.141	0.077	1.83
Three Siblings	0.052	0.094	0.56	0.063	0.108	0.58
Four or more siblings	-0.057	0.119	0.48	-0.004	0.137	0.03
Siblings: missing data	-0.142	0.056	2.52	-0.124	0.064	1.92
Unemployment (lagged)	0.008	0.007	1.04	-0.032	0.009	3.49

Table 4 (BCS) cont'd)

	Without controls for unobserved heterogeneity			*Including controls for unobserved heterogeneity*		
Third Birth	**Coeff**	**Std err**	**t-stat**	**Coeff**	**Std err**	**t-stat**
Factor Loading				14.115	1.655	8.53
Intercept	-2.328	0.332	7.00	-12.377	1.611	7.68
Gamma_1	2.233	0.475	4.70	4.844	0.664	7.29
Gamma_2	-4.230	0.888	4.76	-5.943	1.104	5.38
Education (base, low)						
Medium	-0.024	0.101	0.24	-0.154	0.158	0.97
High	0.331	0.147	2.25	0.045	0.224	0.20
Free School Meals (FSM) at age 10	0.346	0.113	3.06	1.058	0.194	5.46
FSM data missing	-0.042	0.172	0.24	0.410	0.281	1.46
Father's SES						
SES I	0.097	0.355	0.27	-0.050	0.512	0.10
SES II	0.037	0.242	0.15	-0.140	0.352	0.40
SES III	0.080	0.215	0.37	-0.221	0.310	0.71
SES IV	0.225	0.232	0.97	0.203	0.336	0.60
SES data missing	-0.044	0.237	0.19	-0.375	0.346	1.08
Mother's education (base, less than 10 yrs f/t education)						
10 yrs of f/t education	-0.332	0.157	2.11	-0.181	0.246	0.74
11 yrs of f/t education	-0.416	0.197	2.11	-0.251	0.310	0.81
12 or more yrs of f/t education	-0.660	0.212	3.12	-0.696	0.325	2.14
Mother's education data missing	-0.288	0.170	1.69	-0.306	0.269	1.14
Ability Test Score Age 10 (base, lowest quintile)						
Highest quintile	0.167	0.174	0.96	-0.202	0.251	0.80
Second quintile	0.354	0.147	2.40	0.254	0.231	1.10
Third quintile	-0.007	0.149	0.05	-0.369	0.232	1.59
Fourth quintile	0.091	0.136	0.67	-0.270	0.209	1.29
Ability test: missing data	0.413	0.130	3.18	0.234	0.202	1.16
Religion (base, no religion)						
Anglican	0.196	0.112	1.74	0.339	0.170	2.00
Roman Catholic	0.039	0.167	0.24	-0.028	0.257	0.11
Other Christian	0.153	0.131	1.17	0.354	0.199	1.78
Non-Christian religion	0.458	0.204	2.24	0.758	0.332	2.28
Number of Siblings (age 16, base one sibling)						
No siblings	0.023	0.147	0.15	-0.162	0.222	0.73
Two siblings	0.250	0.129	1.94	0.248	0.203	1.23
Three siblings	0.274	0.171	1.60	0.504	0.273	1.85
Four or more siblings	0.258	0.189	1.36	0.474	0.284	1.67
Siblings: missing data	0.018	0.115	0.16	-0.049	0.173	0.28
Unemployment (lagged)	0.133	0.015	9.08	0.027	0.025	1.09
Log likelihood	-7214.04			-7055.63		

Note. *Gamma_1 is the coefficient on the linear term and Gamma_2 on the quadratic term in the hazard specification.*

There was a strong, positive relationship between the experience of poverty in childhood (as measured by receipt of free school meals) and the hazard of the first birth. This applied to both cohorts and regardless of whether the model specification controlled for unobserved heterogeneity. The size of estimated coefficients was similar for NCDS and BCS70 samples. There was little evidence of any relationship of childhood poverty with the hazard of the second birth for either cohort. For NCDS women, the free school meals variable was marginally significant in models which did not control for unobservables, but this effect disappeared once allowance was made for unobserved heterogeneity.

The hazard of the first birth tended to be higher for those cohort members whose fathers were in lower SES categories. This finding applied to both cohorts. As for the second birth, father's SES variables were largely non-significant, but for the younger cohort, there was some evidence of higher hazards of second births for women whose fathers were in higher SES groups. On the whole, the education level of the cohort member's own mother appeared to have little association with birth hazards. Exceptions were that the cohort member's mother's leaving school at age 15 or 16, was associated with higher hazard of first birth for NCDS, while 11 years of mother's completed schooling was associated with a higher hazard of second birth for the BCS70 cohort. Certain coefficients were statistically significant, but there was no clear pattern to these results.

Those cohort members who scored highly on general ability tests in childhood, tended to have a reduced hazard for first births. The magnitude of this association increased once controls for unobservables were included in the models, and it was larger for NCDS women than for BCS70 women. There was less evidence that hazards of the second birth were associated with the ability test scores, but for NCDS it seemed that those in the second or third quintiles of attainment tended to have a higher second birth hazard.

The models also included measures of the religion of cohort members. The hazard of first births was lower for Roman Catholics in both cohorts, and was statistically significant for the NCDS women, but not significant at the 5 per cent level for the BCS70 women. Anglicans also had an increased hazard for the second birth in both cohorts. It may also be worth noting that third birth hazards were higher for Roman Catholic and those of non-Christian religion, which was perhaps more in line with prior expectations.

NCDS and BCS70 cohort members who had a large number of siblings also had a significantly higher hazard for the first birth. The number of siblings was also positively associated with the second birth for NCDS; this finding did not apply consistently for BCS70 women, but those with no siblings had a significantly lower hazard than the base case of one sibling.

Information about the non-parametric heterogeneity distributions estimated in the models for NCDS and BCS70, appears in Table 5. The procedure here was that two of the mass points were fixed at zero and one, and other mass points and all associated probabilities were freely estimated. I began by estimating a distribution with just two mass points, and increased the number of points until two converged on the same location. The outcome of this process was different for NCDS and BCS70. In the case of the models for NCDS, a distribution with six mass points could be estimated, while for BCS70 there were just three mass points.

Table 5. Estimated mass points and probabilities

Location	SD	Cumulative probability	SD
NCDS			
0.00000	0.00000	0.13258	0.00989
0.46785	0.04192	0.25559	0.07704
0.59949	0.07893	0.39177	0.09280
0.74234	0.06201	0.63190	0.17446
0.85187	0.03583	0.89496	0.04325
1.00000	0.00000	1.00000	0.00000
BCS			
0.00000	0.00000	0.33621	0.01462
0.75971	0.02120	0.87491	0.01242
1.00000	0.00000	1.00000	0.00000

Note. *See Model B in each of Tables 3 and 4 for details of the estimated models*

The implication, essentially, is that there were a number of different groups of women in each cohort, six in the case of NCDS and three in the case of BCS70, with differing unobserved characteristics.

6. Conclusion

This research is concerned with the roles of education and labour market conditions in the timing of births. On education, an important finding is that the negative relation of education to the timing of the first birth is still observed – in fact becomes stronger – when one makes allowance for unobserved heterogeneity. Can it then be concluded that education is likely to be causal? While the proper treatment of residual heterogeneity addresses potential biases in parameter estimates, it does not in itself enable the causal effect of education to be identified. The person-specific unobservable component in the model is not correlated with the covariates, so that controlling for unobserved heterogeneity will not address the endogeneity of educational attainment, if the latter is correlated with this unobservable component. However, it can plausibly be argued that the person-specific heterogeneity correlated with education has effectively been removed by the inclusion of other variables associated with educational attainments, particularly own cognitive ability, father's SES, mother's education and the number of siblings. In other words, the inclusion of a number of other variables correlated with education makes a causal interpretation more likely, as the heterogeneity remaining will not be correlated with own education. The results, then, are consistent with an interpretation which sees education as having a causal effect on fertility, rather than there just being an association with education[ii]. The explanation for this would be that, not only is childbearing avoided during studies, but once a woman is on the labour market, earnings reach higher levels than those which might be achieved if first childbearing is not delayed. Women who attain higher levels of education have higher earning potential, and therefore a larger opportunity cost in terms of lost earnings, if time is spent out of the paid labour force giving birth to, and looking after, children. The level of education may influence fertility dynamics for a number of reasons, including skills in effective use of contraception, parenting skills, and knowledge about the responsibilities involved in raising children. Determining the relative importance of such changes in behaviour suggests an agenda for further research.

There were substantial differences between the 1958 and 1970 cohorts in the relationship between fertility and labour market conditions, as measured by an aggregate, time-varying series for the unemployment rate. The hazard of the first birth was negatively and significantly related to the unemployment rate for the 1970 cohort, but not for the 1958 cohort, while the hazard of the second birth was positively related to the unemployment rate for the 1958 cohort, but negatively related for the 1970 cohort (and was statistically significant for both cohorts). What might account for these different results? Now, the observation period for this study covered 1974 to 2004. Over these decades, there were a number of changes in the direction of making employment and motherhood more compatible. There were the gradual improvements in maternity leave, the introduction of paternity leave, the improvement of terms for part-time employment, and with the New Labour Government, elected in 1997, a new emphasis on public support for childcare. Although these changes do not resolve themselves neatly into monthly time series, they do add up to a secular trend, differentiating the environment in which the two cohorts faced early adulthood and the prospect of fertility. They could explain why the estimated fertility reaction of the earlier cohort to the prospect of unemployment was more dominated by substitution effects, and the later cohort, by income or 'added worker' effects, like those observed in Scandinavia.

The effects of other covariates in the models appear to be broadly similar across the two cohorts. For example, NCDS and BCS70 cohort members who had a large number of siblings also had a significantly higher hazard for the first birth. There was also a positive relationship between the experience of poverty in childhood (as measured by receipt of free school meals) and the hazard of the first birth for both cohorts. This confirms that in Britain, women from disadvantaged backgrounds tend to be more likely to make an early entry into motherhood. While the findings in this paper refer to cohorts of women born in 1958 and 1970, Hawkes (2009) shows that they also hold in a survey of more recent origin, the Millennium Cohort Study.

In demographic research, and more generally in the literature on duration analysis, there has been debate on the best way to control for heterogeneity. Heckman and co-authors have advocated a robust, non-parametric approach and this method has been utilised in some studies of fertility (although not previously for UK data). Heckman and Walker (1990) analysed data on the first three births for four cohorts of Swedish women, and actually found that unobserved heterogeneity terms were not statistically significant, concluding that "unobservables correlated across spells are not an important feature of modern Swedish fertility data". In contrast, Merrigan and St-Pierre (1998) conducted a very similar analysis (in terms of explanatory variables and modelling strategy) on Canadian birth history data, and found non-parametric heterogeneity to be important. I have also utilised Heckman and Singer's non-parametric method to control for unobservables. Controlling for heterogeneity improved the fit of the models, in that the likelihood was improved and the factor loading terms in our models were highly significant for all transitions. Moreover, controlling for unobserved heterogeneity made a considerable difference to substantive research findings. Before allowing for unobservables, it appeared that there was a positive association between education and the hazard of second birth for the NCDS cohort. Also, unemployment did not appear to be related to the timing of higher-order births. Once controls for unobservables were incorporated into the models, education was no longer significantly related to second birth hazards, while it became apparent that there was a positive association between unemployment and the hazards of second and third births for the NCDS cohort. These results affirm the importance of including robust controls for unobservables, when modelling the timing of births.

Acknowledgements

This paper is one component of a project on fertility in Britain and the US, led by Professor Heather Joshi and Professor Mark Killingsworth, and I am very grateful to them for advice and comments at all stages of the work on this paper. Helpful comments were also received from Dr Kirstine Hansen, from conference participants at ESPE (London) and EALE (Amsterdam) and from participants at an Institute of Education seminar. Funding from the ESRC UPTAP programme and from the British Academy is also gratefully acknowledged.

References

Aassve A, Burgess S, Propper C and Dickson M. (2006) Employment, family union and childbearing decisions in Great Britain. *Journal of the Royal Statistical Society*, 169, 781-804.

Ahn N and Mira P. (2001) Job bust, baby bust? Evidence from Spain. *Journal of Population Economics*, 14, 505-521.

Allison PD. (1984) *Event history analysis*. Sage, Beverley Hills, CA..

Blossfeld H-P, Golsch K and Rohwer G. (2007) *Event history analysis with stata*. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Boyer GR and Hatton TJ. (2002) New estimates of British unemployment, 1870-1913. *Journal of Economic History*, 62, 643-675.

Bray H. (2008) *National population projections 2006-based*. Office for National Statistics, Basingstoke.

Crafts N. (2007) Living standards. In N Crafts, I Gazeley and A Newell. eds. *Work and pay in twentieth-century Britain.* Pp 11-34. Oxford University Press, Oxford.

Davies RB. (1994) From cross-sectional to longitudinal analysis. In A Dale and RB Davies. eds. *Analyzing social and political change: a casebook of methods.* Pp 20-40. Sage Publications, London.

del Bono E. (2001) Estimating fertility responses to expectations: evidence from the 1958 British Cohort. *Discussion Paper No 80*: University of Oxford, Department of Economics, Oxford.

Dex S, Gustaffson S, Jenkins A, Joshi H, Kenjoh E and Killingsworth MR. (2005) Mothers' changing labour supply in Britain, the USA and Sweden. In R Gomez-Salvador, A Lamo, B Petrongolo, M Ward and E Wasmer. eds. *Labour supply and incentives to work in Europe.* Pp 115-150. Edward Elgar, Cheltenham, UK.

Ermisch J and Pevalin D. (2003) Does a teen birth have a longer-term impact on the mother? Evidence from the 1970 British Cohort Study. *ISER Working Paper 2003-28*. ISER, University of Essex, Colchester.

Ermisch, J and Pevalin D. (2005) Early motherhood and later partnerships. *Journal of Population Economics*, 18, 469-489.

Gustaffson S. (2001) Optimal age at motherhood: theoretical and empirical considerations on postponement of maternity in Europe. *Journal of Population Economics*, 14, 225-247.

Gutierrez-Domenech M. (2002) The impact of the labour market on the timing of marriage and births in Spain. *Discussion Paper 556*: London School of Economics, London.

Hawkes D. (2003) Education, earnings, ability and early child bearing: evidence from a sample of UK twins. Unpublished PhD thesis. Queen Mary, University of London, London.

Hawkes D. (2009) The UK Millennium Cohort Study and the circumstances of early motherhood. In R Edwards. ed. *Researching family, community and generational change*. Routledge, London.

Heckman JJ and Singer B. (1984) A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52, 271-320.

Heckman JJ and Walker JR. (1990) The relationship between wages and income and the timing and spacing of births: evidence from Swedish longitudinal data. *Econometrica*, 58, 1411-1441.

Hoem B. (2000) Entry into motherhood in Sweden: the influence of economic factors on the rise and fall in fertility, 1986-1997. *Demographic Research*, 2.

Hotz V, Klerman J and Willis R. (1997). The economics of fertility in developed countries: a survey. In M Rosenzweig and O Stark. eds. *Handbook of population and family economics*. Elsevier, New York.

Kiefer NM. (1988) Economic duration data and hazard functions. *Journal of Economic Literature*, 26, 646-679.

Killingsworth M and Heckman J. (1986) Female labor supply: a survey. In O Ashenfelter and R Layard. eds. *Handbook of labor economics.* Pp103-204*.* Elsevier Science Publishing Company, New York.

Kravdal O. (1992) The emergence of a positive relation between education and third birth rates in Norway with supportive evidence from the United States. *Population Studies*, 46, 459-475.

Kravdal, O, (2002) The impact of individual and aggregate unemployment on fertility in Norway. *Demographic Research*, 6, 263-294.Kreyenfeld M. (2000) Employment careers and the timing of first births in East Germany, *MPIDR Working Paper 004*: Max Planck Institute for Demographic Research, Rostock, Germany.

Kreyenfeld, M. (2002). Time-squeeze, partner effect or self-selection? An investigation into the positive effect of women's education on second birth risks in West Germany. *Demographic Research*, 7, 15-48.

Lancaster T. (1990) *The econometric analysis of transition data*. Cambridge University Press, Cambridge.

Makepeace G, Dolton P, Woods L, Josh, H and Galindo-Rueda F. (2003) From school to the labour market. In E Ferri, J Bynner and M Wadsworth. eds. *Changing Britain, changing lives: three generations at the end of the century.* Pp 29-70. Institute of Education, London.

Merrigan P and St-Pierre Y. (1998) An econometric and neoclassical analysis of the timing and spacing of births in Canada from 1950 to 1990. *Journal of Population Economics*, 11, 29-51.
Newman JL and McCulloch CE. (1984) A hazard rate approach to the timing of births. *Econometrica*, 52, 939-961.
Office for National Statisics. (2007) *Birth Statistics*, Series FM1 no. 35, Table 10.3
Rendall M and Smallwood S. (2003) Higher qualifications, first-birth timing and further childbearing in England and Wales. *Population Trends*, 111, 18-26.
Santow G and Bracher M. (2001) Deferment of the first birth and fluctuating fertility in Sweden. *European Journal of Population*, 17, 343-363.
Wright RE, Ermisch JF, Hinde PR and Joshi HE. (1988) The third birth in Great Britain. *Journal of Biosocial Science*, 20, 489-496.
Yi K-M, Honore B and Walker J. (1987) *CTM user's manual*: ERC/NORC and University of Chicago, Chicago.

Appendix 1. Policy milestones affecting mothers' employment in the UK since 1974

Date of implementation	Legislation	Statutory provisions
1975	Child Benefit Act 1975	Universal cash benefit paid to mothers
December 1975	Equal Pay Act 1970	Equal pay for equal work
June 1976	Employment Protection Act 1975	Right to reinstatement up to 29 weeks after birth for those who qualified (fulltime for 119 weeks with same employer or 265 weeks part-time)
April 1977	Employment Protection Act 1975	Statutory maternity pay replaced lower flat rate benefit (s.t. qualification conditions), 6 weeks at a rate related to earnings, flat rate allowance still available 12-14 weeks
1983	Equal Pay Amendment Act	Equal pay for work of equal value
1992	Social Security Contributions and Benefits Act 1992	Updated conditions for payment of statutory maternity pay
	Workplace (Health and Safety) Regulations 1992	Employers required to provide for pregnant women to rest and breastfeeding women to express breastmilk at work
April 1990	Finance Act	Introduction of independent taxation for husbands and wives
October 1994	EU Directive 1992 Trade Union Reform and Employment Rights Act 1993	Relaxed conditions of eligibility for leave - a response to EU Directive on Part-time Workers
1998	National Minimum wage 1998	Affects women (part-timers) disproportionately
	Working time Regulation 1998	Regulates the work week, annual leave, rest periods and night work
	National Child Care Strategy	Began a rollout of subsidized places for pre-school children
June 1999	Employment Relations Act 1999	Further improvement of rights to leave. Introduction of unpaid parental leave and leave for family emergences -a response to EU directive on parental leave
July 2000	The Part time Workers (Prevention of Less Favourable Treatment Regulations) 2000	Part-timers should not be treated less favourably in their contractual terms and conditions than comparable full-timers,
April 2003	Maternity and Parental Leave Regulations 1999, Amendments 2002 Employment Act 2002	Paid leave increased to 26 weeks. Paid paternity leave (2 weeks) introduced, also adoption leave. Employers obliged to consider requests for flexible working
October 2008	Work and Family Act 2006	Further increased in maternity leave to 52 weeks

The above table sets out some of the main milestones in policies which affected the compatibility of paid work and motherhood for the two cohorts investigated in this paper.

It focuses particularly on maternity provisions, where statutory benefits are shown. Some mothers would have been entitled to better leave or pay than the statutory, which was offered by some employers, sometimes in response to negotiation by some trades unions. Many others, especially in the period before the mid-1990s, would not have met conditions of service with their employer to be eligible even for statutory benefits. Before the mid-1970s there were flat-rate maternity benefits for women contributing to National Insurance, and a means-tested allowance for those not insured. This reflected the assumption made by the principal architect of Britain's post-war welfare state, Sir William Beveridge, that mothers would not generally return to the labour market after childbirth.

Note also that the table does not include all the relevant legislation. For example, there was a tightening of qualifying conditions in the 1980s to protect small employers, and nor is every detailed change in levels, duration and eligibility for the two strands of statutory paid leave included. Also omitted is the introduction of Working Tax Credits under New Labour and a number of details of public support for child care.

Main source: Earnshaw J. (1999) Maternity rights in the UK : light at the end of the tunnel? *Economic and Labour Relations Review* 10, 196-187.

Endnotes

[i] In some disciplines unobserved heterogeneity may be referred to as unmeasured confounding.

[ii] Controlling robustly for heterogeneity allows us to be much more confident that education is having an effect on fertility behaviour. However, it is possible to think of circumstances in which such results would be consistent with education not being causally linked to fertility; for example, if there are unanticipated shocks which impact on education and hence on fertility.

RESEARCH NOTE

Emotional and behavioural problems in childhood and risk of overall and cause-specific morbidity and mortality in middle-aged Finnish men

Laura Kauhanen
Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland
laura.kauhanen@uef.fi
Janne Leino
Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland
Hanna-Maaria Lakka
Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland
Kuopio National Institute of Health and Welfare, Helsinki
John Lynch
School of Population Health and Clinical Practice, University of Adelaide
Department of Social Medicine, University of Bristol
Jussi Kauhanen
Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland
Prevention Research Center, Pacific Institute of Research and Evaluation, Berkeley, CA, USA

(Received December 2009 Revised April 2011)

Introduction

Little is known of the overall and cause-specific morbidity and mortality of those having emotional and behavioural problems in childhood. In the study by Jokela, Ferrie, and Kivimäki (2008), childhood externalizing and internalizing behaviours were associated with increased risk of premature death. Externalizing problems are characterized by inattention, poor conduct, opposition and defiance. Internalizing problems, such as avoidant and withdrawn temperament, is manifested by low self-esteem, worry, fear, and shyness (Dick et al 2005; Roza et al 2003; Zahn-Waxler, Klimes-Dougan and Slattery 2000). Internalizing problems in childhood have been linked with adult depression and anxiety disorders (Clark et al 2007). In turn, externalizing problems have been related with later antisocial behaviour, delinquency (Simonoff et al 2004), and substance abuse (King, Iacono and McCue 2004). For example, Shepherd, Farrington and Potts (2004) found that antisocial lifestyle in childhood and adolescence increased the risk of injury and psychological illness. In addition, Laub and Vaillant (2000) found that alcohol abuse and poor self-care were associated with subsequent death, in the study of 1,000 delinquent and non-delinquent boys.

It is suggested that psychological stressors may increase the vulnerability to cancer and auto-immune diseases through a deregulatory effect on the immune system (Irwin et al 1990; Stein, Keller and Schleifer 1985). For example, it has been reported that personality, emotional suppression, depression, and social isolation are risk factors for cancer (Edelman 2005; Grossarth-Maticek et al 1997; Penninx et al 1998; Persky Kempthorne-Rawson and Shekelle 1987; Shekelle et al 1981; Shaffer et al 1987) although some studies have found no evidence of such relationship (Price et al 2001; Bleiker et al 1996). In addition, it has been reported that depression, social isolation, and lack of social support are risk factors for coronary heart disease (CHD) (Peach et al 2003). Furthermore, there is some evidence that cynical hostility increases the risk of all-cause and cardiovascular mortality, and incident myocardial infarction (Everson et al 1997), and cancer-related mortality (Tindle et al 2009).

Emotional and behavioural problems among children and adolescents are perceived to be increasing in many countries, sometimes attributed to childhood poverty, increase in the proportion of single parent families, and substance abuse among families. (Kelleher et al 2000; Rimpelä et al 2006; Sweeting and West 1998). Studies report that approximately 7-20% of children and adolescents meet the criteria for a broadly-defined behavioural problem (Horwitz et al 1992; Kaltiala-Heino et al 2005; Kelleher et al 2000; Kumpulainen et al 1998). For example Rimpelä and colleagues (2006) found that, in a Finnish school health survey, 24% of children had psychosomatic symptoms, such as anxiety or physical manifestations (Rimpelä et al 2006).

The purpose of this study was to investigate reports, obtained from nurses, of emotional and behavioural problems during childhood, as predictors of overall and cause-specific mortality and morbidity in later life among participants of the Kuopio Ischemic Heart Disease Risk Factor Study (KIHD). Additionally in this study, we examined the effect of biological, behavioural, and socio-economic factors, on the associations between emotional and behavioural problems in childhood and all-cause, cancer and cardiovascular mortality, morbidity and alcohol-associated diseases.

Methods

Study population

The subjects were participants in the KIHD which is a prospective population-based study designed to investigate risk factors for cardiovascular diseases, including psychosocial and socio-economic factors, in middle-aged and ageing men from Eastern Finland. The original study population consisted of a random age-stratified sample of 2,682 men, who were 42, 48, 54, or 60 years of age at baseline in 1984 (Kaplan et al 1994). The Research Ethics Committee of the University of Kuopio approved the study. The school health records were available for 952 (35.5%) men, because some of the archives, where school health records were stored, were destroyed during World War II and others by fires. There were 72 men who were excluded from the analyses because of the missing data on some of the covariates. The final study sample was therefore 880. A comparison of the historical study sample with the rest of the KIHD cohort revealed that the study participants were on average somewhat younger, their education, occupational, and income levels were higher, they were physically more active and they have smoked cigarettes less than the rest of the KIHD cohort (Table 1).

Table 1. Comparison of the historical study sample (n=880) with the rest of the KIHD cohort (n=1603)

	Mean (SD) or proportion (%)		
Covariates	**Study sample (*n*=880)**	**The rest of the KIHD cohort (*n*=1603)**	**p-values for the difference between groups**
Age group (%)			
1 (42 years)	24.0	5.9	<0.001
2 (48 years)	18.6	9.4	
3 (54 years)	46.3	68.4	
4 (60 years)	11.1	16.3	
Educational level (%)			
1 (low)	7.7	11.1	<0.001
2	43.8	50.5	
3	41.9	31.8	
4 (high)	6.6	6.6	
Occupational level (%)			
1 (farmer)	12.5	20.2	<0.001
2 (blue collar)	46.4	41.6	
3 (white collar)	41.1	38.2	
Smoking history (pack/years)	148.4 (286.9)	180.9 (360.7)	0.014
BMI (kg/m^2)	26.4 (3.7)	26.9 (3.5)	0.681
LDL cholesterol (mmol/l)	4.0 (1.0)	4.1 (1.0)	0.006
HDL cholesterol (mmol/l)	1.3 (0.3)	1.3 (0.3)	0.570
SBP (mm Hg)	133.9 (16.8)	134.4 (17.3)	0.395
Leisure time physical activity (h/year)	123.5 (153.1)	105.8 (130.9)	0.003
Alcohol consumption (g/week)	82.2 (148.3)	73.8 (134.2)	0.150
Income (marks/year)	82,222.5 (52467.6)	74,785.8 (49352.3)	<0.001

Emotional and behavioural problems in childhood

Childhood information was obtained from elementary school health records which were filled out by the school health nurses and doctors in the 1930s to 1950s. The school health records contained data on health status, school attendance, behaviour of the child at school, general hygiene/cleanliness of the child, and socio-economic circumstances at home, based on the personal observations of school health nurses, and doctors at school and during home visits until the children were 13 years of age.

A man was defined as having emotional/behavioural problems in childhood if a school health nurse had reported one or both of the following:

1. Emotional problems
2. Behavioural problems

1. "Emotional problems" was defined as school health nurses reporting a child having withdrawal problems such as nervousness, shyness, fearfulness, and anxiety. 2. "Behavioural problems" was defined as a school health nurse reporting aggressive, antisocial, or delinquent behaviour of the child. These items were scored dichotomously and the scores summed. A summary variable of emotional and behavioural problems variables was made to represent the total emotional/behavioural problems score in childhood. If there was no mention of items 1 or 2, a man was defined as not having emotional/behavioural problems in childhood.

Covariates

Age and examination year Age was categorized into four groups: 42 years, 48 years, 54 years, and 60 years. Examination year was categorized from 1984 to 1989.

Biological factors The gathering of blood specimens (Salonen et al 1992) and the measurement of serum lipids (Salonen et al 1991) have been explained elsewhere.

The ratio of low density lipoprotein (LDL) to high density lipoprotein (HDL) cholesterol and systolic blood pressure (SBP) were included in the analysis.

Adulthood behavioural factors The assessment of alcohol consumption in grams per week with a structured quantity and frequency method using the Nordic alcohol consumption inventory has been described previously (Kauhanen et al 1997). Leisure-time physical activity in hours per year was assessed from a 12-month history questionnaire (Lakka et al 1994). Cigarette smoking was estimated by self-reporting and converted to pack-years (the average number of cigarettes per day times the number of years smoked). Body mass index (BMI) was calculated as the ratio of weight in kilograms to the square of height in metres (kg/m^2).

Childhood socio-economic variables Socio-economic position (SEP) in childhood was a summary variable including poor social conditions at home, poor hygiene, attending a special summer camp for poor children, and attending a school meal programme meant for children in need (Kauhanen et al 2006). Education was also included in the analysis of childhood SEP. It was categorized into four groups: less than elementary, elementary, full or some secondary, and high school or above.

Adulthood socio-economic variables Adult SEP was assessed by the self-report of annual personal income and occupation. Occupation was categorized into three groups: 1=farmer, 2=blue collar, 3=white collar.

Outcomes

Mortality Deaths were ascertained by computer linkage to the national death registry using the Finnish social security number. All deaths occurring between study entry (March 1984 to December 1989) and 31 December 2007 were included. Deaths coded with the Ninth International Classification of Diseases (ICD-9) codes 140-239 and the tenth revision (ICD-10) codes C00-D48 were included in the analysis of cancer deaths. Deaths coded with ICD-9 codes 390-459 and ICD-10 by codes I00-I99, were considered cardiovascular (CVD) deaths. Deaths were coded with ICD-9 codes 410-414, and ICD-10 codes I20-I25 were included in the analysis of coronary heart disease (CHD) deaths. The median follow-up time was 20.7 years (range 0.2 to 24.8 years). There were 72 cancer deaths, 130 CVD deaths and 89 CHD deaths during the follow-up period. Death codes were all validated according to the international criteria adopted by the WHO MONICA (Monitoring of Trends and Determinants of Cardiovascular Disease) Project (Bothig 1989).

Acute coronary events Data on fatal or non-fatal acute coronary events between the study entry and 2004, were collected prospectively and diagnostic classification was made by the FINMONICA coronary registry group (Tuomilehto et al 1992). Since 1 January 2004, the events were obtained by computer linkage to the national computerized hospital discharge registry. Diagnostic information was collected from hospitals and events were classified by one internist using the same diagnostic criteria as in the FINMONICA project. The median follow-up time to the first coronary event was 17.6 years (range 0.1 to 21.8 years). If the subject had multiple non-fatal events during the follow-up, the first one was considered as the endpoint. Data were available up to 31 December 2004, during which period, 209 acute coronary events occurred.

Alcohol-associated diseases All alcohol-associated diseases that occurred between study entry and 31 December 2007 were included. Data on alcohol-associated diseases were obtained by record linkage from the national computerized hospitalization registry, which covers every hospitalization in Finland. Alcohol diseases were coded with the Eighth International Classification of Diseases (ICD-8) or the Ninth revision (ICD-9) or the 10th revision (ICD-10). The median follow-up time to the first alcohol-associated disease was 20.7 years (range 0.04 to 24.8 years). If the subject had multiple non-fatal events during the follow-up, the first one was considered as the endpoint. During the follow-up period, 69 alcohol-associated diseases occurred.

Statistical analysis

The association between emotional and behavioural problems in childhood and the risk of all-cause, cancer, CVD, and CHD deaths, and the risk of acute coronary events and alcohol-associated diseases in later life, were analysed with Cox proportional hazards models[i]. The analysis sample was 880. Emotional problems were reported for 9.5% of men and behavioural problems for 2.3% of men. Men with any emotional/behavioural problems in childhood formed the index group (11.8%) and men without emotional/behavioural problems in childhood were a reference group in the summary problems score analyses.

A sequence of models was carried out to examine the relationship between childhood emotional and behavioural problems and mortality and morbidity in adulthood. Model1 included age and examination year. Model 2 was the same as model 1 and additionally adjusted for SEP in childhood (poor social conditions at home, poor hygiene, attending a special summer camp for poor children, and attending a school meal programme meant for children in need, education). Model 3 was the same as model 1 and additionally adjusted for adulthood SEP (occupation, income). Model 4 was the same as model 1 and additionally adjusted for the biological factors (systolic blood pressure, LDL/HDL), and behavioural characteristics (alcohol consumption, smoking, BMI, physical activity). All analyses were performed using SPSS for Windows 17.0.

Results

Table 2 shows the mean $\pm$ standard deviation or prevalence for the covariates: age, the biological and behavioural factors (systolic blood pressure, HDL and LDL cholesterol, leisure time physical activity, BMI, alcohol consumption, and smoking), and education and occupation, for men with emotional problems (n=84), behavioural problems (n=20), and without emotional/behavioural problems (n=776) in childhood. The educational and income levels were lower, and LDL levels higher in men with behavioural problems in childhood compared to men with emotional problems and without emotional/behavioural problems in childhood. Men with emotional problems in childhood were somewhat younger than others. Table 2 also shows crude mortality rates of all-cause, cancer, CVD and CHD deaths, and incidence density of acute coronary events, and alcohol-associated diseases in men with and without emotional/behavioural problems in childhood.

Table 2. Baseline characteristics of men with emotional and behavioural problems and without emotional/behavioural problems in childhood

Covariates	**Mean (SD) or proportion (%)** **Men with emotional problems in childhood (*n*=84)**	**Men with behavioural problems in childhood (*n*=20)**	**Men without emotional/behavioural problems in childhood (*n*=776)**	**p-values for the difference between groups**
Age group (%)				
1 (42 years)	33.3	0.0	23.6	0.009
2 (48 years)	15.5	10.0	19.2	
3 (54 years)	41.7	85.5	45.7	
4 (60 years)	9.5	5.0	11.5	
Educational level (%)				
1 (low)	8.3	35.5	6.9	< 0.001
2	31.0	50.0	45.0	
3	53.6	15.0	41.4	
4 (high)	7.1	0.0	6.7	
Occupational level (%)				
1 (farmer)	6.0	0.0	13.5	0.002
2 (blue collar)	39.3	85.0	46.1	
3 (white collar)	54.8	15.0	40.3	
Smoking history (pack/years)	167.9 (296.1)	273.6 (340.4)	143.1 (284.0)	0.107
0 (never)	29.8	5.0	26.8	0.076
1 (former)	36.9	40.0	42.9	
2 (current)	33.3	55.0	30.3	
BMI (kg/m^2)	26.9 (3.6)	28.3 (4.4)	26.8 (3.7)	0.191
LDL cholesterol (mmol/l)	3.8 (0.8)	4.7 (1.5)	4.0 (1.0)	0.004
HDL cholesterol (mmol/l)	1.3 (0.3)	1.2 (0.2)	1.3 (0.3)	0.527
SBP (mm Hg)	133.1 (15.7)	140.0 (18.7)	133.8 (16.9)	0.237
Leisure time physical activity (h/year)	136.4 (203.2)	169.6 (178.9)	120.9 (145.9)	0.269
Alcohol consumption (g/week)	76.5 (113.8)	54.2 (70.0)	82.9 (150.0)	0.711
Median	25.3	19.0	39.9	
Range	0-512.5	0-220.8	0 - 2853.0	
0 (abstainers)	13.1	20.0	11.1	0.587
1 (0.1-279.9 g/week)	82.1	80.0	83.2	
2 (280.0-2853.0g/week)	4.8	0.0	5.7	
Income (marks/year)	83,535.7 (46 890.7)	48,800.0 (21 338.4)	82,941.7 (53,348.8)	0.015

Median	72,500	43,500	74,000
Range	5,000-234,000	15,000-90,000	0 – 550,000

(Table 2 cont'd)

Mortality/incidence/100 000 person-years			
All-cause death	1,405	4,289	1,452
Cancer death	550	2,144	355
CVD death	550	1,787	696
CHD death	428	1,072	475
Acute coronary events	1,109	2,946	1,399
Alcohol-associated diseases	318	1,121	365

BMI, body mass index; LDL, low density lipoprotein; HDL, high density lipoprotein; SBP, systolic blood pressure; CVD, cardiovascular disease; CHD, coronary heart disease

Table 3. Relative hazards (RH) of all-cause death, cancer death, CVD death, CHD death, acute coronary events and alcohol-associated diseases in men with emotional, behavioural, and emotional/behavioural problems in childhood, compared with men without emotional, behavioural and emotional/behavioural problems in childhood as a reference group.

	RH (95%CI)					
	All-cause death	**Cancer death**	**CVD death**	**CHD death**	**Acute coronary events**	**Alcohol-associated diseases**
Events /Total n in the model	252/880	68/880	118/880	81/880	193/880	61/880
No emotional problems in childhood	1.0 (reference)	1.0 (reference)	1.0 (reference)	1.0 (reference)	1.0 (reference)	1.0 (reference)
Emotional problems in childhood						
Model 1	0.94 (0.61-1.44)	1.44 (0.71-2.90)	0.78 (0.39-1.53)	0.88 (0.41-1.92)	0.82 (0.49-1.36)	0.83 (0.33-2.06)
Model 2	0.82 (0.53-1.27)	1.44 (0.70-2.99)	0.63 (0.31-1.26)	0.70 (0.42-1.55)	0.70 (0.42-1.18)	0.79 (0.32-2.00)
Model 3	0.90 (0.58-1.38)	1.41(0.69-2.88)	0.73 (0.37-1.44)	0.85 (0.39-1.84)	0.81 (0.48-1.35)	0.86 (0.34-2.16)

Model 4	0.88 (0.57-1.36)	1.39 (0.67-2.88)	0.74 (0.38-1.47)	0.84 (0.39-1.83)	0.85 (0.51-1.43)	0.77 (0.31-1.94)

(Table 3 cont'd)

No behavioural problems in childhood	1.0 (reference)	1.0 (reference)	1.0 (reference)	1.0 (reference)	1.0 (reference)	1.0 (reference)
Behavioural problems in childhood						
Model 1	2.80 (1.57-5.02)	5.31 (2.29-12.36)	2.50 (1.01-6.14)	2.15 (0.68-6.85)	1.96 (0.92-4.20)	2.54 (0.79-8.20)
Model 2	1.88 (1.01-3.50)	5.09 (1.89-13.76)	1.46 (0.56-3.76)	1.23 (0.37-4.12)	1.37 (0.62-3.03)	1.93 (0.54-6.86)
Model 3	2.34 (1.31-4.20)	4.39 (1.89-10.19)	2.08 (0.85-5.13)	1.85 (0.58-5.90)	1.79 (0.84-3.83)	2.09 (0.65-6.77)
Model 4	2.23 (1.24-4.02)	3.85 (1.63-9.10)	1.92 (0.77-4.77)	1.56 (0.48-5.02)	1.40 (0.65-3.01)	2.18 (0.66-7.14)
No emotional/behavioural problems	1.0 (reference)	1.0 (reference)	1.0 (reference)	1.0 (reference)	1.0 (reference)	1.0 (reference)
Emotional/behavioural problems in childhood						
Model 1	1.31 (0.91-1.87)	2.29 (1.29-4.01)	1.10 (0.63-1.92)	1.14 (0.59-2.22)	1.00 (0.64-1.56)	1.17 (0.56-2.46)
Model 2	1.11 (0.76-1.62)	2.41 (1.29-4.50)	0.85 (0.47-1.52)	0.86 (0.43-1.72)	0.82 (0.52-1.30)	1.07 (0.49-2.32)
Model 3	1.23 (0.86-1.77)	2.22 (1.24-3.98)	1.01 (0.58-1.78)	1.08 (0.56-2.10)	0.98 (0.63-1.53)	1.17 (0.55-2.49)
Model 4	1.21 (0.85-1.73)	2.07 (1.16-3.67)	1.03 (0.58-1.80)	1.05 (0.54-2.04)	0.98 (0.63-1.53)	1.09 (0.51-2.30)

Notes.
Model 1 Adjusted for age and examination year
Model 2 The same as model 1 and childhood SEP, educational level
Model 3 The same as model 1 and occupation, income level
Model 4 The same as model 1 and biological and behavioural factors (systolic blood pressure, ratio of low density lipoprotein to high density lipoprotein cholesterol, body mass index, leisure time physical activity, smoking, alcohol consumption)
CVD, cardiovascular disease; CHD, coronary heart disease

Table 3 shows that men who had emotional/behavioural problems in childhood had a 2.29-fold (95% confidence interval (CI) 1.29 to 4.01) age- and examination-year adjusted risk of cancer death. After adjustment for the SEP in childhood and adulthood and for the biological and behavioural factors in adulthood, the association remained unchanged. All-cause, CVD, and CHD death, risk of acute coronary events, and alcohol-associated morbidity showed no associations with emotional/behavioural problems in childhood.

Table 3 also shows that there were no statistically significant relationships between emotional problems in childhood and adult all-cause, cancer, CVD, CHD mortality, acute myocardial infarctions and alcohol-associated morbidity.

Men who had behavioural problems in childhood had a 2.80-fold (1.57 to 5.02) age- and examination-year adjusted risk of all-cause death, a 5.31-fold (2.29 to 12.36) risk of cancer death, and a 2.50-fold (1.01 to 6.14) risk of CVD death. The association between behavioural problems and all-cause and cancer deaths was somewhat attenuated after adjusting for SEP in childhood and adulthood, and biological, and behavioural factors in adulthood, whereas CVD mortality risk was no longer significant after further adjustments. The risk of CHD mortality, acute myocardial infarctions, and alcohol-associated morbidity was also elevated, but the results were not statistically significant.

Discussion

Our findings suggest that behavioural problems in childhood are associated with increased risk of all-cause and cancer mortality in adulthood, even after adjustment for the socio-economic position in childhood and adulthood, and biological and behavioural factors in adulthood. There was also an elevated risk of CVD, CHD death, acute myocardial infarctions, and alcohol-associated diseases, but the results were not statistically significant. Combined emotional/behavioural problems score showed also a relationship with cancer death. This effect is likely to be driven by behaviour problems in childhood, because emotional problems did not show any effect when analysed on its own.

It is hypothesized that risky and self-harmful behaviour, exposure to dangerous environments, and low socio-economic status would explain the increased mortality risk with those having problem behaviours in childhood. The findings by Jokela, Ferrie and Kivimäki (2008) suggested that externalizing behaviours, and possible co-morbidity between internalizing and externalizing behaviours, in addition to adverse family environment in childhood, would cause the increased mortality risk in adulthood. Our results give some support to the hypothesis that behavioural problems in childhood could be manifested in the life course, through long-term risky lifestyle factors, such as smoking, which in turn increase the mortality risk in later life. It is possible, that shy and fearsome children do not engage themselves so easily to risk-taking or self-harmful behaviour, compared to aggressive personality types who may act more recklessly, causing damage to their health.

It is also possible that negative personality type can act as an independent risk factor for all-cause, cancer and CVD mortality. Cynical hostility is known to be associated with perceived stress, coping ability, and social support. Hostility may impair the positive effects of social support on stress, which may in turn produce greater neural, endocrine, or inflammatory physiological responses that facilitate greater disease burden (Tindle et al 2009). For example, Weidner et al (1987) found that Type A behaviour and hostility were linked with elevated levels of plasma and LDL cholesterol. They concluded that Type A and hostile individuals spend a lot of time in a high arousal/attentional state, which could be associated with increased sympathetic nervous system activity, that may affect to the atherosclerotic process. In the present study, men with behavioural problems in childhood had an unfavourable profile of baseline characteristics, including age, socio-economic status, smoking, and LDL cholesterol levels. Nevertheless, the relationship between behavioural problems in childhood and all-cause and cancer mortality, remained after adjustment for the potential confounding factors.

The present study had some noteworthy strengths, such as long follow-up time, the use of several confounding factors, and reliable mortality and morbidity data in the analyses, and the use of historical records in assessing childhood problem behaviours. In the retrospective study design, recall bias can cause underestimation of the true impact of childhood factors, as people may not remember all the details of the past. Limitations of the study are that the sample size is relatively small due to missing data, which leads to imprecise estimates. Another limitation is that although the use of external raters may be more objective than self-report for childhood factors, rater variability may contribute to random or systematic misclassification of the data. For example the prevalence for behavioural problems is low by contemporary standards. It is difficult to know how nurses interpreted these problem behaviours in Eastern Finland in the first half of the 20th century or what the true period prevalence was for behavioural problems, as there is no representative data. In addition, it is a significant limitation to have no information on the period between the childhood behavioural ratings and the adult outcomes. Adult behavioural factors, taken into account in the analyses, were not necessarily sufficient to determine the full lifetime risk of unhealthy behaviour and its changes during the life-course.

Although this study has insufficient statistical power, there is some suggestion that reported behavioural problems in childhood increase the risk of all-cause and cancer deaths in adulthood. The long-term effects of problem behaviours highlight the importance of early intervention of such problems in young children. The developmental route of behavioural problems into cancer remains still unclear, although adult lifestyle behaviour seems to play an important role in the risk association. It is also possible that different emotional and behavioural characteristics may act as risk factors for different kinds of diseases. More long-term epidemiological studies are needed to clarify the relationship between emotional and behavioural problems in early life and subsequent morbidity and mortality.

Acknowledgements

The Kuopio Ischaemic Heart Disease Risk Factor Study was supported by grants 118551, 41471, 1041086, 2041022, and 200942 from the Academy of Finland; 167/722/96, 157/722/98 from the Ministry of Education of Finland; HL44199 from the National Heart, Lung and Blood Institute of the United States and the city of Kuopio. Laura Kauhanen was supported by grants from the Juho Vainio Foundation and Finnish Cultural Foundation. The authors would like to thank Professor Jukka T. Salonen and Professor Rainer Rauramaa for the use of the original survey and Pertti Happonen, MD, Pirjo Halonen, M.Sc. and Kimmo Ronkainen, M.Sc. for statistical assistance. We are also grateful to Ms Eeva Kumpulainen for gathering the childhood data.

References

Bothig S. (1989) WHO MONICA Project: objectives and design. *International Journal of Epidemiology,* 18, 29-37.

Bleiker EM, VanderPloeg HM, Hendriks JH and Ader HJ. (1996) Personality factors and breast cancer development: a prospective longitudinal study. *Journal of the National Cancer Institute*, 88, 1478-1482.

Clark C, Rodgers B, Caldwell T, Power C and Stansfeld S. (2007) Childhood and adulthood psychologicall ill health as predictors of midlife affective and anxiety disorders: The 1958 British Birth Cohort. *Archives of General Psychiatry*, 64, 668-687.

Dick D, Viken R, Kaprio J, Pulkkinen L and Rose R. (2005) Understanding the covariation among childhood externalizing symptoms: genetic and environmental influences on Conduct Disorder, Attention Deficit Hyperactivity Disorder, and Oppositional Defiant Disorder Symptoms. *Journal of Abnormal Child Psychology,* 33, 219-229.

Edelman S. (2005) Relationship between psychological factors and cancer: an update of the evidence. *Clinical Psychology,* 9, 45-53.

Everson S, Kauhanen J, Kaplan G, Goldberg D, Julkunen J, Tuomilehto J and Salonen JT. (1997) Hostility and increased risk of mortality and acute myocardial infarction: the mediating role of behavioral risk factors. *American Journal of Epidemiology,* 146, 142-152.

Grossarth-Maticek R, Eysenck H, Pfeifer A, Schmidt P and Koppel G. (1997) The specific action of different personality risk factors on cancer of the breast, cervix, corpus uteri and other types of cancer: a prospective investigation. *Personality and Individual Differences,* 33, 949-960.

Horwitz S, Leaf P, Leventhal J, Forsyth B and Speechley K. (1992) Identification and management of psychosocial and developmental problems in community-based, primary care pediatric practices. *Pediatrics,* 89, 480-485.

Irwin M, Patterson T, Smith T, Caldwell C, Brown S, Gillin J, and Grant I. (1990) Reduction of immune function in life stress and depression. *Biological Psychiatry,* 27, 22-30.

Jokela M, Ferrie J and Kivimäki M. (2009) Childhood problem behaviors and death by midlife: the British National Child Development Study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48, 19-24.

Kaltiala-Heino R, Rimpelä M, Marttunen M, Rimpelä A and Rantanen P. (2005) Bullying, depression, and suicidal ideation in Finnish adolescents: school survey. *British Medical Journal,* 319, 348-351.

Kaplan G, Wilson T, Cohen R, Kauhanen J, Wu M and Salonen J. (1994) Social functioning and overall mortality: prospective evidence from the Kuopio Ischemic Heart Disease Risk Factor Study *Epidemiology,* 5, 495-500.

Kauhanen J, Kaplan G, Goldberg D and Salonen J. (1997) Beer bingeing and mortality: results from the Kuopio ischaemic heart disease risk factor study, a prospective population based study. *British Medical Journal,* 315, 846-851.

Kauhanen L, Lakka H, Lynch J and Kauhanen J. (2006) Social disadvantages in childhood and risk of all-cause death and cardiovascular disease in later life: a comparison of historical and retrospective childhood information. *International Journal of Epidemiology,* 35, 962-968.

Kelleher K, McInerny T, Gardner W, Childs G and Wasserman R. (2000) Increasing identification of psychosocial problems: 1979-1996. *Pediatrics,* 105, 1313-1321.

King SM, Iacono WG, McCue M. (2004) Childhood externalizing and internalizing psychopathology in the prediction of early substance use. *Addiction*, 99, 1548-1559.

Kumpulainen K, Räsänen E, Henttonen I, Almqvist F, Kresanov K, Linna S-L, Moilanen I, Piha J, Puura K and Tamminen T. (1998) Bullying and psychiatric symptoms among elementary school-age children. *Child Abuse and Neglect,* 22, 705-717.

Lakka T, Venäläinen J, Rauramaa R, Salonen R, Tuomilehto J and Salonen J. (1994) Relation of leisure-time physical activity and cardiorespiratory fitness to the risk of acute myocardial infarction in men. *New England Journal of Medicine,* 330, 1549-1554.

Laub JH and Vaillant GE. (2000) Delinquency and mortality: a 50-year follow-up study of 1,000 delinquent and non-delinquent boys. *American Journal of Psychiatry*, 157, 96-102.

Neeleman J, Wessely S and Wadsworth M. (1998) Predictors of suicide, accidental death, and premature natural death in a general-population birth cohort. *The Lancet,* 351, 93-97.

Peach B, Ruth D, Tennan, C and Tonkin A. (2003) Stress and coronary heart disease: psychosocial risk factors. *Medical Journal of Australia,* 178, 272-276.

Penninx B, Guralnik J, Pahor M, Ferrucci L, Cerhan J, Wallace R and Havlik R. (1998) Chronically depressed mood and cancer risk in older persons. *Journal of the National Cancer Institute,* 90, 1888-1893.

Persky V, Kempthorne-Rawson J and Shekelle, R. (1987) Personality and risk of cancer: 20-year follow-up of the Western Electric Study. *Psychosomatic Medicine,* 49, 435-449.

Price MA, Tennant CC, Smith RC, Butow PN, Kennedy SJ, Kossof MB and Dunn SM. (2001) The role of psychosocial factors in the development of breast carcinoma: Part I. The cancer-prone personality. *Cancer*, 91, 679-685.

Regier D, Kessler L, Burns B and Goldberg I. (1979) The need for a psychosocial classification in primary-care settings. *International Journal of Mental Health,* 8, 16-29.

Rimpelä M, Luopa P, Räsänen M and Jokela J. (2006) Nuorten hyvinvointi 1996-2005 - Eriytyvätkö hyvinvoinnin ja pahoinvoinnin kehityssuunnat [Welfare trends among the youth 1996-2005]. In M. Kautto. ed. *Suomalaisten hyvinvointi 2006 [Welfare trends in Finland 2006]* Pp. 57-77. Stakes, Helsinki.

Roza SJ, Hofstra MB, Van der Ende J and Verhulst FC. (2003) Stable prediction of mood and anxiety disorders based on behavioral and emotional problems in childhood: a 14-year follow-up during childhood, adolescence, and young adulthood. *American Journal of Psychiatry,* 160, 2116-2121.

Salonen J, Nyyssönen K, Korpela H, Tuomilehto J, Seppänen R and Salonen R. (1992) High stored iron levels are associated with excess risk of myocardial infarction in eastern Finnish men. *Circulation,* 86, 803-811.

Salonen J, Salonen R, Seppänen K, Rauramaa R and Tuomilehto J. (1991) HDL, HDL2 and HDL3 subfractions and the risk of acute myocardial infarction. A prospective population study in eastern Finnish men. *Circulation,* 84, 129-139.

Shaffer J, Graves P, Swank R and Pearson T. (1987) Clustering of personality traits in youth and the subsequent development of cancer among physicians. *Journal of Behavioral Medicine,* 10, 441-447.

Shekelle R, Raynor WJ, Ostfeld A, Bieliauskas L, Lui S and Oglesby P. (1981) Psychological depression and 17-year risk of death from cancer. *Psychosomatic Medicine,* 43, 117-125.

Shepherd J, Farrington D and Potts J. (2004) Impact of antisocial lifestyle on health. *Journal of Public Health*, 26, 347-352.

Simonoff E, Elander J, Holmshaw J, Pickles A, Murray R and Rutter M. (2004) Predictors of antisocial personality. *British Journal of Psychiatry*, 184, 118-127.

Stein M, Keller S and Schleifer S. (1985) Stress and immunomodulation: the role of depression and neuroendocrine function. *Journal of Immunology,* 135, 827-833.

Sweeting H and West P. (1998) Health at age 11: reports from schoolchildren and their parents. *Archives of Disabled Child,* 78, 427-434.

Tindle HA, Chang Y-F, Kuller LH, Manson JE, Robinson JG, Rosal MC, Siegle GJ and Matthews KA. (2009) Optimism, cynical hostility, and incident coronary heart disease and mortality in the Women's Health Initiative. *Circulation*, 120, 656-662.

Tuomilehto J, Arstila M, Kaarsalo E, Kankaanpaa J, Ketonen M, Kuulasmaa K, Lehto S, Miettinen H, Mustaniemi H and Palomäki P. (1992) Acute myocardial infarction (AMI) in Finland-baseline data from the FINMONICA AMI register in 1983-1985. *European Heart Journal,* 13, 577-587.

Weidner G, Sexton G, McLellarn R, Connor S and Matarazzo J. (1987) The role of type A behavior and hostility in an elevation of plasma lipids in adult women and men. *Psychosomatic Medicine,* 49, 136-145.

Zahn-Waxler C, Klimes-Dougan B and Slattery M. (2000) Internalizing problems of childhood and adolescence: prospects, pitfalls, and progress in understanding the development of anxiety and depression. *Developmental Psychopathology,* 12, 443-466.

Endnotes

[i] Chi-squared tests and independent samples T-tests were used to assess differences in the study sample and the rest of the KIHD cohort. The differences in baseline characteristics between the three groups were analysed by chi-squared tests and analysis of variance (ANOVA).