Collecting biomedical and social data in a longitudinal survey: a comparison of two approaches

Matt Brown[1]*
Emily Gilbert[1]
Lisa Calderwood[1]
Kate Taylor[2]
Hannah Morgan[2]

[1] Centre for Longitudinal Studies, University College London
[2] National Centre for Social Research

*matt.brown@ucl.ac.uk

Keywords: longitudinal; biomarkers; response rates; data quality

Abstract

The inclusion of the collection of biomeasures within social surveys, and longitudinal surveys in particular, is becoming ever more common. Combining objective measurements of health with detailed information about lifestyles and behaviour collected over long periods of time offers enormous research potential.

Studies which combine an interview with the collection of biomeasures can be conducted in various ways. One model is that field interviewers make initial contact with participants, conduct the interviews and arrange follow-up visits for a nurse to collect the biomeasures. Alternatively, field interviewers can be trained to collect biomeasures, but there remain questions about whether the quality of data collected is comparable to that collected by a nurse. Other studies invite participants to visit clinics, but this can be very costly in a large-scale national study. There is no consensus on the optimal strategy for combining a social survey with the collection of biomeasures.

The 1970 British Cohort Study (BCS70) is a longitudinal birth cohort study which began in 1970. The 11th sweep of the study began in 2016, when study members were aged 46, and included an interview component alongside the collection of a range of biomeasures.

The first phase of fieldwork was conducted using a new approach where nurses conducted all of the data collection. Mid-way through fieldwork BCS70 switched to a two-stage approach where interviews were conducted by interviewers followed by a separate nurse visit. This presented a unique opportunity to evaluate the success of the two approaches.

Introduction

It has become increasingly common for longitudinal social surveys to combine traditional self-reported data with the collection of biomeasures. Surveys often ask participants questions about their health but it is known that responses can be prone to bias (Cannell et al., 1965; Madow, 1967). Objective health measures and biological specimen collection can augment survey data considerably, enabling researchers to discover things that cannot be captured through survey questions. Collecting biomeasures within social surveys allows us to increase the accuracy with which health can be assessed which then allows us to gain deeper understanding of the interplay between social and biological factors, including genetics, in explaining human behaviour (Benzeval et al., 2016).

Combining the collection of social and biomedical data in a large scale survey can be both operationally challenging and costly. A number of approaches are used in different studies, partly depending on the measures being collected. All of the approaches have advantages and disadvantages in terms of feasibility, cost, level of respondent burden and data quality.

Studies with a biomedical focus often invite participants to visit clinics, but this can be very costly in a large-scale national study. Social science focused studies typically collect biomeasures during home visits. Examples of the types of biomeasures collected include anthropometric measurements (height, weight etc.), measuring blood pressure, collection of saliva samples and dried blood spots. A common home visit approach is for field interviewers to make initial contact with participants, conduct interviews and arrange follow-up visits for a nurse to collect the biomeasures, such is the case for the English Longitudinal Study of Ageing (Steptoe et al., 2013) and the Health Survey for England (Mindell et al., 2012). Some studies have trained field interviewers to collect biomeasures so that the interview and biomeasure collection can take place in a single visit, such as the Health and Retirement Study in the US (Sonnega et al. 2014) and The Survey of Health, Ageing and Retirement in Europe (Guyer et al., 2017). In the first half of fieldwork for the 1970 British Cohort Study Age 46 Survey a new

approach was trialed in which nurses were trained to conduct interviews; again allowing the collection of interview and biomeasure data in one visit. The nurse-only approach has not previously been used in a large scale national longitudinal study. In the latter half of fieldwork, the two stage approach involving interviewers and nurses was used, creating a 'natural experiment' by which these two approaches can be contrasted. This paper provides evidence about the impact on response rates and data quality of using a 'nurse only' approach to collecting interview and biomeasure data in a large scale national longitudinal survey.

Background

Introducing the collection of objective health measures into an already complex data collection routine can increase the burden placed on study participants. Whilst a typical data collection visit in a longitudinal social survey might include a face-to-face interview and perhaps a self-completion questionnaire, the collection of biomeasures increases the time commitment required by the respondent because of the request to participate in a number of additional activities, such as anthropometric measurements, blood pressure and/or lung function measurements, and the collection of biological samples such as blood or saliva.

The impact of burden on survey response is complex and interacts with other factors such as topic interest. However, it is generally hypothesised in the survey methodology literature that the higher the perceived burden on a respondent, the more likely they are to refuse to participate in a survey (Collins et al., 1998; Holbrook, Krosnick and Pfent 2008). Longitudinal surveys must also consider the issue of cumulative burden, and the impact on attrition at future waves. Over time, it might be that asking the same people for more and more information, either social or biological, can increase the perceived burden of participating in the study. The consequence of increased burden on longitudinal study members can be attrition – dropping out of the study (Lynn et al., 2005).

Despite the issue of potential additional burden, there may be benefits from collecting biomeasures above the scientific ones. For some respondents, interest in the health measures collected by a study and the opportunity for a 'health check' could motivate participation. When discussing Whitehall II, a longitudinal study of British civil servants, Marmot and Steptoe (2008) suggest that participants can benefit from the medical screening that is part of the study, as health problems can be detected where they might not otherwise have been, and participants often see the clinic visits as a free health check-up.

However, there is also some evidence that health status can cause differential participation in medical aspects of a study. Looking at the 2006 Health and Retirement Study in the US, Sakshaug et al. (2010) find that individuals with physical limitations were less likely to participate in the full set of health-related measurements. However, those who reported having diabetes were more likely to take part.

Given the scientific benefits of collecting biomeasures from respondents in longitudinal social surveys, it is important that studies seeking to do this consider carefully the best approach to combining social and bio-medical data in order to mitigate the potential risks associated with the increased burden and encourage as many people as possible to take part.

UK studies of a medical origin such as the Avon Longitudinal Study of Parents and Children (ALSPAC) and Whitehall II have tended to use a clinic-based approach. A clinic based approach benefits from the study being able to set up biomeasure collection in a uniform way, particularly if there is a single clinic for participants to attend, and allows for the widest range of measures to be conducted including those that could not feasibly conducted in a home setting because of the bulky equipment required. It is also possible to use more expensive, specialist equipment in a clinic setting, because the data collection equipment doesn't have to be purchased for multiple data collectors. ALSPAC is a local area study and Whitehall II has a relatively small sample size; establishing the number of clinics that would be required to meet

the needs of large scale studies like the national birth cohort studies is very challenging and costly, although this has been done by UK Biobank.

There have also been studies who have attempted to collect biological data using a respondent-led approach, where respondents collect their own samples and send them to a lab directly. Examples include the UK Millennium Cohort Study, who asked parents to post their child's shed milk teeth to the study (Parsons and Platt, 2016), the ALSPAC and Whitehall II have asked participants to collect their own saliva, and the National Social Life, Health, and Aging Project has asked female participants to conduct vaginal swabs (O'Doherty et al., 2014). This is an attractive approach due to the potential for reduced costs and possibly higher compliance rates, but more research is required to understand the implications of respondent-led collection on data quality, among other things.

Most national UK longitudinal studies, especially those with a social science focus, such as the English Longitudinal Study of Ageing (ELSA), Understanding Society*:* the UK Household Longitudinal Study (UKHLS) and the 1958 British Cohort Study (National Child Development Study) have collected biomeasures during home visits.

A common approach used in home-visit focused studies is to use a two stage model where participants are first interviewed by an interviewer and then visited by a nurse at a later date who conducts the biomeasures. This approach was used by UKHLS and ELSA. A number of other UK cross-sectional studies e.g. Health Survey for England also use this approach.

The use of two data collection agents and separate visits is more expensive and there will inevitably be a degree of attrition between the two visits (Guyer et al., 2017). Clemens et al. (2012) looked at response rates to biomeasure collection if they are collected on a separate occasion to the interview, and that between 14% and 56% of those interviewed did not provide biomeasures.

In order to increase efficiency and reduce cost, some studies, including the Millennium Cohort Study, the US Health and Retirement Study and the US National Social Life, Health, and Aging Project (NSHAP) have successfully trained interviewers to collect a range of biomeasures so that interviews and biomeasure collection can be conducted in a single visit (Calderwood et al., 2014; Weir, 2008; Jaszczak et al., 2009). A single visit could seem more convenient to participants, as they don't have to make two appointments to complete their participation. This approach also reduces the potential for drop-out between the interview and the biomeasure collection.

Interviewers may not however be able to collect all biomeasures which a study may wish to collect. For example, there are no precedents for standard field interviewers being trained to collect intravenous blood samples. In addition, there may be concerns about the quality of biomeasure data collected by interviewers, although there is no conclusive evidence that this is the case. However, there is evidence that interviewers can take longer to collect biomeasures than nurses (McFall et al., 2014).

An alternative single-visit approach would involve training nurses to conduct interviews so that they can conduct all aspects of data collection. This approach has the same 'efficiency' advantages as the single visit from an interviewer approach, but allows a wider range of potentially higher quality biomedical measurements to be conducted.

However, there are also a number of potential disadvantages to a nurse-only approach. Firstly, given their specialist skills, nurses are typically paid higher wages than interviewers so data collection costs will be higher than an interviewer-only approach. Secondly, interviewers are experienced in making contact with respondents, tracing them if they have moved, and gaining their cooperation, especially among those respondents who may be more reluctant to participate. These are tasks that field nurses do not usually undertake, so there may be risks to response rates by asking them to take on these tasks. However, it is arguably the case that in the context of a well-established longitudinal survey, the job of persuading a participant to take

part should be less difficult than in a cross-sectional study, as prior participation will likely have resulted in at least a degree of investment in the study.

Field nurses are far less numerous than field interviewers, so a nurse only approach would likely involve a smaller pool of data collection agents than an approach using interviewers. As such, this may well have implications on the speed at which fieldwork can be covered.

There is also little evidence about the quality of data and accuracy of measurement for questionnaire data collected by nurses. Nurses do not usually administer standard face-to-face interviews, so there is uncertainty about whether the quality of the data collected by nurses would be comparable to that collected by the more experienced interviewers. One concern relates to levels of item non-response, particularly for sensitive questions. It is well-documented in the literature that financial questions about income, savings, debt, etc. are considered sensitive by respondents, and as such often have higher non-response rates than other types of questions in a survey (Moore et al., 2000; Tourangeau and Yan, 2007). It may be the case that interviewers are more used to encouraging respondents to answer questions that they may be reluctant to provide answers for, so the risk of using nurses is that item non-response will be higher. Added to this, there may be differences in the way that participants respond to particular questions when nurses versus interviewers are conducting the interview. For example, it could be that health-related measures are reported more accurately to a nurse, leading to differences in measurement between interviews conducted by nurses and by interviewers.

A nurse-only approach was used in the first half of the 1970 British Cohort Study (BCS70) Age 46 Survey. This is the first time that this approach has been used in a large scale longitudinal study so there is little evidence about its effectiveness. In the latter half of fieldwork, the more commonly used two stage approach using interviewers and nurses was used. This 'natural experiment' provides an opportunity to contrast the nurse-only model with the more

commonly used two stage model in terms of achieved response rates, measurement issues and cost.

Design and Methods

BCS70 is a nationally representative birth cohort study following 17,000 people born in one week in 1970 in England, Wales and Scotland. There have been 10 follow ups to date, at birth, age 5, 10, 16, 26, 30, 34, 38, 42 and the most recent sweep at age 46.

The Age 46 Survey had a biomedical focus, alongside the more traditional survey data collection. Data was collected by NatCen Social Research, who were appointed via a competitive tender process. Cohort members were asked to complete a face-to-face interview and a self-completion questionnaire, and four cognitive assessments, taking about 50 minutes. Alongside this, they had their height, weight, body fat percentage, waist and hip circumference measured, balance assessed, blood pressure taken, and intravenous blood taken for a range of analyses and DNA extraction, as well as being asked to complete a diet questionnaire for two days and wear and activity monitor for seven days – which took another 50 minutes. Participants were provided with the results of a number of the measurements and also their blood test results (glycated haemoglobin, total cholesterol and HDL cholesterol). The Age 46 Survey was the first time that any biomeasures had been collected from participants in adulthood.

Written consent was collected from participants for the blood sample collection and subsequent analysis and storage, and for the results of the blood analysis and blood pressure measurement to be sent to the cohort member's GP if they wished. The survey was reviewed and approved by the Health Research Authority's London-Central Research Ethics Committee.

As the study wanted to collect whole blood samples for analysis and DNA extraction, a trained nurse (or phlebotomist) had to be used for at least this element, as standard interviewers are

generally unable to take whole blood. In the UK context, none of the fieldwork agencies will allow their standard interviewers to take a whole blood sample.

Fieldwork was split into eight waves, spanning a planned 18 months of data collection. The initial approach to conducting the survey was to use nurses to conduct all elements – interviews and biomeasure collection - in one home visit. The primary motivation for opting for this approach was to save costs relative to a two stage model, but in addition, it was hoped that the approach would bring with it the advantages described above. Nurses attended a 4 day briefing which covered all of the elements of data collection, with a particular focus on the aspects that they would be less familiar with – namely contacting participants, tracing those who have moved, encouraging participation and conducting interviews.

A typical follow-up of BCS70 would use in excess of 200 interviewers, but the nurse panel at NatCen consists of around 120 nurses of whom 108 were briefed to work on BCS70. As such each nurse would have a greater number of cases and would need to travel longer distances to make visits than would typically be the case for an interviewer. To take account of this, an 18 month fieldwork period had been planned which is around six months longer than typical fieldwork periods. Despite the extended timetable, coverage of fieldwork fell behind and by wave three, the data collection was six months behind schedule. Additionally, the interview response rate being achieved by nurses was considerably under-target.

As such, at wave four, the use of a two-stage, interviewer-first approach was piloted, whereby interviewers conducted the interview, self-completion and cognitive assessments with respondents, then nurses visited a short while later to collect biomeasures.. Following the pilot, a judgement was made that the two-stage approach would likely improve response rates and would enable fieldwork to be completed much more swiftly as well as mitigate the risk of further delays. As such this approach was rolled out for the final four waves of fieldwork, creating a natural experiment through which the interviewer-first and nurse-only approaches can be compared.

More specifically, our research questions and subsequent hypotheses related to the two approaches are as follows:

1. Are interviewers or nurses more successful at locating and contacting respondents?
    - Interviewers will be more successful at locating and contacting respondents, given they have more experience of this task than nurses.
2. Are interviewers or nurses more successful at gaining co-operation?
    - Interviewers will be more successful at gaining co-operation from respondents, given they have more experience of this task than nurses.
3. Are interviewers or nurses more successful at gaining participation from those more likely to refuse to participate in the survey?
    - Interviewers will be more successful at gaining participation from those more likely to refuse to participate in the survey, given they have more experience of this task than nurses.
4. Does the two stage approach result in fewer biomeasures than the nurse-only approach?
    - The two-stage approach will result in fewer biomeasures than the nurse-only approach, due to higher drop-out between the interviewer and nurse visit.
5. Are there differences in the quality of interview data, as measured by rates of missingness, collected by interviewers and nurses?
    - The interview data collected by interviewers will be of higher quality (less missingness) than that collected by nurses, given interviewers have more experience of conducting interviews than nurses.
6. Do participants answer questions differently when interviewed by interviewers or nurses?
    - The fact that nurses are medically trained will result in respondents being more likely to report poor health and less socially desirable health behaviours.

The fact that the study is longitudinal allows us to utilise what we already know about study members from previous waves to help answer these questions. Specifically, characteristics of respondents collected in previous waves can be adjusted for in the two different 'experimental' groups, allowing for a fairer comparison of the two approaches. Additionally, for the third research question in particular, it can be determined who is most likely to refuse to participate based on participation history.

The nurse only approach was used for 5681 cases. These were cases allocated to Waves 1 to 3 in England and Wales and all cases in Scotland where there was insufficient interviewer capacity during the fieldwork period to roll out the interviewer first approach. The interviewer-first approach was used for 5050 cases, those allocated to Waves 5 to 8 in England and Wales. Wave 4 is excluded from the analyses as it was the pilot wave. The analysis focuses on interview response rates and data quality, as nurses were used to collect the biomeasures in both models.

Table 1 (below) compares the characteristics of the sample allocated to the two fieldwork models. The characteristics examined are all variables that have been found to be associated with non-response in BCS70 in prior sweeps. In the main, allocation to waves was based on geographical clustering and availability of nurses at the outset of fieldwork. However, it was also the case that 'difficult' cases, which had been identified based on their participation history as being likely to be reluctant or more likely to require tracing, were disproportionately allocated to earlier waves of fieldwork when the 'nurse only' approach was used. Related to this, the cases allocated to the nurse only approach were more likely to have last taken part at age 34 or 38 rather than age 42 (the previous sweep), were less likely to be educated to degree level and less likely to be employed in professional or managerial positions (see Table 1). These differences are all controlled for in the analyses that follow.

Table 1: Composition of sample allocated to the two fieldwork models

	Nurse only	Interviewer first	Diff (Int-Nurse)
Difficult case	19.0%	14.0%	-5.0%**
Last participation – Age 42 (2012)	76.8%	81.0%	4.2%**
Last participation – Age 38 (2008)	7.1%	5.0%	-2.1%**
Last participation – Age 34 (2004)	5.5%	3.2%	-2.3%**
Last participation – Age 30 (2000) or earlier	10.6%	10.8%	0.2%
Male	49.0%	49.0%	0.0%
Fair or poor health (Age 42)	15.3%	14.9%	-0.4%
Degree (last adult sweep)	30.4%	33.9%	3.5%**
Professional / managerial (last adult sweep)	56.0%	60.8%	4.8%**
Base: all issued	5681	5050	

KEY: * denote statistically significant differences at p<0.05
** denote statistically significant differences at p<0.01

The participant communication strategy differed somewhat between the two approaches, and this could potentially have had an impact on participation rates. In the nurse-only model, the advance mailing provided full information about what participating in the interview and the biomeasure collection would involve. It also specified that the entire visit would be conducted by a nurse. In the interviewer-first model the advance mailing fully described the interview phase but provided only brief information about a second visit which would involve a nurse conducting a series of health measurements. Full information about the biomeasures to be collected was provided in a second leaflet which was given to respondents on completion of their interview. Aside from this, the two approaches did not differ.

The other key difference between the two models is the number of data collection agents. As mentioned above, 108 nurses were briefed to work on the project, but the interviewer first approach involved 202 interviewers.

Findings

In this section we explore differences between interviewers and nurses in terms of the location/contact rates, co-operation rates and the interview response rates they achieved. We explore whether interviewers or nurses were more successful at gaining interviews with particular sub-groups, and in particular, sub-groups who are more likely to attrit from a longitudinal survey. We compare the number of biomeasure visits obtained in the two models and finally look at differences in the interview data collected by interviewers and nurses.

Contacting respondents, cooperation rates and response rates

We explore here whether interviewers or nurses were more successful at contacting participants and at gaining their cooperation, and the impact this has on the overall response rate for the interview part of the survey for the two groups.

Successful contact is here defined as making contact with the study via telephone or in person. As shown above, the characteristics of the sample allocated to the two models were different, with more 'difficult' cases being allocated to the 'nurse only' approach. The contact rates shown in Table 2 are therefore estimated marginal means from a regression model which includes the variables listed in Table 1 as controls. This approach is used in all of the analyses that follow (unless indicated). Interviewers were a little more successful than nurses at locating and contacting study members (90% compared with 87%) (Table 2). This is as expected, given interviewers have more experience of locating and contacting respondents than nurses do.

Given a sample member has been located, we then looked at whether cooperation rates differed between interviewers and nurses. Table 2 shows that interviewers were considerably more likely to be able to persuade a sample member to take part in the survey than nurses – 77% for interviewers and 70% for nurses. Again, given interviewers have more experience than nurses of persuading respondents to participate in surveys, this is not an unexpected finding.

These differences in contact and cooperation rates then have an impact on the overall response rates for the interview part of the survey for the two groups, as also shown in Table 2 Interviewers achieved a response rate nearly 7% higher than nurses – 69% for interviewers and 63% for nurses.

Table 2: Contact, cooperation and response rates
Percentages are marginal means adjusted for sample characteristics shown in Table 1

	Nurse only	**Int. first**	**Diff (Int-Nurse)**
Contact rate	87.3%	89.5%	2.2%**
Base: All issued (excluding ineligible)	5644	5007	
Co-operation rate	70.1%	76.7%	6.6%**
Base: All located and contacted	4895	4514	
Response rate	62.5%	69.3%	6.8%**
Base: All issued (excluding ineligible)	5644	5007	

KEY: * denote statistically significant differences at p<0.05
** denote statistically significant differences at p<0.01

The impact of previous-wave characteristics on response

The longitudinal nature of BCS70 allows us to examine whether interviewers or nurses were more successful at gaining interviews with participants with particular characteristics known from previous waves. Table 3 shows how the size of the difference in contact rates, co-

operation rates and overall response rates between interviewers and nurses differed by a range of variables which have all been identified as associated with the likelihood of attriting from longitudinal surveys. Each sub-group is examined one-by-one, and as such simple t-tests are conducted to test the significance of differences between interviewers and nurses, with no controls for other characteristics.

Interviewers were more successful than nurses across the board and there was no sub-group for whom nurses were more successful on any of the three measures of success.

Interviewers were considerably more successful than nurses at making contact with 'difficult' cases but there was no difference between interviewers and nurses when it came to finding 'non-difficult' cases. The difference in co-operation rates gained by interviewers and nurses was also slightly larger for difficult cases than non-difficult cases.

The difference in contact rates and co-operation rates between interviewers and nurses did not differ by the sex of the participant. It had been speculated that those in poorer health (as measured at the previous wave) may be more likely to be persuaded by a nurse to participate but this was not the case. The magnitude of the difference between co-operation rates achieved by interviewers and nurses was not affected by health status.

The difference in contact rates and co-operation rates between interviewers and nurses was greater for those with lower levels of education and those with lower social class occupations.

We have already shown that interviewers are more successful at gaining participation than nurses, but the overall pattern shown in Table 3 suggests that in particular, interviewers are more successful at gaining participation from groups who are more likely to attrit from a longitudinal survey, such as those who identified as likely to be 'difficult' based on their previous participation history, those with lower levels of education and those from lower social class groups.

Table 3: Differences between interviewers and nurses in contact, cooperation and response rates (for first issue cases)

	Location/contact rate – difference between interviewers and nurses	**Cooperation rate – difference between interviewers and nurses**	**Response rate**	**Base**
'Difficult' case	6.8%**	7.8%**	7.5%**	1783
'Non-difficult' case	1.1%	6.6%**	6.9%**	8877
Last int. – Age 42 (2012)	1.1%	7.0%**	7.5%**	8412
Last int. – Age 38 (2008)	6.5%	13.0%**	12.8%**	647
Last int. – Age 34 (2004)	11.7%*	7.5%	8.6%**	472
Last int. – Age 30 (2000) or earlier	6.5%*	4.8%	4.6%*	1128
Male	3.3%**	8.2%**	9.5%**	5256
Female	3.6%**	8.3%**	10.1%**	5395
Excellent to good health – 42	1.1%	7.0%**	7.5%**	7114
Fair or poor health – 42	1.3%	6.9%**	7.3%**	1262
Degree (last adult sweep)	2.2%*	6.1%**	7.6%**	3166
No degree	4.1%**	9.8%**	11.8%**	6692
Prof./ managerial (last adult sweep)	1.5%*	7.0%**	7.8%**	5421
Non-prof. managerial	4.4%**	8.5%**	9.9%**	3891

KEY: * denote statistically significant differences at p<0.05
** denote statistically significant differences at p<0.01

Level of effort put in by interviewers and nurses

We are able to delve further into the differences between interviewers and nurses by comparing the level of effort put in to locating and contacting respondents by each group. The paradata available allows a comparison between interviewers and nurses on the number of face-to-face visits made to each case. Table 4 shows that on average, interviewers made one more visit to each case than nurses and had made a greater made a greater number of visits to both productive and unproductive cases (and non-contact cases in particular).

Table 4: Mean number of face to face visits made to respondents, split by sample type

Figures are marginal means adjusted for sample characteristics shown in Table 1

	Nurses	Interviewers	Diff (Int-Nurse)
All cases	1.8	2.9	1.1**
Productive cases	1.4	2.8	1.4**
Unproductive cases	2.5	3.0	0.5**
Non-contact cases	3.9	4.7	0.8**

KEY: * denote statistically significant differences at p<0.05
** denote statistically significant differences at p<0.01

Biomeasure response rate

Table 5 shows that, as hypothesised, the nurse-only approach lead to a much lower level of attrition between the interview and the biomeasure collection than the interviewer first approach. Almost all participants (99%) who were interviewed by a nurse went on to complete the biomeasure collection. Biomeasure completion is defined as the collection of at least one biomeasure. For those who were interviewed by an interviewer and then offered a follow-up visit from a nurse, the biomeasure response rate was 15% lower at 84%. However, the difference in interview response rate was of a sufficient magnitude to mean that when examining the overall biomeasure response rate – i.e. the rate amongst all issued cases rather than all interviewed cases, the difference was substantially reduced to just under 4%.

Table 5: Biomeasure response rate (for first issue cases)

	Nurse only	Int. first	Diff. (Int-Nurse)
Response rate (All interviewed)	99.3%	84.1%	-15.2%**
Response rate (All issued)	62.0%	58.1%	-3.9%**
Base: All interviewed	3448	3552	

Base: All issued	5644	4979

KEY: * denote statistically significant differences at p<0.05
** denote statistically significant differences at p<0.01

Interview data quality

Now turning to our fifth research question, specifically, whether there are differences in the quality of interview data collected by interviewers and nurses. Item non-response is used as an indicator of data quality. In BCS70 it is typically the case that financial questions are those with highest levels of item non-response. We find no significant differences in item non-response between interviews conducted by nurses and those conducted by interviewers for all four financial items we looked at (see Table 6). Item non-response rates were also examined for a wide range of other questions including those relating to health (not shown). These questions had considerably lower item non-response rates than those for the financial questions and no differences between interviewers and nurses were found.

Table 6: Item nonresponse

	Nurse	**Int.**	**Diff. (Int-Nurse)**
Gross pay	2.1%	2.1%	0.0%
Total income	6.9%	7.9%	1.0%
Savings	6.3%	5.9%	-0.4%
Debt	4.3%	3.6%	-0.7%

KEY: * denote statistically significant differences at p<0.05
** denote statistically significant differences at p<0.01

Our final research question sought to examine whether participants answer questions differently when interviewed by interviewers or nurses. We selected a range of questions to assess, including income, savings and debt, a number of health behaviours (alcohol

consumption, smoking, and whether the respondent is trying to lose weight) and a number of health measures (self-reported health, number of health problems, long standing illness or disability, number of hospital visits in the last five years, reported chest pain, self-reported weight, and difference in self-reported and measured weight). We also looked at the scores obtained in three of the cognitive assessments respondents took part in (details in Table 7).

When looking at the financial questions, we were reassured to find no significant differences between responses to the interview questions when they were asked by nurses versus interviewers. In terms of health behaviours, there were no differences between reports of alcohol consumption or smoking but those interviewed by nurses were more likely to report that they were trying to lose weight. Losing weight could be classified as a socially desirable behaviour, and it may be that the presence of a medically trained nurse encouraged respondents to over-report this.

We had hypothesised that participants may be more likely to report poor health when interviewed by a nurse. There was no difference in the proportion who described their health as 'fair' or 'poor', nor the proportion reporting suffering from a long-term illness or disability but when going through an inventory of 20 health conditions a slightly higher number of conditions were reported to nurses. However, a slightly higher number of hospital visits were reported to interviewers and respondents were also more likely to report suffering from chest pain to interviewers. So, overall no consistent pattern or poor health being more likely to be reported to nurses was found.

During the interview, participants were asked to report their weight. There was no difference between self-reports to interviewers and nurses. Nurses also conducted objective measurements of weight, allowing self-reported weight to be compared with actual weight. The difference between self-reported weight and measured weight was significantly smaller for those who were interviewed by a nurse, suggesting that participants more accurately reported their weight to nurses.

The survey included a series of cognitive assessments. Nurses had not previously administered these assessments but most of the interviewers had experience of conducting these in other studies. Scores recorded by interviewers and nurses did not differ on the two word recall tasks but interviewers recorded slightly higher scores on the animal naming task. This assessment involves the participant naming as many animals as they can in a minute whilst the assessor notes them down and records the total. It is a more difficult test to administer and requires high levels of concentration, so the additional experience interviewers had gained previously likely accounts for this difference.

Table 7: Measurement differences

	Nurse	Interviewer	Diff.
Finance			
Gross pay per week (mean)	£737	£777	
Savings (mean)	£38,964	£42,693	
Debt (mean)	£10,168	£9,705	
Health behaviours			
Beer (no. of pints last 7 days) – beer drinkers (mean)	6.5	6.7	
Wine (no. of glasses last 7 days) – wine drinkers (mean)	8.6	8.6	
Current smoker	21%	21%	
Trying to lose weight	48.3%	46.0%	*
Health			
Self-reported health – fair or poor	18.9%	19.7%	
Number of health problems (mean)	1.4	1.3	*
Long-standing illness/disability	36%	36%	
Hospital visits in last 5 years (mean)	2.6	2.8	*
Reports chest pain	9%	11%	*
Self-reported weight (kg) (mean)	81.8 kg	81.2 kg	
Difference between measured and self-reported weight (mean)	1.6 kg	2.5 kg	**
Cognitive tests			
Immediate word recall (out of 10) (mean)	6.6	6.6	

Delayed word recall (out of 10) (mean)	5.5	5.5	
Animal naming (mean)	23.4	23.9	**

KEY: * denote statistically significant differences at p<0.05
** denote statistically significant differences at p<0.01

Interview length

On average, nurses took 5% longer than interviewers to conduct a complete interview (51.6 minutes compared with 48.8 minutes). The interview was comprised of 11 modules and nurses took longer to administer 9 of them. The modules where the difference in average length differed most greatly were the employment module (16% longer for nurses) which collects details of economic activity history and details relating to current occupation including earnings and the income module (12% longer for nurses) which collects information about household income from sources other than employment. These two modules are arguably the most complex aspects of the questionnaire to administer. The extra time taken by nurses likely results from a comparative lack of experience of asking questions of this nature.

Conclusions

For the BCS70 Age 46 Survey, which combined survey data collection with biomedical measurement, we had the opportunity to compare two different approaches to collecting interview and biomedical data – a nurse-only approach, and a two-stage interviewer-first strategy. This unique opportunity has given us insights into differences between these two approaches in collecting both social survey and biomedical data.

We found, perhaps unsurprisingly, that interviewers were significantly more successful than nurses at locating and contacting respondents and gaining co-operation from study members. Making contact, tracing movers and persuading reluctant respondents to take part are key aspects of an interviewer's role but nurses are far less experienced in these tasks and as such, lower than expected response rates were a known risk when adopting the nurse-only approach.

In the context of a longitudinal study such as BCS70, the 'interviewer' task is arguably easier than in a cross-sectional study, as many participants are well-invested in the study and will require little persuasion to participate again. For this reason it was hoped that nurses' lack of experience would be less problematic, but this was not the case.

Interviewers were particularly more successful at gaining co-operation from cases identified as likely to be more difficult to contact and to persuade to take part, and also those with lower levels of education and those from lower social class groups but they also had significantly more success with those who, based on prior participation, were not expected to be difficult. In order to maintain representativeness it is particularly important for longitudinal studies to retain more difficult cases, and so using nurse only approach does not seem advisable in this regard.

High response rates rely on data collection agents making repeat phone calls and face-to-face visits to study members, including to those who are initially reluctant to take part. A significant number of nurses were reluctant to fully comply with aspects of the contact strategy and some were sufficiently unhappy with performing these tasks that they declined to work further on the study. Examining paradata on number of face-to-face visits makes evident that interviewers made more effort to secure interviews and more visits before accepting an unproductive outcome.

Capacity is another factor that will have affected the differential success of the two approaches. The number of nurses trained to work on the study was significantly smaller than the number of interviewers because research nurses are a scarcer resource. As a result, nurses had to travel larger distances to make visits to study members, making it less feasible to carry out unsolicited visits. As fieldwork progressed the number of nurses willing to continue working on the study declined, and so this issue was exacerbated.

The different participant communication strategies used in the two approaches may also partially explain the differences. 'Nurse only' cases received information about all aspects of the Age 46 Survey upfront including a full description of the interview which would last 50 minutes and all of the planned biomeasures which would take a further 50 minutes. Those contacted using the 'interviewer-first' approach received full detail about the interview upfront and only brief information about the nurse visit. They were told that participating would take 50 minutes. For the 'nurse-only' cases, the initial perceived burden of taking part may have therefore seemed greater than for those who were initially only asked to take part in an interview.

When looking at which model allowed us to collect biomedical data from the largest proportion of respondents, we find that although the drop-off between the interview and nurse visit is much greater in two stage model (as we would expect), the difference in the initial interview response rate is sufficiently large that the difference in overall biomeasure response rate is not huge.

When examining item non-response rates as a measure of data quality, no differences between interviewers and nurses were found. In terms of measurement, no differences were found between interviewers and nurses for income, savings and debt questions. There was some limited evidence that socially desirable health behaviours were more commonly reported to nurses than to interviewers, but this was not consistently the case. Comparing self-reported weight to measured weight suggests that participants more accurately reported their weight to nurses than to interviewers, but there was no consistent evidence that participants were more likely to report poor health to nurses.

It's also worth noting that once the two stage model was implemented there were no further fieldwork delays.

There are of course cost implications to consider. Using a two-stage approach involving interviewers and nurses is more expensive than using nurses alone. The estimated cost of achieving an interview and a biomeasure visit is estimated to be 13% higher using the two-stage approach rather than the nurse only approach. It is estimated that if the two-stage approach had been used throughout the entire fieldwork period, then taking into account the differential interview response rates and biomeasure response rates that would have been achieved, the fieldwork costs would have been 10% higher than had a nurse only approach been used for the whole time.

An alternative approach could be to use interviewers to conduct the interview and all biomeasures except the blood draw which could be conducted in a second visit by a phlebotomist. This approach has not been used in a large scale longitudinal study in the UK, but is potentially worth exploring.

Overall the evidence suggests that nurses can be effectively trained to collect interview data which is of comparable quality to that collected by interviewers, but nurses are much less successful than interviewers at contacting participants and gaining co-operation, particularly for more difficult cases. For this reason, despite the financial implications, it is very unlikely that BCS70 will use the nurse only approach again, and our findings suggest that other similar studies should also be extremely cautious about using this approach.

References

Benzeval, M., Kumari, M. and Jones, A.M. (2016). How Do Biomarkers and Genetics Contribute to Understanding Society?. *Health Economics*, *25*, 1219–1222.

Calderwood L., Rose N., Ring, S. and McArdle W. (2014). Collecting saliva samples for DNA

extraction from children and parents: findings from a pilot study using lay interviewers in the UK. *Survey Methods: Insights from the Field.* Retrieved from https://surveyinsights.org/?p=3723.

Cannell, C.F., Fisher, G., and Bakker, T. (1965). Reporting of Hospitalization in the Health Interview Survey. *Vital and Health Statistics, U.S. Public Health Service, Series 2*, No. 6.

Clemens, S., Given, L. and Purdon, S. (2012). Methods of Collecting Biological Data: Considerations, challenges and implications. Presentation to the 67th annual American Association for Public Opinion Research conference. Orlando, Florida.

Collins, M., Sykes W., Wilson, P. and Blackshaw, N. (1988) Nonresponse: The U.K. experience. In Groves, R.M., Biemer, P.B., Lyberg, L.E., Massey, J.T., Nicholls II, W.L. & Waksberg, J. (Eds.), *Telephone Survey Methodology*. New York: John Wiley & Sons, Inc.

Guyer, H., Ofstedal, M.B., Lessof, C. and Juerges, H. (2017) The benefits and challenges of collecting physical measures and biomarkers in cross-national studies. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. Retrieved from https://hrs.isr.umich.edu/sites/default/files/biblio/Collecting%20PM-Bio%20Data_DocumentationReport.pdf

Holbrook, A.L. Krosnick, J.A. and Pfent, A. (2008). The cases and consequences of response rates in surveys by the news media and government contactor survey research firms. In Lepkowski, J.M., Tucker, C., Brick, J.M., de Leeuw, E., Japec, L., Lavrakas, P.J., Link, M.W. & Sangster, R.L. (Eds.), *Advances in Telephone Survey Methodology*. New York: John Wiley & Sons, Inc.

Jaszczak, A., Lundeen, K. and Smith, S. (2009). Using Nonmedically Trained Interviewers to Collect Biomeasures in a National In-Home Survey. *Field Methods, 21*(1), 26-48.

Lynn, P., Buck, N., Burton, J., Jäckle, A. and Laurie, H. (2005). A review of methodological research pertinent to longitudinal survey design and data collection. *ISER Working Paper Series 508111*.

Madow, W. Ci. (1967). Interview Data on Chronic Conditions Compared with Information Derived from Medical Records. *Vital and Health Statistics, U.S. Public Health Service, Series 2*, No. 23.

Marmot, M. and Steptoe, A. (2008). Whitehall II and ELSA: Integrating epidemiological and psychobiological approaches to the assessment of biological indicators. In Weinstein, M., Vaupel, J.W. & Wachter, K.W. (Eds.), *Biosocial Surveys*. National Academies Press: Washington (DC).

McFall, S., Conolly, A. and Burton, J. (2014). Collecting Biomarkers Using Trained Interviewers. Lessons Learned from a Pilot Study. *Survey Research Methods, 8*(1), 57–66.

Mindell, J., Biddulph, J.P., Hirani, V., Stamatakis, E., Craig, R., Nunn, S. and Shelton, N. (2012) Cohort profile: The Health Survey for England. *International Journal of Epidemiology, 41*(6): 1585-1593.

Moore, J.C., Stinson, L.L. and Welniak, Jr., E.J. (2000). Income measurement error in surveys: a review. *Journal of Official Statistics*, *16*(4), 331-361.

O'Doherty, K., Jaszczak. A., Hoffmann, J.N., You, H.M., Kern, D.W., Pagel, K., McPhillips, J., Schumm, L.P., Dale, W., Huang, E.S. and McClintock, M.K. (2014) Survey field methods for expanded biospeciman and biomeasures collection in NSHAP Wave 2. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 69*(S2), S27-S37.

Parsons, S.J. and Platt, L. (2016). Enhancing social surveys through the postal collection of shed milk teeth: an example of large-scale cost-effective collection on a longitudinal study. *Survey Research Methods*, *10*(1), 1-14.

Sakshaug, J.W., Couper, M.P. and Ofstedal, M.B. (2010). Characteristics of Physical Measurement Consent in a Population-Based Survey of Older Adults. *Medical Care, 48*(1), 64–71.

Sonnega, A., Faul, J.D., Ofstedal, M.B., Langa, K.M., Phillips, J.W.R. and Weir, D.R. (2014) Cohort profile: the Health and Retirement Study (HRS). *International Journal of Epidemiology, 43*(2): 576-585.

Steptoe, A., Breeze, E., Banks, J. and Nazroo, J. (2013) Cohort profile: The English Longitudinal Study of Ageing. *International Journal of Epidemiology, 42*(6), 1640-1648.

Tourangeau, R. and Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *13*(5), 859.

Weir, D. (2008). Elastic powers: The integration of biomarkers into the health and retirement study. In Maxine Weinstein (Ed.), *Committee on Advances in Collecting and Utilizing Biological Indicators and Genetic Information in Social Science Surveys*. Washington, DC: The National Academies Press.

Acknowledgements

The study is core funded by the Economic and Social Research Council. Additional funding for the Age 46 Survey was provided by the Medical Research Council and the British Heart Foundation.