FACULTY OF
BUSINESS &
ECONOMICS

HILDA PROJECT DISCUSSION PAPER SERIES
No. 2/13, December 2013

If at first you don't succeed?
Fieldwork effort, panel attrition, and health-employment inferences in BHPS and HILDA

Stephen Pudney, University of Essex
Nicole Watson, University of Melbourne

The HILDA Project was initiated, and is funded, by the Australian Government Department of Social Services

Acknowledgements

This work was supported by the European Research Council (project no. 269874 [DEVHEALTH]), and an Australian Research Council Discovery Grant (#DP1095497). This paper uses unit record data from the British Household Panel Survey (BHPS) and the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The BHPS was funded by the Economic and Social Research Council (ESRC) and managed by the Institute for Social and Economic Research (ISER, University of Essex). The HILDA Project was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper, however, are those of the authors and should not be attributed to ESRC, ISER, DSS, or the Melbourne Institute.

1 Introduction

By observing people repeatedly over time, panel data make it possible to study changes in health states and the employment consequences that flow from them. Sample attrition is an obstacle to this kind of research. If panel members' agreement to continue participation is influenced by their health and employment status, then analysis based on people who remain in the panel may be biased because observed changes in health and employment are confounded with the survey participation process. Sample attrition bias may in turn be influenced by the design of the fieldwork procedures that are used to control loss of data through attrition. The recent interest in panel attrition and missing data has focused on methods for dealing with attrition in the estimation of statistical models. Less attention has been paid to the link between fieldwork procedures and the nature of attrition bias in statistical modelling.

Survey managers frequently focus on response rates as a measure of success of their study, and this has led to increasing persistence in fieldwork efforts to retain reluctant or hard-to-reach respondents. This "try, try, try again" policy is not necessarily a good one. It increases cost and has an uncertain effect on the quality of research findings since, as Groves (2006) warns, measures to increase response rates can perversely increase bias when people with distinctive values on the survey variable are differentially sensitive to the particular design feature used to raise response propensities. In a meta-analysis of 59 studies examining bias in survey estimates, Groves and Peytcheva (2008) found response rates to be a poor predictor of bias: most of the variation in bias was across estimates within the same survey (with the same response rate) rather than across surveys with different response rates.

The methods commonly used to study non-response bias include using rich sampling frame data or supplementary matched data (eg., Olson, 2006; Kreuter, Muller and Trappmann, 2010), comparing similar estimates from other sources (eg., Keeter et al, 2006), comparing alternative weighting adjustments (eg. Abraham et al, 2006; Hall et al 2013), and studying the variation in response bias among subgroups of respondents defined in terms of the level or type of effort required to gain an interview. Studies have variously used as indicators of fieldwork effort the number of calls made (Curtin, Presser, and Singer, 2000; Carlson and Strouse, 2005), the number of weeks in field until an interview is achieved (Carlson and Strouse, 2005; Qayad et al, 2010), and whether a refusal conversion attempt was required (Curtin et al, 2000; Retzer, Schipani, and Cho, 2004; Carlson and Strouse, 2005; Burton et al, 2006; Keeter et al, 2006; Peytchev et al, 2010). Others have used a mix of these to define subgroups (Lynn et al, 2002; Safir et al, 2002).

Longitudinal data offer rich information about non-respondents following an initial interview, but most of the research using measures of fieldwork effort has focused on cross-section data

and there are few longitudinal studies (we have only identified two: Burton et al, 2006 and Haring et al, 2009). Burton et al (2006) found that the BHPS refusal conversion reduces bias in the regional distribution of the sample, employment status and qualifications, but increases bias in sex, marital status and in the rate of paid employment. They found no effect on self-assessed health status. Similarly, Haring et al (2009) found extended fieldwork effort increases the bias in the employment rate but had no effect on health variables.

Studies using cross-section data typically find some differences in respondent characteristics between early and late respondents, or between easy- and hard-to-recruit respondents, but few differences in substantive variables such as health and employment (Curtin et al, 2000; Retzer et al, 2004; Peytchev et al, 2010; Qayad et al, 2010, Safir et al, 2002). Weighting can help reduce non-response bias, but may not eliminate it completely (Carlson and Strouse, 2005; Peytchev et al, 2010; Hall et al, 2013). In the US Community Tracking Study (CTS), Carlson and Strouse (2005) find late responders have better health, fewer doctor visits and postpone medical care they may need. Similarly, Qayad et al (2010) find late responders to the US Behavioural Risk Factor Surveillance Survey are more likely to have health care coverage, be in good health, be sufficiently physically active, but to have high cholesterol. They are less likely to have a range of health conditions, be obese and eat sufficient fruit and vegetables. Hall et al (2013) analyse three UK surveys and find respondents that are 'hard-to-get' are likely to have lower blood pressure, no long-standing illness, but also to be smokers; they are also more likely to be employed. These findings suggest that high-effort respondents tend to work long hours and be away from their homes, and thus less likely to provide an interview.

In this paper, we use simulation methods to estimate the possible consequences for data analysts of variations in fieldwork policy, and particularly variations designed to reduce cost by imposing limits on the fieldwork effort devoted to eliciting full household response. The use of simulation in this context was introduced by Carlson and Strouse (2005) in the context of access to health services as measured by the CTS. We go beyond that study in several important respects: we take a longitudinal view and allow for potential long-term consequences of allowing response to fall; we work with a primarily face-to-face household interview setting rather than the random-digit dialling telephone interviewing used in the CTS; we allow the cap on fieldwork effort to be adapted to household characteristics (particularly its size); and we examine the impact of changes in the pattern of response on the results of complex longitudinal multivariate modelling, in addition to sample means and proportions.

Simulation is an alternative to randomised controlled trials (RCT), which randomly assign survey participants to groups receiving different fieldwork 'treatments'. The RCT is often described as the 'gold standard' but it is open to objection. Any small-scale trial explicitly

designed as an experiment is necessarily quite different from a routine wave of an established large-scale survey. Trials are generally subject to closer attention from survey managers, often use a special group of interviewers and are temporary, rather than sustained, studies, so the extrapolation of their results to the practical situation of a large-scale continuing survey is uncertain. While the lack of experimental control in our simulation approach is a disadvantage for the causal interpretation of observed effects, there is an offsetting advantage in terms of external validity for surveys used in real-life applied research.

Our aim is to examine the process of panel response in relation to health and employment, which are important interrelated subjects of longitudinal research. We take a comparative approach, using Australian and British household panel surveys to indicate the robustness of our findings. The next section compares the two surveys and summarises the health and employment data that we focus on. Section 3 describes the pattern of non-response in each survey and investigates the relationship between response and fieldwork persistence. Section 4 develops the simulation approach to evaluate the possible impact of alternative fieldwork policies. In section 5, we extend the simulation method to consider the possibility of irreversible long-term impact of fieldwork policy on sample integrity.

2 Two household panels: BHPS and HILDA

Since their origin with the US Panel Survey of Income Dynamics in 1968, household panel surveys have become a major resource for economic and social research and, increasingly, health research. The British Household Panel Survey (BHPS) and the Household, Income and Labour Dynamics in Australia (HILDA) Survey are two leading examples, with similar basic designs, but significant differences of detail and implementation. Here, we outline their design and operational features, define and summarise the health and employment measures that we focus on in this study, and the patterns of attrition in the two surveys. The characteristics of the BHPS and HILDA are explained in detail in Lynn (2006) and Wooden and Watson (2007) respectively; see also Frick et al (2007) for a comparison. Note that the BHPS has been absorbed into the much larger UK Household Longitudinal Survey (also known as *Understanding Society*); we only use data up to 2008 when the BHPS ended. We summarise the two surveys in Table 1 below.

Table 1 Design and operational differences between BHPS and HILDA

Design feature	BHPS	HILDA
Sample	5,538 responding households originally selected by stratified 2-stage random sampling from (most of) Great Britain. Sample expanded with low-income boost for the *European Community Household Panel* and inclusion of Northern Ireland at wave 7 and Scotland and Wales boost at wave 9 and further expansion of the Northern Ireland sample at wave 11	7,682 responding households originally selected by stratified, 3-stage area based sample design excluding very remote parts of Australia and non-private dwellings. Sample expanded with a general top-up at wave 11
Eligibility	All household members aged 16+	All household members aged 15+
Time period	Annual, 1991-2008; Main fieldwork period Sep-Nov each year	Annual, 2001 onwards
Availability of variables	Health variables not available in comparable form for waves 9 and 14. Health satisfaction not available for waves 1-5 or wave 11. Paradata on fieldwork outcomes not available for waves 1-3.	Waves 1-11 available
"Following rules"	Original sample members, their children born/adopted after 1991 and their other parent (if not an original sample member) are followed over time. Information about all household members living with these sample members is collected each wave.	Original sample members, their children born/adopted after 2001 and their other parent (if not an original sample member) are followed over time. Immigrants arriving after 2001 who join the households of these sample members are also followed over time. Information about all household members living with these sample members is collected each wave.
Interview mode	PAPI for waves 1-8, CAPI for waves 9-18	PAPI for waves 1-8, CAPI for waves 9-11. Some interviews conducted by telephone, ranging from 0.5% in wave 1 to 10.1% in wave 8.
Wave 1 household response rate	Partial and full households = 74% Full households = 69%	Partial and full households = 66% Full households = 59%
Wave 1 respondents re-interviewed, excluding out-of-scope cases	At wave 10=70%	At wave 10=70%

2.1 Outcome variables

Any study of the consequences of attrition must focus on specific types of analysis, since no conclusion is likely to have general applicability. We are interested in the important relationship between health and employment, and we consider a range of different types of analysis applied to four specific binary indicators of the respondent's health and employment status:

(*i*) A low ***subjective assessment of general health***, derived from questions with a 5-point response scale. The BHPS question refers to the preceding 12 months, uses a scale 1 (very poor), 2 (poor), 3 (fair), 4 (good), 5 (excellent), and is carried in the main interview. HILDA carries the question in a self-completion questionnaire, refers to health 'in general' rather than at a specific time, and uses a scale 1 (poor), 2 (fair), 3 (good), 4 (very good), 5 (excellent).[1] We use binary indicators of a health assessment at "fair" or worse (1-3 for BHPS; 1- 2 for HILDA).

(*ii*) A ***long-standing health problem*** which limits daily activities. There are differences between BHPS and HILDA in wording and the reference period used to define "long-standing". In both surveys this question appears in the main interview. "Daily activities" can mean any normal activity of daily life, not necessarily work.

(*iii*) An ***exit from employment*** that appears to be linked to poor health. At wave *t*, this event is defined by the requirements that the respondent was employed or self-employed at *t*-1 and not at *t*; and, at *t*, either describes his or her economic activity status as long-term sick/disabled or reports a health condition that limits work somewhat or a lot.[2]

(*iv*) ***Current employment*** status (defined to include self-employment).

Appendix Figure A1 shows the empirical age profiles of these health and employment indicators by birth cohort, using nonparametric smoothing to abstract from sampling variation. They show the expected pattern of slow deterioration over time in the general health with corresponding rising profiles for disability prevalence and occurrence of health-related job loss. Employment status follows the expected humped profile.

2.2 The BHPS and HILDA response processes

Figure 1 outlines the sequential process leading to the interview outcome. To achieve a valid response to the health or employment question of interest, it is necessary to make contact, gain immediate agreement to participate or successfully convert an initial non-response to

[1] For both HILDA and BHPS, this is a recoding of the original question which used the response scale in reverse order. The inclusion of the HILDA question in the self-completion questionnaire introduces an additional element of non-response relative to the BHPS, since this particular survey instrument has an annual response rate of 87-94%.
[2] In HILDA, respondents rate their ability to work on a 0 to 10 scale (0=not at all and 10=unable to do any work), and we assume 6 to 10 is similar to 'somewhat or a lot'.

agreement, and then complete the interview. The HILDA and BHPS response processes differ in a number of respects. The BHPS conversion step is only used for refusals, towards the end of the fieldwork period (typically from December, towards the end of the 3 month fieldwork period). The interviewer's comments are reviewed and a refusal conversion may be attempted from the office by telephone. If the household agrees, an interviewer is sent to conduct an interview in the family home. Refusal conversion was only used in the BHPS from wave 3 onwards and its use was only recorded in the dataset from wave 4.

The HILDA fieldwork period is divided into three distinct phases, with the conversion process occurring earlier than in the BHPS. During the first 9 weeks, all households are issued to field and, if the interviews have not been completed, they may be reissued to field for further attempts (typically by a different interviewer) in the next period, which lasts 8 weeks. The final 5-week period is used to follow up households that required extended tracking or where there is reason to contact the household again (for example a household member may be away, temporarily unwell or busy).

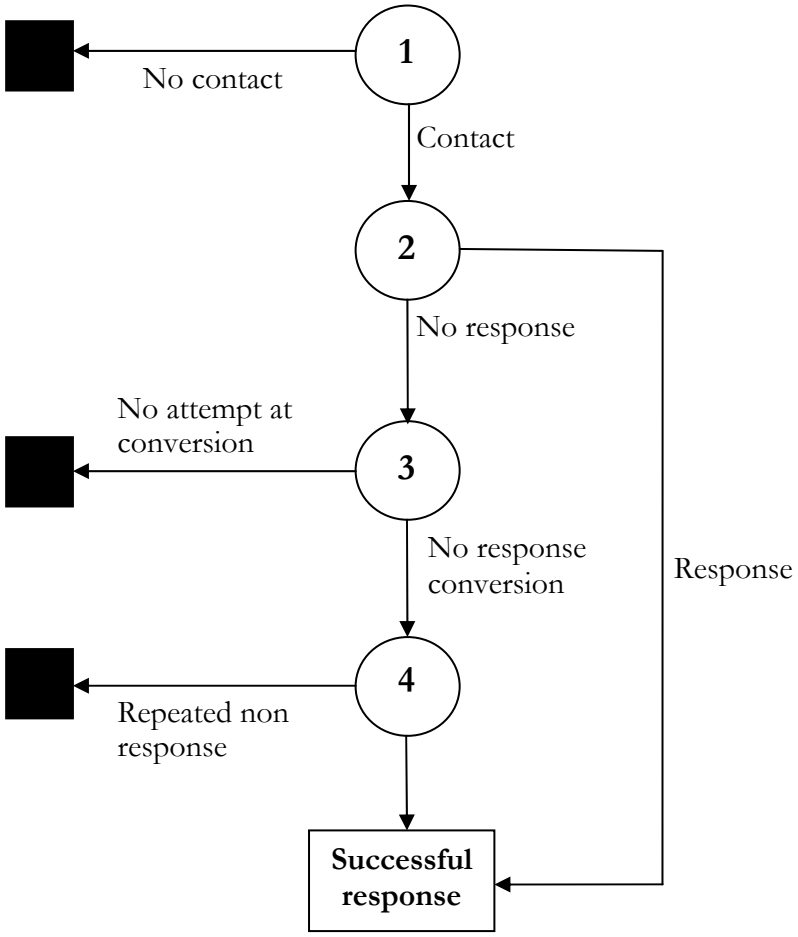
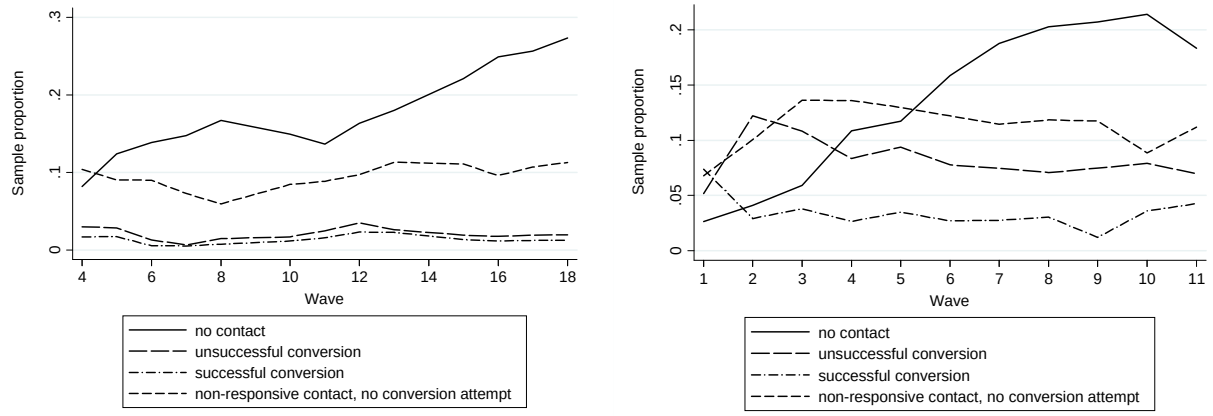


Figure 1 Schematic description of the BHPS and HILDA response process

Figure 2(a) plots the empirical prevalence at waves 4-18 of the four non-straightforward interview outcomes for the BHPS: non-contact; unproductive contact with no attempt at conversion; conversion attempted but with no eventual response; and successful conversion of an initially unproductive contact. The response rates are calculated on an individual basis and 'response' refers here to the existence of a valid interview response to a question on the individual's general health. The base for calculation of these response problems is the set of all individuals known to the BHPS and who are believed to be in-scope at a particular wave (including earlier "adamant refusal" or "long-term non-contact" cases who are never revisited). Where a panel member makes an adamant permanent refusal and is not subsequently re-issued to field, we classify these later missing observations as non-contact, which accounts for the strong rising trend in non-contacts in Figure 2. We cannot observe how long a permanent refusal/non-contact remains in scope after leaving the panel. For our calculations, we assume that all such cases remain in-scope before age 85, except for temporary sample members. This is an approximation, but it is clearly preferable to the extreme assumptions that all permanent sample members either become out-of-scope following permanent exit from the panel, or remain alive and in-scope indefinitely.



(a) BHPS (b) HILDA

Figure 2 Wave-specific rates of non-response to BHPS and HILDA (general health question)

Figure 2 reveals some marked differences between the BHPS and HILDA. The unsuccessful conversion rate is much higher in HILDA, reflecting the practice in HILDA to attempt conversions at an earlier stage of the fieldwork when typically a different face-to-face interviewer re-approaches the household. A second difference is that the non-contact rates in the HILDA Survey are slightly lower than the BHPS. It should be borne in mind that the BHPS sample was expanded twice during the period covered by waves 8-12 and this accounts for some of the rise in non-contact and refusal rates during that period, since attrition rates are

highest soon after joining the panel. HILDA only added additional sample at the last wave considered here, so that effect is largely absent from the HILDA sample.

The response process depicted in Figure 1 implies five possible routes to a response or non-response outcome, whose sample proportions are summarised in Table 2. For the BHPS, we give two sets of figures, one for the whole sample for waves 4-18, the other restricted to individuals who entered the panel after wave 3, for whom their response outcomes are observed for the whole of their panel membership. For the HILDA Survey, we also provide two sets of figures, one for the variables collected in the person interview and the other for the low general health score that is collected in the self-completion questionnaire.

Table 2 Empirical distribution of routes to response

Interview outcome	Sample proportion (%)	No. of interviewer visits	Low general health score (%)	Limiting disability (%)	Health-related employment exit (%)	Employment (%)
		BHPS				
No contact	20.2 [14.5]	1.56 [2.10]	-	-	-	-
Contact →unconverted initial non-response	9.7 [13.8]	2.37 [2.47]	-	-	-	-
Contact → initial non-response → conversion→non-response	2.1 [2.5]	2.87 [2.99]	-	-	-	-
Contact→ initial non-response → conversion→response	1.4 [1.5]	3.17 [3.30]	29.0 [28.9]	17.6 [19.6]	0.79 [0.81]	64.1 [62.2]
Contact→ response	66.6 [67.7]	2.59 [2.74]	30.6 [31.1]	17.5 [19.2]	0.99 [1.13]	56.8 [55.6]
		HILDA				
No contact	14.3 [14.3]	0.89 [0.89]	-	-	-	-
Contact →unconverted initial non-response	4.3 [11.3]	5.35 [6.06]	-	-	-	-
Contact → initial non-response → conversion→non-response	6.7 [8.2]	9.8 [10.42]	-	-	-	-
Contact→ initial non-response → conversion→response	4.9 [3.4]	10.59 [9.52]	16.9	22.4	1.3	68.5
Contact→ response	69.9 [62.8]	4.76 [4.57]	17.0	26.7	1.0	63.4

[1] BHPS: waves 4-18 pooled [subsample of new entrants at or after wave 4 in square brackets]; HILDA: waves 1-11 pooled [estimates for low general health score variable from the self-completion questionnaire in square brackets]. All means are unweighted.

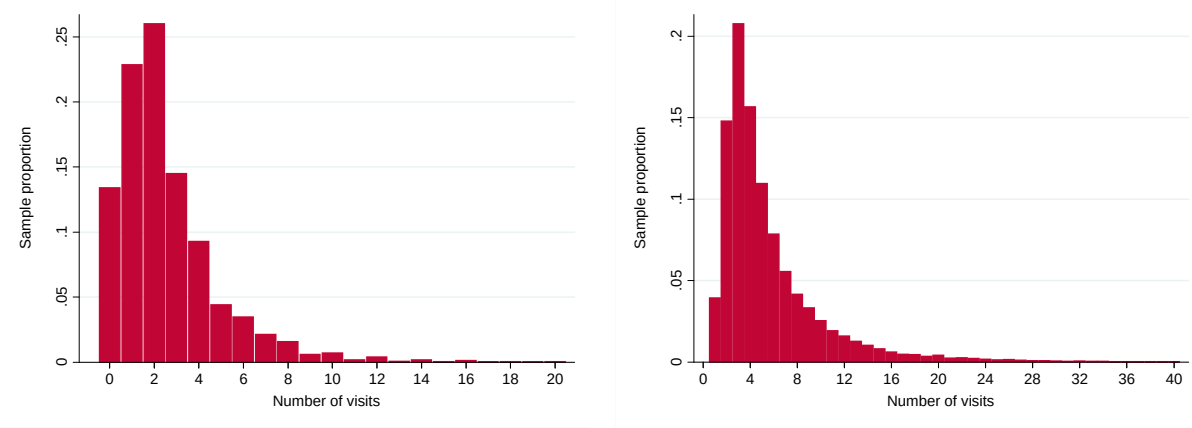
3 Fieldwork effort and sample characteristics

The optimal degree of fieldwork persistence is not obvious, nor is the influence of persistence on the results of analysis based on achieved samples. Both BHPS and HILDA use a high degree of persistence in their attempts to achieve a complete set of interviews for each household.

3.1 Measures of fieldwork effort

Our measures of fieldwork effort are based on counts of contacts between the fieldwork operation and panel households. It is important to realise that these counts are made at the household level. It may be that a particular household member was successfully interviewed during the first visit, but that more calls were needed to gain interviews with all household members, so our measure of fieldwork effort is approximate when applied at the individual level, and tends to over-estimate the average level of effort required to achieve an individual interview.[3] In the BHPS, the count of visits is based on handwritten cover sheet entries made by interviewers, edited by fieldwork managers to ensure that the count covered only visits made in person by an interviewer to the household's address, not telephone calls made in an effort to arrange an appointment for a visit. There is evidence of recording error since some of these entries are clearly invalid – for example zero is recorded for 13.6% of interviewed individuals and there are a few implausible counts in excess of 40 (around 0.01% of the individual sample). In HILDA, for waves 1-8, calls were recorded on the front page of the household form and subsequently recorded within the CAPI program. For HILDA, the count includes personal visits and some telephone calls (where they result in an appointment, interview or some information relevant to coding the outcome of the case). Consequently, the mean number of calls recorded in HILDA (5.5) is roughly double that recorded in the BHPS (2.7). Figure 3 compares the distribution of the number of interviewer calls as recorded in the two surveys. In the BHPS analysis from this point on, all zero counts of interviewer visits to respondent households have been recoded as single visits, giving a modified distribution close to geometric form. Dropping these cases instead makes no perceptible difference to our subsequent findings.

[3] The importance of this depends on household characteristics – for instance, large households are likely to require more visits to achieve a full set of interviews. Using the detailed HILDA call record information available in wave 10, we find that the mean number of additional calls to a multi-adult household after the first interview is 1.8. Some of these calls will be to pick up any remaining Self-Completion Questionnaires from the household (this may account for about 0.5 of the calls) and the remainder will be to contact and interview other household members. The fieldwork for a vast majority (86%) of multi-adult households are completed within 2 additional calls after the first interview.



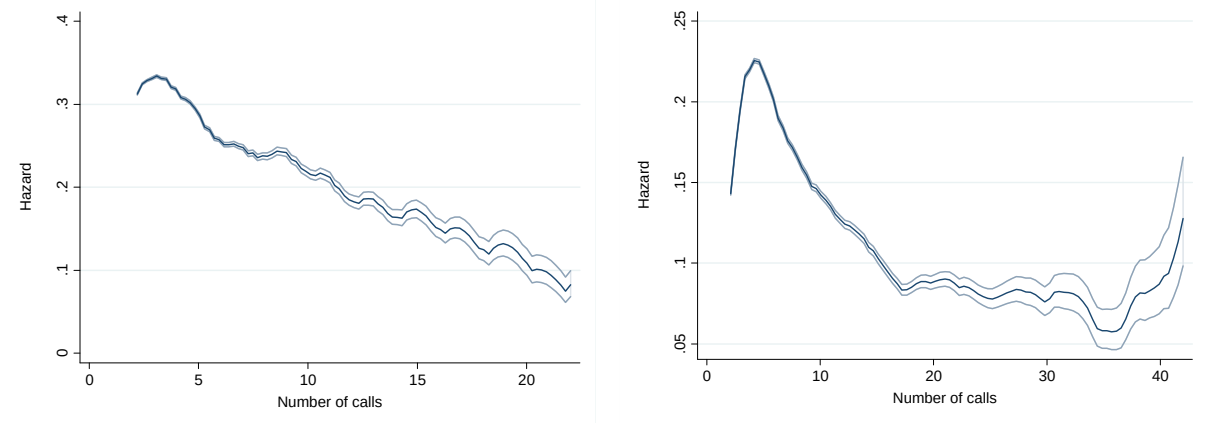
(a) BHPS: waves 4-18 pooled (b) HILDA: waves 1-11 pooled

Figure 3 Distributions of the number of interviewer calls

We characterise the distribution via a quasi-hazard rate for interview with respect to the number of calls, treating cases of non-response as censored at the recorded number of visits:

$h(m)$ = Pr(interview within m visits | household response incomplete after m-1 visits) (1)

Figure 4 shows that, for BHPS, where calls are measured as attempted interviewer visits, there is a smoothly decreasing quasi-hazard, reflecting diminishing returns to fieldwork persistence. For HILDA, the calls measure includes attempts by both telephone and personal visit and the shape of the quasi-hazard is quite different. The initial rise represents initial contacts to set up appointments, followed by a smooth decline , then a levelling-off at around 18 contact attempts,. Beyond 35 call attempts, the HILDA quasi-hazard rises, suggesting that extreme persistence is used selectively in cases where there is particular reason for confidence in eventual success.



(a) BHPS: all waves pooled (b) HILDA: waves 1-11 pooled

Figure 4 Empirical hazard rates for response (to employment question) by number of calls
(kernel smoothed hazard, 90% confidence intervals)

3.2 The relationship between fieldwork effort and outcomes

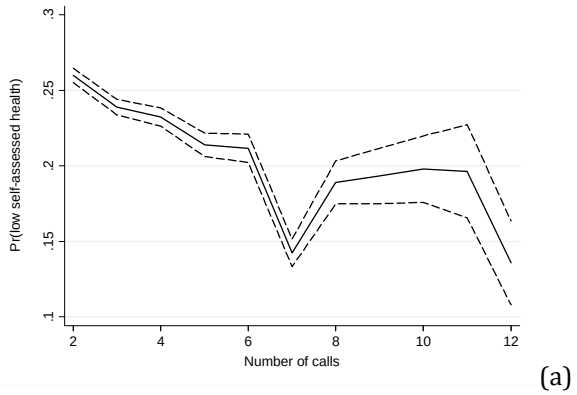
Table 2 shows some signs that response difficulties are associated with employment: those who are interviewed only after going through the conversion process have an employment rate around 7 (BHPS) or 5 (HILDA) percentage points higher than those interviewed directly without conversion. Figure 5 shows the empirical relationship between the number of calls required to achieve an interview and the observed health and employment outcomes. For the BHPS, we find a tendency to pick up relatively healthy respondents as fieldwork effort increases up to 7 visits; beyond that there is no significant relationship with health. The same pattern is apparent for HILDA over the comparable range 0-15 calls. The surveys also show a similar relationship between employment status and the number of contact attempts. In the BHPS, the employment rate rises from 55% to a plateau of roughly 63% at around 5 contact attempts; the pattern is very similar for HILDA, with the plateau reached at about 10-12 calls.

A more systematic way of summarising the relationship between fieldwork effort and the nature of the responses achieved is to use the R-indicators proposed by Schouten and Cobben (2007) and Cobben and Schouten (2008). They define two measures:

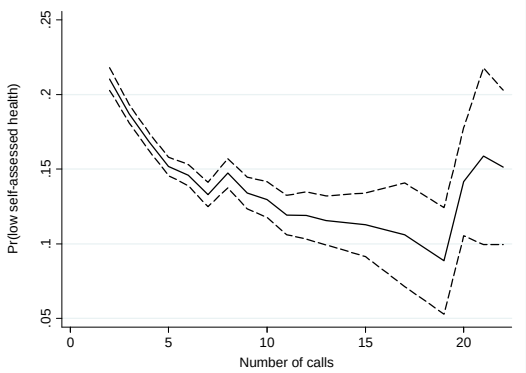
$$R(\rho) = 1 - 2S(\rho) \tag{1}$$

$$B(\rho, y) = [1 - R(\rho)]S(y)/2\bar{\rho} \tag{2}$$

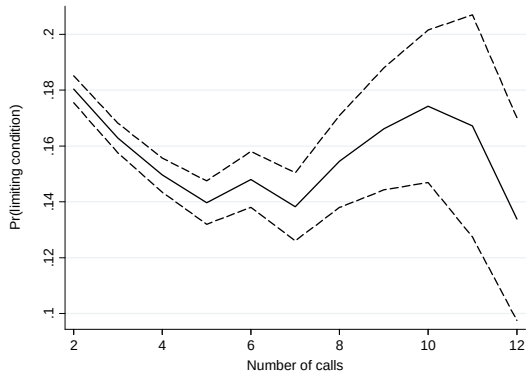
where: $\rho = \Pr(\text{response} \mid x)$ is the estimated propensity score; x is a vector of auxiliary variables; $\bar{\rho}$ and $S(\rho)$ are estimates of the mean and standard deviation of the propensity score, weighted to adjust for non-uniform sampling selection; and $S(y)$ is the weighted standard deviation of any particular binary survey measure of interest. The *R*-indicator (1) measures the extent to which the composition of the responding sample differs from the population composition with respect to x. A value of 1 indicates complete uniformity of response with respect to x; lower values indicate heterogeneity of response and therefore scope for bias. The measure $B(\rho, y)$ is a corresponding upper bound on the bias in the sample proportion of individuals possessing the characteristic measured by y; the bias will necessarily be zero if $R(\rho) = 1$. However, it should be borne in mind that, like all other propensity score methods, this assumes "selection on observables", so bias may still be substantial if there is important unobservable heterogeneity in response, even if $B(\rho, y)$ is very small for a particular vector x.



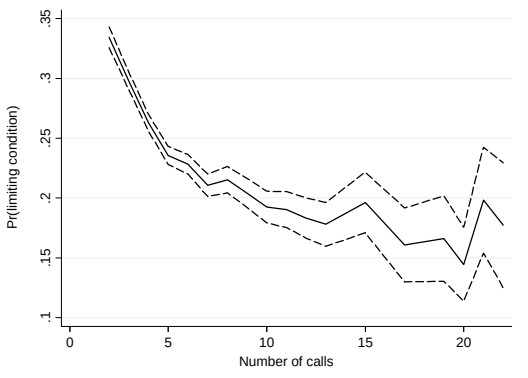
(a) BHPS low health assessment



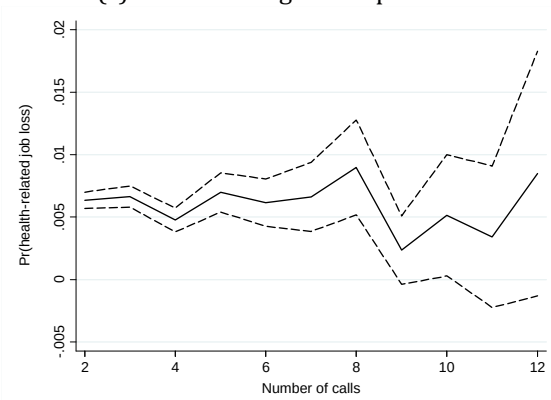
(b) HILDA low health assessment



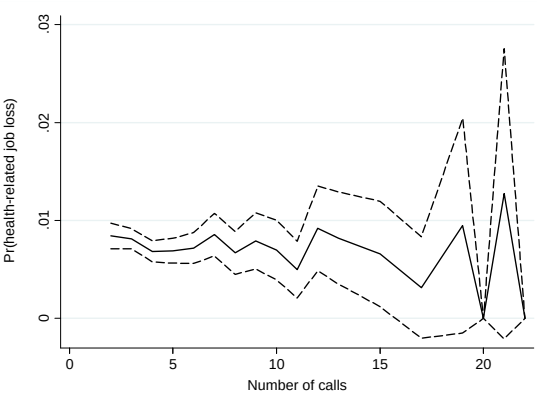
(c) BHPS limiting health problem



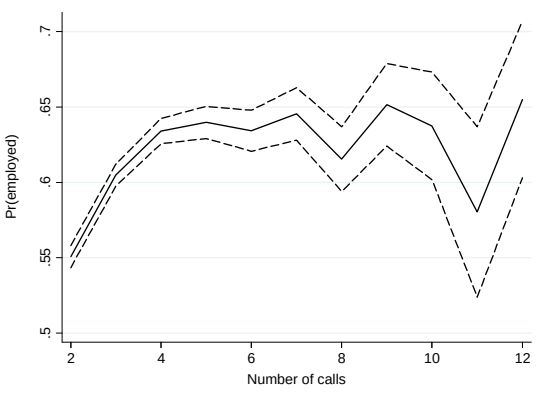
(d) HILDA limiting health problem



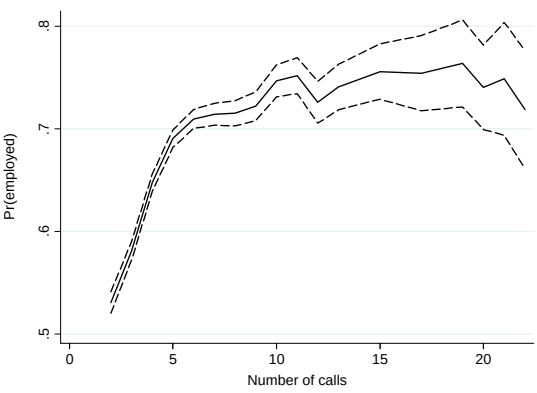
(e) BHPS health-related employment exit



(f) HILDA health-related employment exit



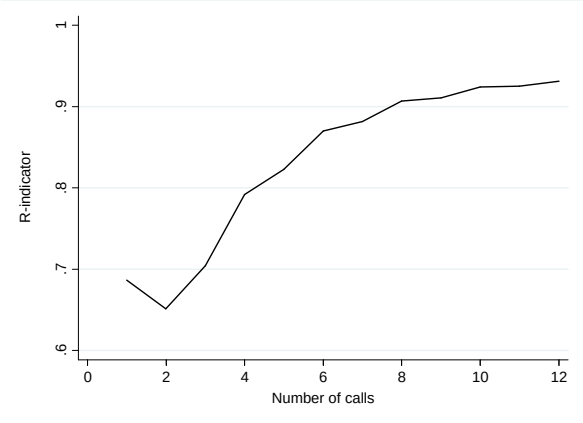
(g) BHPS employment



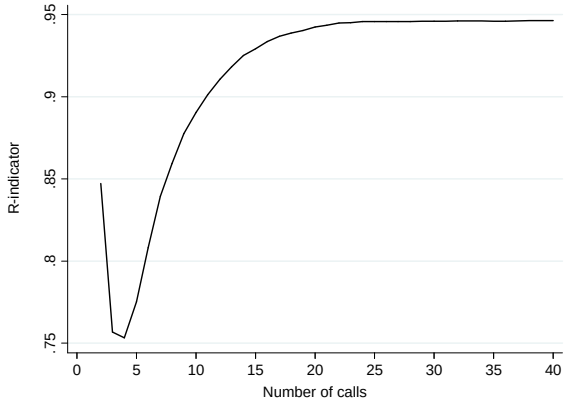
(h) HILDA employment

Figure 5 BHPS and HILDA health and employment measures by number of calls made by interviewer (all available waves pooled)

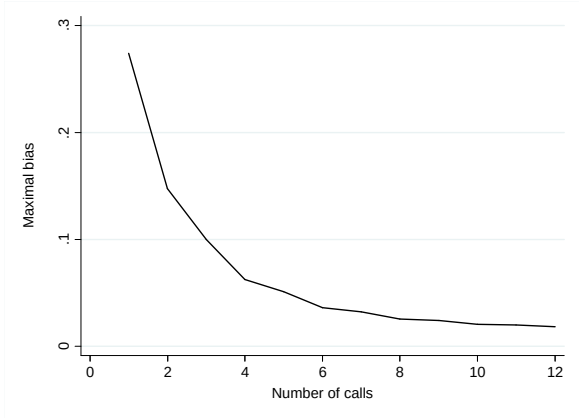
We estimate $R(\rho)$ and $B(\rho, y)$ for subsamples with various limits on the number of fieldwork calls and show the estimated maximal bias, $B(\rho, y)$, for a hypothetical indicator with $E(y) = 0.5$, implying $B(\rho, y) = (1 - R(\rho))/4\bar{\rho}$. The vector x contains variables observed at the previous wave, including age, sex, education level, marital status, number of adults and children in the household, region, whether moved to new address, time in panel, sample member status, low health status, limiting health condition and employment status. Our sample is all previous wave respondents and the selection weights are the response weights from the previous wave, which are designed to produce a representative sample at that wave. The results are presented in Figure 6. These indicators suggest that there are limited gains from the fieldwork in terms of reducing variability in response propensities of reducing the bias after about 10 calls for BHPS and 20 calls for HILDA (at least as far as can be measured by the auxiliary variables).



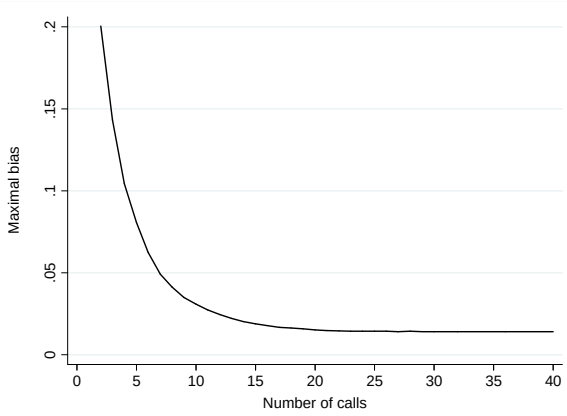
(a) BHPS R-indicator: all waves pooled



(b) HILDA R-indicator: waves 1-11 pooled



(c) BHPS maximal bias: all waves pooled



(d) HILDA maximal bias: waves 1-11 pooled

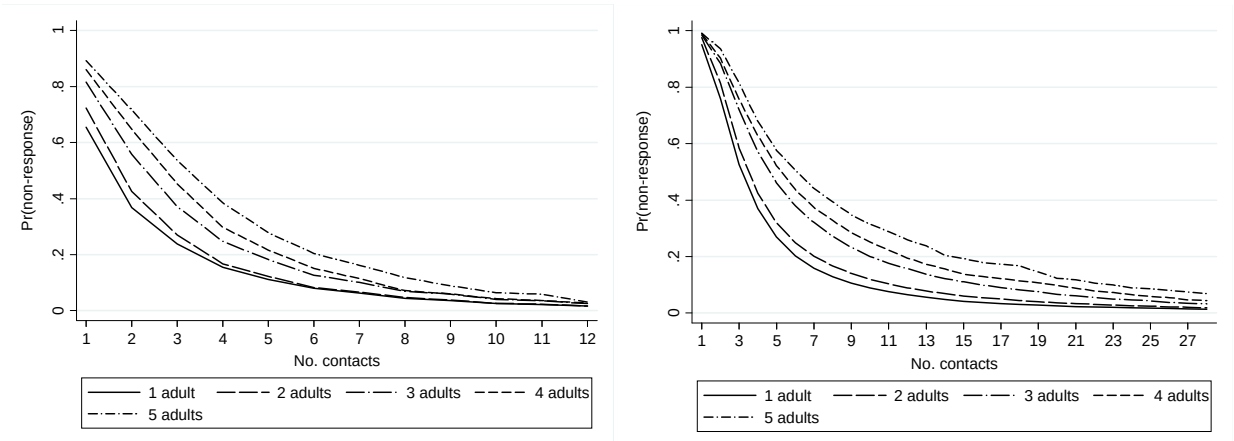
Figure 6 BHPS and HILDA R-indicators and maximal bias by number of calls made by interviewer (all available waves pooled)

4 Simulating the consequences of alternative fieldwork procedures

We simulate the effect of a less persistent fieldwork policy by discarding observations on cases that require more than a given amount of fieldwork effort for achievement of an interview. In this section, we assume that the loss of an interview at any wave has no implications for panel participation beyond the current wave – a strong assumption which we relax in section 5. We consider two forms of fieldwork curtailment. The simpler one uses an absolute limit, M, on the number of fieldwork contacts with any household, irrespective of its characteristics. There are three alternative regimes of M = 4, 6 and 9 for BHPS and 8, 12 and 18 for HILDA, giving simulated loss rates of approximately 2%, 5% and 15% of observations respectively for each survey.

We also use a household size-specific adaptive rule to achieve a more uniform rate of household response across all household types. This implies a higher limit for larger households for whom full response by all household members may take longer to achieve. Figure 7 confirms this, using nonparametrically-estimated survivor functions, $\Pr(r_{it} = 0 \mid M_{it} = m, S_{it})$, where: M_{it} is the number of interviewer contacts with the household containing individual i; r_{it} is the interview outcome (r_{it} = 1 for successful response, r_{it} = 0 for non-response); and S_{it} is the number of household members eligible for the full adult interview.

Our simulations are conservative in the sense that we interpret reductions in aggregate fieldwork effort as a net reduction in resources spent on the survey. If, instead, the resources saved by curtailing fieldwork persistence were used to expand the sample size, sample losses would be smaller than those simulated (although biases would be unaffected by a pure sample expansion).



(a) BHPS: all waves pooled

(b) HILDA: all waves pooled

Figure 7 Empirical survival functions for non-response (to employment question), by number of adults in the household

We choose household size-specific limits on the number of contacts to achieve an approximately equal response probability between surveys and across size groups. Again, we simulate three curtailment regimes, set out in the bottom panel of Table 3, reducing interviews by approximately 2%, 5% and 15%, with callback limits ranging from 3 to 12 (BHPS) and 6 to 30 (HILDA). The consequent savings in terms of calls avoided are also summarised in Table 3. They range from 0.06 to 0.40 (BHPS) or 0.15 to 0.93 (HILDA) per case issued to field. For both surveys, these savings equate to a reduction of around 2.5-17% in the scale of contact effort, with consequent fieldwork cost savings. Note that there would be additional cost savings from the similar reduction in the number of interviews.

Table 3 Alternative limits on fieldwork contacts

	Curtailment of BHPS fieldwork effort			Curtailment of HILDA fieldwork effort		
	Minor	Moderate	Radical	Minor	Moderate	Radical
Fixed fieldwork limit						
	9	6	4	18	12	8
Lost interviews *	3,741 (1.6%)	11,771 (5.2%)	41,278 (17.4%)	2,367 (1.6%)	7,431 (5.0%)	19,918 (13.5%)
Average saved calls per individual issued to field	0.06	0.18	0.40	0.16	0.44	0.93
Adaptive fieldwork limits						
1 adult	8	5	3	13	9	6
2 adults	8	6	4	16	10	7
3 adults	10	6	4	22	15	9
4 adults	10	7	5	26	18	11
5+ adults	12	8	6	30	21	13
Lost interviews *	3,760 (1.7%)	11,855 (5.2%)	41,758 (17.6%)	2,314 (1.6%)	7,880 (5.3%)	19,636 (13.3%)
Average saved calls per individual issued to field	0.07	0.18	0.39	0.15	0.42	0.87

* Loss of person-year interviews (losses to specific variables vary slightly through item non-response). Radical curtailment restricts the number of calls and removes all refusal conversion attempts in BHPS and post-New Year non-response conversions in HILDA. Original sample size = 227,339 (BHPS) and 197,737 (HILDA), including cases not issued to field.

Table 4 summarises the simulation results in terms of observation counts and (unweighted) sample means. There are three main conclusions. First, relative to the full sample, the observations lost through curtailment of fieldwork are from people with significantly fewer health problems, as indicated by a subjective assessment of "fair" or worse, or a report of a limiting condition or disability. This is also true dynamically: the onset rate for each of these adverse states is significantly lower in the cases that would be lost in a substantially curtailed fieldwork regime.

Second, the employment rate is considerably and significantly higher in the subsample of observations that would be lost than in the full sample. However, in the BHPS (but not HILDA), for some groups of special interest, such as employed lone parents and the self-employed, there is no evidence of a large or significant difference between the observations lost through curtailment and the rest of the sample.

Third, the sample sizes available for certain types of analysis may become infeasibly small as a result of the more substantial curtailment regimes. For example, the size of the BHPS balanced sub-panel for waves 4-14 falls from 4,193 individuals to less than half that size for the most radical curtailment, despite the fact that the loss of total person-year observations is under 18%. If we are interested in tightly-focused research, for example on the recent education and employment decisions of young adults, we might select a sample of people reaching the age of 16 sometime in the period 2001-5 who are also observed at least 4 times from age 16 until the 2008 wave (this period has been chosen to give comparable samples for BHPS and HILDA). For the BHPS, there are just over 900 individuals meeting these requirements in the full sample, but the number reduces by 15-20% to 721 (fixed fieldwork limits) or 763 (adaptive limits). In the HILDA sample, the number falls from just under 1000 individuals to 774 or 811 respectively. This loss of sample numbers could be enough to threaten the viability of a marginally feasible study. We investigate this further in section 5, allowing for the possibility of additional permanent attrition induced by fieldwork curtailment.

Table 4 Simulated effect of restricting fieldwork effort on mean health and employment indicators

Outcome variable	Status quo	Curtailment of fieldwork effort [1]			Status quo	Curtailment of fieldwork effort [1]		
	Minor	Moderate	Radical		Minor	Moderate	Radical	
	BHPS				***HILDA***			
	Full sample mean (%)	*Mean in **discarded** observations (%)*			*Full sample mean*	*Mean in **discarded** observations (%)*		
	Absolute fieldwork limits							
Prevalence of low health assessment	20.2	19.2	16.0***	17.7***	17.0	11.5***	11.9***	12.7***
Disability prevalence	17.1	12.9***	13.2***	14.0***	26.5	15.2***	17.0***	18.9***
Employment rate	57.0	64.4***	64.2***	63.6***	63.7	73.4***	74.4***	73.4***
Onset of low health assessment	10.4	12.0***	9.2***	10.2	7.1	6.4	5.7***	5.7***
Onset of disability	5.6	5.3	4.7***	4.9***	8.2	6.1***	6.1***	6.7***
Health-related job loss	1.00	0.91	1.06	1.04	1.0	0.6	0.8	0.9
Lone parents' employment rate	50.4	48.0	47.4	52.6	60.4	45.5	66.1*	70.3***
Lone parents' disability rate	16.3	16.2	17.7	17.0	23.1	18.2	14.4***	18.4***
Proportion of self-employed	7.0	7.4	7.4	7.6**	6.7	6.0	6.1	6.7
Number of panel members present in all waves (%) [2]	4,193 (46%)	3,999 (45%)	3,318 (39%)	1,920 (26%)	7,229 (52%)	6,830 (49%)	5,852 (42%)	3,744 (30%)
Rising 16s: no. present ≥ 4 waves 2001-8 (%) [3]	902 (69%)	892 (69%)	844 (66%)	721 (59%)	996 (80%)	962 (78%)	905 (73%)	774 (64%)
	Adaptive Fieldwork Limits							
	Full sample mean	*Mean in **discarded** observations*			*Full sample mean*	*Mean in **discarded** observations*		
Prevalence of low health assessment	20.2	19.9	17.7***	18.6***	17.0	11.6***	12.8***	14.1***
Employment rate	17.1	14.0***	14.3***	14.9***	26.5	16.1***	18.1***	20.3***
Disability prevalence	57.0	65.2***	63.6***	62.5***	63.7	74.9***	74.7***	73.2***
Onset of low health assessment	10.4	12.5***	10.0	10.6	7.1	5.9	6.1**	6.6*
Onset of disability	5.6	5.4	5.0***	5.1***	8.2	5.7***	6.3***	7.0***
Health-related job loss	1.00	0.77	1.14	1.08	1.0	0.7	0.8	0.9
Lone parents' employment rate	50.4	49.5	50.1	52.1	60.4	48.4**	55.8	58.2
Lone parents' disability rate	16.3	17.0	17.8	17.2	23.1	16.8	19.5	22.1
Proportion of self-employed	7.0	7.5	7.3	7.5*	6.7	6.1	6.7	6.8
Number of panel members present in all waves (%) [2]	4,193 (46%)	3,956 (44%)	3,285 (39%)	1,876 (25%)	7,229 (52%)	6,731 (48%)	5,586 (40%)	3,480 (25%)
Rising 16s: no. present ≥ 4 waves 2001-8 (%) [3]	902 (69%)	895 (69%)	863 (67%)	763 (61%)	996 (80%)	980 (79%)	927 (75%)	811 (66%)

Note: *, **, *** = significant difference between means of discarded and retained observations at 10%, 5% and 1% levels, allowing for clustering within individuals. [1] Fieldwork regimes as defined in Table 3. [2] Out of waves 4-18 for BHPS; waves 1-11 for HILDA; % of those giving an interview in wave 4 (BHPS) or wave 1 (HILDA). [3] Of those interviewed at age 16 in any year 2001-5, the number (and %) who were interviewed at least 4 times by 2008.

We can also use these simulations to examine the impact of fieldwork persistence on the results of statistical modelling. Four illustrative models are used as a testbed: (*i*) a probit model of disability onset (the probability of observing a limiting disability at wave t for the set of respondents who reported no disability at t-1; (*ii*) a dynamic probit model of the occurrence of a low health state ("fair" or below) allowing for state dependence; (*iii*) a probit model of a health-related exit from employment; and (*iv*) a dynamic probit model of employment for people rising 16 during the period 2001-5. All models include covariates representing gender, a cubic function of age, dummies for education and year and, for (*iii*) and (*iv*), the local unemployment rate. We test the null hypothesis that there is parameter stability across the retained and discarded groups by including in a full-sample model additional covariates constructed as interactions between each of the covariates and a binary variable identifying observations that would be discarded under a curtailed fieldwork regime. We use a Wald test of the hypothesis that these interaction effects are zero and show the *P*-values in Table 5, for each of the three hypothetical fieldwork regimes.

For the BHPS, there is little evidence of structural instability for these models. Of the 24 tests, three are significant at the 5% level, all under radical curtailment of fieldwork effort, in the dynamic model of a low health state and the youth employment model. However, if we use a Bonferroni correction for the significance level to allow for multiple testing, the coefficient differences would not be significant at any conventional significance level. For HILDA, the evidence of instability is more extensive but erratic, with structural instability identified for the model of health-related job loss under absolute but not adaptive fieldwork limits; for the youth employment model under the moderate and radical fieldwork curtailment regimes; and for one of the estimates of the model of disability onset.[4]

"Significant" differences are not necessarily substantively important. In Table 5 we also indicate the magnitude of coefficient differences, using the ratio of Euclidean coefficient norms:

$$\Delta = \left(\hat{\beta}_T - \hat{\beta}_F\right)'\left(\hat{\beta}_T - \hat{\beta}_F\right)/\hat{\beta}_F'\hat{\beta}_F \qquad (4)$$

where $\hat{\beta}_T$ and $\hat{\beta}_F$ are the coefficients in the truncated and full samples. The coefficient differences are very small indeed, mainly as a consequence of the small proportion of observations lost, even with radical curtailment. If this robustness of research results is typical, one might ask whether the huge effort devoted by the managers of panels like the BHPS and HILDA to achieving low attrition rates is an efficient use of resources. The important issue for

[4] The full parameter estimates are voluminous and not reproduced here. They are available on request from the authors.

users of the panel is not the loss of data itself but the bias it may generate, and we have not found evidence that additional fieldwork effort has much impact on the results of typical statistical modelling.

Table 5 *P* values and measures of coefficient stability under curtailment of fieldwork with no induced permanent attrition

Survey	Fieldwork curtailment [1]	Disability onset model		Dynamic model of low health state		Model of health-related job loss		Model of youth employment	
		P	Δ	*P*	Δ	*P*	Δ	*P*	Δ
Absolute fieldwork limits									
BHPS	Minor	0.115	.0000	0.115	.0001	0.545	.0000	0.482	.0000
	Moderate	0.155	.0000	0.157	.0003	0.287	.0001	0.290	.0001
	Radical	0.112	.0001	0.113	.0014	0.027	.0002	0.039	.0002
HILDA	Minor	0.361	.0000	0.446	.0000	0.002	.0000	0.216	.0000
	Moderate	0.045	.0000	0.752	.0001	0.008	.0000	0.033	.0005
	Radical	0.114	.0002	0.530	.0013	0.009	.0003	0.015	.0002
Adaptive Fieldwork Limits									
BHPS	Minor	0.336	.0000	0.120	.0001	0.729	.0000	0.558	.0000
	Moderate	0.402	.0000	0.058	.0003	0.356	.0000	0.825	.0000
	Radical	0.173	.0001	0.009	.0014	0.360	.0001	0.462	.0001
HILDA	Minor	0.146	.0000	0.298	.0000	0.648	.0000	0.100	.0002
	Moderate	0.063	.0001	0.770	.0002	0.477	.0001	0.031	.0003
	Radical	0.495	.0002	0.844	.0014	0.872	.0001	0.475	.0004

[1] Fieldwork regimes as defined in Table 3.

5 The risk of permanent attrition

The simulations in section 4 ignore possible long-term effects of missing an interview, and therefore provide a lower bound on the potential damage to the panel from curtailment of fieldwork effort. There are several reasons why such long-term damage to the panel may occur. There may be a *precedence effect*: if respondents miss a wave and observe that no uncomfortable consequences follow, further non-response might become an easier option. There may be a *commitment signalling effect*: if interviewers are seen to abandon their contact attempts quickly, this may appear to potential respondents as a sign that their response is not highly valued, which might induce withdrawal of co-operation. These are in addition to a possible direct *contact effect*: an extended break in communication may make it more likely that information about household moves is missed. We simulate this composite process of permanent sample damage, making strong assumptions so as to provide an upper bound.

In the simulations, we first discard any observation which was achieved with fieldwork effort beyond the assumed limit (either fixed or adaptive). In a further step, with some probability specific to the individual, we also drop some or all later observations from that individual;

otherwise, we assume the fieldwork curtailment is confined to the observation directly affected. Since we do not have a controlled experiment to estimate the long-run impact of variations in fieldwork persistence, it is difficult to construct the appropriate probability to use for generating this additional long-term attrition. If we construct the probabilities on the basis of available data, the empirical association between non-response in one wave and the pattern of non-response in later waves is unlikely to be entirely causal – much of that association will be the result of the underlying unobservable propensity to respond. Our strategy is to use a simple empirical model to generate probabilities of induced permanent attrition, assuming that the empirical association is entirely causal, and therefore giving a pessimistic upper bound on the long-term consequences of fieldwork curtailment.

We model the joint probability of response at t+1 and t+2 conditional on non-response at t, using a sequential conditional decomposition. Each component is specified as a logit conditional on a set of covariates, x_{t-1}, describing the characteristics of the individual and his or her household at the most recent observation:

$$Pr(r_{t+1} = 0 \mid r_t = 0, r_{t-1} = 1, x_{t-1}) = \Phi(x_{t-1}\beta) \tag{3}$$

$$Pr(r_{t+2} = 0 \mid r_{t+1} = 0, r_t = 0, r_{t-1} = 1, x_{t-1}) = \Phi(x_{t-1}\gamma) \tag{4}$$

$$Pr(r_{t+2} = 0 \mid r_{t+1} = 1, r_t = 0, r_{t-1} = 1, x_{t-1}) = \Phi(x_{t-1}\delta) \tag{5}$$

where r_t is the response indicator. The model is estimated using data from waves 2-15 (BHPS) or 2-8 (HILDA) to allow at least three post-estimation periods for non-response to occur.

In the simulation, we treat any generated case where $r_{t+1} = r_{t+2} = 0$ as permanent withdrawal from the sample. This corresponds to the assumption that, in the new less persistent regime, fieldwork managers would abandon their contact attempts completely if a further two waves are lost following abandonment of an interview through the limit on contact attempts. Thus, for an individual who is observed in wave t to have a number of interviewer contacts above the limit M, we discard the wave t observation and further observations sequentially as follows:

(i) $\hat{r}_{t+1} = 0$ with probability $\Phi\left(x_{t-1}\hat{\beta}\right)$

(ii) $\hat{r}_{t+2} = 0$ with probability $\hat{r}_{t+1}\Phi(x_{t-1}\delta) + (1 - \hat{r}_{t+1})\Phi(x_{t-1}\gamma)$

(iii) $\hat{r}_{t+3} = \cdots = \hat{r}_T = 0$ if $\hat{r}_{t+1} = \hat{r}_{t+2} = 0$

The rates of sample loss are summarised in Table 6. Relative to the simulation with no persistent impact on sample integrity, data losses are considerably greater – almost double, under the radical scenario. The proportion of interviews lost rises to 6-34% (BHPS) and 3-23% (HILDA), while the savings in contact effort (of up to 0.6 and 1.1 respectively per BHPS and HILDA panel member issued to field) amount to a proportionate saving of 4-22% (BHPS) and 3-

19% (HILDA) in the total number of contact attempts, on top of the 6-34% (BHPS) and 3-23% (HILDA) savings in interview time. The potential long-term data losses and cost savings are clearly large, but what are the consequences for substantive researchers using the achieved sample?

Table 6 Alternative limits on fieldwork contacts, with induced permanent attrition

	Curtailment of BHPS fieldwork effort			Curtailment of HILDA fieldwork effort		
	Minor	Moderate	Radical	Minor	Moderate	Radical
Fixed fieldwork limit						
Lost interviews*	13,572 (5.8%)	38,507 (15.9%)	87,642 (33.7%)	5,253 (3.4%)	15,332 (9.7%)	36,849 (22.5%)
Average saved calls per individual issued to field	0.11	0.29	0.58	0.19	0.51	1.06
Adaptive fieldwork limits						
Lost interviews *	13,163 (5.7%)	37,730 (15.6%)	87,709 (33.9%)	5,033 (3.3%)	14,937 (9.6%)	33,693 (21.0%)
Average saved calls per individual issued to field	0.11	0.28	0.57	0.17	0.48	0.98

* Loss of person-year interviews. Losses to specific variables vary slightly because of item non-response. Radical curtailment restricts the number of calls as well as removing the refusal conversion attempts in BHPS and the post-New Year refusal conversions in the HILDA Survey. Original sample size = 227,339 (BHPS) and 197,737 (HILDA) , including cases not issued to field.

Comparison of Table 7 and Table 4 shows that induced permanent attrition increases the number of statistically significant differences between mean characteristics in the full sample and the sample diminished by fieldwork curtailment for BHPS but not HILDA. The main source of difference is the sample rate of self-employment which rises more for BHPS under both absolute and adaptive curtailment when induced permanent attrition is allowed for.

Sub-sample numbers are also affected more by fieldwork curtailment when the consequences may be permanent, although this effect is smaller than might be expected. For example, in the BHPS with no persistent losses triggered by fieldwork curtailment, the number of individuals observed for at least 4 waves at age 16 and after falls by 20% (absolute fieldwork limits) or 15% (adaptive limits) as we go from the full sample to radical curtailment. With induced permanent attrition, these loss rates rise by a modest margin of 6 percentage points. For HILDA, the effect is smaller still. Loss rates of 22% (absolute) and 20% (adaptive) worsen by 6 and 4 percentage points respectively. While there are certainly types of subgroup analysis that will be rendered infeasible by fieldwork curtailment, we find no grounds here to be unduly alarmed about the risk of additional permanent sample losses as a contributory factor.

Table 7 Simulated effect of restricting fieldwork effort on mean health and employment indicators, allowing for long-term attrition

Outcome variable	Status quo	Curtailment of fieldwork effort [1]			Status quo	Curtailment of fieldwork effort [1]		
		Minor	Moderate	Radical		Minor	Moderate	Radical
	BHPS				***HILDA***			
	Full sample mean (%)	*Mean in **discarded** observations (%)*			*Full sample mean*	*Mean in **discarded** observations (%)*		
	Absolute fieldwork limits							
Prevalence of low health assessment	20.2	18.9*	18.1***	18.8***	17.0	11.3***	12.2***	13.4***
Disability prevalence	17.1	14.3***	15.1***	15.2***	26.5	15.9***	17.7***	19.8***
Employment rate	57.0	64.8***	64.5***	63.7***	63.7	73.0***	74.3***	73.2***
Onset of low health assessment	10.4	10.5	9.7***	10.3	7.1	6.0	5.6***	6.0***
Onset of disability	5.6	5.7	5.3	5.2***	8.2	6.5***	6.4***	6.9***
Health-related job loss	1.00	1.25	1.20*	1.06	1.0	0.9	0.9	1.0
Lone parents' employment rate	50.4	45.7	45.8	52.1	60.4	50.9	65.6	69.3***
Lone parents' disability rate	16.3	14.4	18.9	17.3	23.1	18.9	16.5***	18.2***
Proportion of self-employed	7.0	8.4*	8.3***	7.8**	6.7	5.8	6.0*	6.7
Number of panel members present in all waves (%) [2]	4,193 (46%)	3,885 (45%)	3,148 (39%)	1,718 (26%)	7,229 (52%)	6,830 (49%)	5,852 (42%)	3,744 (30%)
Rising 16s: no. present ≥ 3 waves 2001-8 (%) [3]	902 (69%)	881 (68%)	816 (64%)	663 (56%)	996 (80%)	958 (77%)	881 (72%)	718 (62%)
	Adaptive Fieldwork Limits							
	Full sample mean	*Mean in **discarded** observations*			*Full sample mean*	*Mean in **discarded** observations*		
Prevalence of low health assessment	20.2	19.8	19.3*	19.6*	17.0	12.5***	13.2***	14.7***
Employment rate	17.1	15.9	16.1*	16.0***	26.5	16.8***	18.9***	21.1***
Disability prevalence	57.0	64.8***	63.4***	62.2***	63.7	74.4***	74.5***	72.8***
Onset of low health assessment	10.4	10.6	10.2	10.5	7.1	5.8	6.2**	6.6**
Onset of disability	5.6	5.8	5.6	5.4	8.2	6.2***	6.6***	7.1***
Health-related job loss	1.00	1.11	1.29**	1.09	1.0	0.8	0.9	0.9
Lone parents' employment rate	50.4	45.8	49.2	52.1	60.4	51.9	56.5	57.1
Lone parents' disability rate	16.3	16.1	17.9	17.2	23.1	17.3	20.4	21.6
Proportion of self-employed	7.0	8.6**	8.0**	7.5*	6.7	5.8*	6.6	6.9
Number of panel members present in all waves [2]	4,193 (46%)	3,861 (45%)	3,115 (39%)	1,680 (25%)	7,229 (52%)	6,731 (48%)	5,586 (40%)	3,480 (25%)
Rising 16s: no. present ≥ 4 waves 2001-8 [3]	902 (69%)	891 (69%)	837 (66%)	710 (59%)	996 (80%)	974 (79%)	906 (74%)	773 (64%)

Note: *, **, *** = significant difference between means of discarded and retained observations at 10%, 5% and 1% levels, allowing for clustering within individuals. [1] Fieldwork regimes as defined in Table 3. [2] Out of waves 4-18 for BHPS; waves 1-11 for HILDA; % of those giving an interview in wave 4 (BHPS) or wave 1 (HILDA). [3] Of those interviewed at age 16 in any year 2001-5, the number (and %) who were interviewed at least 4 times by 2008.

Table 8 summarises the significance and magnitude of change in the coefficients of our illustrative statistical models of health and employment. As in the simulations with purely temporary sample loss, although fieldwork curtailment causes significant coefficient changes in some cases (mainly in the two HILDA employment models with absolute fieldwork limits), their magnitude is very small and would be of no practical concern.

Table 8 *P* values and measures of coefficient stability under curtailment of fieldwork with simulated induced permanent attrition

Survey	Fieldwork curtailment [1]	Disability onset model		Dynamic model of low health state		Model of health-related job loss		Model of youth employment	
		P	Δ	*P*	Δ	*P*	Δ	*P*	Δ
Absolute fieldwork limits									
BHPS	Minor	0.060	.0000	0.274	.0000	0.035	.0001	0.236	.0000
	Moderate	0.016	.0000	0.190	.0005	0.175	.0002	0.606	.0003
	Radical	0.005	.0003	0.092	.0013	0.733	.0001	0.264	.0002
HILDA	Minor	0.458	.0000	0.621	.0001	0.003	.0000	0.335	.0000
	Moderate	0.103	.0000	0.592	.0001	0.010	.0001	0.085	.0001
	Radical	0.051	.0002	0.225	.0014	0.009	.0003	0.037	.0007
Adaptive Fieldwork Limits									
BHPS	Minor	0.127	.0000	0.079	.0001	0.049	.0000	0.000	.0000
	Moderate	0.005	.0001	0.014	.0006	0.006	.0002	0.933	.0001
	Radical	0.015	.0003	0.060	.0012	0.270	.0003	0.673	.0001
HILDA	Minor	0.645	.0000	0.118	.0001	0.167	.0000	0.101	.0000
	Moderate	0.163	.0001	0.494	.0002	0.783	.0001	0.144	.0000
	Radical	0.172	.0002	0.804	.0015	0.620	.0002	0.410	.0006

[1] Fieldwork regimes as defined in Table 3.

6 Conclusions

In practice, the managers of longitudinal surveys have to make difficult decisions about the appropriate degree of persistence to use in achieving interviews with hard-to-reach panel members. Current practice in two of the leading household panels – the BHPS and HILDA Survey – has been to exercise great persistence. There is a potential trade-off here: persistence is costly, so limiting fieldwork effort in difficult cases could free resources to pay for an expansion in the panel size. But there is little research available to tell us what we gain from fieldwork persistence and whether the "try, try, try again" policy is an efficient use of resources.

In this paper, we have used simulation to investigate the possible consequences of switching to a less persistent policy, concentrating on the important related areas of health and employment. Curtailment of fieldwork effort has been simulated by imposing alternative "minor", "moderate"

and “radical” limits on the effort devoted to achieving any given interview, using either crude fixed limits on interviewer calls or more sophisticated adaptive limits taking account of household size. We have used two alternative assumptions about the consequences of losing interviews through fieldwork curtailment, with such losses simulated to be either purely temporary or subject to a high risk of inducing permanent drop-out from the panel.

There are five main conclusions. First, the broad conclusions emerging from BHPS and HILDA evidence are similar, so that our findings appear to have quite broad validity.

Second, the losses from fieldwork curtailment produce clear changes in some important sample characteristics. This is particularly true for employment, where the effect of persistence is to increase substantially the employment rate in the achieved sample and, consequently, to reduce the sample prevalence rate of disability and ill-health. This finding is broadly what we would expect, given the typical findings in the research literature on survey response and panel attrition, where the difficulty of contacting employed individuals is a recurring theme. There is some support here for the policy of maintaining a high degree of fieldwork persistence, although it is a type of bias that is relatively easy to correct using standard weighting methods.

A third, more surprising, conclusion is that multivariate modelling of health, employment and the relationship between them is highly robust. We used a number of representative multivariate models (static and dynamic panel probit models) to investigate the impact of fieldwork curtailment. Although there is statistically significant evidence of a change in the coefficients of a few of these representative models of health and employment, in every case the magnitude of the coefficient changes was so small as to be of no practical importance.

A fourth finding is that, for some subgroup analyses, existing subsample sizes are sufficiently small that a radical curtailment of fieldwork effort could threaten the viability of the analysis. Of course, in a large survey it will almost always be possible to find some marginally-viable subgroup analysis that would be critically affected by a variation in sample design. It should also be borne in mind that, with a fixed survey budget, reducing fieldwork persistence would free resources, making possible a larger panel size, so our simulations overstate the viability threat somewhat.

A final, rather surprising, conclusion is that the threat of increased permanent drop-out from the panel appears less serious than we expected. There are reasons why we might expect some weakening of panel adherence: for example, the lower level of fieldwork persistence could undermine the sense of importance of the survey or the missing of an interview could trigger a ‘habit’ of non-response. Our simulation of a (very strong) causal link between a missed interview and subsequent permanent attrition does worsen most of the adverse consequences

of curtailment of fieldwork effort, but not dramatically so. Assumptions about the strength of this possible causal link do not appear to be critical to the choice of a fieldwork policy.

Overall, since the adverse impact of reducing fieldwork persistence is more modest than we expected, our findings leave open the possibility that there might be some gain to be made by switching from highly persistent fieldwork within a smaller panel to less persistent fieldwork within a larger panel. Given the currently available evidence, we certainly would not want to push this view, but it is one that deserves further research attention, using a wider range of measures than the health-employment relationship we are concerned with here.

References

Abraham, K., Maitland, A., and Bianchi, S. (2006), 'Nonresponse in the American Time Use Survey', *Public Opinion Quarterly*, Vol. 70, No. 5 Special issue, pp. 676-703.

Burton, J., Laurie, H., and Lynn, P. (2006), 'The long-term effectiveness of refusal conversion procedures on longitudinal surveys', *Journal of the Royal Statistical Society Series A*, Vol. 169, No. 3, pp. 459-478.

Carlson, B.L., and Strouse, R. (2005), 'The value of increasing effort to maintain high response rates in telephone surveys', ASA Section on Survey Research Methods.

Cobben, F. and Schouten, B. (2008), 'An empirical validation of R-indicators', Discussion Paper 08006 Statistics Netherlands, Voorburg/Heerlen.

Curtin, R., Presser, S., and Singer, E. (2000), 'The effects of response rate changes on the index of consumer sentiment', *Public Opinion Quarterly*, Vol. 64, pp. 413-428.

Frick, J. R., Jenkins, S. P., Lillard, D. R., Lipps, O. and Wooden, M. (2007). The Cross-National Equivalent File (CNEF) and its member country household panel studies, *Schmollers Jahrbuch* **127**, 627-654.

Groves, R.M. (2006), 'Nonresponse rates and nonresponse bias in household surveys', *Public Opinion Quarterly*, Vol. 70, No. 5 Special issue, pp. 646-675.

Groves, R.M., and Peytcheva, E. (2008), 'The impact of nonresponse rates on nonresponse bias', *Public Opinion Quarterly*, Vol. 72, No. 2, pp. 167-189.

Hall, J., Brown, V., Nicolaas, G., and Lynn, P. (2013), 'Extended field efforts to reduce the risk of non-response bias: Have the effects changed over time? Can weighting achieve the same effects?', *Bulletin de Méthodologie Sociologique*, Vol. 117, pp. 5-25.

Haring, R., Alte, D., Völzke, H., Sauer, S., Wallaschofski, H., John, U., and Schmidt, C.O. (2009), 'Extended recruitment efforts minimize attrition but not necessarily bias', *Journal of Clinical Epidemiology*, Vol. 62, pp.252-60.

Keeter, S., Kennedy, C., Dimock, M., Best, J., and Craighill, P. (2006), 'Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey', *Public Opinion Quarterly*, Vol. 70, No. 5 Special issue, pp. 759-779.

Kreuter, F., Müller, G., Trappmann, M. (2010), 'Nonresponse and measurement error in employment research', *Public Opinion Quarterly*, Vol. 74, No. 5, pp. 880-906.

Lynn, P., Clarke, P., Martin, J., and Sturgis, P. (2002), 'The effects of extended interviewer efforts on nonresponse bias', in R. Groves, D. Dillman, J. Eltinge, and R. Little (Eds.), *Survey nonresponse*, pp. 135-47, New York: Wiley.

Lynn, P. (ed.) (2006). Quality profile: British Household Panel Survey (version 2.0 waves 1-13: 1991-2003). Mimeo: University of Essex.

Olson, K. (2006), 'Survey participation, nonresponse bias, measurement error bias, and total bias', *Public Opinion Quarterly*, Vol. 70, No. 5 Special issue, pp. 737-758.

Peytchev, A., Baxter, R., Carley-Baxter, L.R. (2009), 'Not all survey effort is equal', *Public Opinion Quarterly*, Vol. 73, No. 4, pp. 785-806.

Qayad, M.G., Chowdhury, P., Hu, S., and Balluz, L. (2010), 'Respondent differences and length of data collection in the Behavioural Risk Factor Surveillance System', *Survey Methodology*, Vol. 36, No. 2, pp. 223-227.

Retzer, K.F., Schipani, D., Cho, Y.I. (2004), 'Refusal conversion: Monitoring the trends', American Association for Public Opinion Research Section on Survey Research Methods.

Rubin, D. B. (1976). Inference and missing data, *Biometrika* **63**, 581-592.

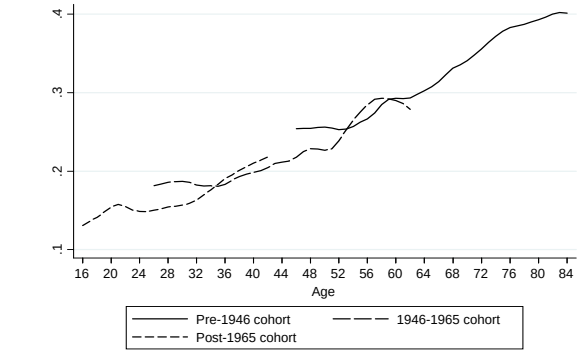
Safir, A., Steinbach, R., Triplett, T., and Wang, K. (2002), 'Effects on survey estimates from reducing nonresponse', in *Proceedings of the Conference: Strengthening our Community*, *AAPOR 2002*, 3024-3029.

Schouten, B. and Cobben, F. (2007), 'R-indexes for the comparison of different fieldwork strategies and data collection rules', Discussion Paper 07002 Statistics Netherlands, Voorburg/Heerlen.

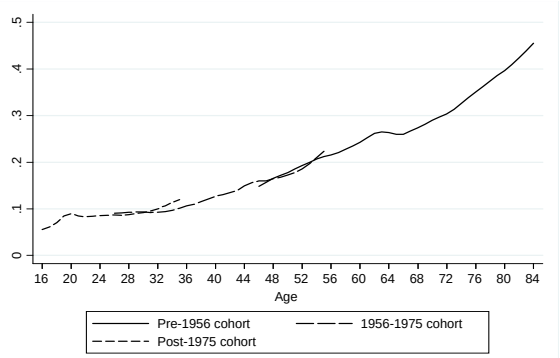
Uhrig, S. C. N. (2008). The nature and causes of attrition in the British Household Panel Survey. University of Essex: ISER Working Paper no 2008-05.

Wooden, M., and Watson, N. (2007). The HILDA Survey and its contribution to economic and social research (so far), *The Economic Record* **83**, 208-231.

Appendix 1: Additional figures

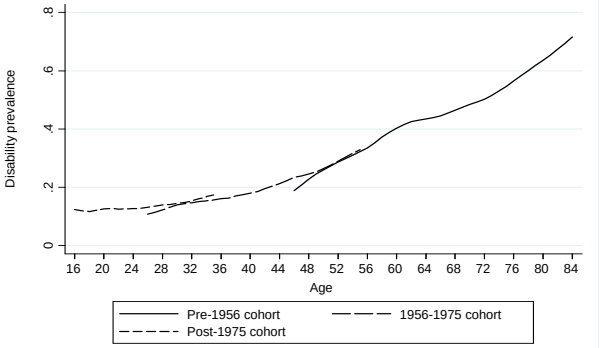


(a) BHPS low general health assessment

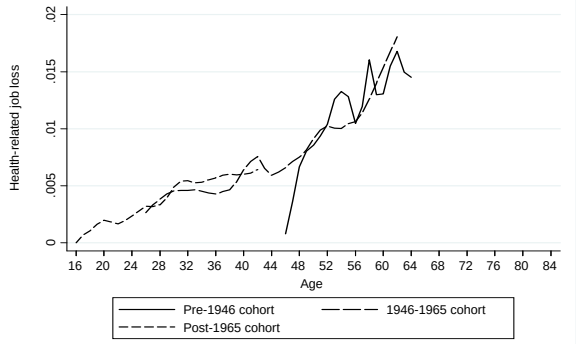


(b) HILDA low general health assessment

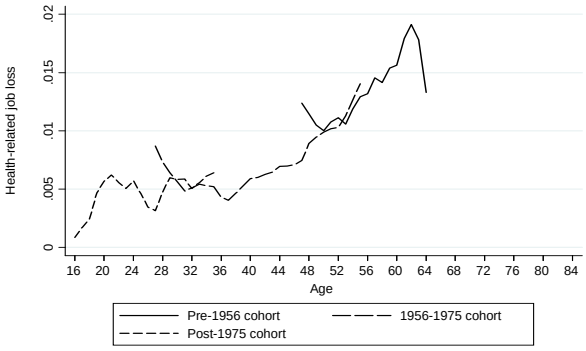
(c) BHPS limiting health problem



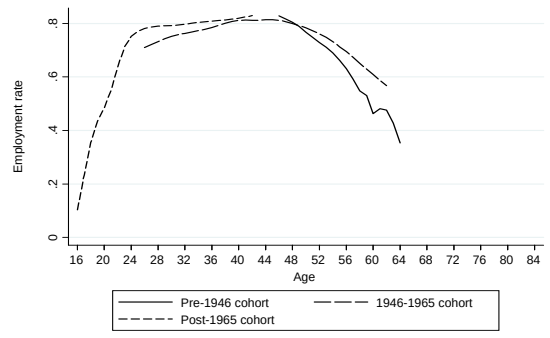
(d) HILDA limiting health problem



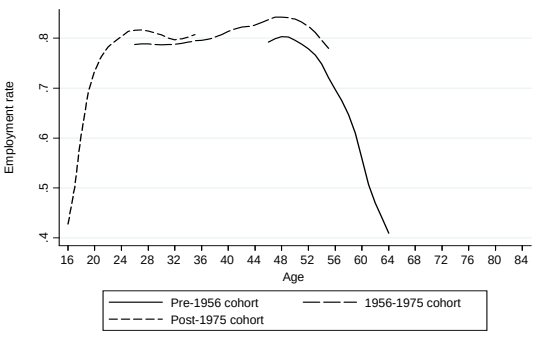
(e) BHPS health-related employment exit



(f) HILDA health-related employment exit



(g) BHPS employment



(h) HILDA employment

Figure A1 Age profiles for health and employment in BHPS and HILDA by birth cohort
(Nonparametric LOWESS estimates; bandwidth = 0.2)